International Journal on Advances in Life Sciences

2014 vol. 6 nr. 3&4

The *International Journal on Advances in Life Sciences* is published by IARIA.
ISSN: 1942-2660
journals site: http://www.iariajournals.org
contact: petre@iaria.org

International Journal on Advances in Life Sciences
Volume 6, Number 3 & 4, 2014

Robert Reynolds, Wayne State University, USA
Joel Rodrigues, Institute of Telecommunications / University of Beira Interior, Portugal
Alejandro Rodríguez González, University Carlos III of Madrid, Spain
Nicla Rossini, Université du Luxembourg / Università del Piemonte Orientale / Università di Pavia, Italy
Addisson Salazar, Universidad Politecnica de Valencia, Spain
Abdel-Badeeh Salem, Ain Shams University, Egypt
Åsa Smedberg, Stockholm University, Sweden
Chitsutha Soomlek, University of Regina, Canada
Monika Steinberg, University of Applied Sciences and Arts Hanover, Germany
Jacqui Taylor, Bournemouth University, UK
Andrea Valente, University of Southern Denmark, Denmark
Jan Martijn van der Werf, Utrecht University, The Netherlands
Liezl van Dyk, Stellenbosch University, South Africa
Lisette van Gemert-Pijnen, University of Twente, The Netherlands
Sofie Van Hoecke, Ghent University, Belgium
Iraklis Varlamis, Harokopio University of Athens, Greece
Genny Villa, Université de Montréal, Canada
Stephen White, University of Huddersfield, UK
Levent Yilmaz, Auburn University, USA
Eiko Yoneki, University of Cambridge, UK

International Journal on Advances in Life Sciences
Volume 6, Numbers 3 & 4, 2014

CONTENTS

Energy Harvesting Nanostructured Porous Silicon Scaffolds for Enhanced Efficiency Biofuel Cells

Mohamad Hajj-Hassan, Hussein Hajj-Hassan, Hassan M. Khachfe
Dept. of Biomedical Engineering
Lebanese International University
Beirut, Lebanon
mohamad.hajjhassan@liu.edu.lb, houssein.hajjhassan@liu.edu.lb, hassan.khachfe@liu.edu.lb

***Abstract*— An Enzymatic biofuel cell is a specific type of fuel cell which uses enzymes as catalysts to oxidize its fuel. Because of their efficient size and immobility, they pose as a great promise in terms of their relatively inexpensive components and fuels, as well as a potential power source for bionic implants. Here, we present the use of dry-etched nanotextured porous silicon scaffolds as a basis for new biofuel designs. Such an architecture increases the contact surface area of silicon with surrounding biofuel to enhance the process of harvesting of energy, and consequently, the efficiency of the cell.**

Keywords-nanotechnology; porous silicon; biofuel cell

I. INTRODUCTION

Recent advances in micro and nanotechnologies allow the development of implantable, portable, and miniature devices for a broad range of applications, including biomedical fields [1]. Powering implantable medical devices necessitates the development of lightweight, non-toxic and stable sources of energy with long life spans. In fact, the number of battery charging cycles in micro-energy harvesting methods is a major source of limitation [2]. Several micro-energy harvesting sources have been already identified in previous research, namely, low and high frequency electromagnetic Radio Frequency (RF) signal harvesting, conversion of vibration into energy, thermal and pressure gradients energy harvesting in addition to the latest attempts towards organic energy generation directly within the human body using fuel cells [3, 4].

Harvesting energy using ambient vibration has been the focus of various projects [5-7]. Devices made for this purpose are mechanically modeled with a base excitation of an elastically mounted seismic mass moving past a coil. Optimal architecture for maximal power generation under different operating conditions has also been shown [8]. Various applications of this principle have manifested in systems integrated in footwear to harvest energy from walking [9], while in other designs piezoelectric and electromagnetic generators convert pressure variations into electricity [10]. The power generated using these methods ranges from tens to hundreds of milliwatts [4, 7, 8]. On the other hand, several studies have focused on energy harvesting from low frequency vibrations [7, 11]. This concept was made viable by creating a generator that converts low-frequency environmental vibrations to a higher frequency by employing the frequency up conversion technique [12, 13]. One major limitation of this technology is encountered with patients that are not able to perform any physical activities in order to power the generator and, hence, produce the necessary charging current.

Energy harvesting using RF inductive coupling is a very promising technology, particularly in the presence of such a wide variety of RF signals in our everyday environment. Additionally, this technology can also be used to send data back to a base station, thus creating a two-way link. The system consists of a power generating circuit linked to a receiving antenna in order to capture the RF signal and convert it to a DC voltage [14]. The main challenge in this technology is in the receiver's capacity to read various frequencies, as well as the use of efficient power rectifiers. Several interesting studies have reported either the use of multiple energy harvesting antennas in one area [15], which has shown that an increase of 83% in area results in 300% increase in power, or the design of a high efficiency, ultra-low voltage active rectifiers [16].

This article covers the use of porous silicon scaffolds for biofuel cells. The next section presents and compares different types of biofuel cells. Section III introduces the immobilization and electrodes configurations for energy harvesting. Section IV presents porous silicon technology. Section V discusses existing porous silicon fabrication techniques. Section IV details the fabrication process of the porous silicon scaffolds using XeF_2.

II. BIOFUEL CELLS

The first enzyme based glucose/O_2 fuel cell to generate electricity was introduced in 1964 by Yahiro et al., aiming at using this concept to power an artificial heart [17]. While the field of fuel cell research has flourished in various industrial and environmental arenas, biomedical applications started making use of the technology only after 2001, with recent successes in micro fuel cell technology [18-20]. The two most dominant classifications of fuel cells are enzymatic, illustrated in Fig. 1, and microbial, based on the catalyst used to oxidize or reduce the fuel used in the design [21]. While microbial catalysts offer more longevity to the fuel cell, microbial fuel cells require a barrier between the cathode and the anode and between the fuel cell and its surrounding environment [22]. Such a design increases its size and decreases the current density since the fuel cell lacks direct contact with the fuel. Most importantly, when it comes to the use of microbial fuel cell for implantable devices, long term infections, thrombosis and other types of complications raise serious concerns [23, 24]. Therefore, it is natural that the use of microbial fuel cells was limited to few studies, one of them suggesting its use within the intestinal environment inside the transverse colon [25]. On the other hand, enzymatic fuel cells have lower stability and shorter lifespan because the longevity of enzymes is in the range of 10 days [26]. This has driven research in enzymatic fuel cells towards short term uses such as glucose sensors, post-op temperature measurement or as a power supply for pressure sensors indicating blockage of fluid flow in the nervous system [23]. However, since enzymes are selective in nature, the design of enzymatic fuel cells can be made into microscopic sizes without the need for a separating membrane to regulate the flow of the fluid and enzymes used in its design, thus achieving higher power densities due to the direct contact between the probes and the fuel [26]. Continuous attempts to increase the lifetime of the enzymes exist using immobilization techniques or using magnetic iron nanoparticles that shield the enzymes from getting oxidized or self-digested [27].

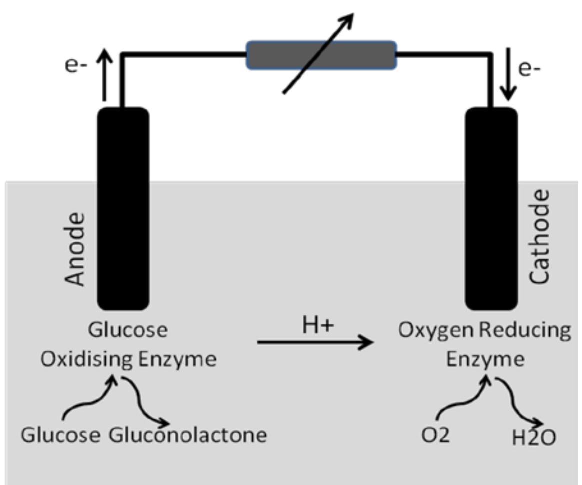


Figure 1. Illustration of an enzymatic biofuel cell using Glucose and Oxygen.

Another important factor in the design of the fuel cell is the target fuel. Although most implantable fuel cell studies have focused so far on the use of blood stream glucose, some studies have considered other alternatives such as the use of white blood cells based on their ability to generate electron current across their cellular membrane [28-30].

Most importantly, the complex environment inside the human body, such as the amount of glucose and oxygen available in addition to the neo-vascular build up that can hinder the exposure of the fuel cell to body fluids, represent important obstacles that any fuel cell design have to overcome in order for it to become a viable one [29]. Based on the first in-vivo study conducted by Cinquin et al., an enzymatic fuel cell was built by adjusting the types of enzymes used in order to account for the specific PH, concentration, and the effect of urea presence on the fuel cell [31]. This was implanted inside the peritoneal cavity of a rat, and has proven to provide a stable power of more than 7.52 μW/mL for a period of three months [31].

Here, we are interested in increasing the efficiency of energy harvesting in enzymatic fuel cells by increasing the contact surface area between the harvesting probes and the surrounding fuel. This can be achieved by (1) using porous interface to provide a large surface to volume ratio and consequently lager area of contact with enzymes (2) increasing the area of electrodes collecting the resulting amount electrons by using an array of electrodes. Doped porous silicon represents a good candidate due to the fact that it combines both biocompatibility and electrical conductivity [32, 33].

III. IMMOBILIZATIONS AND ELECTRODES PLATFORM

The proper functioning of an enzyme-based biofuel cell relies on both the chemical and physical properties of the immobilized enzyme layer. Physical and chemical methods can be used for immobilization of enzymes. Physical methods include: (1) Gel entrapment wherein enzymes are entrapped in a gel matrix, such as gelatine and polyacrylamide, as well as dialysis tubing [34]. (2) Adsorption where no additional reagents are required but only weak bonding involved between enzymes and electrode surface. Chemical methods are the main methods used for developing enzyme-based biofuel cells. The methods include covalent immobilization and immobilizing enzyme in polymer matrix.

Although enzymes are highly efficient catalysts they are difficult to incorporate into fuel cells. Low catalytic efficiency and stability of enzymes have been seen as barriers for the development of large-scale operations to compete with traditional chemical processes. This can be tackled with the use of nanostructured materials possessing large surface areas leading usually to high enzyme loading, resulting in improvement of power density of the biofuel cells.

Fig. 2 is the schematic representation of the novel enhanced porous silicon biofuel cell. It consists of an array of inter-digitized fingers made out of silicon covered by a layer of gold. The area between the inter-digitized fingers is made porous for entrapping and better immobilization of enzymes. The energy harvesting process is based on oxidation-reduction reactions taking place between the two arrays of inter-digitized fingers so that one array can be the anode and the second array can be cathode. Using microfabrication technology, it is possible to obtain identical anode and cathode electrodes. The enzymes get attached and immobilized between the two electrodes which in turn collect the resulting electrons and relay them to electronic circuit.

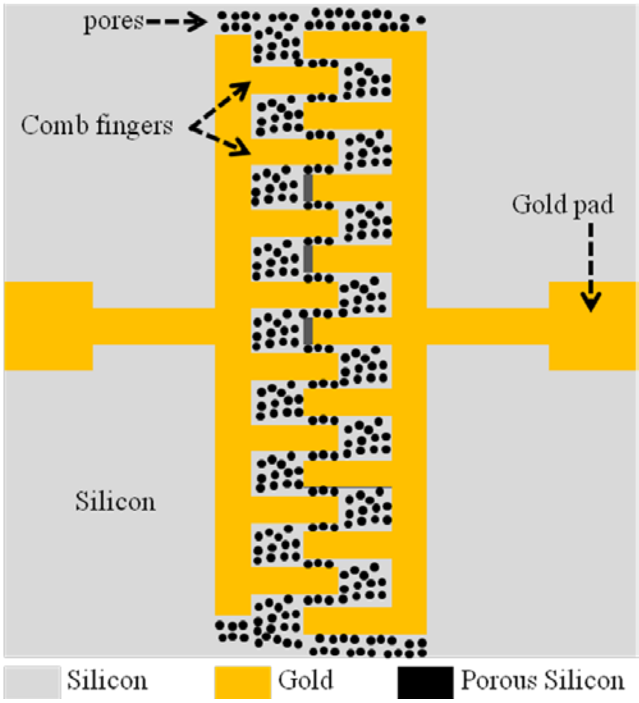


Figure 2. Schematic representation of the enhanced enzymatic biofuel cell.

IV. Porous Silcion Technology

Implantable biomedical devices built from bulk silicon have been available for biosensing and actuating applications for several years. However, this form of silicon is not biocompatible and so far this has prevented its use in vivo. Bulk silicon-based devices need coating or packaging in a biocompatible material, if they are to be used in and linked to living tissues [32, 33]. The majority of today's medical devices are coated with materials such as Polyvinylchloride (PVC), polypropylene, polycarbonate, fluorinated plastics and stainless steel. These materials are tolerated by the human body and are described as bioinert. An effective biomaterial, however, must bond to living tissue and is known as bioactive.

Nanostructured porous silicon (PS), whose particular texture can be described as a network of pores interconnected by solid nanocrystalline silicon, has properties that make it a very promising bioactive biomaterial [35, 36], in particular for devices that need to be linked to the biological system such as implantable devices [37]. Porous silicon material is useful and attractive for a wide variety of applications to develop biological sensors [37-39] and biomedical devices [40, 41]. This has significantly increased the interest in using porous silicon in biofuel cells.

An essential requirement for fabricating porous silicon in different applications is to have the ability to vary the size and configuration of the pores by choosing the appropriate fabrication parameters and conditions. For instance, for photonic bandgap filters, the pores are designed to the on the order of the wavelength of the light to retain and tune the optical reflectivity of the porous silicon [42, 43]. For biological sample filters, the pore size has to be large enough to allow the desired biomolecules to be filtered and cross through the pores freely [44].

V. Fabrication of Porous Silicon

Many previous reports have shown that porous silicon can be prepared through a galvanostatic, chemical, or photochemical etching procedures in the presence of hydrofluoric (HF) acid solutions or through stain etching [45-47]. Other methods such as pulsed anodic etching [48] and magnetic-field assisted anodization [49] were also employed for porous silicon preparation. In these techniques, the pore characteristics such as diameter, geometric shape and direction of the pores not only depend on the composition of the etching solution, but they also depend on temperature, current density, crystal orientation, dopant and doping density of the silicon substrate [45, 47, 50]. Moreover, porous silicon produced on large surface areas along with high porosity and/or thickness leads to a systematic cracking of the layer during the evaporation of the etching solvent. The origin of the cracking is the large capillary stress associated with evaporation from the pores. During the evaporation process, a pressure drop occurs across the gas/liquid interface that forms inside the pores [51].

A. Gas-Based Etching Technique

Gas etching method provides a suitable solution for the usage of the wet etching methods [52]. Due to the use conventional integrated circuit technology, the wet etching methods are not compatible with the widespread use of gas cluster tools. Moreover, wet etching techniques generate large quantities of dangerous waste in the manufacturing environment.

A mixture of oxygen (O_2) and nitrogen dioxide (NO_2) gases are combined with hydrogen fluoride (HF) and water vapors to produce photo-luminescent porous silicon layers as depicted in Fig. 3. The processes that were taken into account in the selection of these gases could be represented by a combination of the following chemical reactions.

Formation of nitric acid:

$$4NO_2 + 2H_2O + O_2 \rightarrow HNO_3$$

$$2NO_2 + H_2O \rightarrow HNO_2 + HNO_3$$

$$3HNO_2 \rightarrow HNO_3 + 2NO + H_2O$$

$$2NO + O_2 \rightarrow 2NO_2$$

Oxidation of silicon:

$$4HNO_3 + Si \rightarrow SiO_2 + 2H_2O + 4NO_2$$

$$Si + O_2 \rightarrow SiO_2$$

Etching of silicon dioxide:

$$SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O$$

The gas etching technique consists of exposing silicon samples to a mixture of O_2 and NO_2 gases in addition to HF and water vapors.

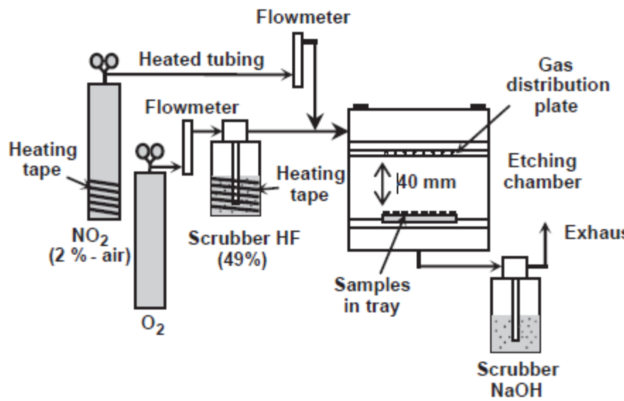


Figure 3. Schematic of the gas etching setup[52].

The experimental details were listed as follow:

- Silicon samples were loaded onto a tray that was mounted at the bottom of a chamber.
- The chamber was sealed after installing a gas distribution plate which aims to improve the uniformity of the gas flow. Noted that the chamber, tray, and distribution plate were made of chemically inert Teflon.
- Pure oxygen (99.995%) was flown through a scrubber containing HF (47-51%). The HF chamber could be kept at room temperature or heated up to 70℃.
- The scrubber HF merges with a flow of diluted nitrogen dioxide (2%) before entering the etching chamber. The NO_2 cylinder was heated at its base to a temperature of 40℃ to avoid accumulation of nitrogen dioxide at its bottom and enhance the mixing of NO_2 and air.
- The stainless steel tubing connection the NO_2 cylinder to the chamber was heated to a temperature of 30℃ to avoid the condensation of NO_2 on the tubing wall.
- The outlet of the chamber is connected to a scrubber containing sodium hydroxide (NaOH) solution which neutralizes the HF.
- The flow rates of O_2 and NO_2 could be varied by a flow-meter. The flow of O_2 and NO_2 varied between 10-50 ml/min.
- Samples are rinsed using ethyl alcohol (95%)
- Substrates are dipped in ethyl alcohol for 5 minutes and then removed.
- Substrates are left to dry in a high purity nitrogen environment (99.95%).
- Silicon samples were obtained from the dicing of boron-doped p-type wafers whose electrical resistivity was 20Ω.
- Samples were cleaned using RCA-type hydrogen peroxide mixtures, etched in a 5% HF solution for 2 minutes, rinsed in deionized water for 5 minutes, and then oxidized at room temperature in a SLM flow rate of ozone (O_3) gas with the presence of nitrogen for 5 minutes.
- Note that is cleaning sequence allowed a stringent control of the sample surface, yielding a hydrophilic surface.
- At the end of oxidization step, samples were loaded in the chamber and etching was performed for a time of 30 minutes.
- The morphology of the porous layers was investigated by a scanning electron microscopy. And the photo-luminescence properties of the porous layers were investigated using a photo-detection system (Fluorescence, PDS).

B. Stain Etching Technique

The formation of porous silicon by the strain etching process was conducted on p-type and n-type silicon wafers having different doping concentrations. Different porosity gradients were conducted since strain etching is a wet etching method which attacks the pore wall.

Experimental Procedure:

- Porous silicon layers were prepared on p-type and n-type wafers with doping concentrations of 2 x 10^{15} atom/cm³.
- More layers were prepared on p+ type and n+ type with doping concentration of 5 x 10^{18} atom/cm³
- Noted that the doping materials could be boron or phosphorus.
- The solutions for strain etching contained concentrated hydrofluoric acid and nitric acid with ratios between 50:1 and 500:1 [53, 54].
- Two additives were added into the solution, sodium-nitrite with a concentration of between 0.1 and 0.6 g/l in order to reduce the incubation time on PSL formation [55] and a surface-active substance to ensure that the evolving bubbles do not stick to the silicon surface [56].
- The mass of dissolved Si was 1 x 10^{-4} g.
- The as-grown porous layer was characterized by spectroscopic ellipsometrical(SE) measurements at 75° nominal angle of incidence in the spectral range from 280 to 840 nm.

By means of the gravimetrical and the spectroscopic ellipsometrical measurements the formation process of stain-etched PSL reveals continuous dissolution of the top surface of the layer and simultaneous formation of porous Si at the porous-crystalline interface. As a result, the stain-etched PSLs have self-limiting thickness when n-type substrates or low doped p-type substrates are used. The structural and optical characterization proved the existence of a porosity gradient in the layers, which stems from the partial dissolution of the pore walls and the top surface during formation. The morphology of the final structure is characterized by a random pore propagation direction in the case of low doped p-type and low and highly doped n-type silicon.

C. Electro-Chemical Etching Technique [11]

Electro-chemical etching is simply presented by the 'AMMT Porous Silicon Fabrication System'. This system is specifically designed for the fabrication of porous silicon using hydrofluoric acid as illustrated in Fig. 4. The user must wear appreciate safety gear and follow the safety guidelines set by the UCLA (University of California, Los Angeles) Nanolab using Hydrofluoric Acid.

The system contains:

1. Porous Silicon Bath (PSB): two chamber of HF-bath for porous silicon formation which are electrically isolated.
2. Waste Jug: located at the lower level of the PSB and is connected to the two HF drain valves.
3. Water Holders: two different water holders are available, one for the wafer and the other for the small samples.
4. Power Supply (Galvanostat): provide up to 24 A of electrical current.
5. Porous Silicon Galvanostat Software: Installed on the computer and enables the user to set the parameters and time etching.
6. Light source: 20 Watt halogen lamp located in front of the illumination windows.

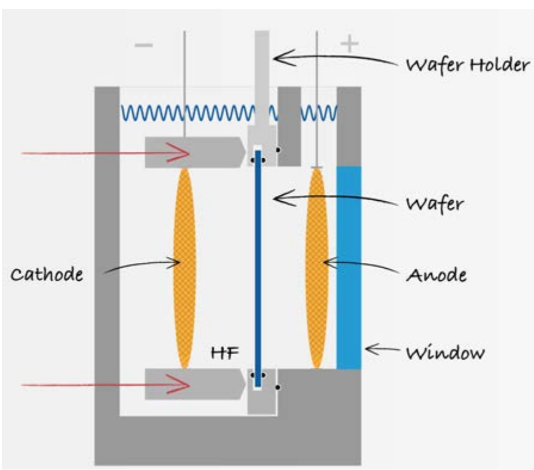


Figure 4. Schematic schematically setup of anodic experiment.

Preparation Process:

1. Cleaning: water should be cleaned from any impurities to ensure optimal porous etching result.
2. HF Preparation: wear the complete safety gear and check the drain valves.
3. Sample Mounting: mount the wafer in one of the wafer holders and place the wafer into the gap between the bayonet and the separator plate.
4. Electrode Positioning: place the electrodes in position with respect to the mounted wafer.

Noted that the electric contact of the electrode and the wafer is made through the HF electrolyte where no physical contact is required.

Fabrication Process:

1. Connect the PSB to the power supply.
2. Open the software and set the etching parameter.
3. To ensure illumination during etching, place the lamp close to the side window.
4. Set the timer for the desired time of etching.
5. After etching is over, turn off the power supply.

D. Photo-Chemical Etching Technique

The anodization is an easy method by which to form a luminescence layer on single crystalline Si. There are many difficulties in forming porous silicon on silicon on insulator (SOI) structure or on multilayered integrated circuit, since the anodization method requires electrodes in electrolyte solutions and on the back surface of a Si wafer. This becomes an obstacle for the applications of the PS for visible luminescence layers. A photochemical etching method that requires no electrode to form a visible luminescence layer on a single crystalline Si wafer is studied [57, 58].

Fig. 5 illustrates the following experimental procedure:

- An n-type (0.22-0.38 and 35-45 Ohm-cm) silicon wafer (100) was set at the bottom of the vessel filled with an etchant.
- A mixture of hydrogen fluoride acid solution (HF) and hydrogen peroxide (H 2O 2) as an oxidant were used for the etchant.
- Noted that the etchant concentration varied with HF: H_2O_2 = 100:17 ~ 250 for volume ratio.
- The silicon wafer surface was irradiated by a He-Ne laser (633 nm, 18.4 mW /cm²) as a visible laser for 5-45 minutes in order to form photochemically etched silicon.

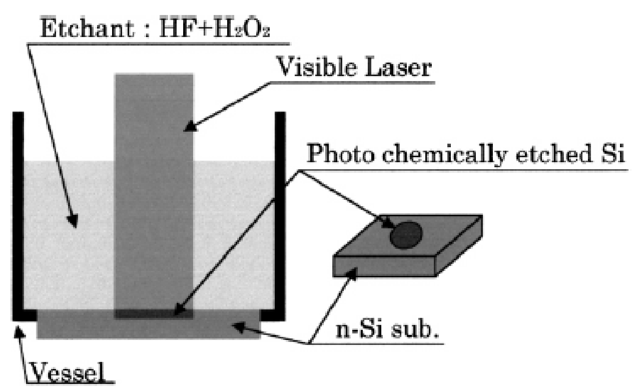


Figure 5. Schematic set up for a photochemical etching method [59].

The following reaction model of photochemical etching process of silicon atoms is depicted in Fig.6.

- He-Ne laser irradiation forms an electron-hole pair as a carrier in the Si substrate.
- H_2O molecular attacks to the wafer surface.
- Silicon atom is oxidized by H_2O and holes.

$Si + 2H_2O + nh^+ \rightarrow SiO_2 + 4H^+ + (4 - n)e^-$

h^+: hole, e^-: electron, $n \leq 4$

1. The SiO_2 region is solved by HF; therefore this means that a silicon atom is etched from the wafer.
2. H_2O_2 as an oxidant removes electrons left in the substrate, and H_2O_2 molecular and H^+ ions change into water molecules.

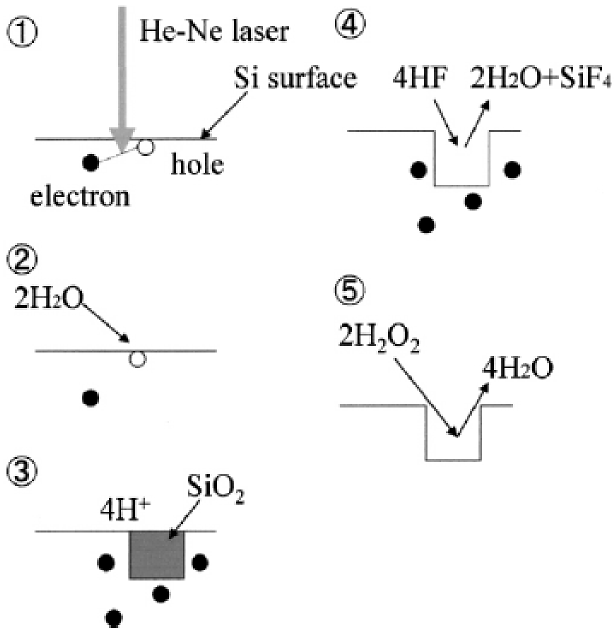


Figure 6. Schematic chemical reaction model of photochemical etching process of silicon atoms [58].

E. Chemical-Based Fabrication Technique

Porous silicon is usually fabricated under anodic polarization in an electrochemical cell. Another technique is introduced to form porous silicon without the use of any external source. Etching will occur by the formation of a galvanic cell, with the silicon acting as local anode and the metal as local cathode [60].

Experimental Procedures

1. Use of <100>-oriented n-type or p-type silicon with a resistivity of 2-5 Ω cm.
2. Etching of the sample with a dilute of HF solution with the use of some ethanol to prevent the formation of hydrogen bubbles.
3. Oxygen is the solution served as an oxidizing agent for the galvanic cell.
4. H_2O_2 may be added to increase the concentration of the oxidizing agent knowing that these agents are not reduced at the p-type silicon and do not cause the semiconductor to be etched chemically.

Two types of experiments could be performed.

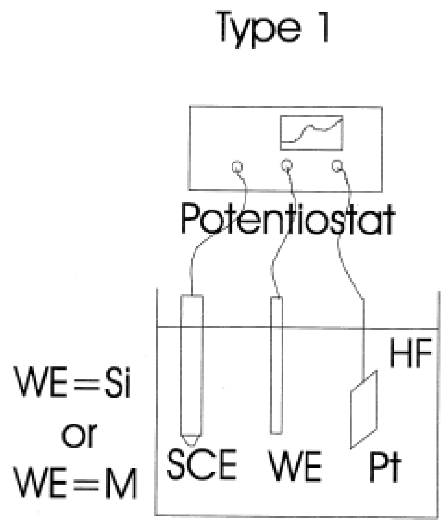


Figure 7. Schematic set up of type 1 experiment [60].

The experiments of type 1, illustrated in Fig. 7, were performed with either a Si working electrode (WE) or a metal (M) WE. A potentiostat the potential of the working electrode was regulated with respect to a standard calomel electrode (SCE). The current was measured between the WE and the Point counter electrode (Pt).

1. POS703 Bank potentiostat resulted in the formation of the current-potential curves of silicon and metal electrodes in HF solution (scan rate of 1mv /s).
2. The potential is always cited with respect to the standard calomel electrode (SCE) and the current is measured between a working electrode (WE) and a Point counter electrode (Pt).
3. The area of the silicon electrode was 0.5 cm² and the edges were protected from the solution by a HF resistant O-ring.
4. The galvanic cell formation is obtained when the Si electrode is connected in short circuit to the metal electrode.

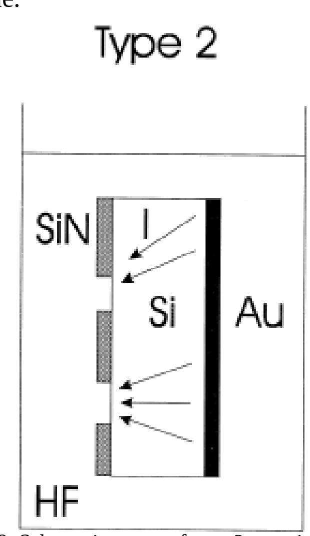


Figure 8. Schematic set up of type 2 experiment [60].

Experiments of type 2, illustrated in Fig. 8, were performed with an Au electrode on chip, noted that the potentials or currents could not be measured. This experiment demonstrates the formation of porous silicon without external contacts.

- A silicon nitride etch mask was deposited on one side of the wafer.
- An inert metal electrode was formed on the other side of the wafer by evaporation a film of Au /Cu.
- The metal /exposed silicon are ratio was typically 16.

The main advantage of galvanic porous formation technique is that a special sample holder to contact the Si is not required. This makes the technique suitable for batch fabrication of porous silicon devices. The contact between the silicon sample and a layer of noble metal is mandatory. The etching rate may be controlled by the metal /Si area ration and the concentration of oxidizing agent in solution.

F. Pulsed Current Etching

This technique for porous silicon formation is based on pulsed current anodic etching. The technique offers the possibility of fabricating luminescence material with selective wavelength emission depending on cycle time (T) and pause time (Toff) of pulsed current during the etching process as depicted in Fig 9.

Pulse current anodization of porous silicon is applied by a sequence of current pulses. During the pause period of anodic current, H_2 bubbles will desorb. Desorption of the H_2 bubbles allows fresh HF species inside the pores to react with silicon wall that sustains the etching process at appreciable rate. This process will increase the thickness of the porous silicon layer thus, enhancing the porous layer intensity [61].

Experimental Procedures

- Porous silicon samples were prepared by electrochemical etching of p-type silicon, boron doped, and 0.75-1.25 Ω cm wafers.
- The electrolytic cell is described in.
- Aluminum film was deposited on the back side of the samples to improve the uniformity of the anodic current.
- The electrolyte solution was a mixture of hydrofluoric acid (HF 49%) and ethanol (95%), 1:4 by volume.
- Anodization process was carried out for 30 min for all samples.
- An output signal from a pulse current generator was used to feed the current through the anodic etching circuit.

Note that both the cycle time (T) and the pause time (T off) were adjusted.

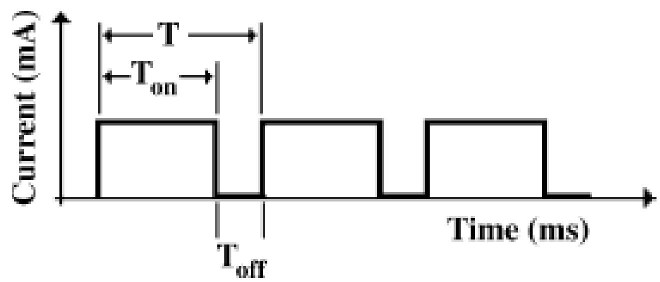


Figure 9. Schematic diagram of wave form of the pulse current used in the etching process [62] .

The PS formation sequence according to the current burst model is given as follows:

- Direct dissolution of silicon.
- Oxidization of silicon.
- Silicon oxide is dissolved.
- A slow surface passivation by H_2 starts at the clean surface.

To start the cycle again, each current burst has to overcome this H-passivation of the surface.

This process shows that there is significant freedom of choice available in peak spontaneous emission wavelength.

In this paper, we employ a novel and simple fabrication technique which employs Xenon Difluoride (XeF_2)-based dry isotropic etching to selectively form porous silicon in bulk single crystal silicon wafers [63]. XeF_2 is plasma-less etching technique and is based on the reaction of the fluorine ions, which represents the main etchant, with the solid silicon to produce – at room temperature – the volatile gas SiF_4. In a XeF_2-based etching process, a standard hard baked layer of photoresist can serve as a masking layer. In addition to its etching process simplicity, XeF_2 has a high etch selectivity to silicon. It reacts readily with silicon, but is relatively inert to photoresistance, silicon dioxide, silicon nitride and aluminum, which allows this technique to be used in the presence of CMOS integrated circuits as a post processing step. This is not the case when HF-based etching is used, as this latter will etch or damage the circuitry without a very hard mask followed by complex post-processing to remove the mask.

VI. METHODS

We utilized XeF_2 dry etching to create porous silicon surfaces on single crystalline silicon wafers. We used 3 inch diameter, 381± 20 μm thick <100> boron-doped (5–10 ohm cm) silicon wafers. The wafer was cut into 1.3 X 1.3 cm^2 that were then loaded in the XeF_2 etching chamber. The XeF_2 etching process does not depend on the silicon crystal orientation or its dopant content.

The fabrication process is achieved in a sequence of steps. First, undissociated gaseous XeF_2 is adsorbed onto the exposed areas of bulk silicon. The adsorbed gas is then dissociated into xenon and fluorine, after which the fluorine ions react with silicon to produce SiF_4 gas. Dissociation of the gas phase at room temperature leaves behind a porous silicon surface. In this process, increasing the etching process time increases the overall size of the pores and the thickness of the porous silicon film. The chemical reaction for the etching of silicon by XeF_2 is: Si + $2XeF_2$ → SiF_4 + 2Xe. As a dry etching technique, there is no post-fabrication drying step required, thus reducing the risk of damage to the newly formed porous surface.

XeF_2 leaves behind porous silicon surfaces on top of the remaining bulk silicon with porous silicon layer thickness on the order of several hundreds of nanometers (600 to 700 nm). The obtained porosity depends on the etching time. Fig. 10 shows a representative Scanning Electron Microscope image of porous silicon sample prepared using XeF_2.

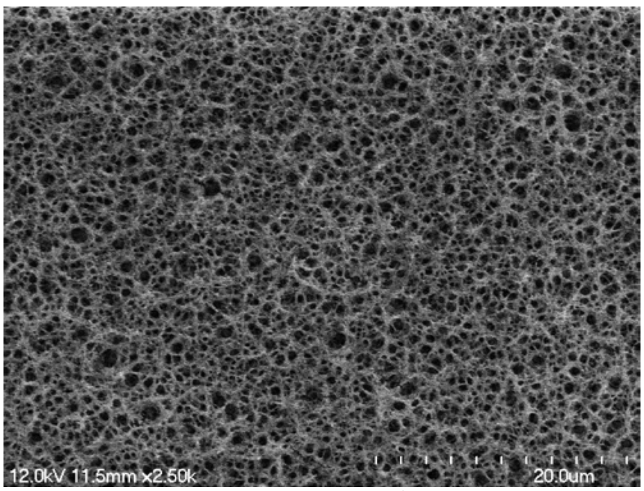


Figure 10. Scanning electron micrograph of a nanostructured porous silicon etched with XeF2.

VII. Fabrication of Enzyme Electrodes

The fabrication process of the comb capacitor starts with dicing 3 inch 381+20-mm-thick <100> boron doped (5–10 ohm-cm) silicon wafer into small 3x3 cm squared pieces. Metal layers of titanium (adhesion layer, 500 nm thick) and gold (conducting layer, 750 nm thick) were deposited by sputtering on the silicon wafer. A 1.4 micron thick layer of photoresist is then spun on and photolithographically patterned to define the inter-digitized fingers. The gold and titanium are then wet etched, with 1:2:10 I_2:KI:H_2O and 20:1:1 H_2O:HF:H_2O_2, respectively. Acetone was then used to remove the remaining photoresist thus exposing the gold layer covering the inter-digitized fingers. Fig. 11(a) is a scanning electron micrograph of the array of fingers. Fig. 11 (b) is a magnified view of the scanning electron microscope (SEM) picture of the network of pores fabricated selectively in between the silicon comb fingers covered with titanium-gold. Porous silicon layers with different porosities can be obtained by changing the etching recipes in the XeF2 system.

VIII. Conclusion and Future Work

Nanostructured doped porous silicon is a promising material for Biofuel cells. It offers several advantages, including the use of silicon in microelectronics, biocompatibility, and simplicity in tailoring porosity and conductivity. Dry etching of porous silicon using XeF_2 allows, due to its compatibility with integrated circuits, allows an easy integration of porous silicon scaffolds with the microelectronic harvesting integrated circuit. Future work will focus on testing porous silicon samples in complete enzymatic fuel cell setup.

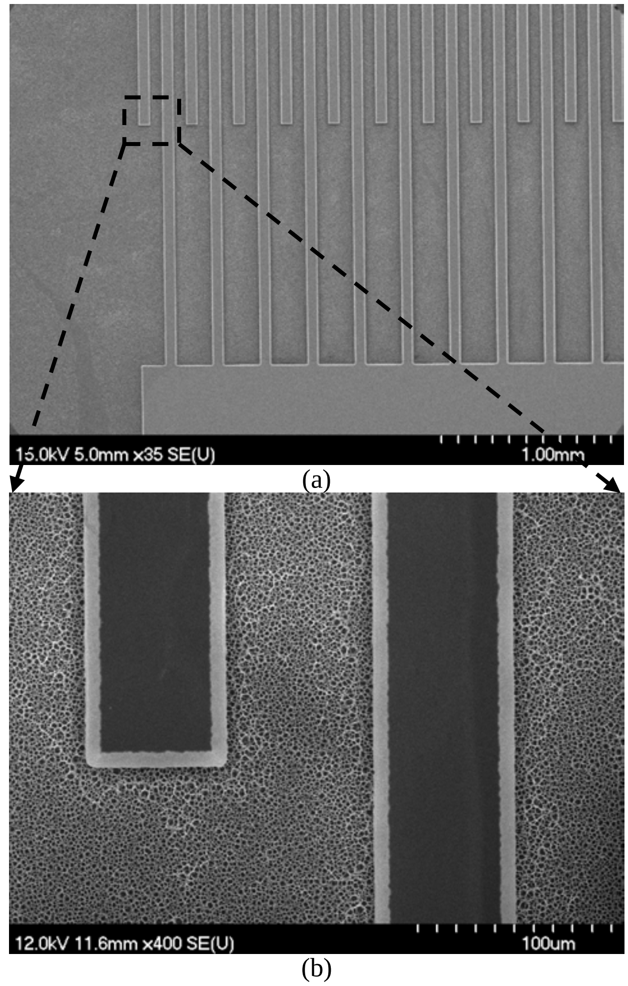


Figure 11. Scanning electron micrographs (a) novel inter-digitized energy harvesting electrodes (b) magnified view of two fingers with the porous area in between.

Acknowledgment

The current project is supported by the National Council for Scientific Research, Lebanon. The authors would like to acknowledge the assistance of the McGill's Nanotools and Microfabrication Laboratory, Montreal, in preparing the porous silicon samples, and the Lebanese International University for logistic assistance.

References

[1] M. Hajj-Hassan, Hajj-Hassan, H., and H. M. Khachfe, "Nanostructured Porous Silicon Scaffolds for Biofuel Cells," The Second

International Conference on Global Health Challenges, Lisbon Portugal, pp. 38-42, 2013.

[2] P. D. Mitcheson, G. K. Rao, A. S. Holmes, and T. C. Green, "Energy Harvesting From Human and Machine Motion for Wireless Electronic Devices," Proceedings of the IEEE, vol. 96, pp. 1457-1486, 2008.

[3] S. Heydari Nasab, M. Asefi, L. Albasha, and N. Qaddoumi, "Investigation of RF Signal Energy Harvesting," Active and Passive Electronic Components, vol. 2010, 2010.

[4] H. Chun-Ching, S. An-Shen, and C. Jing-Chih, "Improvement of Pyroelectric Cells for Thermal Energy Harvesting," Sensors, pp. 534-548, 2012.

[5] N. G. Stephen, "On energy harvesting from ambient vibration," Journal of Sound and Vibration, vol. 293, pp. 409-425, 2006.

[6] H. A. Sodano, D. J. Inman, and G. Park, "Comparison of piezoelectric energy harvesting devices for recharging batteries," Journal of Intelligent Material Systems and Structures, vol. 16, pp. 799-807, 2005.

[7] S. P. Beeby, RN. Torah, MJ. Tudor, P. Glynne-Jones, T. O'Donnell, CR. Saha, and S. Roy, "A micro electromagnetic generator for vibration energy harvesting," Journal of Micromechanics and Microengineering, vol. 17, p. 1257, 2007.

[8] F. Cottone and H. Vocca, L. Gammaitoni, "Nonlinear energy harvesting, "Physical Review Letters, vol. 102, p. 080601, 2009.

[9] J. Q. Liu, H. B. Fang, Z. Y. Xu, X. H. Mao, X. C. Shen, H. Liao, and B. C. Cai, "A MEMS-based piezoelectric power generator array for vibration energy harvesting," Microelectronics Journal, vol. 39, pp. 802-806, 2008.

[10] H. B. Fang, Liu, J. Q., Xu, Z.Y., Dong, L., Wang, L., Chen, D., Cai, B.C., Liu, Y., "Fabrication and performance of MEMS-based piezoelectric power generator for vibration energy harvesting," Microelectronics Journal, vol. 37, pp. 1280-1284, 2006.

[11] J. A. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," Pervasive Computing, IEEE, vol. 4, pp. 18-27, 2005.

[12] H. Kulah and K. Najafi, "Energy scavenging from low-frequency vibrations by using frequency up-conversion for wireless sensor applications," Sensors Journal, IEEE, vol. 8, pp. 261-268, 2008.

[13] O. Zorlu, E. T. Topal, and H. Kulah, "A vibration-based electromagnetic energy harvester using mechanical frequency up-conversion method," Sensors Journal, IEEE, vol. 11, pp. 481-488, 2011.

[14] B. Jiang, J. R. Smith, M. Philipose, S. Roy, K. Sundara-Rajan, and A. V. Mamishev, "Energy scavenging for inductively coupled passive RFID systems," in Instrumentation and Measurement Technology Conference, 2005. IMTC 2005. Proceedings of the IEEE, 2005, pp. 984-989.

[15] M. Mi, M. H. Mickle, C. Capelli, and H. Swift, "RF energy harvesting with multiple antennas in the same space," Antennas and Propagation Magazine, IEEE, vol. 47, pp. 100-106, 2005.

[16] A. J. Stratakos, R. W. Brodersen, and S. R. Sanders, "High-efficiency low-voltage DC-DC conversion for portable applications," University of California, Berkeley, Fall 1998.

[17] A. T. Yahiro, S. M. Lee, and D. O. Kimble, "Bioelectrochemistry: I. Enzyme utilizing bio-fuel cell studies," Biochimica et Biophysica Acta (BBA)-Specialized Section on Biophysical Subjects, vol. 88, pp. 375-383, 1964.

[18] B. C. H. Steele and A. Heinzel, "Materials for fuel-cell technologies," Nature, vol. 414, pp. 345-352, 2001.

[19] R. Hahn, S. Wagner, A. Schmitz, and H. Reichl, "Development of a planar micro fuel cell with thin film and micro patterning technologies," Journal of Power Sources, vol. 131, pp. 73-78, 2004.

[20] T. J. Yen, N. Fang, X. Zhang, G. Q. Lu, and C.Y. Wang, "A micro methanol fuel cell operating at near room temperature," Applied Physics Letters, vol. 83, pp. 4056-4058, 2003.

[21] S. K. Chaudhuri and D. R. Lovley, "Electricity generation by direct oxidation of glucose in mediatorless microbial fuel cells," Nature biotechnology, vol. 21, pp. 1229-1232, 2003.

[22] B. E. Logan and J. M. Regan, "Microbial fuel cells-challenges and applications," Environmental science & technology, vol. 40, pp. 5172-5180, 2006.

[23] S. Calabrese Barton, J. Gallaway, and P. Atanassov, "Enzymatic biofuel cells for implantable and microscale devices," Chemical Reviews, vol. 104, pp. 4867-4886, 2004.

[24] B. E. Logan, B. Hamelers, R. Rozendal, U. Schröder, J. Keller, S. Freguia, P. Aelterman, W. Verstraete, and K. Rabaey, "Microbial fuel cells: methodology and technology," Environmental science & technology, vol. 40, pp. 5181-5192, 2006.

[25] Y. Han, C. Yu, and H. Liu, "A microbial fuel cell as power supply for implantable medical devices," Biosensors and Bioelectronics, vol. 25, pp. 2156-2160, 2010.

[26] I. A. Ieropoulos, J. Greenman, C. Melhuish, and J. Hart, "Comparative study of three types of microbial fuel cell," Enzyme and microbial technology, vol. 37, pp. 238-245, 2005.

[27] J. D. Keighron and C. D. Keating, "Enzyme: nanoparticle bioconjugates with two sequential enzymes: stoichiometry and activity of malate dehydrogenase and citrate synthase on Au nanoparticles," Langmuir, vol. 26, pp. 18992-19000, 2010.

[28] G. A. Justin, Y. Zhang, M. Sun, and R. Sclabassi, "An investigation of the ability of white blood cells to generate electricity in biofuel cells," in Bioengineering Conference, 2005. Proceedings of the IEEE 31st Annual Northeast, 2005, pp. 277-278.

[29] A. E. Franks and K. Nevin, "Microbial fuel cells, a current review," Energies, vol. 3, pp. 899-919, 2010.

[30] G. A. Justin, Y. Zhang, M. Sun, R. Sclabassi, "Biofuel cells: a possible power source for implantable electronic devices," in Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, 2004, pp. 4096-4099.

[31] P. Cinquin, C. Gondran, F. Giroud, S. Mazabrard, A. Pellissier, F. Boucher, J.P. Alcaraz, K. Gorgy, F. Lenouvel, and S. Mathé, "A glucose biofuel cell implanted in rats," PLoS One, vol. 5, p. e10476, 2010.

[32] S. C. Wang, F. Yang, M. Silva, A.W.Y. Zarow, and Z. Iqbal, "Membrane-less and mediator-free enzymatic biofuel cell using carbon nanotube/porous silicon electrodes," Electrochemistry Communications, vol. 11, pp. 34-37, 2009.

[33] C. Liu, S. Alwarappan, Z. Chen, X. Kong, and C.Z. Li, "Membraneless enzymatic biofuel cells based on graphene nanosheets," Biosensors and Bioelectronics, vol. 25, pp. 1829-1833, 2010.

[34] D. Schumacher, J. Vogel, and U. Lerche, "Construction and applications of an enzyme electrode for determination of galactose and galactose-containing saccharides," Biosensors and Bioelectronics, vol. 9, pp. 85-89, 1994.

[35] W. Sun, J.E. Puzas, T.J. Sheu, and P.M. Fauchet, "Porous silicon as a cell interface for bone tissue engineering," Physica Status Solidi A-Applications and Materials Science, vol. 204, pp. 1429-1433, 2007.

[36] M. Hajj-Hassan, M. Khayyat-Kholghi, H. Wang, V. Chodavarapu, and J.E. Henderson, "Response of murine bone marrow-derived mesenchymal stromal cells to dry-etched porous silicon scaffolds," J Biomed Mater Res Part A, pp. 269-274, 2011.

[37] S. Chan, P. M. Fauchet, Y. Li, L. J. Rothberg, and B. L. Miller, "Porous silicon microcavities for biosensing applications," Physica Status Solidi a-Applied Research, vol. 182, pp. 541-546, Nov 2000.

[38] R. A. Williams and H. W. Blanch, "Covalent Immobilization of Protein Monolayers for Biosensor Applications," Biosensors & Bioelectronics, vol. 9, pp. 159-167, 1994.

[39] A. Janshoff, K. P. S. Dancil, C. Steinem, D. P. Greiner, V. S. Y. Lin, C. Gurtner, K. Motesharei, M. J. Sailor, and M. R. Ghadiri, "Macroporous

p-type silicon Fabry-Perot layers. Fabrication, characterization, and applications in biosensing," Journal of the American Chemical Society, vol. 120, pp. 12108-12116, Nov 25 1998.

[40] J. L. Coffer, M. A. Whitehead, D. K. Nagesha, P. Mukherjee, G. Akkaraju, M. Totolici, R. S. Saffie, and L. T. Canham, "Porous silicon-based scaffolds for tissue engineering and other biomedical applications," Physica Status Solidi A-Applications and Materials Science, vol. 202, pp. 1451-1455, Jun 2005.

[41] E. J. Anglin, L. Y. Cheng, W. R. Freeman, and M. J. Sailor, "Porous silicon in drug delivery devices and materials," Advanced Drug Delivery Reviews, vol. 60, pp. 1266-1277, Aug 17 2008.

[42] G. Lammel, S. Schweizer, S. Schiesser, and P. Renaud, "Tunable optical filter of porous silicon as key component for a MEMS spectrometer," Journal of Microelectromechanical Systems, vol. 11, pp. 815-828, Dec 2002.

[43] M. G. Berger, M. Thonissen, R. Arensfischer, H. Munder, H. Luth, M. Arntzen, and W. Theiss, "Investigation and Design of Optical-Properties of Porosity Superlattices," Thin Solid Films, vol. 255, pp. 313-316, Jan 15 1995.

[44] C. C. Striemer, T. R. Gaborski, J. L. McGrath, and P. M. Fauchet, "Charge- and size-based separation of macromolecules using ultrathin silicon membranes," Nature, vol. 445, pp. 749-753, Feb 15 2007.

[45] H. Foll, M. Christophersen, J. Carstensen, and G. Hasse, "Formation and application of porous silicon," Materials Science & Engineering R-Reports, vol. 39, pp. 93-141, Nov 1 2002.

[46] E. A. de Vasconcelos, E.F. da Silva, B.D. Santos, W.M. de Azevedo, and J.A.K. Freire, "A new method for luminescent porous silicon formation: reaction-induced vaporphase stain etch," Phys. Status Solidi A—Appl. Mat, vol. 202, pp. 1539–1542, 2005.

[47] R. L. Smith and S. D. Collins, "Porous Silicon Formation Mechanisms," Journal of Applied Physics, vol. 71, pp. R1-R22, Apr 15 1992.

[48] J. Escorcia-Garcia, O. Sarracino Martínez, J. M. Gracia-Jiménez, and V. Agarwal, "Porous silicon photonic devices using pulsed anodic etching of lightly doped silicon," Journal of Physics D: Applied, vol. 42, p. 145101 (7pp), 2009.

[49] T. Nakagawa, H. Koyama, and N. Koshida, "Control of structure and optical anisotropy in porous Si by magnetic-field assisted anodization," Applied Physics Letters, vol. 69, pp. 3206-3208, Nov 18 1996.

[50] Z. C. Feng and R. Tsu, Porous Silicon: World Scientific Pub Co Inc, 1994.

[51] M. Bouchaour, A. Ould-Abbas, N. Diaf, and N.C. Sari, "Effect of drying on porous silicon," J Therm Anal Calorim, vol. 76, pp. 677–684, 2004.

[52] S. Boughaba and K. Wang, "Fabrication of porous silicon using a gas etching method," Thin Solid Films, vol. 497, pp. 83-89, 2006.

[53] J. Xu and A. J. Steckl, "Stain ‐ etched porous silicon visible light emitting diodes," Journal of Vacuum Science & Technology B, vol. 13, pp. 1221-1224, 1995.

[54] A. Hidekazu, M. Akira, K. Akira, A. Tomoyoshi, and S. Akinobu, "A Comparative Study of Visible Photoluminescence from Anodized and from Chemically Stained Silicon Wafers," Japanese Journal of Applied Physics, vol. 32, p. L1, 1993.

[55] M. T. Kelly, J. Chun, K. M. Bocarsly, and B. Andrew, "High efficiency chemical etchant for the formation of luminescent porous silicon," Applied Physics Letters, vol. 64, pp. 1693-1695, 1994.

[56] V. Lehmann, H and Föll, "Formation Mechanism and Properties of Electrochemically Etched Trenches in n-Type Silicon," J. Electrochem. Soc. , vol. 137, pp. 653-659, 1990.

[57] S. Shih, K. H. Jung, T. Y. Hsieh, J. Sarathy, J. C. Campbell, and D. L. Kwong, "Photoluminescence and formation mechanism of chemically etched silicon," Applied Physics Letters, vol. 60, pp. 1863-1865, 1992.

[58] Y. Naokatsu and T. Hiroshi, "Blue Luminescence from Photochemically Etched Silicon," Japanese Journal of Applied Physics, vol. 38, p. 5706, 1999.

[59] N. Yamamoto and H. Takai, "Formation mechanism of silicon based luminescence material using a photo chemical etching method," Thin Solid Films, vol. 388, pp. 138-142, 2001.

[60] A. M. Hynes, H. Ashraf, J. K. Bhardwaj, J. Hopkins, I. Johnston, and J. N. Shepherd, "Recent advances in silicon etching for MEMS using the ASE™ process," Sensors and Actuators A: Physical, vol. 74, pp. 13-17, 1999.

[61] A. A. Aziz, "Fabrication and characterization of uniform quantum size porous silicon," Materials science forum, vol. 517, pp. 232-236, 2006.

[62] N. K. Ali, M. R. Hashim, A. Abdul Aziz, I. Hamammu, "Method of controlling spontaneous emission from porous silicon fabricated using pulsed current etching," Solid-State Electronics, vol. 52, pp. 249-254, 2008.

[63] M. Hajj-Hassan, M. Cheung, and V. Chodavarapu, "Dry etch fabrication of porous silicon using xenon difluoride," Micro & Nano Letters, IET, vol. 5, pp. 63-69, 2010.

Design, Creativity and Human Computer Interaction Design Education

Alma L. Culén
Department of Informatics
University of Oslo
Oslo, Norway
almira@ifi.uio.no

Henry N. Mainsah
Centre for Design Research
School of Architecture and Design
Oslo, Norway
Henry.Mainsah@aho.no

Sisse Finken
Department of Informatics
University of Oslo
Oslo, Norway
finken@ifi.uio.no

Abstract—**In this paper, we investigate ways to engage computer science students, majoring in design, use, and interaction (with technology), in design practices through an advanced graduate course in interaction design. We take a closer look at how prior embodied and explicit knowledge of HCI that all of the students had before their enrollment in the course, combined with better understanding of design and design practice, and in particular the emergence of creativity on both individual and team levels, shapes them as human-computer interaction designers. We evaluate the results of the effort in terms of increase in creativity, novelty of ideas, body language when engaged in design activities, and in terms of perceptions of how well this course prepared the students for the work practice outside of the university, usually, in multidisciplinary settings.**

Keywords—HCI education; interaction design; creativity; studio; design education; multidisciplinary teamwork.

I. INTRODUCTION

There is an increased movement towards informing and embedding education practices from other disciplines into Human Computer Interaction (HCI). We have discussed in our paper [1] how design practice and design pedagogy may contribute to HCI education.

Many authors have stressed a need for considering new pedagogical approaches to HCI education that creatively synthesize HCI theory and methods with design thinking-in-action (see, for example, [2]–[5]). Faiola has argued for development of pedagogical models intended for teaching HCI that *"provide students with knowledge domains that can account for understanding design, social context, and business strategies in addition to computing"*, [6, p. 30].

Winograd and Klemmer, discussing the reasoning behind opening of the now famous d.school at Stanford, an innovation hub with a core in human computer interaction design, state: "*The basic premise of the d.school is that students need two complementary kinds of training. The disciplinary training provided by conventional departments provides them with depth in the concepts and experience of a specific field. This gives them intellectual tools, but often misses the larger context of relevance and integration with other kinds of knowledge, which are required to innovate effectively in the 'real world'"*, [7, p. 1]. Such multidisciplinary and effective learning arenas are not easy to create. They represent innovative thinking and innovative education, which has not yet been able to prove itself worthy over time. Thus, embedding innovative educations into traditional educational institutions is difficult. However, the evidence is there that the multidisciplinary approach, such as that of the d.school, has its merits. In line with how Bannon argues why HCI needs to change in the 21st century [8], we argue that the HCI educations needs to change in order to accommodate for new technologies, new interaction forms, new practices, and new areas of research. One practice outside the traditional HCI field, which has a strong influence on changes taking place within HCI, is the design practice. Many scholars have explored the relation between HCI and design. Some of the notable results of these explorations are: a proposition to consider HCI as research through design, see [9]–[14], a proposition to consider Human Computer Interaction Design (HCID) as a radically interdisciplinary dialogue [15], convergent - divergent questioning [16], HCI design studio [17], models, theories and frameworks toward a multidisciplinary science [18].

Two of the authors, of this paper, work within department of informatics, teaching traditional HCI and qualitative research methods. The third author works at a traditional design institution, the school of architecture and design. Over the past few years, the two schools have cooperated and run a graduate course in interaction design together. The course took place at the school of architecture and students from both institutions worked on design projects in multidisciplinary teams. The cooperation recently came to an end, as the design school faculty felt that the differences in traditions and practices between the two schools were too far away from each other. This situation was the immediate motivator for exploring different venues and different approaches to teaching design practice and design thinking within the department of informatics.

In this paper, we present the teaching approach that we have chosen and the results of applying the design oriented practices, more specifically, design thinking and design pedagogy in the context of an advanced HCID course in the department of informatics. Our goal was not to educate designers, but to teach HCI students about design practice through direct experience and reflection. Similar approaches have been advocated by other scholars, e.g., [3][10][19]. Our approach differs from those also in that we really wanted to keep the multidisciplinary in focus. Our intention was to prepare HCID students for better collaboration and participation in multidisciplinary teams, to bridge some of the differences in traditions, cultures, to learn, inspired by the d.school, about design thinking and design practice, and to understand a reflexive practice. In order to evaluate the success of our approach we have chosen the following criteria: emergence of creativity, novelty of generated ideas, body language when engaged in design activities and perception of how well the course prepared students for the work practice outside the university.

The students, as mentioned, were master degree students who are about to graduate from computer science department with degree in design, use and interaction. They are soon to be considered as professionals specialized in HCI and with a choice of research, education or interaction design practice as their future work. Regardless what they choose, as Churchill, Bowser and Preece point out in [20], they would have to be *progressionals* – people who follow closely and persistently with technological progress and master new technical competencies, new design and evaluation methods, while keeping a solid base in HCI.

Creativity is something that both scientists (also HCI practitioners) and designers need in their work. However, it is cultivated and expressed differently within practices of science and design. According to Owen [21], creative people tend to work in one of the two ways: by invention (makers) or by discovery (finders), see Fig. 1.

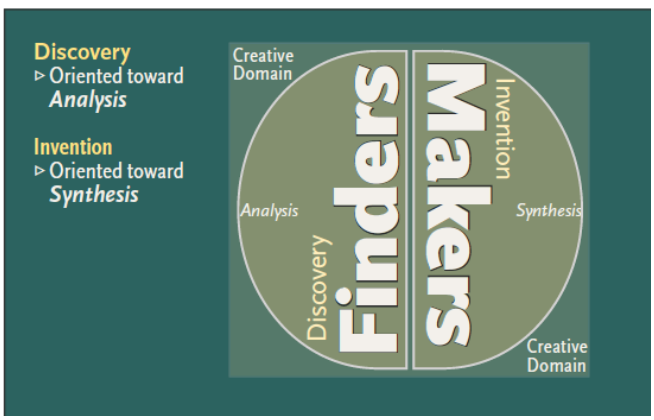


Figure 1. Creativity model according to Owen, from [21].

In HCI practices, the insight is often confused with ‘scientific’ creativity. Similarly, within makers’ practices, originality is frequently identified with creativity; we find such identifications problematic, or worthy of further scrutiny. Both insight and originality come about rarely, while, we believe, creativity is something that may be learned and cultivated [22][23]. In learning and cultivation of creativity, the environment plays an important role. As Csikszentmihaly points out, *“It is easier to enhance creativity by changing conditions in the environment than by trying to make people think more creatively. And a genuine creative accomplishment is almost never the result of a sudden insight, a light bulb flashing in the dark, but comes after years of hard work [22, p. 7] ”*.

The modern study of creativity is making advances, [24]. Creativity, just like the HCI has moved through three waves. The first wave of creativity research focused on personalities of exceptional creators. The second wave investigated internal mental processes that occur when people are engaged in creative activities and behavior. The third, current wave is concerned with socio-cultural, interdisciplinary approaches and relates to social systems and groups of people performing acts of creativity together. It is this brand of creativity studies that best applies to HCI design students, and we tried to better understand it in the context of the course described in this paper. It is also the one that fits the third wave of HCI the best [25][26], the wave that addresses humans, their values, emotions, everyday lives and the role of technology in it.

The paper is structured as follows: in Section II, we briefly describe the design model used by the school of architecture and design to teach interaction design and then the HCI pedagogical model that the students were familiar with. We proceed to explain in Section III, our case, where the new pedagogical model was applied to teach a course in HCID. In Section IV, we discuss our findings, and sum up the paper in the concluding Section V.

II. PEDAGOGICAL MODELS: DESIGN AND HCI

A. Design model

Hoadley and Cox state: *Design is an important class of human activity because it links theory and practice, bridging scientific activities with creative ones in order to deal with ill-structured, open-ended problems*, [27, p. 20]. To solve problems for real-life contexts, designing combines formal knowledge, experience, practice and judgment, both through and in action. Schön proposes an *“epistemology of practice implicit in the artistic, intuitive processes, which some practitioners bring to situations of uncertainty, instability, uniqueness, and value conflict,”* that he characterized as a “reflective practice”, [28]. Design pedagogy can be understood through a socio-cultural perspective on learning that is centered on developmental aspects occurring between cultural and socially mediated actions in contemporary and legacy contexts [29].

Figure 2. The studio classroom gives opportunities for creative development, discussions, feedback and rapid prototyping.

Design studio learning (see Fig. 2) and teaching relies on the integration between people. The studio learning happens usually in small groups of students. Schadewitz and Zamenopoulos say: *"The studio model has fostered the type of enculturation into practice that modern schemes for distributed situated learning are just coming to understand,* [30, p. 1]*"*. Shaffer [31] presents the academic design studio as a coherent system where surface structures, pedagogy, and epistemology interact to create a unique learning community. Surface structures refer to components of the learning environment such as the space, furniture, assignments and so forth. Pedagogical activities include activities such as iterative design cycles, field research, and group discussions of work in progress. Epistemological understanding describes the beliefs and the nature of design knowledge and how it is constructed [3]. Brandt and colleagues draw upon Lave and Wenger's concept [32] of "communities of practice" that describes learning communities where novices are first introduced as legitimate peripheral participants and integrated more centrally into the community through their participation in increasingly more complex tasks. Learning-in-practice is contextual and situated in time and space and is shaped by the historical dimensions of institutions and participants' own life experiences that contribute to shape the manner in which the learning environment is enacted. In this way, it is similar to an "ecological approach" to understanding learning in a more holistic way that brings together a focus on practice, tools, learning environments, and social context; see [3, p. 336]. The teaching rarely involves research articles, books to teach from, or regular formal lectures.

B. HCI model

Many HCI pedagogical approaches include a mix of user-centered requirement analysis, design, implementation and evaluation [19]. This mix is exactly what our students have received through three HCI courses, which they had taken prior to the graduate class described here. All three courses are project based and in all three the students do work in teams. Their first course covers material from the book [33]. In the second course, students gain theoretical and practical knowledge on how to study situated use of technology and how such studies can inform design of technology.

Reimer and Douglas [34] point out that such study program often falls short of teaching students good design of real-world artifacts, while engaging in real-world design processes. In order to address the real-world settings, the third course in HCI, using [35] as a course book, defines projects based on the needs of local companies and organizations, thus bringing real life project experience into the classroom. However, they use a classical teaching model consisting of two hour-long lectures, in a lecture hall (see Fig. 3), and two hour-long sessions in smaller groups. The later provides help with exercises from the book, questions around the material covered during the lectures, or issues related to the project work. The third course in HCI also offers an hour-long design feedback session with the instructor and a representative of a company that students are designing for. The projects are carried out in project teams of 3-4 students. Although the third HCI course addresses the issue of real-life problems, there is still a gap between multidisciplinary teamwork in professional circles and what students can experience in terms of teamwork in the context of this HCI course.

Figure 3. Regular classroom lectures are still common when learning HCI.

Another important aspect of learning, present in some design disciplines and often lacking within HCI, is related to approaches that emphasize speculative and inductive ethos. Lewis argues in [36] that technology education nowadays needs to promote more than simply knowledge of materials, mastery of special technical skills and techniques, or correct use of tools or instruments. It should move beyond these to pursue "more subjective and elusive goals", [36, p. 35]. Among these goals he includes creative insight. According to Lewis, the teaching of design is ideally suited to uncover students' creative potentials, because design allows open-endedness [36, p. 45]. Design problems are ill-structured, solutions are not defined in advance, and pathways to the solution are open. Cropley, in [37], identifies these issues precisely as conditions that promote creativity. Creativity can be nurtured through a pedagogical framework that builds on an open-ended problem solving, using design processes for real-life contexts [38].

While *"HCI specialists still focus on the issues that gave birth to the field: Are technologies learnable, usable, useful, reliable, comprehensible, ethical? We are still concerned with assessing whether technologies serve, engage, and satisfy people and extend their capabilities, or frustrate, thwart, and confound them"*, as the authors state in [20, p. 44], so does most of HCI education as well. In order to answer questions about users and technology, the focus is, naturally on users and what they do with the technology. The students of HCI are thus also trained in seeking the input from users, whether it is for research or design purposes. However, as Bødker [26] points out, the so-called third wave of HCI includes broader consideration of cultural and historical embeddedness of technology, also in non-work contexts, where emotion and aesthetics play a much larger role. The third wave of HCI has, therefore, comes closer to traditional design disciplines, not only through aesthetics, but also focus on solving real-life problems through design of technological solutions.

We are now in a position to present our case and describe challenges and lessons learned from introduction of design studio practices into teaching an advanced HCI class.

III. THE CASE: DESIGN PRACTICES IN HCI TEACHING

A. *Course Organization and Structure*

In implementing a design practice in the context of the class we chose to work in the lab, where it was possible to implement practice-based learning. The students had access to materials such as scissors, paint, fabric, paper, tools like sawing machine, hammers, pliers and like, as well as electronic components, such as Lilypads, GPS sensors, LED lights, wires, welding station, etc. The number of students was restricted to ten. They were all advanced graduate students with three prior courses in HCI, as described in the previous section, including experience with user-centered and participatory design approaches. The teaching team consisted of the two in-house teachers and one teacher from the school of architecture and design. The later attended the class approximately every third week, providing feedback on the students' design projects. The authors of this article are the three teachers who have an insatiable curiosity about creativity and how it emerges. Specifically, we are curious about what happens when HCI practitioners cannot rely on the usual ways of thinking and working – that is, when they do not have the support of users in the design process. It is commonly considered that HCI designers should design interactive products *"to support the way people communicate and interact in their everyday and working lives"*, [33, p. 9]. In order to design such products, HCI designers rely on user participation and user studies. These studies inform the design, but also split the responsibility for "good" design between designers and users who informed it. It could be said that users are a great support to HCI designers in, at least, the following ways: helping in testing products and prototypes, informing design processes, participating in them, allowing designers to observe them interacting with technology and last but not least, by making HCI designers feel that designs processes are not dependent on the mystery of creativity and creative processes.

Throughout the course, the in-house teachers have uploaded literature of relevance to a dropbox. This literature covered a range of different subject matters such as: design thinking, design anthropology, differences between interaction design practices within design and HCI, service design, participatory service design, design research, and an article concerning design of wearable technology. Some of the papers were chosen in response to students' project ideas and others aimed at explaining the differences between practices of Interaction Design (ID) as thought at design schools and HCID.

Instead of having the traditional lectures according to the earlier described model, which the students are used to, the shared time between the teachers and the students was spend on discussing various topics, design ideas, and on providing feedback on designs in progress. Some, perhaps unusual forms of stimulating students to be more open and creative were used. For example, in order to increase the energy level and engagement, we would form a circle, and in turn, everyone had to "design" a move that the whole circle then repeated for a while, and then the next person got engaged, see Fig. 4. Other times, we encouraged new ways of exploring the world [39]. For example, we brought artists with interesting ideas and products into the classroom; see Fig. 5. Altering the 'lecture set-up' in this manner was a part of the pedagogical aim of introducing new ways of conducting HCI teaching, with intention to increase the D(esign). In doing this, we aimed at making the students step outside of their comfort zone, as well as encourage bodily engagement and hands-on design practices.

Figure 4. Students and a faculty member standing in a circle and "creating" new body movements.

B. *The Assignment*

In order to further support the bodily engagement and hands on practices, the students were asked to complete two projects during the semester. The aim of the first project was to design an exhibit addressing the activities and research interest of the group for design. The exhibit was shown at a Student Faire (held annually at the department of informatics), presenting the work of different research groups. The Faire also featured representatives from many local IT companies.

The second project was to design an installation for the library. The interactive installation had as a goal to bring forward those resources and services, available through the department's library, which usually remain hidden or under-used. This second project is not described in this paper.

Additionally, we handed out briefs with targeted questions concerning creativity, work effort in class, expectancies of outcomes from the course, and addressed issues concerning multidisciplinary work.

C. *The data collection methods*

During the first eleven weeks of the course, we carefully observed and documented the students' work on their first design projects. In documenting the process we took photographs and collected Post-it notes, which we used to quickly note input (aim, what, how, do-ability) during a feedback session. Further, we took notes during conversations with students, or when they presented their work.

Figure 5. Amanda Steggell, on the far left, an artist who made the energy bank, showed her product (in the red square, see also [40]).

Both in-house teachers and students have taken photographs as part of learning how to use visual methodologies, amounting to well over 300 images, which were shared in the aforementioned dropbox. The dropbox was made specifically for this class to share photos, presentations of the design projects, and other class related material, such as the literature. The photographs, used in this paper, are from the shared pool in the dropbox. From a teacher's perspective the photographs have been a way of documenting [41] the process from the first drafting of ideas to the materialization of the designs. In addition, the photographs have served as information, beyond mere documentation, and have been used as entrances to gain understandings about increase in creativity, novelty of ideas, and body language when students engaged in design activities, or presentation of design outcomes, see Fig. 4 and Fig. 7 – Fig. 11. Furthermore, a number of targeted questions were asked and answered either orally or in writing during the semester. Finally, each student has filled out a questionnaire at the end of the semester, where questions targeted student's perception of their own creativity, ability to work in the multidisciplinary teams, learning outcomes and opportunities for the improvement of the course.

IV. NURTURING CREATIVITY AND BLENDING MODELS

Blending of pedagogical models was carried by introducing desk crits, allowing for bricolage and assemblages of skills and practices, where the usual HCID practice was naturally one of the components (reading academic papers, as well as actively using whatever knowledge from HCI was appropriate for the task at hand).

A. "Re-cycling"ideas and materials

It was clear that the students used "re-cycled" ideas, by taking ideas from one domain and applying them to another. This reflects a creative practice common in design processes often referred to as *bricolage*. The term *bricolage* (French for "tinkering") is commonly associated with the French anthropologist Claude Levi Strauss [42], and it refers to the construction and creation of a work from a diverse range of things that happen to be available in order to create something new.

In Fig. 6 we can see a picture of a white board with images on paper representing ideas of different concepts juxtaposed together. We see on the image how the students in the group took an old idea, "faces in places" and integrated it into a new one, a heated glove bicycle glove, to create a concept for warming the hands of cyclists during winter. Finding "face in places" (finding images that resemble faces, Fig. 6 upper right corner) was suggested as an extra activity during the exhibit. However, neither the faculty nor the students could find any faces in the building, making the task really hard and little fun as an activity on the day of the exhibit. The idea itself was not novel and neither was a way of using it [39].

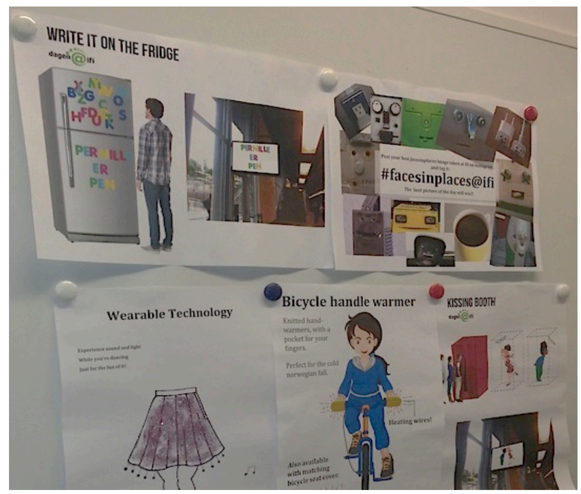


Figure 6. Among the initial ideas one can see many "re-cycled" ones, such as the "faces in places", mixed with new ones, such as heated bicycle gloves for biking in harsh nordic weather conditions.

In this context, the old objects carry a meaning, given to them by their past uses, and the creators' experience, knowledge and skill, a meaning that can be modified by the requirements of the project and the creator's intentions.

In this case, the practice of mixing the old and the new fits well with the ethos of the theme of the students' project, which was on sustainaible uses of energy. By "re-cycling" old ideas into new ones, the students are creatively "re-cycling" ideas in the process of creating an artefact based on "re-cycled" materials.

B. The desk crits

In the course we drew on one of the most central component of design studio pedagogy, the desk crit (constructive critique). The desk crit is simply an extended and loosely structured interaction between the designer (in this case the HCI student) and a critic (teachers and fellow course mates) involving discussion and collaborative work on a design process, Fig. 7.

In the course, this process involved the students displaying their work, presenting their plans, and getting feedback from their fellow students, teachers, and guest teachers.

Figure 7. The body language of those present during the first feedback session shows little excitement or passion.

The students had never experienced studio work before and needed time to understand surface structures, pedagogical activities and epistemological beliefs. The HCID students' work was never publically criticized before, nor was it ever exhibited for others to see. In addition, they have little practice in speaking about their work. The experience with shifting from humbleness underway to pride in ownership of the work they did and the ideas that led to the final product, was also new.

During the course, critiques, especially in the beginning, were often taken personally, as the students were not familiar with this practice. However, making things, such as the wearable technology embedded skirts, or cushions for the "iConfess" booth, slowly placed smiles on students' faces, see Fig. 8. The act of making helped students to start unfolding some of internal processes leading towards increased creativity and willingness to learn new skills, or use the existing ones, in order to further the processes that the group was engaged in. We have reported on the emergence of creativity in this context, see [43], and used assemblages of skills as a framework for analysis.

Figure 8. The process of making things for the exhibit.

C. *Final projects*

After the first six weeks, there was a breakthrough. The first assignment was solved and organized as exhibit consisting of three parts.

Figure 9. The students working on the exhibit site, the exhibits representative of research on sustainable design, privacy and wearables.

The first part built on the idea of sustainability, producing the energy while biking, to power blinkers built into the glove, as well as to heat gloves, light up the wheels, etc. The presentations now included sketches, material choices, and hand knitted gloves; see Fig. 10.

Figure 10. Sketches of the prototype on the left, palette of material choices, as well as very handmade gloves, connected to a small dynamo.

The second part was related to design group's research projects regarding design for people with dementia. The skirt for dement ladies was the result, as as a counter-balance a "blinky" party skirt that lights up when the proximity sensor is activated; see Fig. 11 and [44][45].

Figure 11. On the left, the skirt with a proximity sensor. Remaining images are of the skirt for dementia: comfort balls, the GPS and a QR code.

The third part was an iConfession booth, a tool for exploring the anonymity and willingness of people to disclose a secret; see Fig. 9 and [46][47].

Finally, all three parts were put together. Fig. 12 and Fig. 13 show making of the exhibit and the final result.

Figure 12. Organizing the exhibit. The exhibit addresses sustainable design, anonymity and meaningful wearables.

Figure 13. Students recruited participants for the part of the exhibit wearing Guy Fowler masks, a symbol of anonymity, as part of the exhibit activity.

The ideas implemented, although inspired by one thing or another, were novel. The audience received the exhibit very well, and students have expereinced the sense of pride and satisfaction with the result.

We now reflect over techniques used to nudge the creativity during the period of 9 weeks and the exhibit design, see Fig. 14.

Photo documenting, as already mentioned, was used to document what was going on in the class. It then also became a practice that students learned and started implementing in their own projects. They learned to express through images processes that they participated in, objects they were working with as well as how the group distributed the work. Furthermore, it was easy to see, watching the pictures they took, through body postures and facial expressions how well they thrived while performing their tasks.

We have tried to encourage convergent and divergent thinking. It was not easy and natural for the HCI students to come up with many prototypes and many ideas, it was much easier to converge and pay attention to details. Fig. 14 does not give a detailed picture. It shows the predominant way of thinking for the period, they, of course, could not be separated so neatly.

In conjunction with divergent thinking, exploration and brainstorming were used. Body-storming was still to difficult to really engage in, but the students were encouraged to use it alongside other techniques to generate new ideas.

Design thinking was very helpful when explaining and making students understand concepts of empathy and rapid prototyping. To a certain extent, abductive thinking was also accepted and the students learn to keep the best parts of diverse solutions and combine them into a single one. This was most evident with iConfess booth, which has gone through several prototypes and where abductive thinking proved itself to be useful. As the start, the iConfess booth was envisioned as a relaxation room, transitioned through a match-making box and a kissing-booth, with the best ideas from each of these concepts still visible in the final solution.

Inspiration from reading about similar work, or local exhibits, was strongly encouraged. Some students have taken this part in, while others did not.

We have tried to push them outside of their comfort zone, and here, we believe to have been successful. We can also state that this experience was mutual.

D. *Assemblages of skills and practices*

The last part, assemblages of skills and assemblages of practices was the most interesting part. We utilized the students' other existing skills (i.e., skills learned outside the university campus, such as knitting and sewing) and made them play a crucial role for the very unfolding of creativity during the realization of their design ideas. In his book *Making*, [48], the anthropologist Ingold observes that such creative practices of making highlight what is often obscured in much of discussions in visual culture and in material culture. In the study of material culture, the main focus is on finished objects. In the study of visual culture, the focus has been on relations between objects, images and interpretations. What is lost, in both cases, is the creativity of the production processes that brings the artefacts themselves into being, through activities such as knitting and sewing: on one hand, the generative currents of the materials that the skirt or the knitted gloves are made, on the other the sensory awareness of those who make them. In this process the conduct of goes along with, and continually answers to the fluxes and flows of the materials with which the students work.

This coming-together of skills that different people have, as well as types of skills they have, we refer to as *'assemblage of skills'* in design efforts. Another coming-together was facilitated and nurtured by the teaching staff. That is the *'assemblage of practices,'* which entailed introducing the students to design practice, design thinking, makers' practices, and reflective practice. Thus, through 'assemblage of skills' and 'assemblage of design practices' the students needed both experiences of understanding (e.g., understanding new practices or the research interests of the Design-group), and 'making' experiences (e.g., producing both presentable and conceptually good physical representations).

E. Students' impressions

The feedback on how it all worked for students was collected using a questionnaire. We were interested in collecting feedback on emergence of both individual and group creativity, whether students felt better prepared for multidisciplinary work and other relevant experiences.

In order to get feedback directly from students around their perception of readiness to partake in multidisciplinary teams, we asked a direct question about their understanding of the difference between the usual HCI classes and this one. One of the students said: *The course focused more on design thinking, and to get a finished product to exhibit. It was less literature, testing, and report writing than other HCI courses.* All students expressed that they have better understanding of the design practice, and perceive this as something that prepares them better for future work together. In a short survey after the course was completed, we asked them if they thought that articles were helpful; half of the students answered that they thought so, while the other half thought that for this course the articles were not so important. When asked if they thought that design oriented practices (making things), have given them new skills, as HCID designers, six students agreed strongly, one did not reply and three were neutral. As for the perception of how well they are prepared after the course for work with designers, six students answered that they strongly feel that they are better positioned for such cooperation, and four did not have strong opinion about the issue, but were not negative.

Our understanding is that the experience they gained may help them discover who they are as HCID designers [49] [50], both by understanding the difference between the HCI practice and the design practice, and by direct experience of the design practice.

V. Conclusion and Future Work

The lessons learned and discussed in this paper show that HCID students could adopt and understand design practices, in spite of a rather long experience and a strong sense of being rooted in the HCI tradition. The teachers, the students, and the audience at the exhibit have all been satisfied with the exhibit in terms of adequateness of concepts in relation to design task, prototypes developed and organization of those into an exhibit. The students' body language has changed from indifferent and closed to engaged and open. They all perceive this piece of learning to prepare them for the work as professional interaction designers better than the HCI courses alone could do. Thus, we conclude that this approach warrants further exploration. As future work, we would like to follow these students into their work life and see if this experience had an impact when working in real multidisciplinary teams with designers, on real-life projects.

Acknowledgment

Thanks go to all of our students: Andrea Gasparini, Henrik Kjersem, Lena Risvik, Paria Tahaee, Anja Simonsen, Sylvia Saxlund, Rita Johnsen, Ingrid Arnesen, Agnethe Heggelund and Rebekka Castro, and our guest lecturers Tom Igoe, Amanda Steggell and Lavrans Løvlie.

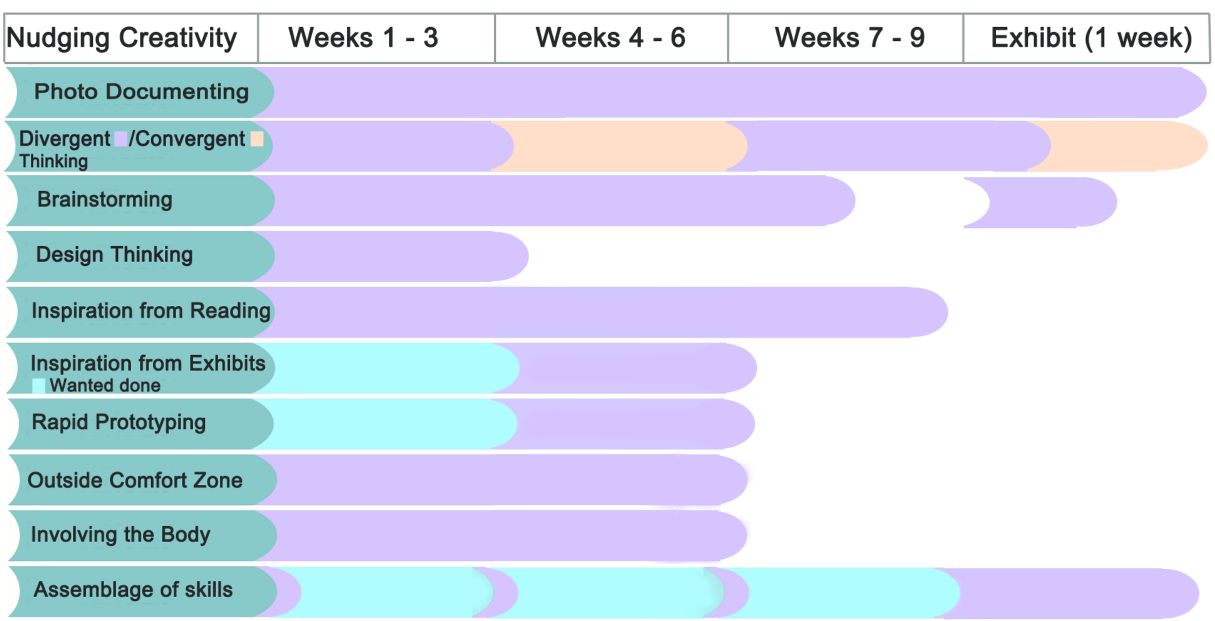


Figure 14. Figure explaining the main techniques used during the teaching to nudge students creativity.

REFERENCES

[1] A. L. Culén, H. N. Mainsah, and S. Finken, "Design Practice in Human Computer Interaction Design Education," The Seventh International Conference on Advances in Computer-Human Interactions, 2014, pp. 300–306.

[2] E. Blevis, "Design Challenge Based Learning (DCBL) and Sustainable Pedagogical Practice," Interactions, vol. 17, no. 3, pp. 64–69, May 2010.

[3] C. B. Brandt, K. Cennamo, S. Douglas, M. Vernon, M. McGrath, and Y. Reimer, "A theoretical framework for the studio as a learning environment," Int. J. Technol. Des. Educ., vol. 23, no. 2, pp. 329–348, May 2013.

[4] J. Kolko, Thoughts on Interaction Design, 1st edition. Morgan Kaufmann, 2010.

[5] S. E. Lundgren, E. Hallnäs, L. Ljungstrand, and P. Torgersson, "Teaching Interaction Design: Matters, Materials and Means", WonderGround-2006 Design Research Society International Conference, Lisbon, 2006, pp. 1–4.

[6] A. Faiola, "The design enterprise: Rethinking the HCI education paradigm," Design Issues, vol. 23, no. 3, pp. 30–45, 2007.

[7] T. Winograd and S. Klemmer, "HCI at Stanford University," Interactions, vol. 12, no. 5, pp. 30–31, September 2005.

[8] L. Bannon, "Reimagining HCI: Toward a More Human-centered Perspective," Interactions, vol. 18, no. 4, pp. 50–57, July 2011.

[9] D. Fallman, "Design-oriented human-computer interaction," in Proceedings of the SIGCHI conference on Human factors in computing systems, 2003, pp. 225–232.

[10] N. True, J. Peeters, and D. Fallman, "Confabulation in the Time of Transdisciplinarity: Reflection on HCI Education and a Call for Conversation," in Human-Computer Interaction. Human-Centred Design Approaches, Methods, Tools, and Environments, M. Kurosu, Ed. Springer Berlin Heidelberg, 2013, pp. 128–136.

[11] J. Forlizzi, J. Zimmerman, and S. Evenson, "Crafting a place for interaction design research in HCI," Design Issues, vol. 24, no. 3, pp. 19–29, 2008.

[12] J. Zimmerman, E. Stolterman, and J. Forlizzi, "An Analysis and Critique of Research Through Design: Towards a Formalization of a Research Approach," Proceedings of the 8th ACM Conference on Designing Interactive Systems, 2010, pp. 310–319.

[13] J. Zimmerman and J. Forlizzi, "Research Through Design in HCI," Ways of Knowing in HCI, J. S. Olson and W. A. Kellogg, Eds. Springer New York, 2014, pp. 167–189.

[14] E. Goodman, E. Stolterman, and R. Wakkary, "Understanding interaction design practices," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 1061–1070.

[15] P. Wright, M. Blythe, and J. McCarthy, "User Experience and the Idea of Design in HCI," Interactive Systems. Design, Specification, and Verification, S. W. Gilroy and M. D. Harrison, Eds. Springer Berlin Heidelberg, 2006, pp. 1–14.

[16] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, "Engineering design thinking, teaching, and learning," Journal of Engineering Education, vol. 94, no. 1, pp. 103–120, 2005.

[17] P. Koutsabasis and S. Vosinakis, "Rethinking HCI Education for Design: Problem-Based Learning and Virtual Worlds at an HCI Design Studio," International Journal of Human-Computer Interaction, vol. 28, no. 8, pp. 485–499, 2012.

[18] J. M. Carroll, HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science. Morgan Kaufmann, 2003.

[19] K. Cennamo, S. A. Douglas, M. Vernon, C. Brandt, B. Scott, Y. Reimer, and M. McGrath, "Promoting creativity in the computer science design studio," in Proceedings of the 42nd ACM technical symposium on Computer science education, New York, NY, USA, 2011, pp. 649–654.

[20] E. F. Churchill, A. Bowser, and J. Preece, "Teaching and Learning Human-computer Interaction: Past, Present, and Future," Interactions, vol. 20, no. 2, pp. 44–53, Mar. 2013.

[21] C. Owen, "Design thinking: Notes on its nature and use," Design Research Quarterly, vol. 2, no. 1, pp. 16–27, 2007.

[22] M. Csikszentmih, Creativity: Flow and the Psychology of Discovery and Invention. London: Harper & Row, 1997.

[23] A.-G. Tan, "Psychology of Cultivating Creativity in Teaching and Learning," Creativity, Talent and Excellence, A.-G. Tan, Ed. Springer Singapore, 2013, pp. 27–42.

[24] R. K. Sawyer and R. K. Sawyer, Explaining Creativity: The Science of Human Innovation. Oxford University Press, 2012.

[25] S. Harrison, D. Tatar, and P. Sengers, "The three paradigms of HCI," Alt. Chi Session at the SIGCHI Conference on Human Factors in Computing Systems San Jose, California, USA, 2007, pp. 1–18.

[26] S. Bødker, "When second wave HCI meets third wave challenges," Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, New York, NY, USA, 2006, pp. 1–8.

[27] C. Hoadley and C. Cox, "What is design knowledge and how do we teach it?" Educating learning technology designers: Guiding and inspiring creators of innovative educational tools, pp. 19–35, 2009.

[28] D. A. Schön, The reflective practitioner: How professionals think in action, Basic books, 1983.

[29] J. V. Wertsch, Mind as action. Oxford University Press, 1998.

[30] N. Schadewitz and T. Zamenopoulos, "Towards an online design studio: a study of social networking in design distance learning," Iternational Association of Societies of Design Research (IASDR) Conference, 2009, pp. 18–22.

[31] D. W. Shaffer, "Learning in design," in Foundations for the future: The need for new mathematical understandings & abilities in the 21st century. Hillsdale, NJ: Lawrence Erlbaum Associates, 2007.

[32] J. Lave and E. Wenger, Situated learning: Legitimate peripheral participation. Cambridge University Press, 1991.

[33] H. Sharp, Y. Rogers, and J. Preece, Interaction Design: Beyond Human-Computer Interaction, 2nd ed. Wiley, 2007.

[34] Y. J. Reimer and S. A. Douglas, "Teaching HCI design with the studio approach," Computer science education, vol. 13, no. 3, pp. 191–205, 2003.

[35] D. J. Lazar, D. J. H. Feng, and D. H. Hochheiser, Research Methods in Human-Computer Interaction. John Wiley & Sons, 2010.

[36] T. Lewis, "Coming to terms with engineering design as content," Journal of Technology Education, vol. 16, no. 2, pp. 37–54, 2005.

[37] D. Cropley and A. Cropley, Engineering Creativity: A Systems Concept of Functional Creativity. Lawrence Erlbaum Associates Publishers, 2005.

[38] A. M. Hill, "Problem solving in real-life contexts: An alternative for design in technology education," International journal of technology and design education, vol. 8, no. 3, pp. 203–220, 1998.

[39] K. Smith, How to be an explorer of the world: portable life museum. New York: Perigee, 2008.

[40] "Energy Bank," 2012. [Online]. Available from: http://www.energy-bank.org/. [Accessed: 21-Jan-2014].

[41] M. Crang and I. Cook, Doing ethnographies. Los Angeles; London: SAGE, 2007.

[42] C. Lévi-Strauss, La Pensee Sauvage, Paris: Pocket, 1990.

[43] S. Finken, A. L. Culen, and A. A. Gasparini, "Nurturing Creativity: Assemblages in HCI Design Practices," Proceedings of DRS 2014, Umeå, 2014, pp. 1204–1217.

[44] R. Johnsen, S. Saxlund, and A. Simonsen, "Skirts with Meaning," skirtswithmeaning, 2013. [Online]. Available from: http://skirtswithmeaning.wordpress.com/. [Accessed: 30-Oct-2013].

[45] A. L. Culén and S. Finken, "A Skirt for Well Aged Ladies with Cognitive Loss," in UAHCI/HCII 2014, 2014, pp. 325–336.

[46] A. Gasparini, R. Castro, L. Risvik, and A. Heggelund, "iConfess at dagen@ifi," 2013. [Online]. Available from: http://folk.uio.no/andrega/iConfess/. [Accessed: 31-Oct-2013].

[47] A. L. Culén, S. Finken, and A. Gasparini, "iCONFESS," in Social Computing and Social Media, G. Meiselwitz, Ed., Springer International Publishing, 2014, pp. 430–439.

[48] T. Ingold, Making: Anthropology, Archaeology, Art and Architecture by Ingold, Tim. Routledge, 100AD.

[49] A. L. Culén, S. Joshi, and A. Atif, "HCID: Who is an interaction designer?" Proceedings of the 2nd International Conference for Design Education Researchers, Oslo, Norway, 2013, vol. 4, pp. 1924–1937.

[50] A. Culén and M. Kriger, "HCI in IT-facilitated Business Innovation: a Design Thinking Perspective," HCI in Business. F. F. H. Nah, Ed., Springer, 2014, pp. 492–503.

An Interactive Web Tool to Facilitate Shared Decision Making in Dementia: Design Issues Perceived by Caregivers and Patients

Marijke Span, Carolien Smits,
Leontine Groen-van de Ven, Jan Jukema
Research group Innovation of Care of Older Adults and Social Work,
Windesheim University of Applied Sciences
Zwolle, The Netherlands
m.span@windesheim.nl, chm.smits@windesheim.nl,
lm.groen-vande.ven@windesheim.nl,
js.jukema@windesheim.nl

Marike Hettinga, Ruud Janssen
Research group IT Innovation in Health Care,
Windesheim University of Applied Sciences
Zwolle, The Netherlands
m.hettinga@windesheim.nl, tjwm.janssen@windesheim.nl

Myrra Vernooij-Dassen
Department of Primary Care, IQ HealthCare,
Radboud University Nijmegen;
Kalorama Foundation, Radboud Alzheimer Centre
Nijmegen, The Netherlands
myrra.vernooij-dassen@radboudumc.nl

Jan Eefsting
Department of Nursing Home Medicine and
EMGO Institute for Health and Care Research,
VU University Medical Centre
Amsterdam, The Netherlands
j.eefsting@zgijv.nl

***Abstract*—The aim of this study is to identify design issues (both weaknesses and strengths) that should be considered for designing a user-friendly design of an interactive web tool that facilitates shares decision making in care networks of people with dementia. Our research questions are: 1) What design issues can be identified? and 2) What is the unique contribution of people with dementia to the design? This study, with its iterative participatory design based on the CeHRes roadmap included six separate focus group sessions with people with dementia, informal caregivers, and case managers; a cognitive walkthrough with researchers; and usability tests with case managers, older adults, informal caregivers, and people with dementia. The design issues were: a screen design based on pleasant and harmonious colors, the use of clear and uniform buttons throughout the interface, the use of multiple-choice questions with smileys as answering options, a foldable menu bar that is closed (for people with dementia) or open (for caregivers) by default, and the incorporation of a chat function that specifically keeps all end users involved in a conversation. The specific and detailed contribution of the participants with dementia dealt with their focus on the present, accuracy of language, and the graphical layout. However, other participants doubted whether the tool would be useful and usable for people with dementia. Designing a user-friendly, interactive web tool for people with different capacities, interests, and perspectives is challenging. People with dementia contributed uniquely to the design of the web tool. A pilot study will show whether the doubts of some of the participants are valid.**

Keywords- dementia, shared decision making, participatory design, assistive technology, interactive web tool.

I. INTRODUCTION

Designing an interactive web tool to facilitate shared decision making (SDM) in care networks of people with dementia is complex because of the progressive character of dementia and the multitude of people involved in the decision making process namely, the person with dementia, informal caregivers, and case managers [1][2]. In a prior study we identified user requirements for such a tool: decision-making topics that the tool should address and additional needs and preferences of all concerned [3][4]. In the present study, we focus on the design of such an interactive web tool – a design that aims to do justice to people with dementia in decision making.

Dementia is characterized by progressive cognitive decline. Decreasing abilities address memory loss, route planning, behavior change, and orientation problems. Dementia is a disease that is affecting increasingly more people worldwide: the prediction is from 66 million in 2030 to 115 million in 2050 [5][6]. The life expectancy of people with dementia after diagnosis is 6 to 8 years. During this period, people with dementia and their loved ones are faced with many decisions related to care and well-being, e.g., can or should the patient drive a car, walk alone outdoors, start day care, or be admitted to a nursing home [7][8]. These decisions, often overloaded with emotions, change the situation of people with dementia continuously. Dementia is not a linear process, and the change in capacities of people with dementia influences their position in decision making [9][10].

The decision making of many people challenges the traditional view of shared decision-making. Shared decision making (SDM) has its origin in the clinical encounter between the clinician and the mentally able patient who decide about one medical question. The decision is mainly based on analytical thinking [11]. Shared decision making is an approach that involves patients in making medical decisions to the extent that they want to and that emphasizes the collaboration between professionals and patients [11][12]. Shared decision making increases patient autonomy and empowers the patient [13]. This could also benefit people with dementia who are capable of expressing their needs [13] and preferences [14]. Unfortunately, their participation in decision-making processes [13][15], research [16] and IT development trajectories [17] is not common.

In designing an interactive web tool to facilitate SDM in care networks of patients with dementia, we have to take into account the two-fold complexity just described, i.e., the progressive character of dementia and the multitude of people involved, all of whom have different capacities and interests in decision making. Because of this, we included all the groups in the designing to create a shared perspective; not only the groups of informal caregivers and case managers, but also the group of people with dementia [18]. The last group is the most vulnerable group in this context. The views of people with dementia are in danger of being overlooked. Caregivers tend to shield them or speak for them rather than with them [19]. Spending time with people with dementia and confirming that their contribution is worthwhile helps to include them in research [16]. This is necessary to better understand their needs and preferences [13].

An additional reason for involving them in the design is the progressive character of dementia that sets them apart from average web users. They have problems using a "one size fits all" computer design [20]. Several researchers have provided evidence-based design criteria for designing for people with dementia: creating easy orientation [21][22]; using familiar cues; making everything legible and distinctive [23]; choosing touch screens, large-format screens, and large font sizes; keeping text to a minimum; assuring a hypermedia structure with limited options for selection and an attractive design [24]; and using tablets [25]. These criteria ensure a dementia-friendly design of IT tools in general.

Thus, the current body of knowledge includes knowledge of design criteria that the final product must meet. It does not, however, provide knowledge of how to design an interactive SDM web tool or what the challenges will be. This design process has to deal with the two-fold complexity as well. Therefore, we aim to identify design issues (strengths and weaknesses) that have to be taken into account in a design for a user-friendly interactive web tool for SDM in care networks for people with dementia. The research questions read:

1) What design issues can be identified for a user-friendly interactive web tool that helps people with dementia with shared decision making?

2) What is the unique contribution of people with dementia to the design?

II. METHODS

In our study with its iterative, participatory design, we consider the involvement of all types of end users, but particularly people with dementia [17], as one of the key factors for developing a user-friendly and usable interactive web tool. We used the Center for eHealth Research and Disease Management (CeHRes) roadmap for the tool because this approach connects a human centered design with eHealth business modeling and emphasizes the importance of involving all those concerned to develop sustainable innovations [26]. The roadmap helps developers structurally integrate interactive web tools in health care and involves the participants in all phases of the development. The CeHRes roadmap offers a holistic framework consisting of five phases:

1) Contextual inquiry: gathering information from the environment where the technology will be implemented

2) Value specification: defining requirements based on participants' values

3) Design: translating these values and requirements into technical specifications and requirements for communicative and lucid prototypes in order to enable the participants to give feedback

4) Operationalization: implementing the technology in practice

5) Summative evaluation: determining the effects of the technology on behavior, health and organization.

This paper describes how the third phase of the CeHRes roadmap was applied to the design of an interactive web tool facilitating SDM in care networks of people with dementia: the *DecideGuide*. We used focus group sessions, a cognitive walkthrough, and usability tests to address both research questions.

A. *DecideGuide*

The DecideGuide is an interactive tool for people with dementia, informal caregivers and case managers to communicate with each other in making shared decisions. The design principles of the DecideGuide are transparency, open communication and information, and giving voice to people with dementia. The DecideGuide incorporates three perspectives: those of people with dementia, informal caregivers, and case managers. The case manager deliberates with the person with dementia and the informal caregivers whether to use the DecideGuide. All participants have an individual login and use the tool on their own or after they are alerted by the case manager. The DecideGuide has three pillars. The first pillar, *Messages* is a chat function that enables users (at a distance) to communicate with each other. The second pillar, *Deciding together* supports decision making step by step. The third pillar, *Individual opinion* enables users to give their individual opinions about dementia-related topics and their circumstances. It particularly supports giving voice to the person with dementia (Fig. 1). The DecideGuide, a safe and shielded web tool, is available for tablets, laptops, and computers.

The DecideGuide was developed in a process consisting of four iterations (Fig. 2). Fig. 3 illustrates the iterative development with the changes made in the chat function. These included using buttons (forward and back) instead of arrows, simplifying the menu bar, changing the position of menu bar, using colors, adding notifications, and providing an option to fold or expand the menu bar. All these changes were made in accordance with the feedback from users.

B. *Focus group sessions with mock-ups (first iteration)*

Firstly, separate focus group sessions were organized with intended end users, including people with dementia,

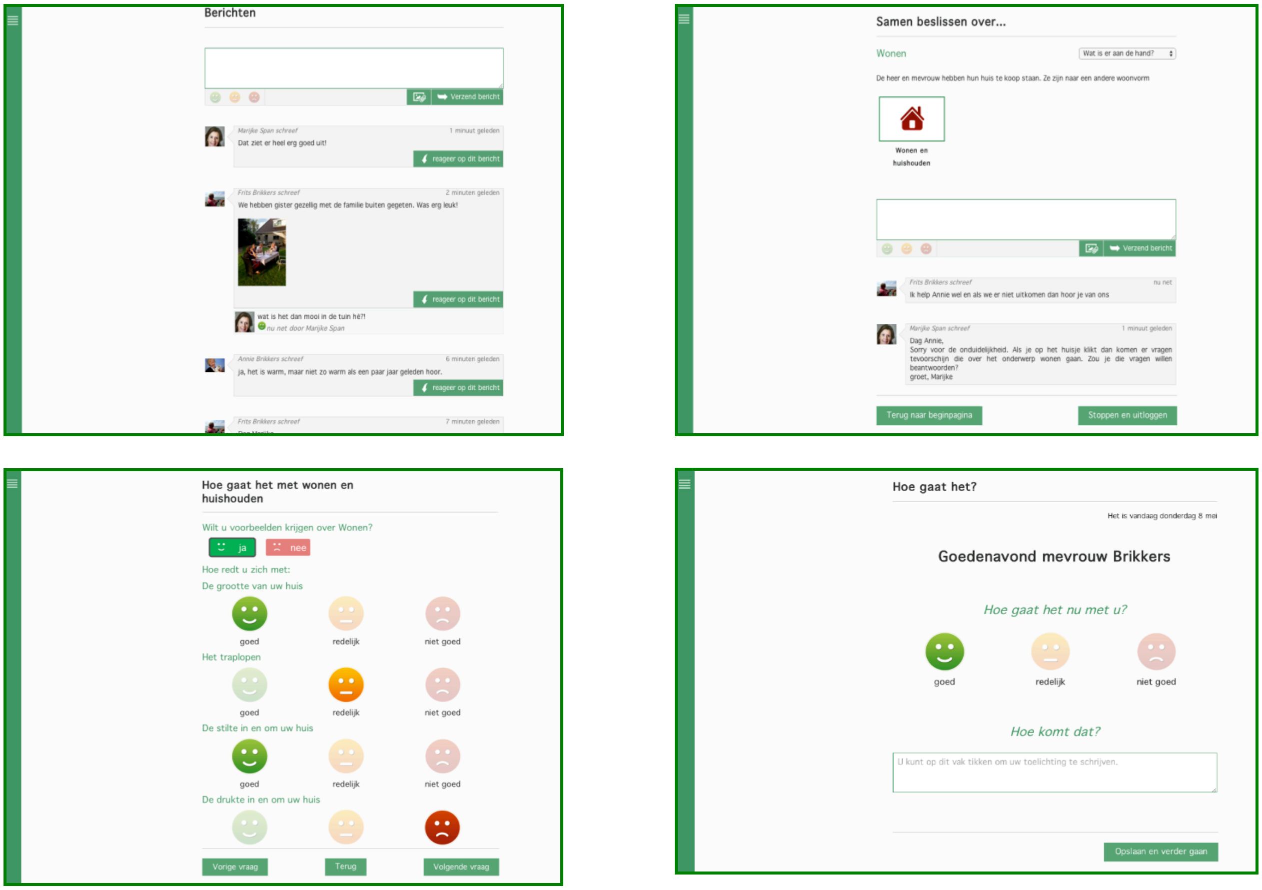


Figure 1. Final layout of three pillars of the DecideGuide (screen view for the person with dementia). Clockwise starting top left: chat function (belonging to the first pillar), deciding together (second pillar), individual opinion "How are you right now?" and individual opinion in questionnaire with examples (both belonging to the third pillar).

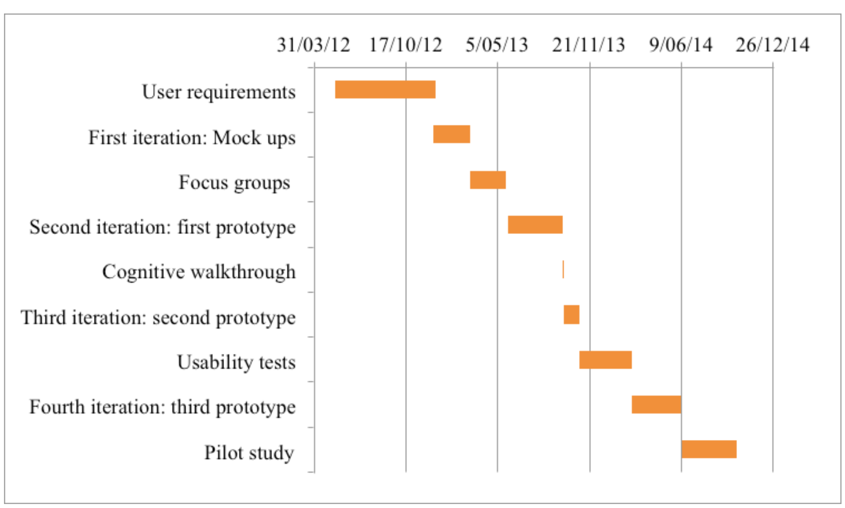


Figure 2. Timeline of designing the DecideGuide.

informal caregivers, and case managers. The goal of these focus group sessions was to receive comments and feedback about the first mock-ups, that is sketches on paper of the user interface of the DecideGuide.

People with mild to moderate dementia were recruited from two daycare centers. Informal caregivers were recruited from residential homes and the Dutch Alzheimer Association. Case managers were recruited from regional case managers' networks. All participants gave their written informed consent. We paid special attention to the informed consent of people with dementia because of their vulnerability. To be sure of their voluntary participation we invested in an ongoing consent [16]. We invested in time for social talk and getting to know each other, checked the consent of people with dementia during their participation after their initial consent, and emphasized the importance of their participation. The investigators watched over any signs, non-verbal or otherwise, of discomfort or restlessness of people with dementia. In such a case, the participant was given ample opportunity to quit.

Twenty-seven end users participated in the six focus group sessions (Table I). The two separate focus group sessions with people with dementia and informal caregivers took place in two rounds. They consisted of six and four participants, respectively, with different participants in each round. The participants of the second-round focus groups commented both on the mock-ups and the feedback of the first round of focus groups. Two focus group sessions with case managers took place with the same group of seven participants. The principal researcher, assisted by another researcher or designer, moderated the focus groups.

The mock-ups of the DecideGuide included 11 sketches on paper. The mock-ups, based on user requirements [3][4], were presented in the focus groups [26]. End users were asked to comment on the different sketches in general (What is your first impression?) content (Do you think what this tool offers is relevant? or Could this tool be helpful in making decisions?), user-friendliness (What do you think of the usability of the tool?) and the attractiveness of the design (What do you think of the look and feel of the tool?).

The focus group sessions lasted 1 to 2 h and were audio taped and transcribed verbatim, except for one focus group with people with dementia in round 1. Although the participants agreed to recording the interview before the meeting, they refused when the meeting was about to begin. They felt they could speak more freely without the session being recorded. Field notes of this interview were taken.

The five steps of framework analysis were used to analyze the focus group interview transcriptions: becoming familiar with the data, identifying a thematic framework (identified themes showed similarities to an existing framework), indexing (coding with thematic framework), charting (rearranging thematic framework if necessary), and mapping and interpretation (explanation of the findings) [27]. We used the CeHRes assessment of design quality which was adapted from the CeHRes roadmap. It consisted of three levels for assessing the quality of design: system quality (user-friendly, safe technology), content quality (understandable and meaningful content) and service quality (adequately provided service) [28]. We used the CeHRes assessment of design quality because it addresses various aspects of eHealth design quality that are based on several studies.

C. *Cognitive walkthrough (second iteration)*

Secondly, the first interactive prototype was built on the basis of the feedback for the mock-ups: the DecideGuide. This prototype was tested in a cognitive walkthrough session with the researchers to identify possible user problems and evaluate the usability of the DecideGuide [18][28][29].

Three of the research team tested the DecideGuide on a tablet in a 2-h role-playing session in a usability lab. First the participants tried the tool on their own, without any instruction. Then they performed tasks in a think-aloud session [29]. The session was video and audiotaped, and field notes were taken. The principal researcher and developer supervised the session. The analysis focused on the three levels of the CeHRes assessment of design quality: system, content, and service quality.

D. *Usability tests (third iteration)*

Thirdly, on the basis of the results of the focus group interviews and the cognitive walkthrough, we tested the usability with the adapted interactive prototype of the DecideGuide on a tablet [28][29]. The goal of these usability tests was to further refine the DecideGuide into a prototype that was robust enough to be used in a pilot study [29]. Three to five usability tests are required to identify most of the bugs [30][31]. The total number of usability tests was 12: three for the case managers, three for the informal caregivers, and three extra usability tests with older adults before the three usability tests with people with dementia were done. The participants were asked to perform tasks (e.g., log in with their user names and passwords, send a message, respond to a message, and fill in a questionnaire). The think-aloud method was used to identify their thoughts and feelings while they used the prototype [32]. The usability tests were video and audiotaped, and field notes were taken. The principal researcher moderated the usability tests with the assistance of another researcher.

Firstly, three case managers who participated in the development tested the DecideGuide together in a session that lasted 2 h. They tried out the DecideGuide on their own and then tried some tasks. Secondly, after adjusting the DecideGuide on the basis of the usability tests with case managers, three older adults recruited from a sounding board of the department *Care innovation for older adults* at a university of applied sciences tested the DecideGuide individually. They tested the DecideGuide with the login of people with dementia in a 1-h session in a usability lab. We chose this approach to find out whether the DecideGuide was "bug free", user friendly, and suitable for people with dementia to test. Thirdly, three informal caregivers recruited by participating case managers tested the DecideGuide. These individual sessions also took place in a usability lab and lasted about 1 h. Fourthly, three community-dwelling people with dementia (Reisberg score: 2-4) recruited by participating case managers tested the DecideGuide on a

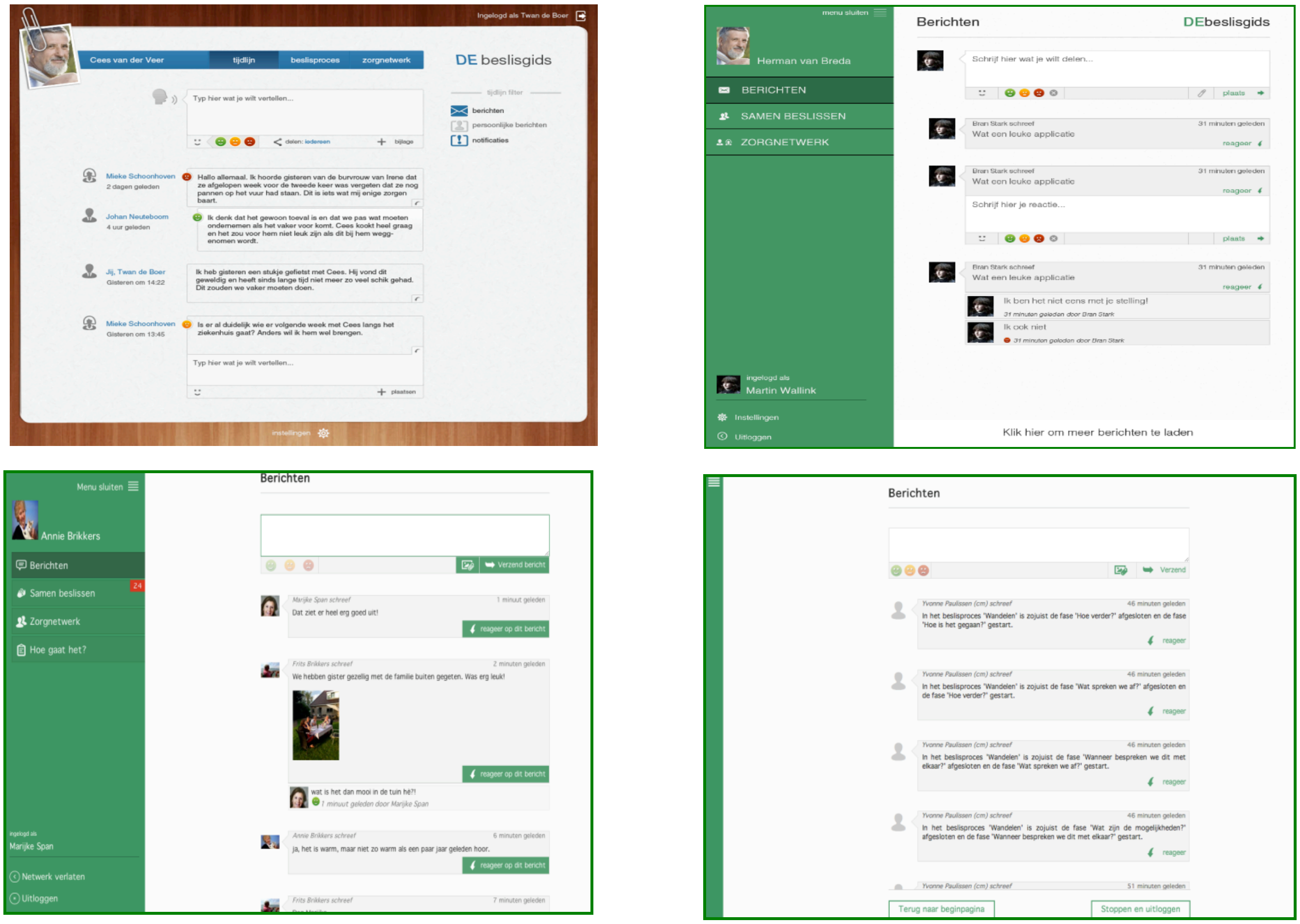


Figure 3. Iterative development of the chat function of the DecideGuide.
Clockwise starting top left: mock-ups, 1st prototype, 2nd prototype, 3rd prototype.

TABLE I. CHARACTERISTICS OF FOCUS GROUP SESSIONS

Focus group session	***Round 1***	***Round 2***
PWD[a)]	6 people commented on 11 sketches (screen views of PWD).	
PWD		6 people commented on 6 sketches and gave feedback about the first focus group
ICs	4 people commented on 11 sketches (screen views of ICs)	
ICs		4 people commented on 11 sketches and gave feedback about the first focus group
CMs (same people in both rounds)	7 people commented on 4 sketches (screen views of CMs)	7 people commented on 15 sketches (screen views of CMs, PWD, and ICs)

[a)]PWD= people with dementia, CMs= case managers, ICs= informal caregivers

tablet at home. These sessions lasted between 30 and 60 min. Based on the feedback from those who participated in the usability tests, a fourth and final prototype was developed that will be evaluated in a pilot study. Analysis focused on the three levels of the CeHRes assessment of the design quality of system, content, and service. The principal researcher analyzed the transcripts to identify design issues that could be added to the CeHRes assessment. The researcher who assisted with the usability tests verified the analysis.

III. RESULTS

All 39 respondents who agreed to participate completed their participation (Table II). There were no dropouts. The respondents were motivated and enjoyed their participation.

A. Research question 1: design issues

1) Focus group sessions

The focus group sessions resulted in feedback that addressed mainly the system quality, particularly the user-friendliness (e.g., too many screens, with too much information per screen) and design style (e.g., presentation of information and use of icons; Table III). Regarding the quality of the content, unclearness, and difficulty of the terms were reported, particularly in the "deciding together" part. Textual refinements were made: wording of decision-making themes and questions, as well as spelling and accuracy of the date. Only case managers commented on the quality of service. They described the tool as very useful in facilitating SDM in in care networks of people with dementia, but they doubted whether the tool would be useful for people with dementia. They thought that using the tool was too difficult for them to be able to participate adequately.

2) Cognitive walkthrough

The cognitive walkthrough with three researchers resulted in feedback mainly about system quality. Many bugs were found (buttons that did not react or were missing and partly black screens). Moreover, navigation issues and user-friendliness issues came up. Feedback about the content quality focused on the presentation of content (unclear icons and questions). The researchers emphasized the usefulness of the tool for informal caregivers and case managers but they doubted whether the tool would be useful for people with dementia. Moreover, the researchers doubted whether people with dementia would benefit from the open communication in the chat function. They felt that communication in the chat mode was too intrusive for people with dementia. The cognitive walkthrough focused mainly on the interaction between users.

3) Usability tests

Design issues arising from the usability tests were mainly about system quality: user-friendliness (e.g., operating a touch screen, unclear interface, navigation within a screen and between screens, purpose of buttons, and software bugs) and design style (e.g. smileys too small, questionnaires too long, and way of addressing people). The design issues concerning the content quality were: accuracy (e.g., unclear use of language, missing words), relevance (e.g., the content of the tool was too difficult for people with dementia and phases of decision making were irrelevant in some user interface elements), and comprehensibility (e.g., questionnaires were too difficult and confrontational for people with dementia). The design issues identified in service quality addressed mainly the usefulness of the tool. The only way the tool can be useful for people with dementia in the informal caregivers' and case managers' view, is by starting to use the tool early in dementia.

The strengths were the possibility of future extensions ("nice to haves" e.g., a skype function, a personal calendar, and alerts), the monitoring of informal caregivers' well-being, colored smileys as answer options, the use of the green for the main interface color, and the examples that are given (Table III).

4) Differences in participants' opinions about design issues

The participants agreed about many design issues (Table III). Nevertheless, there were some differences of opinion. Firstly, in the focus group sessions, informal caregivers and case managers were of the opinion that the tool should consist of fewer screens and fewer examples with smileys. The people with dementia agreed about the screens but not about the examples – they liked them and wanted even more examples to choose from. We reduced the number of the screens and expanded the examples (that is, we added more examples per theme and also examples for possible solutions).

Secondly, in the focus group sessions, some informal caregivers and people with dementia called into question the use of smileys. They were afraid the smileys would be too childish. Others liked them; they found them as easy, clear, and appealing. We decided to continue the use of smileys because of their simplicity and clarity. Moreover, we wanted to design the screens as similarly as possible for both informal caregivers and people with dementia in order to rule out causing possible feelings of inferiority among people with dementia when designing two different views. The usability tests later showed that all participants indeed liked the use of examples and smileys.

Thirdly, from the beginning, informal caregivers and case managers said that they needed a separate communication channel without the person with dementia where they could speak freely about the situation. The people with dementia did not mention such a need. We decided not to create such a channel and adhere to the design rationale of the DecideGuide based on principles of SDM: open communication and transparency. The results of the upcoming pilot study will show whether or not this transparency can be maintained.

Fourthly, informal caregivers and case managers said that the DecideGuide was too difficult and the chat function too intrusive for people with dementia. They questioned transparency as design rationale for the tool. They were convinced that the design rationale was very nice and desirable, but also very confrontational for people with dementia. They said that transparency tends to decrease the distrust of people with dementia, but may increase their restlessness. In the role playing during the cognitive walkthrough, researchers had similar considerations. The researcher who played the role of the "person with dementia" disliked what others said via the chat function, took this as a personal attack, and even wanted to stop using the tool. The

TABLE II. CHARACTERISTICS OF PARTICIPANTS

Total number of participants of focus groups and usability tests (n = 39)			
Characteristics of focus group participants (n = 27)			
	People with dementia (n = 12)	***Informal caregivers (n = 8)***	***Case managers (n = 7)***
Gender	8 Male 4 Female	1 Male 7 Female	6 Female 1 Male
Age in years	69-85 (M = 80.0) 6 Unknown	53-83 (M = 67.6)	38-60 (M = 48.8)
Educational level	0 Low 4 Medium 2 High 6 Unknown	0 Low 2 Medium 6 High	6 High
Type of dementia	3 Alzheimer's disease 1 Front temporal dementia 2 Mild cognitive impairment/dementia 6 Unknown		
Reisberg scale	2-4		
Marital status	4 Married 2 Widowed 6 Unknown		
Relation to person with dementia		6 Spouse 1 Daughter 1 Friend	
Caregiving was experienced as		3 Heavy 4 Medium	
Experience as case manager			1 <1 year 3 1-5 years 2 >5 years

[a]Low = primary or secondary school graduate, medium = high school graduate, high = college graduate

Characteristics of usability test participants (n = 12)				
	Case managers (n = 3)	***Informal caregivers (n = 3)***	***Older adults (n = 3)***	***People with dementia (n = 3)***
Gender	3 Female	2 Female 1 Male	2 Female 1 Male	1 Female 2 Male
Age	42, 50, and 62 years	61, 65, and 74 years	62, 63, and 67 years	72, 79, and 82 years
Type of dementia				3 Alzheimer's disease
Reisberg scale				3–4
Educational level[a]	3 High	2 Medium 1 High	1 Medium 2 High	1 Medium 2 High
Electronic equipment *(computer, laptop, tablet, smartphone)*	3 Computer 1 Tablet 3 Smartphone	2 Computer 3 Laptop 2 Tablet	3 Computer 1 Tablet	2 Computer 1 Tablet 1 nothing
Experience with computers *(years)*	10–15 years	7–15 years	5–25 years	1–5 years
Software and networks used *(Word, Excel, Power Point, Email, Internet, Social media)*	3 Email 3 Internet 3 Word 3 Power Point 3 Excel 2 Social media	3 Email 2 Internet 2 Word	2 Email 2 Internet 3 Word 1 Excel 1 Power Point	2 Email 1 Internet 1 Word 1 Nothing
Assessment of one's own IT capacities *(excellent, good, moderate, or poor)*	2 Good 1 Excellent	2 Moderate 1 Good	3 Moderate	1 Moderate 2 Poor

[a] Low = primary or secondary school graduate, medium = high school graduate, high = college graduate

informal caregivers believed that the only chance for successfully using the tool was starting to use it in the early stages of dementia. The people with dementia did not comment on how other participants would view the tool, except for one remark. When the person filled in the questionnaire about one of the eight dementia related themes, "daily activities", he wondered how his answers could be relevant for the case manager. The pilot study will show whether the participants' concerns are justifiable.

5) CeHRes assessment of design quality

Data analysis showed that most of our design issues fit into the structure of the CeHRes assessment of design quality (Table IV). These design issues did not cover all the items of the CeHRes assessment. The items that did not seem applicable to our interactive web tool included efficient search functionality (A1e), and design persuasiveness (A3c,d) addressing system quality; evidence-based information (B2), language and ethnicity (B6), and disclosure (B7) addressing content quality; and responsiveness (C2), reliability (C5), and credibility (C6) addressing service quality. Other items were simply not mentioned by the participants: technical support (A1h), and safety and technical security (A2b-d). Moreover, some design items that were mentioned did not fit into the CeHRes framework: items lacking in the system such as "absence of extra answer option button" and benefits (usefulness) of using an interactive web tool such as "learning from the tool", and "pleasure in using the tool". We added them as 1j *Completeness* (system quality) and 7 *Perception*: 7a *Learning* and 7b *Pleasure* (service quality) respectively (Table IV).

B. Research question 2: the unique contribution from people with dementia to the design of the interactive web tool

All end users contributed to one or more items of the system, content, and service quality of the design (Tables III and IV). Informal caregivers contributed to most items; the case managers and people with dementia, to fewer items. The people with dementia were very well able to give their opinions in the focus group sessions and the usability tests. Their feedback given for the two focus group sessions addressed mainly the system and content quality (Table III). We honored their feedback in the design in as far as it did not conflict with the design rationale of the interactive web tool. The people with dementia liked participating: it gave them pleasure and they liked learning new things such as using an iPad and the DecideGuide. Without their feedback, we would have missed their focus on the present (i.e., the "here and now" of their perspective) and their preferences about careful use of language and a pleasant graphical layout. We discuss each of these unique contributions below.

1) Focus on the present

The feedback of the people with dementia addressed concrete items in the present. They did not reflect on possible future items, or on the web tool in a more abstract way, or on the perspective of other participants the way informal caregivers and case managers did. They just focused on the assignments researchers asked them to do and to comment on. This focus on the "here and now" is reflected in their specific feedback.

When answering the first question of the web tool, "How are you today?" in the focus group, someone commented: "Do you know how long today is? I cannot tell you how I feel today. Today has so many moments. I only can tell you how I feel right now. But that can be totally different in a few hours. So please ask me: How are you right now?" We changed the formulation of this question according to their proposal.

A fictive case was used in the sketches of the mock-ups for the focus group sessions because we thought that a fictive case would be less confrontational for people with dementia. This was a misjudgment: it was apparently difficult for people with dementia to answer the questions: "I cannot answer this question because I do not know this person. He is not familiar to me. I can only answer for myself". All other participants in the focus group sessions also had difficulties with the mock-ups. They found it difficult to imagine what the interactive web tool could look like.

The people with dementia were relaxed and very well able to accomplish the tasks in the usability tests and to give their opinions. Moreover, for the tasks of "send a message to the network members" and "respond to the message of the case manager" in the chat, they were the only participants who accomplished these tasks without assistance.

2) Careful use of language

The people with dementia gave very specific and detailed feedback about the tool's user-friendliness and presentation of content of the tool. Fine-tuning and accuracy of text were important to them. They gave feedback about the wrong date on the screen shots and proposed synonyms for some themes (e.g., "family and friends" rather than "social contacts"). They did not like the name of the theme "future": "future…. future?....there is no future….". We therefore changed this theme to "important now and later". Sometimes, the people with dementia prevented researchers from oversimplifying wording. When we were searching for a simple synonym for the theme "mobility" only the people with dementia did not agree. "Mobility" was fine and clear, but should be augmented with "transport". This resulted in the theme "mobility and transport".

The people with dementia were the only people who commented on form of addressing them in the tool. It was important to them that they were called by their first names rather than their surnames: "I am not a sir/madam…just call me by my first name. That's who I am".

They wanted to answer questions very precisely. Several times they discussed the three answering options with smileys: "good", "don't know or neutral", and "not so good". Although they liked the limited options available, the meaning of these options did not reflect their answers. They preferred an extra button with "moderate". Their reactions included: "I don't know what to answer; no smiley reflects my opinion well enough. I would like another button".

3) Pleasant graphical layout

The informal caregivers, case managers and older adults argued that the screens had to be as simple as possible for people with dementia. They liked the green menu bar at the left side of the screen rather than the white buttons on the homepage. However, in their opinion, the homepage without the menu bar was enough for people with dementia. They liked the green color of the design, but emphasized their wish for more contrast. Buttons should be distinctive enough. They liked green buttons with white letters rather

TABLE III. CEHRES ASSESSMENT OF DESIGN QUALITY: WEAKNESSES AND STRENGTHS

Mock-ups of the focus group sessions			
CeHRes assessment of design quality		***Weaknesses***	***Strengths***
System quality	User-friendliness	*Navigation structure and ease of use* Too much screens for people with dementia. CM and PWD Too much information on screens. IC) Screens are too lively for people with dementia. PWD and IC Letters too small (not all PWD agreed) Messages in timeline with chat become a big mess. CM	*Nice to haves* Adding things such as a personal calendar (personalized part in tool for case managers, linking with home technology, skype function). CM and IC Alerts for daily activities (taking medication, eating, etc). IC and CM
	Design persuasiveness	*Lens for design* All network members view all messages because the tool is based on transparency and open communication. This is not always advisable for people with dementia. CM Too many examples with too many colors. IC Use of smileys is a bit childlike. IC *Presentation of content* Use of smileys is clear but not really nice. PWD Attention to use of red in tool. Red smiley is similar to feeling unwell. Using red for a theme can also suggest danger. IC Use of colors in messages is not distinct enough. CM and IC	Monitoring well-being is important. IC Use of red, orange, and green for smileys is nice. IC
Content quality	Accuracy	Date and year are incorrect. PWD The terms are not specific enough: How are you today? PWD	
	Comprehensibility	Use of some terms is not clear enough and too difficult e.g., options and pros and cons of options. PWD	
	Relevance	Open questions are less attractive than questions that also offer examples. Tool is too directive. PWD	
Service quality	Perceived usefulness	The tool is very useful for facilitating SDM in care networks of people with dementia, but how useful will it be for people with dementia? CM	Tool is directive: easy to use because you do not have to invent answers by yourself. IC

Cognitive Walkthrough with researchers			
CeHRes assessment of design quality		***Weaknesses***	***Strengths***
System quality	User-friendliness	*Navigation structure/ease of use* Unclear login Use of smileys in chat function is unclear Adding a smiley is impossible Unclear whether message or smiley has been sent Unclear navigation structure Questionnaires are incomplete and unclear “Deciding together” is unclear Giving one’s individual opinion is unclear to PWD Too difficult for PWD Notification for new messages is lacking	*‘Nice to haves’* Notifications for new activities in tool Separate communication possibility for IC and CM?
	Design persuasiveness	*Lens for design* All network members view all messages because the tool is based on transparency and open communication. This is not always advisable for PWD *Presentation of content* Use of gray in “Deciding together” does not contrast enough Use of icons is unclear Content of questionnaires differs a lot	Monitoring well-being is important for IC Overall color (green): nice and restful Use of red, orange and green for smileys is nice
Content quality	Accuracy	Questionnaires are incomplete	
	Comprehensibility	Use of some terms is not clear enough and too difficult, e.g., options and pros and cons of options Use of decision phases is too difficult for PWD	
	Relevance	Use of decision phases is too difficult for PWD	
Service quality	Rerceived usefulness	Tool is very useful in facilitating SDM in care networks of PWD but how useful will it be for PWD?	

PWD = People with dementia; IC = informal caregiver; CM = case manager

TABLE III (CONTINUED)

Individual usability tests			
CeHRes assessment of design quality		***Weaknesses***	***Strengths***
System quality	User friendliness	*Navigation structure and ease of use* Call into question: tool user friendly for PWD? CM, IC, OA Operating the touch screen (tapping; scrolling; keyboard). PWD, IC, OA Log in difficulties for PWD and OA Interface is not clear enough (buttons too small; too close together, use of color not distinctive enough, font size too small, not enough answer options. PWD, IC, OA Meaning of some buttons and UI elements is unclear (text on buttons, length of questionnaires, automated messages, adding smileys to messages. PWD, IC, OA Software bugs (buttons do not react, black surface). PWD, IC, CM, OA Navigating within the screen (meaning of buttons; where to start? PWD, A Navigating between screens (partly absence of "back" button and "go on", going back in tool is unclear. PWD, IC, CM, OA Absence of extra answer option button A1j. PWD, IC	Combination of icons and buttons is nice Notifications for new messages
	Design persuasiveness	*Lens for design* Call into question: design tool suitable for PWD? (use of color [less white, more green; more contrast]; too much text; drop down menu). CM, IC *Presentation of content* Smileys in chat too small; mismatch icon and text? PWD, IC Presentation of information (dosed offering of questionnaires; type of questions does not suit every user). IC, CM, OA Addressing users (PWD: I am not Sir, but just John. Just call me by my first name). PWD	Green color: restful and nice. All Predominantly quiet screens: nice. All Use of smileys is good (all): it says more than a number of sentences. PWD Monitor question for IC as start question is appreciated a lot by IC. IC
Content quality	Accuracy	Use of language is unclear: sentences are incorrect; words are missing. IC	
	Relevance	Called into question the relevance and difficulty of content for PWD. CM, IC Doubts about relevance of UI elements (mentioning phase of decision-making for PWD and IC? CM	Appreciation of monitor question for IC. IC Questionnaires are relevant: both open and closed questions as well as examples with smileys. IC, CM, OA Examples are nice. PWD, IC, OA
	Comprehensibility	Questionnaires are too difficult and too confrontational for PWD? IC, CM Unclear UI elements ("Deciding together"; drop down menu). IC, OA	Questionnaires and smileys are clear. IC, OA
	Completeness	UI element ("Messages" screen too complicated. IC	
Service quality	Usefulness	Called into question: the usefulness of tool for PWD (starting early in dementia process). CM, IC Called into question: personal contact versus digital contact. IC Called into question: transparency in tool versus confrontation for PWD. CM	Tool is fun. IC Examples in tool are supporting. PWD, IC, OA Tool helps thinking about things; offers handles for discussing things. IC, OA Reducing difference in information of IC nearby and at distance. IC Tool contributes to appreciating IC.
	Social dynamics	Unclear what happens with information in tool. IC	
	Psychological influence	Tool is too confronting for PWD. CM Careful way of writing is necessary. IC, OA, CM	
	Perception		Learning from tool (and iPad). PWD Pleasure. PWD, IC, CM, OA

PWD = people with dementia; IC = informal caregiver; CM = case manager; OA = older adults; UI = user interface

than the opposite that other participants liked. Moreover, some people with dementia liked the menu bar. We honored the preferences of the people with dementia in coloring all buttons into green and giving them the choice of viewing the menu bar folded or expanded (Fig. 4).

IV. DISCUSSION

In this study, we have identified design issues (weaknesses and strengths) for an interactive web tool facilitating shared decision making in care networks of people with dementia and its twofold complexity: the progressive character of dementia and the multitude of people involved.

The weaknesses were mainly in the quality of system: user friendliness (too many screens and too much information), unclear navigation (in screens and between screens), and design style (use of colors, smileys, and graphical layout). The weaknesses in the content of the design were the relevance of the content and the accuracy. The strengths

TABLE IV. PARTICIPANTS' CONTRIBUTIONS TO THE DESIGN OF THE INTERACTIVE WEB TOOL, THE DECIDEGUIDE

CeHRes assessment of design quality			**Focus group sessions and usability tests**				**Cognitive walkthrough**
			Case managers	***Older adults***	***Informal caregivers***	***People with dementia***	***Researchers***
A. System quality	1. User friendiness	a.Ease of acces	x	x	x	x	x
		b.Ease of use	x	x	x	x	x
		c.Absence of technical errors	x	x	x	x	x
		d.Clear navigation structures	x	x	x	x	x
		e.Efficient search functionality					
		f.Efficient feedback channels	x	x	x		
		g.Push factors	x		x	x	x
		h.Technical support					
		i.Readability of text		x	x	x	x
		j.Completeness		x	x	x	
	2.Safety & technical security	a.Privacy& confidentiality					x
		b.Encryption					
		c.Authentification					
		d.Interoperability					
	3.Design persuasiveness	a.Lens for design	x	x	x	x	x
		b.Presentation of content	x	x	x	x	x
		c.Observation					
		d.Conditioning					
B. Content quality	1. Accuracy			x	x		
	2. Evidence based						
	3. Relevance		x	x	x	x	x
	4. Comprehensibility		x	x	x	x	x
	5. Completeness			x	x	x	x
	6. Language and ethnicity						
	7. Disclosure						
C. Service quality	1. Usefulness		x	x	x	x	x
	2. Responsiveness				x		x
	3. Social dynamics				x		x
	4. Psychological influence		x		x		x
	5. Reliability						
	6. Credibility						
	7. Perception	a. Learning			x	x	
		b. Pleasure		x	x	x	

An X means that the target group commented on this item; an empty cell means that the target group made no comment included possible future extensions, monitoring informal caregivers' well-being, the use of smileys, and the green interface color. The participants' disagreements about the designing issues of the DecideGuide included the numbers of screens and examples, the use of smileys, the design rationale of the SDM-based DecideGuide open communication, transparency, and giving voice to people with dementia. The people with dementia gave detailed and unique feedback that focused on the present, careful use of language, and a pleasant graphical layout.

A. Design process

We used an iterative participatory approach to develop the DecideGuide. In a prior study user requirements were identified for the interactive web tool [3][4][33]. These user requirements all have been given a place in the final interactive prototype.

We conducted usability tests with older adults to check whether the prototype was robust enough for people with dementia to do the test. We did not want them to drop out because of an unreliable prototype. The older adults' feedback was relevant, but the feedback of the people with dementia was the most important for the DecideGuide. This is in line with Riley and colleagues' [35] conclusion. In contrast to Riley and colleagues [35], we used paper prototyping for all end users; they did not because it was less effective for older adults and therefore not useful for people with dementia. In our study, paper prototyping seemed to be difficult for all the participants. Nevertheless, it led to useful feedback in the design process. Nygård and Starkhammar [36] identified difficulties in the use of every day technology for people with dementia (e.g., in handling the technology and limitations of knowledge and personal condition). In our study, we recognized the knowledge limitations; nevertheless, this difficulty was not a problem thus far.

The CeHRes assessment of design quality was helpful. We used it as a checklist afterwards and for categorizing the issues. Nevertheless, some categories seem to show an overlap (e.g., ease of use/A1b and clear navigation structure/A1d) and/or are difficult to distinguish from each other (e.g., relevance/B3 and usefulness/C1). Since no descriptions were offered, we had to interpret the subcategories of the framework ourselves. Furthermore, not all subcategories were relevant to an interactive web tool like the DecideGuide (e.g., efficient search functionality and interoperability). Some categories and subcategories could not be filled because the participants did not give feedback about them. These empty cells in Table IV suggest that such items were irrelevant, but items such as "technical support" and "safety and technical security" are certainly important for a web tool like the DecideGuide. Did we have a blind spot for these empty cells? Were the questions we asked good enough? We used the CeHRes assessment of design quality evaluative, after all iterations were finished. Using this assessment in advance and as a formative checklist during the development might have been helpful in the timely identification of a blind spot and the meaning of the empty cells. It enables researchers to reflect on the result of an iteration and add leading questions for the next iteration to be sure all items are discussed. Further, there are no items about the perception of end users in the current version of the CeHRes assessment of design quality. More and more helpful IT applications for vulnerable older adults are being developed due to the growth of this target group [37][38]. In order to ensure that the perspective of vulnerable people has been taken into account, a criterion could be added to the CeHRes assessment: perception.

B. Participation of people with dementia

In a prior review, Span and colleagues [17] concluded that participation of people with dementia in developing assistive technologies is not self-evident. The involvement of all end users, and particularly people with dementia themselves, is important in developing a useful and user-friendly tool for people with dementia. A recent study about the European Rosetta project (designing assistive technology for people with mild to severe dementia) confirms this [34]. In our study, the caregivers were positive about including people with dementia, but most caregivers had a biased view of the ability of people with dementia to use an interactive web tool and to participate in research. Nevertheless, the caregivers' assumptions that the usability tests would be too difficult for people with dementia proved to be wrong. Moreover, most caregivers said that web tools would be more appropriate for the coming generation of people with dementia than the current one. Only a few participants had no opinion and said, "first ask people with dementia". Deciding for people with dementia instead of asking them to participate excludes people with dementia. This approach seems to be in line with studies that exclude people with dementia – whether deliberately or not – on the basis of "shielding" (caregivers tend to shield people with dementia from participating because they are afraid of exposing them to possible stressful situations) [39][40] or "difficulty" (caregivers think that people with dementia cannot participate because it is too difficult for them) [41][42]. In both cases, caregivers decide for people with dementia rather than with them.

We included people with dementia in the design process just as we did the other participants. The people with dementia were asked to participate and give feedback in the same way as the informal and formal caregivers. This resulted in a tool that takes account of the wishes of people with dementia: use of the first name, asking how they feel "right now", use of examples to hit upon an idea, and the use of smileys that are not childish at all, but nice and handy. Awareness of the importance of involving people with dementia in the design is crucial in order to develop a useful and user-friendly web tool. This is in line with Hanson and colleagues [43] and Robinson and colleagues [44], who emphasize the importance of collaborating with people with dementia in the development of an IT application from the beginning.

In our study, the people with dementia were very well able to give detailed comments about the interactive web tool. In verbalizing their comments they related them directly to their personal views, not to those of the other participants. Our people with dementia seemed to live and comment more in the "here and now", unlike the other participants, who took the future into account. This is illustrated by the question “How are you today?” that we changed to “How are you right now?” on the basis of their comments. The other participants seemed to think more abstractly about and for people with dementia. Replacing people with dementia with others to represent their perspective is therefore unsatisfactory. None of the other participants made comments that were identical or similar to those of the people with dementia. The “here and now” comments of the people with dementia therefore have added value. Their contribution is unique and requires careful inclusion of them in designing a web tool for them.

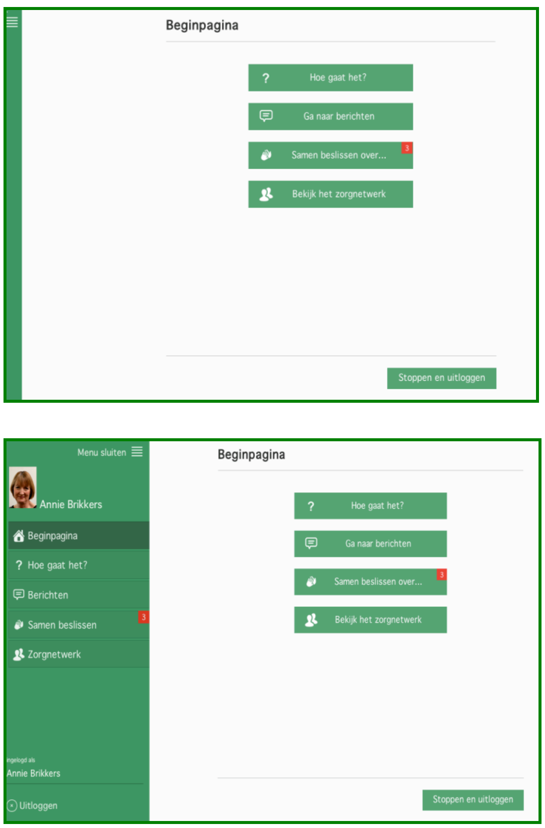


Figure 4. Folded and expanded menu bar

C. *Limitations and strengths*

This study has some methodological limitations. The first limitation concerns the participating people with dementia. Most of them had a high level of education. Further, the case managers recruited most of them; in other words, although they could have registered on their own, they did not. These two facts may have influenced the findings in this study positively.

The second limitation lies in the location of the usability tests. Older adults, informal caregivers, and case managers took the usability test in a skills lab at the university. Particularly the older adults and informal caregivers showed some stress and hasty behavior at the beginning of the test. They overlooked things on the screens. The people with dementia took the usability tests at home. They all were relaxed and showed no signs of stress. This discrepancy between the people with dementia and the others was unexpected, but may have been caused by the differing environments.

The strength of this study lies in its iterative, participatory approach. The interactive web tool was developed step by step with maximum participation of all end users. Different methods at the individual and group levels were used to enable end users to speak for themselves and also to challenge them in encounters with others. We explored the end users’ views thoroughly, and we listened carefully to the people with dementia.

V. CONCLUSIONS

Designing an interactive web tool that facilitates SDM in care networks of people with dementia, for participants who have different capacities and interests, is challenging. Design issues included a screen design based on pleasant and harmonious colors, the use of clear and uniform buttons throughout the interface, the use of multiple-choice questions with smileys as answering options, a foldable menu bar that is closed (for people with dementia) or open (for caregivers) by default, and the incorporation of a chat function that specifically keeps all end users involved in a conversation. All viewpoints were included in the design process, with special attention to the most vulnerable participants – the people with dementia. Their specific and detailed contribution was their focus on the present, the accuracy of language, and the graphical layout. Their feedback about the design was therefore unique and very valuable. However, other participants doubted whether the tool would be useful and usable for people with dementia. A pilot study will show whether these doubts about the value of the tool for people with dementia are valid.

ACKNOWLEDGMENT

The authors would like to thank all participants of the individual interviews and focus group interviews for their contribution.

This work was supported by The Dutch Foundation Innovation Alliance (SIA RAAK [Regional Attention and Knowledge Circulation] PRO), Zorgpalet Hoogeveen (Residential care organization for older adults) and Windesheim University of Applied Sciences.

VI. REFERENCES

[1] M. Span, M. Hettinga, C. Smits., L.M. Groen-van de Ven, J. Jukema, M. Vernooij-Dassen, et al., "Developing a Supportive Tool to Facilitate Shared Decision Making in

Dementia, Involvement of End Users in the Design Phase," The Sixth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2014) IARIA, March 2014, pp. 253-256, ISBN: 978-1-61208-327-8

[2] C. A. G. Wolfs, M. E. de Vugt, M. Verkaaik, M. Haufe, P. J. Verkade, F. R. J. Verhey, et al., "Rational decision-making about treatment and care in dementia: A contradiction in terms?," Patient Education and Counseling, vol. 87, pp. 43-48, 2012.

[3] M. Span, C. Smits, L. M. Groen-van de Ven, A. Cremers, J. S. Jukema, M. J. F. Vernooij-Dassen, et al., "Developing eHealth technology for people with dementia: towards a supportive decision tool facilitating shared decision making in dementia," The Fifth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2013) IARIA, Feb 2013, pp. 69-72, ISBN: 978-1-61208-252-3

[4] M. Span, C. Smits, L. M. Groen-van de Ven, A. Cremers, J. Jukema, M. Vernooij-Dassen, et al., "Towards an Interactive Web Tool that Supports Shared Decision Making in Dementia: Identifying User Requirements," International Journal On Advances in Life Sciences, vol. 6, no 3&4 2014a, in press.

[5] M. Prince, R. Bryce, and C. Ferri, "World Alzheimer Report 2011," 2011. Available at: http://www.alz.co.uk/research/WorldAlzheimerReport2011.pdf. Last access date 20-10-2014.

[6] C. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, et al., "Global prevalence of dementia: a Delphi consensus study," Lancet, vol. 366, pp. 2112-2117, 2005.

[7] G. Livingston, G. Leavet, M. Manela, D. Livingston, G. Rait, E. Sampson, et al., "Making decisions for people with dementia who lack capacity: qualitative study of family carers in UK," BMJ, vol. 341, 2010.

[8] H. G. van der Roest, F. J. M. Meiland, H. C. Comijs, E. Derksen, A. P. D. Jansen, H. P. J. van Hout, et al., "What do community-dwelling people with dementia need? A survey of those who are known to care and welfare services," International Psychogeriatrics, vol. 21, pp. 949-965, 2009.

[9] C. G. Lyketsos, O. Lopez, B. Jones, A. L. Fitzpatrick, J. Breitner, and S. DeKosky, "Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: Results from the cardiovascular health study," JAMA, vol. 288, pp. 1475-1483, 2002.

[10] M. Pinquart and S. Sörensen, "Correlates of Physical Health of Informal Caregivers: A Meta-Analysis," The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, vol. 62, pp. P126-P137, March 1, 2007.

[11] G. Elwyn, A. Edwards, and P. Kinnersley, "Shared decision-making in primary care: the neglected second half of the consultation.," British Journal of General Practice, vol. 49, pp. 477-482, 1999.

[12] G. Elwyn, A. O'Connor, D. Stacey, R. Volk, A. Edwards, A. Coulter, et al., "Developing a quality criteria framework for patient decision aids: online international Delphi consensus process," BMJ, 2006.

[13] M. von Kutzleben, W. Schmid, M. Halek, B. Holle, and S. Bartholomeyczik, "Community-dwelling persons with dementia: What do they need? What do they demand? What do they do? A systematic review on the subjective experiences of persons with dementia," Aging & Mental Health, vol. 16, pp. 378-390, April 2012.

[14] J. Whitlatch, Menne, H, "Don't forget about me. Decision making y people with dementia," Journal of teh American Society on Aging, vol. 33, pp. 66-71, 2009.

[15] S. L. Dupuis, J. Gillies, J. Carson, and C. Whyte, "Moving beyond patient and client approaches: Mobilizing 'authentic partnerships' in dementia care, support and services," Dementia, vol. 11, pp. 427-452, 2011.

[16] K. Murphy, F. Jordan, A. Hunter, A. Cooney, and D. Casey, "Articulating the strategies for maximising the inclusion of people with dementia in qualitative research studies," Dementia, January 8, 2014.

[17] M. Span, M. Hettinga, M. J. F. J. Vernooij-Dassen, J. Eefsting, and C. Smits, "Involving People with Dementia in the Development of Supportive IT Applications: a Systematic Review," Ageing Research Reviews vol. 12, pp. 535-551, 2013.

[18] H. Beyer and K. Holzblatt, Contextual Design. Defining Customer-Centered Systems. : Morgan Kaufmann Publishers, 2010.

[19] L. Nygärd, "How can we get access to the experiences of people with dementia? Suggestions and reflections.," Scandinavian Journal of Occupational Therapy, vol. 13, pp. 101-112, 2006.

[20] P. Zaphiris, M. Ghiawadwala, and S. Mughal, "Age-centered Research-Based Web Design Guidelines," in Proceedings of CHI 2005 Conference on Human Factors in Computing Systems, Portland, USA, 2005.

[21] K. Day, D. Carreon, and C. Stump, "The therapeutic design of environments for people with dementia: a review of the empirical research," The Gerontologist, vol. 40, pp. 397-416, 2000.

[22] L. Mitchell, E. Burton, S. Raman, T. Blackman, M. Jenks, and K. Williams, "Making the outside world dementia friendly: design issues and considerations," Environment and Planning B: Planning and Design, vol. 30, pp. 605-632, 2003.

[23] E. Burton and L. Mitchell, "Urban design for longevity," Urban Design Quarterly, vol. 87, pp. 32-35, 2003.

[24] N. Alm, R. Dye, A. Astell, M. Ellis, G. Gowans, and J. Campbell, "Making software accessible for users with dementia. In J. U. Lazar (Ed.), Niversal usability: Designing computer interfaces for diverse users ", ed New York: Wiley, 2007, pp. 299-316.

[25] F. S. Lim, T. Wallace, M. A. Luszcz, and K. J. Reynolds, "Usability of tablet computers by people wih early-stage dementia," Gerontology, vol. 59, pp. 174-182, 2013.

[26] J. E. W. C. Van Gemert-Pijnen, N. Nijland, M. A. H. Van Limburg, S. M. Kelders, B. J. Brandenburg, H. C. Ossebaard, et al., "Introducing a holistic framework for eHealth technologies," Journal of Medical Internet Research, vol. 13 (4):e111, 2011.

[27] J. Ritchie and L. Spencer, Eds., 'Qualitative data analysis for applied policy research', in Bryman and Burgess, eds., Analysing Qualitative Data. London: Routledge 1994, pp. 173-194.

[28] N. Nijland, "Grounding eHealth: towards a holistic framework for sustainable eHealth technologies," University of Twente, Enschede, 2011. ISBN: 9789036531337

[29] J. Brender, Handbook of Evaluation Methods for Health Informatics: Elsevier, 2006.

[30] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability," in Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems: 24–29 April 1993, Amsterdam, Netherlands: Association for Computing Machinery pp. 206-213, 1993.

[31] R. A. Virzi, "Refining the test phase of usability evaluation: How many subjects is enough?," Human Factors, vol. 13, pp. 457-468, 1992.

[32] M. W. M. Jaspers, T. Steen, C. v. d. Bos, and M. Geenen, "The think aloud method: a guide to user interface design," International Journal of Medical Informatics, vol. 73, pp. 781-795, 2004.

[33] M. Span, C. Smits, L. M. Groen-van de Ven, A. Cremers, J. Jukema, M. Vernooij-Dassen, et al., "Towards an Interactive Web Tool that Supports Shared Decision Making in

Dementia: Identifying User Requirements," International Journal On Advances in Life Sciences, submission nr 31054, vol. 6, 2014.

[34] F. J. M. Meiland, B. J. J. Hattink, T. Overmars-Marx, M. E. de Boer, A. Jedlitschka, P. W. G. Ebben, et al., "Participation of end users in the design of assistive technology for people with mild to severe cognitive problems; the European Rosetta project," International Psychogeriatrics, vol. 26, pp. 769-779, 2014.

[35] P. Riley, N. Alm, and A. Newell, "An interactive tool to promote musical creativity in people with dementia," Computers in Human Behavior, vol. 25, pp. 599-608, May 2009.

[36] L. Nygård and S. Starkhammar, "The use of everyday technology by people with dementia living alone: Mapping out the difficulties," Aging & Mental Health, vol. 11, pp. 144-155, 2007.

[37] L. Magnusson, E. Hanson, and M. Borg, "A literature review study of information and communication technology as a support for frail older people living at home and their family carers," Technology and Disability, vol. 16, pp. 223-235, 2004.

[38] S. Lauriks, A. Reinersmann, H. v. d. Roest, F. J. M. Meiland, R. J. Davies, F. Moelaert, et al., "Review of ICT-based services for identified unmet needs in people with dementia," Aging Research Reviews, vol. 6, pp. 223-246, 2007.

[39] R. Orpwood, A. Sixsmith, J. Torrington, J. Chadd, and G. Chalfont, "Designing technology to support quality of life of people with dementia," Technology and Disability, vol. 19, pp. 103-112, 2007.

[40] P. Topo, O. Maki, K. Saarikalle, N. Clarke, E. Begley, S. Cahill, et al., "Assessment of a music-based multimedia program for people with dementia," Dementia (14713012), vol. 3, pp. 331-350, 2004.

[41] H. G. van der Roest, F. J. Meiland, T. Haaker, E. Reitsma, H. Wils, C. Jonker, et al., "Finding the service you need: human centered design of a Digital Interactive Social Chart in DEMentia care (DEM-DISC)," Studies in health technology and informatics, vol. 137, pp. 210-224, 2008.

[42] J. Boger, J. Hoey, K. Fenton, T. Craig, and A. Mihailidis, "Using actors to develop technologies for older adults with dementia: A pilot study," Gerontechnology, vol. 9, pp. 450-463, 2010.

[43] E. Hanson, L. Magnusson, H. Arvidsson, A. Claesson, J. Keady, and M. Nolan, "Working together with persons with early stage dementia and their family members to design a user-friendly technology-based support service," Dementia: The International Journal of Social Research and Practice, vol. 6, pp. 411-434, 2007.

[44] L. Robinson, K. Brittain, S. Lindsay, D. Jackson, and P. Olivier, "Keeping In Touch Everyday (KITE) project: developing assistive technologies with people with dementia and their carers to promote independence," International Psychogeriatrics, vol. 21, pp. 494-502, 2009.

The State of Video-Based Learning: A Review and Future Perspectives

Ahmed Mohamed Fahmy Yousef, Mohamed Amine Chatti, Ulrik Schroeder
Learning Technologies Research Group (Informatik 9), RWTH Aachen University
Aachen, Germany
{Ahmed.fahmy; Schroeder}@cil.rwth-aachen.de; Chatti@informatik.rwth-aachen.de

***Abstract*— The pedagogical strength of Video-Based Learning (VBL) is presenting knowledge in consistent and attractive manner. In recent years, the new forms and technologies of VBL such as flipped classrooms, and most prominently MOOCs, have had a remarkable impact on teaching and learning methodologies. A significant number of academic publications have investigated and analyzed VBL environments from different perspectives, including potential usage, effects on learning outcomes, satisfaction levels, and effectiveness. This study provides a critical analysis of the current research in VBL conducted from 2003 until today. We aim to help educators in building deeper understanding about the educational benefits of VBL. In this study, 76 peer reviewed papers are identified through journals and academic databases and they are categorized into four main dimensions: effectiveness, teaching methods, design, and reflection. In the scope of this analysis, we also provide future visions and research opportunities in VBL that support self-organized and network learning.**

Keywords-Video-Based Learning; VBL; MOOC; Review of Research; Blended Learning; Video Design; Flipped Classroom; Technology-Enhanced Learning.

I. INTRODUCTION

Video-based learning (VBL) is now recognized by Technology-Enhance Learning (TEL) researchers as a powerful learning resource in online teaching activities. This paper presents an extended and more detailed version of our paper presented at the sixth international conference on mobile, hybrid, and online learning (eLmL 2014), where we reviewed the existing methodologies of VBL research [1]. VBL has unique features that make it an effective learning method that can enhance and partly replace traditional classroom-based and teacher-led learning approaches. VBL can change the way we learn as well as how we teach [2]. Videos can help students by visualizing how something works [3] and show information and details difficult to explain by text or static photos [4]. In addition, videos can attract students' attention, thus motivating them and engaging them to increase their collaboration. Using videos thus can lead to better learning outcomes [5]. Moreover, video can support different learning styles, specifically students who are 'visual learners' [6].

Indeed, VBL has a long history as a learning tool in educational classes. First experiments started during the Second World War. Soldiers were then trained with a combination of audio and film strips [7]. As a result, the static film strips helped to increase their skills while saving a lot of time as well. By the late 1960s, educational television was used as an extra tool in classrooms. Also teachers were confronted with videos of their own lessons to reflect on their teaching methods and improve their performance [8]. In the 1980s, VHS videotapes meant a quantum leap as it became much easier to use video in classrooms. But, still, learners were rather passive and could only watch the video. This changed with the rise of digital video CDs in the mid-1990s. Teachers could now add multimedia control and assessment tools by using the video on a computer. Thus, learners became much more active than before. By the 2000s, classrooms got connected to the internet and interactive digital video as well as video conferences became possible. Since then, new technologies such as smartphones and tablets in combination with social media such as YouTube have contributed to increasing social interaction and have made it easier as ever to integrate video applications in education [9][10]. In recent years, VBL publications have increased in order to discuss how VBL can facilitate learning and enhance learner's outcome as well as teacher's performance. Thus, there was a need to collect existing research, document the benefits of video in improving learning, and explore the design and teaching methods in VBL environments. In this study, we critically analyze the research on VBL to answer the following research questions:

1. What are the educational benefits that VBL has on teaching and learning?
2. How VBL technologies enhance students' learning outcome?
3. How educators and researchers design VBL environments?
4. How is VBL used to improve teacher's and learner's reflection?
5. What are possible applications of VBL in open and networked TEL environments?

In order to answer these questions, this paper will discuss different angles of VBL. The remainder of this paper is structured as follows: Section II is a review of the related work dealing with the systematic review of research on VBL in the past ten years. Section III describes the research methodology, how we collected the research data, and how we categorized the VBL literature. In Section IV, we review and discuss the current research based on several dimensions. In Section V, we present recent implementations in VBL with a focus on the MOOCs and flipped classroom models. Finally, Section VI gives a summary of the main findings of this paper and highlights new research opportunities for future work with some guidelines for practitioners.

II. RELATED WORK

This section surveys the previous work most closely related to the current study and place our contributions in the proper context.

Tuong et al. [11] conducted a systematic review of 28 VBL studies in order to examine the effectiveness of the instructional videos in modifying health behaviors. The main findings of this review show that instructional videos interventions appear to be effective in the general self-care testing (e.g., breast self-examination, heart failure and treatment adherence).

Greenberg and Zanetis [12] reported the positive impact of video broadcast and streaming in education. As a result of their study, the authors encourage teachers and educators to use interactive video training materials in classes especially with children.

Borgo et al. [13] conducted a study to provide an overview of the major advances in automated video analysis and investigate some techniques in the field of graphic design and visualization.

Tripp and Rich [14] reviewed 63 studies in order to understand the ability of teachers to reflect on their teaching through video recording. The result of this study was that teachers prefer to use video recording for reflection in collaboration with colleagues than reflecting individually. Also, teachers report that the use of a guiding framework (e.g., rubric, checklist, teaching principles) helps to focus on their reflection by focusing their attention on certain tasks.

Although these studies asserted that the video is a powerful tool in TEL and that videos enable teachers to reflect on their teaching, they do not take into account the teaching methodologies, design approaches, and the impact of teachers' reflections on their students' learning outcomes. As compared to the above studies, our study adds a wide range of peer-reviewed studies that have been conducted between 2003 and 2014 and provides a quantitative as well as qualitative analysis of the VBL literature. Moreover, we apply a cognitive mapping approach to categorize the VBL publication into several dimensions. The study further provides critical discussion according to each dimension and suggests new opportunities for future work.

III. METHODOLOGY

The research methodology was carried out in two main phases including identification of eligible studies followed by a cognitive mapping approach to categorize the VBL literature into several dimensions.

A. *Identification of Eligible Studies*

The significant research method of identifying papers from Internet resources was applied to collect data in this study [15]. This method was carried out in three rounds. Firstly, we conducted a search in 7 major refereed academic databases. These include Education Resources Information Center (ERIC), JSTOR, ALT Open Access Repository, Google Scholar, PsychInfo, ACM publication, IEEEXplorer, and Wiley Online Library.

Secondly, we searched 23 academic journals in the field of educational technology and TEL indexed by Journal Citation Reports (JCR) including Australasian Journal of Educational Technology, British Journal of Educational Technology, Canadian journal of learning and technology, CITE Journal, Computers in Human Behavior, The Electronic Journal of e-Learning (EJEL), European Journal of Open, Distance and E-Learning (EURODL), Instructional Science, Interactions Journal, The International Journal of Instructional Technology and Distance Learning, International Review of Research in Open and Distance Learning (IRRODL), Journal of Asynchronous Learning Networks, Journal of computer assisted learning (JCAL), Journal of Computing in Higher Education, Journal of distance education, Journal of Interactive Media in Education, Journal of Interactive Online Learning, Journal of Learning Design, Journal of Online Learning and Teaching (JOLT), Journal of Technology, Learning, and Assessment , Learning, Media and Technology, and Turkish Online Journal of Distance Education (TOJDE), using the keywords (and their plurals) "Video-based learning", "VBL", "teaching with interactive video", and "Video Instruction". As a result, 127 peer-reviewed papers were found.

Thirdly, a set of selection criteria were identified as follows:

1. Studies must focus on VBL in educational development. Studies on video coding and semantic retrieval of video were excluded.
2. Experimental or empirical case studies on how learners learn with and from videos were included. Studies of video recording strategies were excluded.
3. Studies that focus on ability of teachers to reflect on their teaching via video recording were included.
4. Studies evaluating the VBL activities and effectiveness in education were included. Studies that focused on video-games and video conferencing tools were excluded.

This resulted in a final set of 76 peer-reviewed studies, which met the selection criteria above. Fig. 1 shows the number of VBL publications between 2003 and 2014, which were found to be relevant for this study.

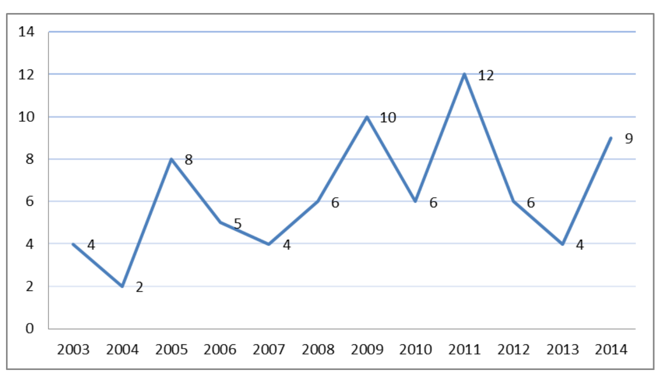


Fig. 1. VBL studies by publication year.

B. Cognitive Mapping Approach

Cognitive mapping approach is a method enabling the researchers to clarify and categorize the research literature conceptions into several dimensions regarding to the research questions. These dimensions are recorded in graphic flowchart to show the hierarchy of VBL terms [16]. We applied the cognitive mapping approach as a classification technique for dividing the VBL literature into four dimensions relevant to the research questions, namely effectiveness, teaching methods, reflection, and design (see Fig. 2).

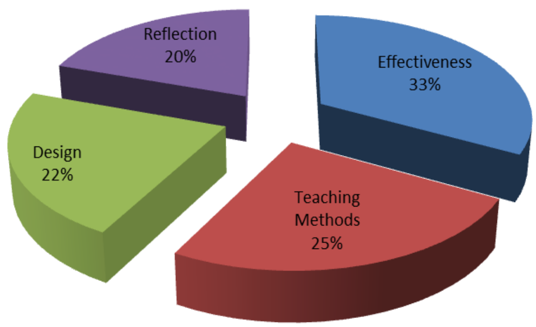


Fig. 2. Visual representation of the VBL dimension.

In order to capture the information gained from the literature analysis, we created a VBL field diagram (see Fig. 3), which has been partitioned into four categories and thirteen sub-categories.

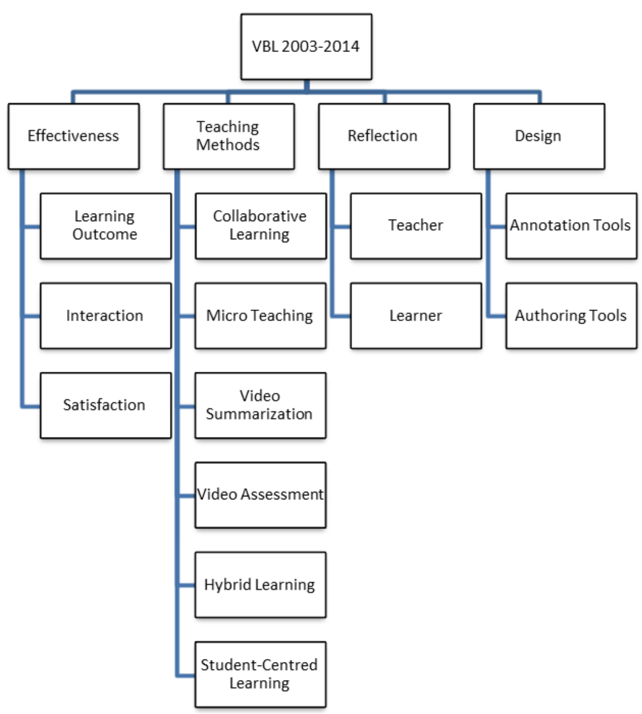


Fig. 3. VBL cognitive map

IV. Discussion

In this section, we critically discuss in details the VBL literature based on the cognitive map dimensions that have been identified in Section III, namely effectiveness, teaching methods, reflection, and design. For the critical discussion part, we apply the meta-analysis method, which aims to contrast and combine results from several studies into a single scientific work [15].

A. Effectiveness

Effectiveness of VBL has received a great deal of attention from academic scientists. 33% of the studies reviewed in this paper examined the effectiveness of VBL. Most of the reviewed case studies asserted the efficacy and usefulness of VBL as a powerful medium used in education. We analyzed each study for the following characteristics: research goal, subject, target group, sample size, and summary of results. In the following sections, we discuss the effectiveness of VBL in terms of learning outcome, interaction, and learners' satisfaction.

1) Learning Outcome: A learning outcome (or achievement) can be described as knowledge, skills, and abilities that learners have to achieve as a result of the learning process [2]. Many TEL scholars believe that VBL has the potential to promote the learning outcome. VBL can, for instance, present knowledge in an attractive and consistent manner [5][17]. Further, Kay and Edward [18] and Balslev et al. [19] compared VBL supported by a cognitive approach with text-based learning. The results showed statistically significant differences in improving learners' skills. Moreover, the authors reported that learners liked the followed cognitive approach in which knowledge was generated through step by step learning in video lectures.

In addition, Lin and Tseng [20] and Hsu et al. [21] conducted two studies to investigate the effect of different VBL designs to improve English language skills of K-12 pupils. The findings indicated that the groups which used VBL outperformed the other groups. Other studies reported the invaluable impact of using VBL in improving teachers' performance. The results asserted that using videos as educational tools improved teaching methods and increased the learning outcome [6][8][22][23].

On the other hand, some studies indicated that there were no statistically significant differences between teaching with videos and other methods, thus making them equivalent [24][25][26]. Moreover, Chuang and Rosenbusch [27] stressed the importance of the pedagogical aspect for an effective VBL experience. The authors pointed out that only using videos without pedagogical approach does not make sense. The authors emphasized that video technology should go side by side with pedagogy, and provided a constructivist framework to engage learners to learn with videos. Equally important, Giannakos et al. [28] highlighted the importance and benefits of applying learning analytics to support teachers and students. Learning analytics will help in guiding the learners to the appropriate learning materials for improving the use of their courses. This can be achieved by aggregating and analyzing learners' interactions with other available learners' data. Learning analytics opens new research directions on VBL courses about accessing recommendations for future learning activities. This means, that issues related to data privacy, ownership, sharing, and access need to be resolved [29].

In sum, the reviewed studies indicated that there were conflicting results of using VBL in educational environments as some found it valuable while others reported no significant results. There was, however, an agreement among researchers that VBL in conjunction with appropriate pedagogical methods has the potential to improve the learning outcome.

2) Interaction: Improved interaction and communication among participants is another effectiveness aspect in VBL. DeLoache and Korac [30] reviewed some case studies of using videos with infants. The authors pointed out that video stories indeed improved communication between children. Hakkarainen and Vapalahti [31] investigated learning with video in the forum-theatre. This study showed that VBL can enhance interaction among learners and improve the ability to solve every day social problems. Recently, Shen [32] evaluated the effects of VBL in nursing simulation practice using the "experimental group and control group" method. The results of this investigation showed that, nurses in the experimental group received significantly higher scores in the final evaluation of catheterization, communication skills, and satisfaction than the nurses in the control group.

On the contrary, Muhirwa [33] investigated VBL in TEL environments in Africa and pointed out that VBL had a lesser role in increasing interaction among learners. This was due to the fact of poor internet connectivity, limited access to computers, and lack of trained instructors in Africa. Additional obstacle that might prevent learners from Africa to actively participate in VBL is the poor technology infrastructure, only 25% of Africa has access to electricity [34].

3) Satisfaction: The level of learning satisfaction is important in evaluating the effectiveness of VBL environments. Zhang et al. [5] examined the level of satisfaction through interactive VBL in a study involving 138 students. As a result, students who used a TEL environment that provides interactive instructional video reported higher levels of satisfaction than those in the control group without video.

Moreover, it has been shown that interactive videos have an impact on the emotional side of the learners' behaviour (e.g., real-life interaction, incorporate the different sound and musical effects that can fit the emotional contents of the learning subject) and that videos can improve the attention to the subject of the lecture in addition to the positive impact on the learners' motivation level [35][36][37].

B. *Teaching Methods*

Dale's cone of experience presents how information is understood, processed, transferred, and maintained as knowledge within the learning process [38]. Fig. 4 shows what learners will be able to do at each level of the cone.

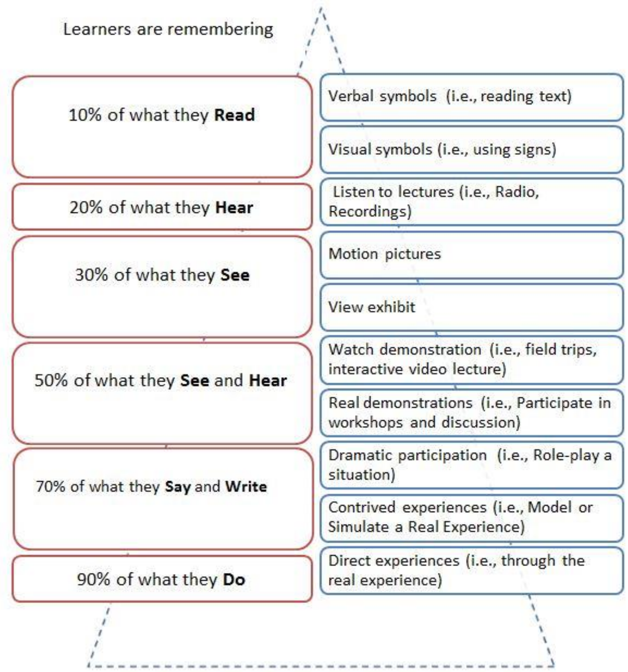


Fig. 4. Cone of experience.
Adapted from E. Dale, Audiovisual Methods in Teaching, 1969, NY: Dryden Press [38].

According to Dale's cone, the most effective methods stand at the bottom. These methods involve direct experience, practical and hands-on workshops, which compel learners to better remember their activities. Interactive videos belong to this category as they enable learners to interact with the video materials through annotations, discussions, and assessment. Educationists and scholars use a broad range of teaching methodologies in VBL environments in order to increase the value of interactive videos. In this literature review, collaborative learning is a key aspect which underlies most of the studies. Other methods involve micro teaching, video summarization, video assessment, hybrid learning, and student-centered learning.

1) Collaborative Learning: In video-based collaborative learning, which focuses on developing, discussing, exploring alternatives rather than directions, learners are able to share responsibilities for their learning [5][39][40]. Most of the reviewed studies validate the efficacy and usefulness of collaborative VBL, where learners can develop their problem-solving abilities via collaboration with others [12]. These studies reported various educational benefits for learners working cooperatively in teams such as shared goals, ideas, resources, activities, and supporting each other [41][42][43][44]. For instance, Pea and Lindgren [45] investigated which collaboration design patterns are used by learners when they have access to a Web-based video collaboration platform. Five collaboration patterns were identified, namely collective interpretation, distributed design, performance feedback, distributed data coding, and video-based prompting. These patterns support teacher-centred learning by providing knowledge and allowing learners to discuss and find solutions.

2) Micro Teaching: The micro teaching method was used in some studies as a teaching practice with a smaller class size and time (e.g., four to nine learners in a class that is held for five to ten minutes). Educators are able to give learners some quick and easy feedback on their learning performance through video podcasts [46]. Finlay et al. [47] reported that learners' responses on micro teaching with video podcasts are very positive. The authors, however, noted that the video of 10 minutes length was too long for many learners and found that the shorter video podcasts (4-5 minutes) have the advantage of giving greater flexibility in micro teaching lessons. Woodruff [48] investigates video lectures with a small group of students with autism in a series of art lessons. The main result is the following: Students with autism spectrum develop their artistic skills and retain more art content knowledge with highest grades than through traditional teaching classes. Other studies showed that micro teaching provides a friendly and supportive learning environment [49][50].

3) Video Summarization: Video summarization technique extracts important information and provides short but informative summary of the lecture content [51][52]. Chang et al. [53] designed a keyword-based video summarization learning platform (KVSUM) which provides a keyword cloud as a textual surrogate to support learners to organize information of videos and enhance them to follow the videos and reducing the learning time.

4) Video Assessment: A video assessment is short video that simulates real life activities and provides possible responses to the several daily problems. Learners are asked to select which of the responses they would take in these circumstances. Afterwards, teachers discuss each response and evaluate learner's responses [54][55].

5) Hybrid Learning: Hybrid learning has become one important TEL model, by integrating online learning and traditional face-to-face classroom together [56][57]. Pang [58] conducted a study by following a hybrid learning approach that uses video-based learning materials in a Physical Education course. In this course, the trainer can review the learner's actions video, pick out the wrong actions, and provide feedback. Then, students can reflect, find out mistakes. The experiment shows that 80.9% out of learners think that the video review indeed improved their physical skills.

In other studies, Shih [59] and Kırkgöz [60] investigated a hybrid learning approach supported by video lectures for an English speaking course. The study showed that the learners made noticeable improvement in their oral communication skills, and that they were satisfied with the blended learning model.

6) Student-Centred Learning: Most of the reviewed VBL studies followed a teacher-centred approach. Only 15% of studies have focused on student-centred learning [61][62]. These studies don't depend on teachers as content providers. They aimed at providing the space for students to be active participants in their learning environment, interact to build and construct knowledge, and get mutual support to make decisions using reflection and critical judgement.

C. Design

Several researchers in TEL have explored how to design effective VBL environments. Annotation and authoring tools are the most used design tools in the reviewed VBL literature.

1) Annotation Tools: Annotation means adding note, comment, explanation, and presentational mark-up attached to a document, image, or video [63]. In VBL, annotation refers to the additional notes added to the video without modifying the resource itself, which help in searching, highlighting, analysis, retrieval, and providing feedback [64]. Moreover, video annotation provides an easy way for indexing, discussion, reflection, and conclusion of content [65][66].

Colasante [3] examined the integration of a video annotation tool (MAT) into the learning and assessment activities of a third year class "Physical Education" course at RMIT University. This tool allowed learners to select and annotate parts of a video. These annotations are then used by students and teachers to discuss, receive feedback, reflect, and evaluate their learning and teaching practice. The results showed that MAT was effective for receiving feedback form teachers and peers. But, some issues regarding the quality of the collaborative input from peers were noted.

Moreover, feedback in VBL is recommended for several resons, it provides an easy way for discussion and reflection on the video content, provides scaffolds for learners to support self-reflection and self-assessment [3][29].

2) Authoring Tools: A number of studies have developed a wide range of authoring tools for VBL content. The primary function of these authoring tool is to increase the interactivity with the VBL environment, thus engaging learners in the learning processe [67]. The following tools were used in various VBL environments:

- Synchronize lecture note: The aim of this tool is to synchronize a video stream with the presentation slide by means of video clip timing [67].

- Content summarization tool: This tool is able to extract summary information from lecture videos and provide it to the learners automatically [68] [69].
- Digital Video Library: This tool uses indexing to enable content-based search for a particular information of a video lecture [70].
- Discussion forum: A space integrated in the VBL environment where learners can discuss and share common interests or goals on a learning topic [71][72].

As an illustration, the College of Engineering at the University of California, Berkeley has launched an online Master's program in integrated circuits. This project embeds VBL modules for library research methods. In this program, the library plays a significant role in providing the teaching resources and instruction to help learners succeed in their studies. The results manifested a positive impact on the university library and encouraged the development of facilities and services, such as using digital video library to enhance personalized interaction with learners [73].

D. Reflection

There is a general interest among researchers and educators in using VBL to support teachers' and students' reflection on their teaching and learning activities [14][74][75].

1) Teacher Reflection: Video recording of the classroom lessons enables teachers to reflect on their teaching [76]. Teachers can record their own teaching, watch what they did in the classroom, think about it, and reflect on the performance using both individual and collaborative reflection [77][78].

Studies examined both individual and collaborative reflection. 85% of the studies on reflection in VBL noted that teachers prefer to reflect on their teaching performance with colleagues [4][76][78]. Similarly, Calandra et al. [78] and Calandra et al. [79] stressed that the teacher's reflective process should be collaborative where groups of teachers provide comments or feedback to each other. Several reflection methods were used, e.g., daily reflection, weekly reflection, and end of semester reflection [80][81].

Only 15% of studies examined self-reflection where teachers reflected individually on their teaching. Teachers used video-taped lesson analysis and wrote comments for self-reflection [82]. Likewise, Gainsburg, [61] implemented video annotation tools to scaffold, structure, and transform teacher reflection.

Recently, video reflection has been used for pre-service teacher education. Blomberg et al., [83] explored the use of two VBL courses, on to determine pre-service teachers' ability to reflect on classroom video. The study found that the video recording distinctly impacts on the pre-service teachers' reflection patterns. On the contrary, Cho and Huang [84] investigated the mutual relationships between pre-service teachers' beliefs and video-based reflection activities in wiki. The authors found that cognitive beliefs partially influenced reflective writing and questioning activities in wikis.

2) Learner Reflection: Recording classroom activities is also important for learners to reflect on their own learning experience, evaluate their performance, and get a clearer overview of their learning progress. Video recordings further help learners in revision prior to exams [75][85].

Dalgarno et al., [86] discussed three common methodologies in which learners are helped to reflect and make connections between their academic learning and their own practical learning. These methodologies are work-integrated learning programs, inquiry-based learning designs, and simulation. The authors recognized the role of rich media technologies such as videoconferencing, web conferencing and mobile videos in learners' self-reflecting and connect university classrooms to sites of professional practice.

V. Future Perspectives

In this section, we present the future perspectives carried out from the critical analysis of the VBL literature. In the last few years, the expansion of new open VBL models, such as Massive Open Online Courses (MOOCs) and flipped classrooms has changed the TEL landscape by providing more opportunities for learners than ever before.

A. MOOCs

The term "openness" has received a great deal of attention from the higher education institutions, due to the growing demand for lifelong learning opportunities. Open Educational Resources (OER) represent a first implementation of openness in higher education. The concept of OER describes any educational materials that can be used and re-used in teaching and learning. These materials are openly available and free of charge [87]. They have been widely used by educators and students as rich and powerful learning resources. OER, however, have two main limitations: they lack human interaction and do not reach massive numbers of learners.

In 2001 the Massachusetts Institute of Technology (MIT) introduced the term of Open CourseWare (OCW) as a TEL platform in order to provide their curricula material for everyone at no cost. The key difference between OCW and OER is that OCW are more specific and structured as courses than the public OER library. OCW succeeded in assisting self-organized learners who do not meet the MIT admission requirement but are interested in an OCW course. [29][88]. The criticism against OCW mainly focuses on the customization necessary to match each institute curriculum requirements and the lack of direct feedback due to the one-way design of interaction.

In 2008, Massive Open Online Courses (MOOCs) have offered a whole new perspective for openness by providing unlimited learning opportunities for a large-scale participation for free. MOOCs represent an evolution of the OER and OCW movements.

1) MOOC Definition: MOOCs are leading the new revolution of TEL by providing new opportunities to a massive number of learners to attend free online courses from anywhere all over the world [89]. Fig. 5 describes the characteristics of the four words included in the MOOC acronym.

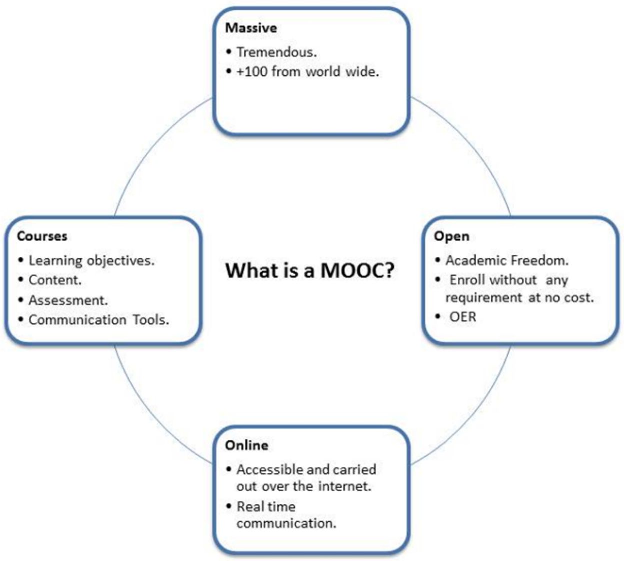


Fig. 5. MOOCs ideation

- **Massive** refers to the necessary size of course participants. But what is massive? Regarding to university campus courses 1000 learners is really a huge amount of participants. Many MOOCs have less than a hundred users, while some courses reached over 150,000 registrations. Basically, any online class that has a higher number of students than regular university courses (+100 participants) can be considered as a MOOC [34].
- **Open** refers to the academic freedom to expand access to participant regardless of their ideological, political, and cultural background [88]. Moreover, open is used in the sense of free reuse, revise, remix, and redistribute of the learning material e.g., learning objects, video lectures, quizzes, textbooks, any other tools [34], [89].
- **Online** requires the MOOC environment to be accessible and carried out over the internet. The hybrid MOOCs model (i.e., blended with face-to-face interaction and support) encourages participants to meet physically and work together on their studying projects [34].
- **Courses** are related to the structure and organization of the learning curriculum. A MOOC includes OER, learning objectives, collaboration tools, assessments, and learning analytics features [34].

Due to the nature of MOOCs environment, we strongly believe that the original definition of MOOCs will change as a result of the various challenges and rapid developments in this field.

2) MOOC Categories

Different forms of MOOCs have been introduced in the MOOC literature. Siemens [90] characterize MOOCs into cMOOCs based on the theory of connectivism, and xMOOCs by virtue of behaviorism and cognitivist theories with some social constructivism aspects as more institutional model, e.g., Coursera, edX, and Udacity. Hills offers a diagram of the evolution of MOOCs over the last few years [91].

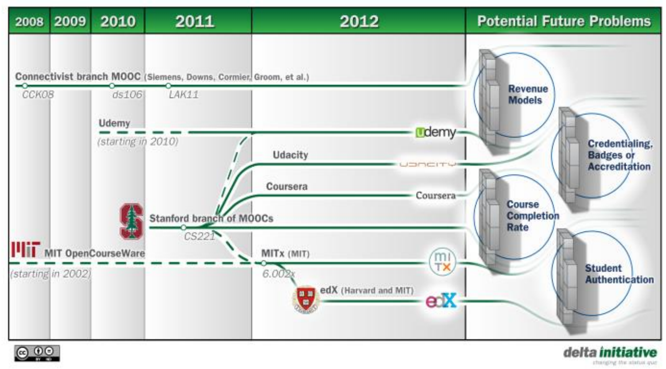


Fig. 6. The evolution of MOOCs [91]

The first design of cMOOCs was established in 2008, based on the connnectivist pedagogy approach. That enables learners to build their own networks via blogs, wikis, Google groups, Twitter, Facebook, and other social networking tools outside the learning platform without any restrictions from the teacher [92]. In xMOOCs, by contrast, learning objectives are pre-defined by teachers who impart their knowledge through short video lectures, often followed by simple e-assessment tasks (e.g., quiz, eTest) [93]. Recently, new forms of MOOCs have emerged. These include smOOCs as open online courses with a relatively small number of participants and blended MOOCs (bMOOCs) as hybrid MOOCs including in-class and online video-based learning activities [34]. The key characteristics of MOOC forms are summarized in Table I [34][94].

The majority of existing MOOCs that have been delivered at higher education institutions are xMOOCs. These university style platforms were developed by different elite institutions and usually delivered via a third party platform provider. For example, Coursera has been developed by Stanford University and currently partnered with top universities and organizations worldwide. In addition, edX was founded by the MIT and Harvard University in May 2012. There are more than 40 high ranked universities co-operated and offered courses on the edX platform [94].

Table I. Characteristics of MOOCs

Compare Item		cMOOCs	xMOOCs	bMOOCs	smOOCs
Learning theory	Connectivism	√	-	-	-
	Behaviorism	-	√	-	-
	Cognitivist	-	√	-	-
	Social constructivism	-	(√)	√	√
Structure	Pre-determined	-	√	√	√
	weekly sequences structure	-	√	(√)	√
	Self-organized	√	-	(√)	-
	Short video lectures	(√)	√	√	(√)
	Fluid structure	√	-	-	-
Teacher role	Teacher-Based	-	√	-	√
	Facilitator	√	-	(√)	-
	Co-organizer with course participants	√	-	√	(√)
Interaction	Open network via social tools e.g., Blogs, forums, live chat, social media	√	(√)	√	-
	Face-to-Face	-	-	√	-
	Daily or weekly meeting	-	-	-	√
	Limited interaction among participants and course teacher	-	√	-	-
Assessment	E-Assessment i.e., automatically grading	-	√	√	√
	Self-Assessment i.e., short quizzes to help participants formatively assess their own learning	-	√	√	-
	Peer-Assessment	√	(√)	√	(√)
	Open Assessment	√	-	-	-

√Completely (√) Partly - Not supported

3) *MOOC Goals*

The question is how and why are higher education institutions engaging with MOOCs. Through interviews with administrators, faculty members and researchers from 29 different institutions that were already offering or using MOOCs, Hollands and Tirthali [94] identified six major goals for MOOC initiatives:

- Massiveness: to extend the reach and access of education to a wider audience.
- Building and maintaining brand.
- Improving economics by reducing the costs of education or using MOOCs as a potential source of revenue i.e., business models.
- Improving learning outcomes.
- Innovation in teaching and learning.
- Research purpose i.e., conducting studies on MOOC design and methodologies.

4) *MOOC challenges*

Much has been written on MOOCs about their design, effectiveness, case studies, and the ability to provide opportunities for exploring new pedagogical stategies and business models in higher education [34][89]. MOOCs are still in a pilot form till now. A variety of concerns and criticisms in the use of MOOCs have been raised [29][94]. In this part we discuss several pedagogical and technological crucial challenges that should be considered in the development of the future MOOC environments.

a) Free against business models: The original idea of MOOCs is to offer learning content to a massive number of participants for free. In reality, however, some providers view MOOCs as a potential source of revenue and offer certificates and teaching assistance for additional fees [29][34].

b) Openness against licensing: Although MOOCs are open for massive number of participants without any entry requirements, they are not open from a copyright perspective. For instance, Coursera does not permit users to reproduce, retransmit, distribute, or publish any material from its platform[1].

c) Massiveness against drop-out rates: MOOCs have reached thousands of learners at a time. However, only few of them have completed the courses [33]. A possible reason for high drop-out rates is the lack of academic guidance for participants to select courses which are suitable for their interest as well as their knowledge level [95].

d) Lack of human interaction: The lack of human interaction is a critical issue in MOOCs, both for learners and professors. In MOOCs, It is not easy to provide direct feedback to a massive number of participants [95]. Moreover, learners in these open courses come from all over the world. They speak English in different levels and have different cultural believes. To address this challenge, integrating social media tools to increase the interaction among MOOC participants can be helpful [96].

[1] https://www.coursera.org/about/terms

e) Certificates: Another important challenge is how to assess the learners and certify their activities. In fact, many learners enrolling in MOOCs are looking for certification to promote their career or complete post-graduate studies. Some MOOC providers already provide certification possibilities, e.g., through test centres.

f) Plagiarism: Scientific integrity is an important factor for the success of online learning, especially MOOCs. The main challenge is how to validate participants' original work and prevent plagiarism? A technical solution can be a plagiarism-detection software but this can be expensive and time-consuming. Peer-reviews can be an option to solve this problem but still quality criteria and indicators are needed to ensure the effectiveness of the peer-review [29][34].

In general, the future of higher education and the potential role of MOOCs require key stakeholders to address these challenges, including questions about the lack of human interaction, plagiarism, certification, completion rates, and innovation beyond traditional learning models. These challenges need to be addressed as the understanding of the technical and pedagogical issues surrounding MOOCs evolve.

B. *Flipped Classrooms*

The flipped classroom is an instance of the VBL model that enables teachers to spend more time in discussing only difficulties, problems, and practical aspects of the learning course [35][97]. In flipped classrooms, learners watch video lectures as homework. Each video lecture comes with a short online quiz as a formative feedback. The class is then an active learning session where the teacher use case studies, labs, games, simulations, or experiments to discuss the concepts presented in the video lecture [6].

Bishop and Verleger [98] define the flipped classroom as interactive learning technique that includes: a) Group learning activities inside the classroom time and b) computer-based learning outside the classroom, as presented in Fig. 7.

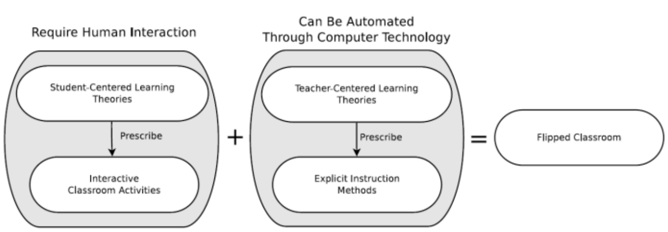


Fig. 7. The Flipped Classroom [98]

We define the flipped classroom as a pedagogical strategy which encompasses several teaching and learning practices split into homework and on-campus activities. Some practices, such as watching video lectures, fall into the home activities. On campus, learners are supposed to conduct their collaborative project or laboratory work and engage in discussions with their peers and teaching staff. On the other hand, teachers plan learning activities, give feedback, and evaluate learners' work. Fig. 8 illustrates the activities in the flipped classroom in more detail.

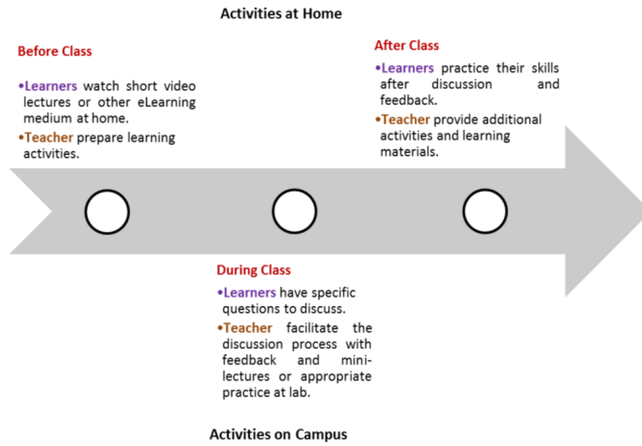


Fig. 8. Flipped classroom activities

1) *The flipped classroom in action*

The flipped classroom model has been successfully applied in the higher education context. This section outlines two case studies that investigated the impact of flipped classrooms on student achievement and engagement.

a) The University of Western Sydney: The flipped classroom has been examined in the first year management accounting unit at the University of Western Sydney in autumn semester 2013. It consists of two main parts: individual instruction outside of the classroom by assigning learners weekly reading of selected chapters (*offline*) and a variety of online activities which are developed to assist students in better understanding the learning topic (*online learning*). The in-class time was devoted to in-depth discussions, problem solving, demonstration, tutorials, and mastering the material through collaborative learning exercises and direct feedback (*face-to-face*). This course had 259 formal learners who were enrolled and have completed the learning course. The most interesting finding was that the majority of learners reported that they have received sufficient instructions and feedback. In addition, they appreciated the quality of the learning material, flexibility, time saving and online activities with the formative feedback. However, the unexpected finding was that some learners did not like the course design because it required learners to complete too many assignments, which was time-consuming [99]. This study, however, did not report on the impact of flipped classroom on learning outcomes.

b) Capital University: Wilson [100] investigated the potential of the flipped classroom model for enhancing learning outcomes in an undergraduate statistics course for social science majors at Capital University in Ohio. The author designed a flipped classroom environment, in which the majority of learning materials were moved out of the classroom and lectures focusing on real-world practices of statistics were given during in-class time. Quizzes were used to measure the learning outcome.

- The quizzes accounted 10% of a learner's overall grade.
- In-class assessments constituted 15% of a learner's grade and were conducted daily.
- Collaborative learning in form of group homework to be completed outside the class accounted 20% of the final grade.
- Final exam accounted 55% of a learner's overall grade.

Learners were asked to evaluate the learning activities that are most helpful for their learning objectives. The students' evaluations of these activities fell into the "somewhat helpful" to "very helpful" categories and resulted in 48% for reading quizzes, 96% for in-class activities, and 91% for group homework. Moreover, the study showed that learners' performance was better in the flipped classroom compared to the traditional class from the previous year. Furthermore, the participants had a higher level of satisfaction with the flipped classroom approach [100]. The limitation of this experiment is that, the number of course participants was only 25 learners.

2) Flipped classroom pros and cons

The flipped classroom approach involves a range of advantages for learners including:

- Flexibility: The flipped classroom helps learners to meet a diverse range of their needs by doing several activities outside the classroom [100][101].
- Student-centred learning: This learning model provides a variety of opportunities for learners to be self-organized and self-independent [8]. Teachers are no longer the only source of knowledge.
- Scaffolding: In flipped classrooms, learning occurs in small learning groups. The teacher's role has been shifting towards facilitating the learning experience by supporting learners in discovering the tools that they need for learning and providing them with the needed guidance and feedback [98][101].

The flipped classroom model, however, suffers from several limitations. These include:

- Lack of motivation: Learners with low motivation or bad learning habits do not pay full attention to out-class activities, such as watching videos, reading materials, or completing assignments at home [102]. As a solution, educators recommended assigning a pre-class quiz on the video material in order to increase the learners' motivation.
- Class structure: Most of the studies that examined flipped classrooms mentioned that the separation between in-class and out-of-class activities is not clearly understood by the learners. Bishop and Verleger [98] recommended that the various learning activities in a flipped classroom should be clearly described at the beginning of the learning process.
- Assessment and feedback: The flipped classroom model emphasizes the role of problem-based learning and project-based learning. This requires creative assessment methods beyond traditional multiple-choice examinations in order to effectively gauge the learner's performance in both individual tasks and group projects [98][100].

VI. Conclusion

In the past few years, there has been an increasing interest in video-based learning (VBL) as a result of popular forms of online education, such as Massive Open Online Courses (MOOCs) and flipped classrooms. VBL is a rich and powerful model used in TEL to improve learning outcomes as well as learner satisfaction. In this paper, we analysed the research on VBL published in 2003-2014. 76 peer reviewed papers were selected in this review. A cognitive mapping approach was used to map the conducted research on VBL into four main dimensions namely, effectiveness, teaching methods, design, and reflection. Most of the reviewed VBL studies still follow a conventional learning approach where the teacher is as the centre of the learning process. Moreover, there is a focus on traditional assessment methods, such as eTests and quizzes.

The following is a summary of the main findings in our study as well as aspects of VBL that need further research, according to each dimension.

A. Effectivness

The analysis of the VBL research showed mixed results in terms of learning outcomes in VBL environments. There is, however, a tendency that users of VBL environments rate interaction and learner satisfaction significantly higher than in traditional classroom environments. Despite these possible advantages, several aspects concerning effectiveness in VBL need further investigation: (1) What are the positive and negative attitudes towards using video lectures? (2) How can VBL motivate learners? (3) How can a MOOC as VBL environment personalize the learning experience for learners? This would enable learners to select the educational resources and the learning style that meet their characteristics best, thus increasing the effectiveness of the learning experience.

B. Teaching Methods

Educators use a broad range of teaching methodologies in VBL environments. These include collaborative learning, micro teaching, video summarization, video assessment, hybrid learning, and student-centered learning. Most of VBL implementations so far still follow a top-down, controlled, teacher-centered, and centralized learning model. Only, 15% of the reviewed research papers describe attempts to implement bottom-up, student-cantered approaches. Additional research is needed to investigate the benefits of new ways of VBL based on new learning concepts such as personal learning environments [103] and network learning [104].

C. Design

Several tools were used in VBL to increase interactivity, collaboration, and learners' satisfaction with the VBL environment. Annotation tools are utilized in searching, highlighting, analysis, retrieval, and providing feedback. To increase interactivity a number of authoring tools were used. These include lecture note synchronization and content summarization tools as well as video libraries and forums. Future research needs to find out how to design more open models of VBL such as MOOCs and flipped classrooms.

D. Reflection

VBL facilitates teachers' as well as learners' reflection. Our study showed that teachers prefer to reflect on their teaching performance with colleagues rather than individually. And, learners think that videos have the potential to be used as a reflection tool. Future research is needed to investigate how learning analytics can help to better understand and improve reflection and awareness in VBL environments, such as MOOCs.

MOOCs and flipped classrooms represent promising implementations of the VBL model. Further work is still needed to investigate how to personalize the learning activities in these environments [103]. Learners are learning at different paces and have different aptitudes. Thus, curriculum, pedagogy, and assessment should be customized in order to fit each learner needs and perspectives [104].

REFERENCES

[1] A. M. F. Yousef, M. A. Chatti, and U. Schroeder, "Video-Based Learning: A Critical Analysis of The Research Published in 2003-2013 and Future Visions", in eLmL 2014, The Sixth International Conference on Mobile, Hybrid, and On-line Learning 2014, pp. 112-119.

[2] M. Merkt, S. Weigand, A. Heier, and S. Schwan, (2011). Learning with videos vs. learning with print: The role of interactive features. Learning and Instruction, 21, pp. 687-704.

[3] M. Colasante, "Using video annotation to reflect on and evaluate physical education pre-service teaching practice," Australasian Journal of Educational Technology, 27(1), 2011, pp.66-88.

[4] M. G. Sherin, and E. A. van Es, "Effects of video club participation on teachers' professional vision," Journal of Teacher Education, 60(1), 2009, pp. 20-37.

[5] D. Zhang, L. Zhou, R. O. Briggs, and J. F. Nunamaker Jr, "Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness," Information & Management, 43, 2006, pp. 15-27.

[6] B. Calandra, L. Brantley-Dias, and M. Dias, "Using digital video for professional development in urban schools: A preservice teacher's experience with reflection," Journal of Computing in Teacher Education, 22(4), 2006, 137.

[7] C. I. Hovland, A. A. Lumsdaine, and F. D. Sheffield, Experiments on mass communication. Studies in social psychology in World War II, 3, 1949.

[8] R. Santagata, "Designing Video-Based Professional Development for Mathematics Teachers in Low-Performing Schools," Journal of Teacher Education, 60 (1), 2009, pp. 38-51.

[9] C. Snelson, K. Rice, and C. Wyzard, "Research priorities for YouTube and video-sharing technologies: A Delphi study," British Journal of Educational Technology, 43 (1), 2012, pp. 119-129.

[10] J. McCarthy, "Blended learning environments: Using social networking sites to enhance the first year experience," Australasian Journal of Educational Technology, 26(6), 2010, pp.729-740.

[11] W. Tuong, E. R. Larsen, and A. W. Armstrong, "Videos to influence: a systematic review of effectiveness of video-based education in modifying health behaviors" Journal of behavioral medicine, 37(2), 2014, pp. 218-233.

[12] A. D. Greenberg, and J. Zanetis, "The impact of broadcast and streaming video in education," Ainhouse Research, CISCO, 2012.

[13] R. Borgo, et al., "State of the art report on video-based graphics and video visualization," Computer Graphics Forum, 31(8), 2012, pp. 2450–2477.

[14] T. Tripp, and P. Rich, "Using video to analyze one's own teaching," British Journal of Educational Technology, 43(4), 2012, pp. 678-704.

[15] A. Fink, Conducting research literature reviews: from the internet to paper (2nd ed.). Thousand Oaks, California: Sage Publications, 2005.

[16] S. McDonald, K. Daniels, and C. Harris, "Cognitive mapping in organizational research," In C.Cassell and G.Symon (Eds.), Essential Guide to Qualitative Methods in Organizational Research, Sage Publications, 2004, pp. 73–85.

[17] N. Fearing, S. Bachman, M. Holzman, D. Scott, and M. Brunt, "Evaluation of a video-based curriculum for laparoscopic biliary surgery: a pilot study from the SAGES MIS Web Learning Center," Surgical Endoscopy, 24, 2010, pp. 3141-3143.

[18] R. Kay, and J. Edward, "Examining the Use of Worked Example Video Podcasts in Middle School Mathematics Classrooms: A Formative Analysis," Canadian Journal of Learning and Technology, 38 (2), 2012.

[19] T. Balslev, W. S. de Grave, A. M. M. Muijtjens, and A. J. J. Scherpbier, "Comparison of a text and video case in a postgraduate problem-based learning format," Medical Education, Vol. 39(11), 2005, pp. 1086-1092.

[20] C. C. Lin, and Y. F. Tseng, “Videos and animations for vocabulary learning a study on difficult words. Turkish Online Journal of Educational Technology, 11 (4), 2012, pp. 346-355
[21] C. K. Hsu, G. J. Hwang, Y. T. Chang, and C. K. Chang, “Effects of Video Caption Modes on English Listening Comprehension and Vocabulary Acquisition Using Handheld Devices,” Educational Technology & Society, 16 (1), 2013, pp. 403–414.
[22] N. B. Kersting, K. B. Givvin, B. J. Thompson, R. Santagata, and J. W. Stigler, “Measuring Usable Knowledge: Teachers’ Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning,” American Educational Research Journal, 49 (3), 2012, pp. 568-589.
[23] S. Kuter, G. Z. Altinay, and A. F. Altinay, “Examination of Co-construction of Knowledge in Videotaped Simulated Instruction,” Educational Technology & Society, 15 (1), 2012, pp. 174–184.
[24] F. Donkor, “The Comparative Instructional Effectiveness of Print-Based and Video-Based Instructional Materials for Teaching Practical Skills at a Distance” International Review of Research in Open and Distance Learning, 11(1), 2010.
[25] K. R. Comeaux, Cognitive Memory Effects on Non-Linear Video-Based Learning. Louisiana State University, Master thesis, 2005.
[26] R. Lindgren, R. Pea, S. Lewis, and J. Rosen, “Learning from digital video: An exploration of how interactions affect outcome. CSCL, 2007, pp. 447-449.
[27] H. H. Chuang, and M. H. Rosenbusch, “Use of digital video technology in an elementary school foreign language methods course,” British Journal of Educational Technology, 36(5), 2005, pp. 869-880.
[28] M. Giannakos, K. Chorianopoulos, M. Ronchetti, P. Szegedi, and S. Teasley, “Video-Based Learning and Open Online Courses” International Journal of Emerging Technologies in Learning (iJET), 9(1), pp 4-7.
[29] A. M. F. Yousef, M. A. Chatti, U. Schroeder, and M. Wosnitza “What Drives a Successful MOOC? An Empirical Examination of Criteria to Assure Design Quality of MOOCs” in ICALT 2014, The 14th IEEE International Conference on Advanced Learning Technologies, 2014, pp. 44-48.
[30] J. S. DeLoache, and Korac, N., “Special Section: children and television Video-based learning by very young children,” Developmental Science, 6(3), 2003, pp. 245-246.
[31] P. Hakkarainen, and K. Vapalahti, “Meaningful Learning through Video-Supported Forum-Theater,” International Journal of Teaching and Learning in Higher Education, 23 (3), 2011, pp. 314-328.
[32] W. Shen, “Using Video Recording System to Improve Student Performance in High-Fidelity Simulation” In Frontier and Future Development of Information Technology in Medicine and Education, 2014, pp. 1753-1757. Springer Netherlands.
[33] J. M. Muhirwa, “Teaching and Learning Against all Odds: A Video-Based Study of Learner-to-Instructor Interaction in International Distance Education,” International Review of Research in Open and Distance Learning, 10 (4), 2009.
[34] A. M. F. Yousef, M. A. Chatti, U. Schroeder, M. Wosnitza and H. Jakobs, “MOOCs - A Review of the State-of-the-Art” in CSEDU 2014, The Sixth International Conference on Computer Supported Education, 2014, Vol. 3, pp. 9-20.
[35] A. R. Montazemi, “The Effect of Video Presentation in a CBT Environment,” Educational Technology & Society, 9 (4), 2006, pp. 123-138.
[36] P. N., Smyrni, and C. Nikopoulos, “Evaluating the impact of video-based versus traditional lectures on student learning,” Educational Research, 1(8), 2010, pp. 304-311.
[37] R. Verleur, A. Heuvelman, and P. W. Verhagen, “Trigger videos on the Web: Impact of audiovisual design,” British Journal of Educational Technology, 42 (4), 2011, pp. 573-582.
[38] E. Dale, Audiovisual Methods in Teaching, 1969, NY: Dryden Press.
[39] C. Zahn, F. Hesse, M. Finke, R. Pea, M. Mills, and J. Rosen, “Advanced video technologies to support collaborative learning in school education and beyond," In D. Suthers & T. Koschmann (Eds.), Proceedings of the International CSCL-05 Conference, Taipei, Taiwan. Mahwah, NJ: Erlbaum Associates, 2005.
[40] J. Goulah, “Village Voices, Global Visions: Digital Video as a Transformative Foreign Language Learning Tool,” Foreign Language Annals, 40 (1), 2007, pp. 62-76.
[41] D. Hung, S. C. Tan, W. S. Cheung, and C. Hu, “Supporting Problem Solving with Case-Stories Learning Scenario and Video-based Collaborative Learning Technology,” Educational Technology & Society, 7 (2), 2004, pp. 120-128.
[42] D. Petko, K. Reusser, C. Noetzli, K. Krammer, and I. Hugener, “Collaborative video based teacher training in a virtual learning environment,” 10th European Conference for Research on Learning and Instruction, 2003.
[43] H. j. Choi, and S. D. Johnson, “The Effect of Context-Based Video Instruction on Learning and Motivation in Online Courses,” The American Journal of distance education, 19(4), 2005, pp. 215–227.
[44] J. M. Wiecha, R. Gramling, P. Joachim, and H. Vanderschmidt, “Collaborative e-learning using streaming video and asynchronous discussion boards to teach the cognitive foundation of medical interviewing: a case study,” Journal of Medical Internet Research, 5(2), 2003.
[45] R. Pea, and R. Lindgren, “Video Collaboratories for Research and Education: An Analysis of Collaboration Design Patterns,” IEEE Transactions on Learning Technologies, 1, (4), 2008.
[46] E. H. van Zee, “Using Web-Based" Snapshots of Practice" to Explore Science Learning and Teaching in a Course for Prospective Teachers Issues,” Teacher Education, 14(1), 2005, pp.63-79.
[47] J. Finlay, et al., “Video-based learning objects for teaching human-computer interaction at different levels” 9th Annual Conference of the Subject Centre for Information and Computer Sciences, 2008, pp. 194-198.
[48] A. W. Woodruff, “Using video based instruction to teach art to students with autism spectrum disorder” Theses and Dissertations, Art & Visual Studies. Paper 5, 2014, http://uknowledge.uky.edu/art_etds/5
[49] L. B. Dias, M. Dias, J. K. Frisch, and G. Rushton, “The role of digital video and critical incident analysis in learning to teach science,” American Educational Research Association. New York City, Annual, Meeting, 2008, pp. 24-28.
[50] T. Seidel, G. Blomberg, and A. Renkl, “Instructional strategies for using video in teacher education,” Teaching and Teacher Education, 34, 2013, pp. 56-65.

[51] C. H. Fu, K. T. Wang, S. C. Cheng, and T. W. Hou, “Building Video Concordancer Supported English Online Learning Exemplification,” In Y.-M.R. Huang et al. (Eds.): PCM LNCS 5353, 2008, pp. 731-737.
[52] W. H. Chang, Y. C. Wu, and J. C. Yang, “Webpage-based and video summarization-based learning platform for online multimedia learning,” In Edutainment Technologies. Educational Games and Virtual Reality/Augmented Reality Applications, 2011, pp. 355-362.
[53] W. H. Chang, J. C. Yang, and Y.C. Wu, “A Keyword-based Video Summarization Learning Platform with Multimodal Surrogates,” 11th IEEE International Conference on Advanced Learning Technologies, 2011, pp. 37-41.
[54] F. Donkor, “Assessment of learner acceptance and satisfaction with video-based instructional materials for teaching practical skills at a distance,” The International Review of Research in Open and Distance Learning, 12(5), 2011, pp.74-92.
[55] Q. Qiao, and P. A. Beling, “Classroom Video Assessment and Retrieval via Multiple Instance Learning,” In G. Biswas et al. (Eds.): AIED 2011, pp. 272–279.
[56] G. Karlsen, “From thumbs to fingertips: introducing networked digital video to online learning,” journal of distance education, 20(2), 2005, pp. 85-87.
[57] R. J. Chenail, “YouTube as a qualitative research asset: Reviewing user generated videos as learning resources,” The Qualitative Report, 16(1), 2011, pp. 229-235.
[58] Y. Pang, “Improving hybrid learning of physical education by video review,” In Advances in Web-Based Learning–ICWL, 2010, pp. 230-239.
[59] R. C. Shih, “Blended learning using video-based blogs: Public speaking for English as a second language student,” Australasian Journal of Educational Technology, 26(6), 2010, pp.883-897.
[60] Y. Kırkgöz, “A Blended Learning Study on Implementing Video Recorded Speaking Tasks in Task-Based Classroom Instruction,” TOJET, 10(4), 2011.
[61] J. Gainsburg, “Creating Effective Video to Promote Student-Centered Teaching,” Teacher Education Quarterly, 2009, pp. 163-178.
[62] R. Smyth, “Enhancing learner–learner interaction using video communications in higher education: Implications from theorising about a new model,” British Journal of Educational Technology, 42(1), 2011, pp. 113-127.
[63] P. J. Rich, and M. Hannafin, “Video Annotation Tools Technologies to Scaffold, Structure, and Transform Teacher Reflection,” Journal of Teacher Education, 60(1), 2009, pp. 52-67.
[64] K. Khurana, and M. B. Chandak, “Study of Various Video Annotation Techniques,” International Journal of Advanced Research in Computer and Communication Engineering, 2(1), 2013.
[65] M. Wang, X. S. Hue, Y. Song, R. Hong, and L. R. Dai, “Lazy Learning Based Efficient Video Annotation,” In Multimedia and Expo, 2007 IEEE International Conference, 2007, pp. 607-610.
[66] R. Schroeter, J. Hunter, and D. Kosovic, “Vannotea: A collaborative video indexing, annotation and discussion system for broadband networks,” In Knowledge capture ACM Press Association for Computing Machinery, 2003, pp. 1-8.
[67] S., Chunwijitra, A. J. Berena, H. Okada, and H. Ueno, “Authoring tool for video-based content on WebELS learning system to support higher education,” In Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference, 2012, pp. 317-322.
[68] J. C. Yang, Y. T. Huang, C. C. Tsai, C. I. Chung, and Y. C. Wu, “An automatic multimedia content summarization system for video recommendation,” Educational Technology & Society, 12 (1), 2009, pp. 49–61.
[69] P. Wouters, H. K. Tabbers, and F. Paas, “Interactivity in video-based models,” Educational Psychology Review, 19(3), 2007, pp. 327-342.
[70] M. Milrad, P. Rossmanith, and M. Scholz, “Implementing an educational digital video library using MPEG-4, SMIL and Web Technologies,” Educational Technology & Society, 8 (4), 2005, pp. 120-127.
[71] H. J. So, H. Lossman, W. Y. Lim, and M. J. Jacobson, “Designing an online video based platform for teacher learning in Singapore,” Australasian Journal of Educational Technology, 25(3), 2009, pp. 440-457.
[72] H. T. D. Huang, and S. T. A. Hung, “Exploring the utility of a video ‐ based online EFL discussion forum,” British Journal of Educational Technology, 44(3), 2013, pp. 90-94.
[73] J. L. Loo, L. T. Ngo, C. K. Hennesy, B. D. Quigley, and J. McKenzie, “Embedding Video-based Learning Modules for Library Research Methods in an Online Graduate Engineering Degree Program. In 121st ASEE Annual Conference & Exposition Proceedings, 2014.
[74] H. Borko, J. Jacobs, N. Seago, and C. Mangram, “Facilitating video-based professional development: Planning and orchestrating productive discussions” In Transforming Mathematics Instruction, 2014, pp. 259-281. Springer International Publishing.
[75] S. C. Kong, R. H. Shroff, and H. K. Hung, “A web enabled video system for self-reflection by student teachers using a guiding framework,” Australasian Journal of Educational Technology, 25(4), 2009, pp. 544-558.
[76] E. A. van Es, and M. G. Sherin, “How different video club designs support Teachers in learning to notice,” Journal of computing in teacher education, 22(4), 2006, pp.125-135.
[77] B. Calandra, L. Brantley-Dias, J. K Lee, and D. L. Fox, “Using video editing to cultivate novice teachers’ practice,” Journal of Research on Technology in Education, 42(1), 2009, pp. 73–94.
[78] C. C. Wu, and H.C. Kao, “Streaming Videos in Peer Assessment to Support Training Pre-service Teachers,” Educational Technology & Society, 11 (1), 2008, pp. 45-55.
[79] B. Calandra, R. Gurvitch, and J. Lund, “An Exploratory Study of Digital Video Editing as a Tool for Teacher Preparation,” Journal of Technology and Teacher Education, 16(2), 2008, pp. 137-153.
[80] R. Santagata, and G. Angelici, “Studying the impact of the lesson analysis framework on preservice teachers’ abilities to reflect on videos of classroom teaching,” Journal of Teacher Education, 61(4), 2010, pp. 339-349.
[81] R. Santagata, and J. Guarino, “Using video to teach future teachers to learn from teaching,” Zdm, 43(1), 2011, pp. 133-145.
[82] C. P. Halter, “The reflective lens: The effects of video analysis on preservice teacher development,” Dissertation Abstracts International, 67, 03, 2006, (UMI No. 3211280).
[83] G. Blomberg, M. G. Sherin, A. Renkl, I. Glogger, and T. Seidel, “Understanding video as a tool for teacher education:

investigating instructional strategies to promote reflection" Instructional Science, 42(3), 2014, pp. 443-463.

[84] Y. H. Cho, and Y. Huang, "Exploring the links between pre-service teachers' beliefs and video-based reflection in wikis" Computers in Human Behavior, 35, 2014, pp. 39-53.

[85] H. Odhabi, and L. N. McCaleb, "Video recording lectures: Student and professor perspectives," British Journal of Educational Technology, 42(2), 2011, pp.327-336.

[86] B. Dalgarno, G. Kennedy, and A. Merritt, "Connecting student learning at university with professional practice using rich media in practice-based curricula" In Curriculum models for the 21st century, 2014, pp. 213-233. Springer New York.

[87] R. Schuwer, and B. Janssen, "Trends in business models for open educational resources and open education" In Trend report: open educational resources 2013, pp. 60-66.

[88] S. Brown, "Back to the future with MOOCs?" in ICICTE 2013 Proceedings, pp. 237-246.

[89] T. R. Liyanagunawardena, S., Williams, and A. Adams, "The impact and reach of MOOCs: A developing countries' perspective. eLearning Papers, ISSN: 1887-1542, Issue 33, 2013.

[90] G. Siemens, "MOOCs are really a platform. eLearnspace. Retrieved March 15, 2013 from http://www.elearnspace.org/blog/2012/07/25/moocs-are-really-a-platform/

[91] P. Hill, "Four Barriers That MOOCs Must Overcome To Build a Sustainable Model" Retrieved 1st August, 2014 from http://mfeldstein.com/four-barriers-that-moocs-must-overcome-to-become-sustainable-model/

[92] G. Siemens, "Connectivism: A learning theory for the digital age," International Journal of Instructional Technology and Distance Learning, 2(1), 2005, pp. 3-10.

[93] S. J. Daniel, "Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility," Journal of Interactive Media in Education (JIME).

[94] F. M. Hollands, and D. Tirthali, "MOOCs: expectations and reality", Center for Benefit-Cost Studies of Education Teachers College, Columbia University, 2014.

[95] B. Dasarathy, K. Sullivan, D. C. Schmidt, D. H. Fisher, and A. Porter, "The past, present, and future of MOOCs and their relevance to software engineering" In Proceedings of the on Future of Software Engineering, 2014, pp. 212-224. ACM.

[96] A. M. F. Yousef and G. Rößling, "How to Design Good Educational Blogs in LMS?" in CSEDU 2013, The Sixth International Conference on Computer Supported Education, 2013, pp. 70-75.

[97] B. Tucker, "The flipped classroom," Education Next, 12(1), 2012, pp.82-83.

[98] J. L. Bishop, and M. A. Verleger, "The flipped classroom: A survey of the research" In ASEE National Conference Proceedings, Atlanta, GA, 2013.

[99] X. Du, and S. Taylor, "Flipped classroom in first year management accounting unit – a case study" In H. Carter, M. Gosper and J. Hedberg (Eds.), Electric Dreams. Proceedings 30th ascilite 2013 Sydney, pp.252-256.

[100] S. G. Wilson, "The flipped classroom: A method to address the challenges of an undergraduate statistics course". Teaching of Psychology, 2013, 40, 193-199.

[101] C. F. Herreid, and N. A. Schiller, "Case studies and the flipped classroom," Journal of College Science Teaching, 42(5), 2013, pp.62-66.

[102] A. Wallace, "Social learning platforms and the flipped classroom" In ICEEE, 2013 Second International Conference on e-Learning and e-Technologies in Education, 2013, pp. 198-200. IEEE.

[103] G. Attwell, "Personal Learning Environments - the future of eLearning?," eLearning Papers, vol. 2, (1) 2007.

[104] M. A. Chatti, Personalization in Technology Enhanced Learning: A Social Software Perspective. Shaker Verlag, November 2010, Dissertation, RWTH Aachen University.

The Inertia of the Status Quo:

A Change Management Analysis of Technological Innovation

Anthony P. Glascock
Department of Culture and Communication
Drexel University
Philadelphia, PA 19104, USA
glascock@drexel.edu

***Abstract*—A reanalysis of the lack of acceptance of an ehealth data records system utilizing the tenets of change management reveals fundamental issues surrounding the challenges facing the introduction of new technologies in health care. Even though the design of the system was largely driven by the end-users themselves, ultimately these same individuals rejected the system once a series of pilot studies ended. The findings from these studies suggest that the key element for the successful introduction of a new technology is the necessity to overcome the inertia of people not wanting to change how they do their jobs, not the technology itself.**

***Keywords*—acceptance of innovative technologies, change management, an ehealth data records system**

I. INTRODUCTION

This article is an example of the reason that one presents papers at professional meetings: being pushed into expanding one's analysis because of questions being asked that one cannot answer. In this case, it was a question posed in Barcelona at eTELEMED 2014: "isn't your work an example of the process of change management?" Although unable to answer the question at the time, the most cursory research into change management allowed the answer to be an emphatic yes. However, that simple answer led to a rethinking of the process on which the eTELEMED paper, "If they designed it, why don't they want it? The lack of acceptance of an ehealth data records system" [1], was based with the result being a much expanded analysis and a greater understanding of what on the surface was a largely inexplicable failure of adoption of a wireless ehealth data records system for use in the delivery of care and services in the residences of at-risk elderly individuals.

This particular wireless ehealth data records system was developed as a response to the desire on the part of governments, in most industrial countries, to introduce electronic health care records systems and to the dramatic increase in the amount of care and services being delivered in residences. Although the use of electronic health records systems has been one of the major initiatives in the United States as well as in most European countries for much of the 21st century, the actual adoption of ehealth records systems has been much slower than anticipated. This is the case even though the United States "government set aside $27 billion for an incentive program that encourages hospitals and providers to adopt electronic records systems" [2]. Although there is some debate over the reliability of ehealth records systems [3], there appears to be little doubt that the greater the ease of storage, access and cost savings will eventually achieve close to universal usage within institutions and among physicians [4][5]. If the pace of adoption in hospitals and physician practices has been slow, the pace of adoption of ehealth data systems for use to chronicle care and services in the home has been even slower. There is little debate that the delivery of care and services in the home has increased significantly and will accelerate at an ever more rapid pace over the coming several decades because of two basic factors: demography and cost. The demographic trends are well known: dramatic increase in the number of elderly, especially the oldest-old; a concomitant increase in chronic diseases associated with aging; and a decline in the number of family members who can provide care. Cost projections are similarly daunting: increasing cost for care delivery within the institutional setting; insufficient number of institutions and insufficient funds to build the large number of additional institutions; and prohibitive costs to government to provide care and services through current care delivery models.

It is generally agreed that the only way to meet the increasing needs brought about by the demographic trends, while at the same time not bankrupting national treasuries, is to provide more care and services in the home [6]. As more care and services are being delivered in the home, several issues have emerged that raise serious concerns. In the first place, the care being delivered had steadily become more extensive. Whereas a decade ago rehabilitation after a serious illness or accident would have been undertaken in a specialized facility; presently many of these services are being provided in the home on an outpatient basis. But rehabilitation is just one of an escalating number of care services being provided in the home: nutritional counseling; wound care; psychological therapy; and medication adherence, to name several of the major ones. Additionally, the range of products and non-care services supplied to individuals in their own homes has increased significantly: oxygen; specialized beds; monitoring; meals; housekeeping; shopping; companion services. As the care and services have multiplied, so have the number of people providing the services. These companies/agencies are, of course, in addition to any services provided by informal carers. Thus, the need for a means of recording and tracking the care and services provided in the home are essentially the same as for institutions: increased reliability; better coordination; appropriate level of care; and cost savings.

As more care and services are delivered by more people representing different companies and agencies, just keeping track of who is providing what becomes increasingly difficult, especially if the individual receiving the care lives alone in her home and is experiencing cognitive decline or other impairments. Scheduling of visits and deliveries, ensuring the correct product or service, avoiding duplication all become difficult if records are scattered among various agencies, companies and individuals and are rarely, if ever, shared. It is also extremely difficult to evaluate if the care and services are having the desired impact on the individual if there is no systematic way to track the outcomes of the care and services. The lack of systematic and comprehensive records also makes it difficult for other care providers to make informed care decisions, since the reliance on the patient to remember specifics about the care and services in the home has proven to be suspect at best. In addition, if ehealth records of care and services in the home do not exist, it is obvious that they cannot be linked with the records that have been created in the hospital and the physician practice. Finally, even though delivering care and services in the home is more economical than in institutions, it still costs money and someone has to pay for it. As a result, from the point of view of the client receiving the care and services, as well as the insurance company and the government, there is a discernible need to track the care and services to ensure that what is paid for is provided and that everyone was paid appropriately.

The problem is that there has been a general reluctance on the part of many care organizations to adopt such systems. Research on various ehealth technologies indicate that there are many explanations for this reluctance, among the most important being: unwillingness of care providers to change how they do their work; fear of a loss of professional autonomy on the part of care providers; insufficient training; inadequate investment in IT infrastructure; and a lack of a supportive culture within the care organizations [7][8][9][10]. The remainder of this paper reports on a series of problems encountered when implementing a particular ehealth data records system and uses ideas from the field of change management to offer an explanation for the system's lack of acceptance.

II. DESIGN PHILOSOPHY

The eTELEMED paper reported on the development of a wireless ehealth data records system: the Home Care Informatics System (HCIS) that attempted to fulfill the need for a means of recording and sharing details on the ever increasing care and services delivered to elderly individuals in their own residences. The system was developed and tested over an eight year period in a total of seven care organizations in three countries. The article's main focus was on the lack of acceptance of the HCIS on the part of the care providers in all seven locations, even though these very same carers had played a major role in the design of the system. Several possible reasons for the lack of acceptance were presented and, although they were consistent with the findings overall, the reasons were fairly unconvincing [1]. So it was not surprising that someone listening to the paper would suggest a different explanation for the lack of acceptance of the HCIS, specifically one rooted in the theory of change management.

However, before reinterpreting the lack of acceptance of the HCIS based upon the theory of change management, some background information is required as to the approach taken to its design and implementation. It was decided from the very beginning that the design of the HCIS would be driven by the end-users—the carers who would actually use the system to record and share information, e.g., visiting nurses, geriatric social workers, care managers. This decision was based upon previous experience gained from the development of another innovative ehealth project [11]. Even though seven patents were issued, a start-up company invested heavily in its commercialization and the company and patents were eventually sold to General Electric, acceptance of the product in the market-place was less than anticipated. One explanation for this lack of success in the market-place was that almost all design decisions were made by management, software developers and marketing personnel with little, if any, input from the people who would be using the system in the delivery of care. Thus, there was little "buy-in" on the part of the end-users and a general reluctance on the part of these care providers to make changes in their actions in order to incorporate the system into their normal care delivery model. This was the case even though the upper administrators at the care organizations were, for the most part, enthusiastic champions of the technology. The problem was that because the decision to use the system was driven by these administrators without any input from the people who actually delivered the care, this high-level enthusiasm was never duplicated among the end-users.

Based upon this experience, it was decided that in the development of the Home Care Informatics System as much control as possible over the design of the system would be ceded to those individuals who were actually going to be using the system to aid in care delivery. Not only did this approach appear to make sense because of the previous experience, but from the very beginning, and throughout the design process, the care providers at all seven care organizations were anything but shy in making suggestions about both the design of the system and how it should be used. In fact, as will be illustrated in the next section, it could be argued that the final iteration was so unlike the one originally conceived that it was actually a different product. Whether this was the case or not, the final design certainly reflected the input of the end-users much more than the developers.

In the next section, the actual design process is described, with emphasis placed on the role played by the care providers; in section four the reasons presented in the previous paper for the lack of acceptance on the part of the carers are briefly presented; in section five the reasons for the lack of acceptance are analyzed based upon the tenets of change management; in section six the two sets of explanations are compared; and in the last section future implications for the development and implementation of electronic records systems, as well as other innovative ehealth technologies, are explored.

III. Three Stages of Development

As stated previously, the development of the HCIS spanned an eight year period, involved seven care organizations and took place in three countries—the United States, the United Kingdom and the Netherlands. The first iteration of the HCIS was actually a research tool designed to work in conjunction with a behavioral monitoring system. This prototype was designed to systematically record the alerts generated by the monitoring system, as well as the care delivered based on the alerts and there was no thought at this stage of developing an electronic health records system. It was the reaction of the end-users, in this case geriatric social workers that began to shift the research tool to an ehealth application.

A. Stage 1—2006-2007

This process took place during a pilot study at Selfhelp Community Services, Inc. in Queens, New York and involved eleven geriatric social workers who provided care management services to over 200 residents [12][13][14]. Twenty-seven of these clients agreed to have the system installed for a six month period and to have the social workers use the resultant data in care management decisions. The problem was that there was no way to systematically collect information on the care actions that they took in response to an alert. The instrument created was labeled the **TAO**: **Trigger**, the system's alert; **Action**, the care action taken by the social worker in response to the alert; and **Outcome**, the health or care outcome brought about by the care action. A brief example illustrates the initial design of the TAO:

> **Trigger:** The system sends an alert to the geriatric social worker indicating an increase in overnight toileting for a particular client;
> **Action:** The social worker phones the client to inquire about the client's behavior;
> **Outcome:** Finding out that the client was frequently in the bathroom because of stomach flu, the social worker contacts the client's physician to obtain a prescription for medication.

Initially, the social workers filled out a paper form with the relevant information, which was then entered into a computer data base in order to allow analysis. This worked well for about two weeks, but at the first care review meeting several social workers complained that filling out the paper version of the TAO was time-consuming. They suggested that, since they were on their desk-top computers throughout the day, that it would be easier for them to enter the information directly into the spreadsheet themselves. In response to these suggestions, a computerized web-based version of the TAO was created and this new version became the first design change driven by its users.

This new Web-TAO form took about five minutes to fill out, could be easily shared with others and, most importantly, could be updated as more actions and outcomes occurred. In the short run, this last feature proved beneficial for the social workers as they could quickly and almost effortlessly update the Web-TAO records for individual clients. In the long run, the need to have an update capability proved essential in the development of the HCIS. This is because, although the alert is a discrete event, care actions and health outcomes are not discrete, but instead roll out over time. The previous example of the TAO narrative has all three elements as discrete events—one Trigger, one Action, one Outcome—and this example corresponds to approximately 40% of the TAOs. However, a majority of the TAOs corresponded more to the following example:

> **Trigger:** The system sends an alert to the geriatric social worker indicating an increase in overnight toileting for a particular client;
> **Action:** The social worker phones the client to inquire about the client's behavior;
> **Outcome:** Finding out that the client was frequently in the bathroom because of a stomach flu, the social worker contacts the client's physician to obtain a prescription for medication;
> **Second Action:** The social worker phones the client's daughter to report that her mother has the flu;
> **Third Action:** Daughter visits her mother the next day finding out that her mother is no better**;**
> **Fourth Action:** Daughter phones social worker reporting on mother's condition;
> **Fifth Action:** Social worker visits client, determines that she is dehydrated, phones physician;
> **Sixth Action, Second Outcome:** Physician decides to have client admitted to hospital;
> **Seventh Action, Third Outcome:** Client is discharged after two days in hospital.

All of the above actions and outcomes were the result of the single alert and could now be entered into the Web-TAO as the events rolled out in real time. As a record of care provided and outcomes generated, the Web-TAO proved extremely helpful to the geriatric social workers as they could more systematically track the progression of care and outcomes. However, the realization of how multiple care actions and outcomes could be gathered together in a single record proved invaluable for the future development of the informatics system that eventually became the HCIS.

Once this alteration was made, the carers had other suggestions: add auto-populated fields; use check-boxes whenever possible; allow for easier follow-up entries; and allow access to individual records by other social workers and supervisors. At the first care review meeting, after these changes were made, two issues that would drive much of the development of the TAO surfaced. Since the objective for these meetings was to review what had happened to each of the clients over the previous month in order to assess how the monitoring system had impacted the delivery of care, it was not surprising that the TAOs were the focus of the discussions, but it was surprising how the social workers utilized the TAOs. They placed the TAOs for each of the clients together and then worked their way chronologically through the TAOs. By their actions, the social workers were constructing an on-going record for each of the clients by putting TAOs for the particular client together into a single "pile". This "piling up" was the first care record and would drive much of the future development of the ehealth system.

The second issue raised by the geriatric social workers concerned the ability of the Web-TAO in the evaluation of performance. For the supervisors, the Web-TAO provided an objective basis on which to evaluate the work performed by the social workers; for the social workers, the Web-TAO allowed supervisors to question their actions and professional conduct using information that had not been available previously. These issues were not resolved before the study ended, but, as discussed subsequently, it remained a vexing problem for the future development of the ehealth system.

B. Stage 2—2007-2008

As the Selfhelp pilot was ending, a pilot study in London was getting underway. Unlike the Selfhelp study in which all clients lived independently and had their care managed by a single care organization, the London study involved several residential types and more than one care organization. All residents lived in Southwark, an area of Central London south of the Thames, and were provided services from one of three care organizations—Southwark Falls, Oasis and Hyde Housing—all of which operated under the broad umbrella of the Southwark Local Authority. Thus, the work undertaken by "carers" in these organizations was much more coordinated than would be found in the independent organizations in the United States. However, even though these organizations were "independent" and served distinct populations, for this discussion it makes sense to view them as a single entity, the Southwark Study, and to aggregate their 97 clients.

Based on the development work undertaken at Selfhelp, the Southwark Study began with a fully operational Web-TAO that had the ability to easily update a report as care actions and outcomes rolled out over time. Within the first six weeks of the study, it became apparent from the analysis of the material being entered into the Web-TAO that the carers were using the system much differently than the social workers at Selfhelp. This was primarily due to the fact that the culture at Southwark was extremely collaborative and, although particular carers had primary responsibility for specific clients, all carers engaged with all clients in some fashion, and thus, the Web-TAO was conceived as a tool to allow for easier sharing of information among all carers rather than just a record of responses to triggering alerts. Therefore, the ability for all members of the care team to not only view the information, but to contribute to the information stream became paramount. The cultural imperative to share and contribute to the information of clients resulted in a modification that allowed for much longer narratives to be entered into the system which, as a result, took on the appearance of "blogs" in which numerous carers listed their actions and the subsequent outcomes for particular clients. Fig. 1 is an example of a typical "blog" for a single client.

Trigger	-Wake up
Actions	-Phoned client -Visited client -Spoke to care professional Care Coordinator, GP -Contacted other person Spoke to OASIS Support Worker -Other action taken: Support worker spoke to client face to face, spoke to the surgery concerning the medical health of the client, GP to call back.
Outcomes	10/25 10:04) Client has been complaining of hip pain for the last two days but on prompting to attend GP surgery or to have home visit, she declined. When support worker visited, she found out that the client appeared unwell and movement was very slow.. Client had not eaten since last night so Support Worker prompted nutrition and medication and asked the client's permission to call GP to look at her hip. Client has agreed and a call has been made to book for a GP to examine the client. GP is aware of the needs, we have left a telephone message on the Next of kin's mobile number and Care Coordinator has been informed. GP visited and assessed Mrs B yesterday. She prescribed paracetamol for pain relief as it was found out that the arthritis in her hip was causing her so much pain. Client is still not able to get out of bed earlier but we hope that the pain will subside. Plan: Monitor the effect and report to GP as the condition changes. Care Coordinator to note. (10/26 10:47) GP stated that Mrs B appeared confused when examined and advised her to increase her fluid intake and contact the Specialist Mental Health Care Coordinator to assess the situation. (10/30 11:42) Following GP's prescription for pain relief, QuietCare showed that Mrs B visited the bathroom at 4.53 am and got out of the bedroom at 9.59 am which was unusual from recent data. She has been on pain relief since Friday 26th October and there appears to be a marked improvement in her health.

Figure. 1. Web-TAO blog narrative

On the surface, this change appeared to be trivial, but in actuality it altered much of the design of the structure of the Web-TAO going forward. The Web-TAO had already mutated from a research tool to a care provision tool that tracked responses to the system's alerts, and now it had transformed again from a limited record of what transpired when an alert occurred, to a more comprehensive ehealth record of all care being delivered to a specific client over time. Fig. 1 not only shows the comprehensive nature of the information recorded, but also illustrates how many carers became involved in contributing care for this client.

The members of the newly formed Smart Team had other suggestions for the Web-TAO. One was to be able to send the "blog" to a client's physician prior to an appointment in order for the physician to have all relevant care information. This required the creation of a new security function that limited who could send and what could be sent to individuals outside the Southwark Smart Team. A second suggestion was to allow the "blogs" to be sorted by alert, particular carer, type of care actions and date of entry. Although technically not a complex undertaking, the challenge was to understand the uses to be made of such a sorting feature, before creating it. This change took time and the requested feature only became fully operational near the end of the study.

C. Stage 3—2007-2012

Work in the Netherlands began in late 2007 as part of a demonstration project to evaluate the role of behavioral monitoring in the delivery of care in both a residential and institutional setting [15]. During the first stage (2007-2008) of the project the behavioral monitoring system was installed in the residences of 12 individuals living independently and 13 individuals living within a sheltered housing facility, while in the second stage (2008-2012) the system was installed in the residences of an additional 230 individuals living independently throughout the largely rural Limburg Region served by two care organizations. Similarly to how the three London organizations were combined, it makes sense to view these two organizations, as well as the demonstration project and larger study, as a single entity and to aggregate the 255 clients into a single Dutch Study.

Since the demonstration project in the Netherlands began as the London Study was winding down, it was possible to provide the Dutch with an enhanced Web-TAO which had the ability to produce "blogs", which we renamed the "Client's Journal". Of course, the content of the Web-TAO, e.g., check-boxes, auto-populated fields, instructions, had to be translated into Dutch. The care delivery model at the two care organizations—Proteion and Zorgroep—required that their care workers spend a considerable amount of each day traveling to and from clients' residences. Thus, they spent little time at the two organizations' administrative headquarters, limiting their ability to both access the Web-TAO and to enter information on computers. This problem was solved by developing the capability for the Web-TAO, renamed the Home Care Informatics System (HCIS), to be accessed on any smart mobile device. This change in the structure of the HCIS to a wireless mobile service raised several design challenges. First, everything had to be reformatted so that it could fit the small screen of the mobile devices. This led to an even greater reliance on check-boxes and auto-populated features and to the development of more efficient scrolling features. Second, there was the challenge of making the HCIS display properly on the different smart devices used by the care workers.

Working directly with the carers during the demonstration project allowed for a series of other suggestions to be incorporated into the HCIS: 1) the Client's Journal feature allowed entries by any authorized personnel; 2) the Journal could be sorted by alert, date, care worker, type of care delivered and outcome; 3) there was a new feature that allowed additions to a previous entry, but not the elimination of the original entry; 4) a series of pop-up prompts helped the user navigate through functions and avoid common errors; 5) additional security features were developed to ensure that only authorized individuals could access and contribute to a client's record; and 6) a read-only feature was added. Even with these modifications, two issues remained unresolved. The first issue concerned how the HCIS was used during care review meetings at which time the care delivered to specific clients was discussed and decisions on future care made. These meetings included both individuals who had knowledge of and access to the HCIS and others who had neither. Since the client reviews were more thorough when everyone at the meeting had access to the information stored in the HCIS record, questions arose as to who should have access, how should they obtain access and who was in charge of making access happen? Although this issue does not directly concern the technical development of the HCIS, it certainly impacts the implementation of the HCIS and its long term use. The second issue concerned whether the information stored in the HCIS could be used by supervisors and administrators in the evaluation of work performance. On the surface, the concern expressed by the care workers in the Netherlands was similar to those raised by the social workers at Selfhelp. It was believed that these concerns could be fairly easily resolved by discussions of interested parties. This was not the case, and this issue remained unresolved at the end of the pilot.

D. Summary of Changes

Below is a summary of the changes made to the HCIS in response to the expressed needs of the people using it. A review of these changes brings into focus how much the TAO/HCIS changed during the six years studies in response to the wishes of the users.

<u>Selfhelp</u>

1. The TAO was put on the web;
2. A feature that allowed the sharing of TAOs was created;
3. Check-boxes were added;
4. Auto-populated fields were added;
5. A feature that allowed follow-up entries was developed.

<u>Southwark</u>

1. Changes made to conform to British English;
2. Enhanced sharing capabilities were developed to allow multiple team members to enter data;
3. Security features added to allow sharing of data beyond the Smart Team;
4. A blog structure was created that allowed data to be entered by multiple carers;
5. An added feature that allowed the sorting of the blogs by alert, carer, care action taken, health outcome and date.

The Netherlands

1. It was translated into Dutch;
2. The blogs became the Clients' Journals;
3. The system was made to be operational on any mobile device;
4. A feature that allowed additional carers to enter data was developed;
5. More auto-populated fields and check-boxes were added;
6. A feature that allowed entries to be corrected without erasing the original was developed;
7. Additional search features were added to the system;
8. A read-only feature was created.

The above summaries support the argument that the final HCIS was actually a different product from the one that was created as a research tool for the pilot study at Selfhelp. Given that the development of the electronic health records system spanned eight years and was used by over one hundred care providers at seven different care organizations, it should not be that surprising that the final product was significantly different from the original. However, what is surprising is that all the significant changes brought about over this period were made at the request of the end-users. Possibly even more surprising is that almost all of the suggestions made by the developers for the "improvement" of the system were rejected by the end-users. Some of these suggestions concerned the way material was displayed on the smart phones, i.e., the number and order of check-boxes, while others were more fundamental, i.e., an automatic notification when a client was discharged from the hospital. Following the design policy outlined previously, all changes suggested by the developers were made operational in order to be tested in the actual provision of care. After one month of use, the end-users were surveyed in order to assess the usefulness of the changes. If the results indicated that over 75% of the care providers rejected the change, it was made non-operational. For example, every one of the end-users found the automatic notification of a client being discharged from the hospital to be both unnecessary and inappropriate, thus, it was immediately made non-operational.

IV. Lack of Acceptance

Even with all of these accommodations to meet the needs of the end-users, ultimately they stopped using the HCIS to record details of the care that they delivered even before the scheduled end date of the pilot studies. Most telling, the pattern of a gradual decline in the use of the HCIS leading to its abandonment, took place in all seven pilot study locations regardless of the care delivery model: care management; visiting nurse; coordinated intensive care management and residential setting: independent living; sheltered housing; high rise NORCS. Although frustrating, and initially perplexing, this result led to an analysis, and after the original paper was presented in Barcelona, a reanalysis of the reasons for why this lack of acceptance occurred. In order to assess the reasons for this lack of acceptance, questionnaires were administered to all eleven social workers at Selfhelp, while individual carers were interviewed in the Southwark and Dutch studies. In London, 70% of the carers were interviewed by phone and in the Netherlands approximately one-third of the carers were interviewed in person, while another third were interviewed by phone. Initial analysis of the findings derived from the questionnaires and interviews indicated that there were three main reasons for the lack of acceptance: 1) fear on the part of the carers that the information would be used to evaluate work performance; 2) the nature of the pilot study model and its impact on the carers' commitment; and 3) unwillingness of the carers to change their work routine.

A. Evaluation of Performance

Analysis of the carers' answers from the seven locations showed that the carers in the United States and the Netherlands were fearful that the information contained in the HCIS would be used by supervisors to evaluate their job performance. Two brief examples illustrate this ability of the information contained in the HCIS to evaluate the carers' performance. Each alert generated by the monitoring system is time stamped, as is every care action taken by a specific carer, and as a consequence, there is a concrete record of whether the carer responded to the alert and how long it took the carer to respond. In addition, the HCIS contains specific information on the type of response and the health outcome for each client over time, allowing supervisors to compare the work of different carers. It is this ability to compare the work of different carers which appeared to disturb the carers the most. And, it is indisputable that the HCIS allows this type of comparison to be made and for performance reviews to be based upon the information contained in the system. Not surprisingly, supervisors viewed this ability as an advantage because it documents performance, whereas, carers viewed it as an intrusion into their professional decision making.

B. Problems with the Pilot Study Model

The deficiencies of the pilot study model employed in the testing of the behavioral monitoring system have been detailed elsewhere [11], but some of these issues relate directly to the lack of acceptance of the HCIS. There appear to be four problems with the pilot study model. First, only a small number of carers were involved in the pilots at each of the organizations resulting in the studies being marginalized. At Selfhelp, only eleven of almost 200 social workers were involved in the study; while in London fewer than a dozen carers within the entire Southwark Local Authority had any role in the study and in the Netherlands, fewer than 5% of carers at the two organizations were involved in the study. Second, in no case was the HCIS used by the carer for all of her clients; instead it was always used for a small fraction of clients—on average no more than 20% and in only two cases over 50%—meaning that the carer was employing two different systems to record care. Third, information derived from the questionnaires and interviews showed that, because the carers knew when the pilot was to end, many carers put little effort into using the HCIS because they knew when it would go away. Finally, the HCIS was never part of "normal" care, but was always viewed as something that was just being "tested". As a result, the majority of carers responded logically by putting less and less effort into its use as the pilot progressed.

C. Unwillingness to Change Routine

Even though information from the questionnaires and interviews showed the importance of the previous two factors in the lack of acceptance of the HCIS by the carers, an even more important reason was their unwillingness to change their normal routine that did the most damage. There is no doubt that the HCIS required carers, at least initially, to do more work and undertake tasks which were unfamiliar. For example, in the Netherlands, carers, instead of just writing a couple of lines on a piece of paper kept in the client's residence, were expected to type in information on their smart phones before driving to their next appointment. They were also expected to update this information as additional care was delivered and even track and record health outcomes over time. These tasks were viewed as especially egregious since not all of their clients were in the pilot study and, therefore, they had to employ two different recording systems. Finally, it was difficult for the carers to see the value in this extra effort, because the benefits of better and more coordinated care were in the future, whereas the extra work had to be done every day.

D. Limited Generalizability

These three reasons appeared, prior to the Barcelona presentation, to explain the lack of acceptance of the HCIS in the care organizations, but there was still a serious unresolved issue: other than somewhat similar reasons for the lack of acceptance of another electronic health care records system in Germany [16], there were no other examples in which these particular reasons for the lack of acceptance of innovative health care technologies. This was disturbing enough, but when combined with the fact that the HCIS actually worked as envisioned, the records contained in it were readily shared, there was documentation that appropriate care based on the records was delivered and the system had been largely developed based upon input from the carers using the system, the three explanations just appeared insufficient to explain the consistent pattern of non-acceptance. And the pattern was eerily consistent. Within four months at each of the organizations the information the carers entered was increasingly uninformative and by the end of the pilots a majority of the carers at each of the seven locations had stopped using the HCIS altogether. There just had to be a better explanation for this lack of acceptance than that the pilot study model was limiting or that the carers were worried about their supervisors using the information contained in the HCIS to evaluate their work. As it turned out, there was; it just took the question about change management to elicit it.

V. CHANGE MANAGEMENT

An examination of the field of change management quickly provided an answer to the apparently simple question asked in Barcelona: the findings from the seven pilot studies were an example of the process of change management. The introduction of the HCIS required fundamental changes in the way that the seven organizations structured their business and care models. This did not take place with the result being a lack of acceptance on the part of the end-users which led to the non-adoption of the product. However, to fully understand this process, a reanalysis of the findings using the tenets of change management was necessary. This reanalysis not only brought into focus the inadequacy of the previous explanations, but, more importantly, highlighted a series of challenges facing the rapid and smooth introduction of new innovative technologies into existing health care systems.

A. Key Features of Change Management

This is not the forum for a detailed discussion of the field of change management, but several key elements of the approach must be briefly outlined before the reanalysis of the findings from the seven pilot studies can proceed. The main problem with providing such a summary is that change management is not a theory, but instead an approach or a process that is used in almost any type of business environment to manage almost any form of change, from the development of a new product, the introduction of a new technology, to the reorganization of a company's management team and everything in-between. There appear to be as many definitions as there are practitioners, consultants and gurus, selling, lecturing about and implementing change management. Any number of certificate programs are offered by any number of entities (companies, associations, universities) and an ever increasing number of books and journals devoted to specialized types of change management within the business community—IT, Human Resources, supply chain—which are easily available. Thus, change management can be almost anything making a single coherent definition difficult.

However, there are some common features upon which almost all practitioners agree. First, the field of change management can be traced to the 1962 publication of *Diffusion of Innovation* by Everett Rogers [17] in which he divided people into four categories: early adopters; early majority; late majority; and laggards to illustrate the way change flows through organizations. Rogers' work proved so influential that by the 1980's and 1990's the field was expanding exponentially as businesses and organizations faced the need to incorporate new technologies. It was during this period that a series of "principles" of change management were developed by different practitioners, usually consulting firms which offered to guide organizations through the process of change management, for a price. The number of principles, as well as the emphasis given different principles, varied, but, most of the lists stressed that successful change management is about people and thus, success came when the focus is placed on the human side of the equation, not the technological side.

Although difficult, given the large number of approaches available, it was necessary to select a single approach to guide the reanalysis of the lack of acceptance of the HCIS in the seven care organizations. The selection of this particular approach was based on the combination of the degree of recognition within the field and suitability to the reanalysis [18]. As in most of the approaches, the authors provide a list of principles, but equally important, they emphasize that long-term structural transformation is based upon four characteristics: scale; magnitude; duration; and strategic importance, without which change will not be successful. It was the insight gained from applying these characteristics, along with the ten principles the authors put forward, to the findings from the seven care organizations which allowed for a much more robust understanding of why the HCIS failed to be adopted.

B. The Characteristics

This and the next sub-section read as a litany of mistakes, mistakes that now appear obvious and avoidable, but at the time of the studies, were hidden by day-to-day challenges of modifying the HCIS, incorporating it into the existing care model and handling technological glitches. The first mistake concerns scale—the need for the change to impact all or most of the organization. As discussed in Section IV, the pilot study model was set up deliberately to restrict the HCIS to a small portion of the organization's care providers. This was the case in all seven of the pilot studies and was really the fundamental mistake made and repeated over and over again, since technological innovation that is limited to a small portion of the organization cannot be successful [14]. Therefore, it was not the pilot study model per se that was one of the reasons for the HCIS's lack of acceptance, but instead its use by only a small percentage of the care providers within the organizations that doomed it to failure.

Likewise, from the very first pilot study the issue of magnitude—a significant alteration in the status quo—was ignored. In many ways, the seven studies were set up to maintain, rather than alter the status quo. First, the small number of carers who were using the HCIS were isolated from their peers within the organization and little, if any, information, other than complaints, flowed from the carers in the pilots to carers who were not. Second, the pilot studies were viewed by top management all the way down to the end-users as tests of a new technology, not as a restructuring of the way people did their jobs. If anything, since the carers participating continued to provide care in the "normal" manner to clients who were not in the pilot studies, even their status quo was not truly altered.

The negative consequence of maintaining the status quo was compounded by the time limits set on the pilot studies because changes in long term transformations must last months if not years in order for the changes to be sustainable [14]. Everyone knew at each of the studies how long they would last. On day one, the carers who were employing the HCIS knew that on a certain date they would quit using it as the pilot study would be over. There were no plans to continue the use of the new system beyond the end date and thus, no understanding of the importance of duration in the success of incorporating a new technology into the organization.

In retrospect, the management of the seven care organizations was telegraphing a lack of strategic importance of the HCIS—the fourth characteristic—by the absence of commitment to the key characteristics of successful change. The scale of implementation was in all cases small; there was no plan to change the way that the care providers did their job and the duration was short with the endpoint always known. The question is why if the management of the care organizations was serious in its desire to introduce the electronic records system, did it not take steps that ensured its successful adoption?

C. The Ten Principles

The answer to this question is complex, but much insight into what went wrong can be gained by comparing the process undertaken in the pilot studies to the ten principles of change management outlined by Jones [18]. A detailed consideration of each of the ten principles in each of the seven pilot studies is beyond the scope of this paper. Therefore, in order to make the main points without the material being overwhelming and redundant, the seven pilot studies have been collapsed into one composite study. This collapsing is justified because, even though the location and care delivery model varied from one location to another, structurally the pilot studies were very similar: the same behavioral monitoring system; one of the several iterations of the HCIS; elderly at-risk clients; and the delivery of care in the clients' residences.

Principle 1—"Address the human side" is important because the change will involve people doing their jobs differently after the introduction of the new technology. This principle was completely ignored in the pilot studies because they were viewed as a test of a new technology without any consideration of how the use of this technology would impact people at all levels in the organization. Only when the issue of the material contained in the HCIS being used in the job evaluation of the carers was there any discussion of the change in the way people did their jobs and this was solved by deciding, both in the United States and the Netherlands, that the material would not be used and nothing would change.

*Principle 2—"*Get buy-in at the top" for the adoption of the new technology appears to have been followed in the pilot studies. Top management and even members of the Board of Directors were champions of the introduction of the new technology, but the problem came from a lack of follow-through. How top management viewed the pilots and what they wanted to gain professionally from the studies has been discussed previously [11] but briefly, the managers were more desirous of personal and professional gain than transforming the way care records could be used to provide improved care. This disconnect between the goals at the top and the work at the level of care provision, turned out to be one of the biggest hurdles to the effective use of the HCIS.

*Principle 3—"*Involve essential personnel" was ignored at all levels and there was lack of leadership during the pilots. No champions were identified and given leadership roles within the implementation. Several champions at the level of care provision emerged during the pilots, but they were not encouraged to play a leadership role and suggestions to compensate these champions were routinely rejected. Perhaps most devastating was that supervisors, who were responsible for assigning clients to particular carers and supervising the performance of the carers, played only marginal, if any, role in the roll out of the technology. Even when several supervisors tried to become involved in the pilot, their efforts were rejected and their suggestions ignored. The result was that these supervisors were the most negative of all employees toward the introduction of the technology.

Principle four—No "formal vision statement" was issued during any of the pilots. Instead of explaining why there was a need to adopt the new technology and spelling out in detail the corporate goals associated with its adoption, the managers left employees to figure out for themselves why they needed to change what they were doing and how the technology fit with the vision of the organization's future. The result was no common vision and no real reason for anyone to buy into the changes necessitated by the introduction of the technology, and certainly no *Principle five*—"feeling of ownership" over the process developed. No one accepted responsibility and, as a result, no one stepped in when things went wrong. Management blamed the carers for not taking the pilots seriously, carers blamed the managers for making them do more and different work without any justification of compensation, the IT department believed that they had not been sufficiently involved in the planning of the pilot and the supervisors just blamed everyone for the disruption in scheduling.

Principle six—"Communicate the message" was violated primarily because no one took ownership of the process. Management's only message was, "test this technology". It was pretty much left up to the end-users to figure out how to use the technology in care provision and it was their responsibility to fix any problems that emerged as the pilot study proceeded. There were no channels of communication created by which information could flow to upper management primarily because upper management did not want to know about any problems that could interfere with their claims of being at the cutting edge in the use of technology in care provision. If any message was sent, it was that no one was taking the pilot studies seriously.

Principle seven—"Assess the cultural landscape" was neglected because of the lack of seriousness conveyed by upper management. In a perverse way, the pilot studies actually assessed the culture and peoples' behavior at each level of the organization by documenting the overall lack of desire on the part of almost everyone in the organization to change how they did their jobs. People were happy doing what they were doing because they firmly believed that the existing care model worked, so why change. Since there was no clear message about the need to change from upper management, there was no reason for people to alter their behavior and change the culture of contented status quo. As a result, since there was no effort to assess the culture, there was no attempt during the pilot studies to follow *Principle eight*—"address the culture explicitly" in order to bring about the effective integration of the technology into the existing care model. Even when it was clear that carers were not using the HCIS, no one in management did anything about it. Instead of explicitly building a new culture that would have encouraged carers to make a greater effort to use the new technology, managers and supervisors allowed the existing culture to continue with the result that, over time, there was no incentive for anyone to change her/his behavior.

Principle nine—Given that upper management took no action when the carers stopped using the HCIS illustrates the degree to which individuals at all levels did not "Prepare for the unexpected". There were no contingency plans, no committees, no individual administrator or carer who was responsible to deal with the unexpected. Everything was ad hoc and problems were not addressed until someone—a carer, a client, a family member of a client—made a fuss. Then someone took an action. The problem was that it was arbitrary who that person was and the action taken, even if effective, was never institutionalized so that it could be used when a similar or different problem arose. As a result, it was the problems that became institutionalized and over time these problems were used as a justification for discontinuing the use of the HCIS.

Principle ten—"Speak to all individuals involved" was perhaps the most neglected of all the principles. Since no one wanted to admit that there were problems, no one spoke about these problems. The concept of having a well thought out system of rewards for carers who used the system was rejected, even though such an idea was continually suggested by the systems developers. Even when carers just stopped using the system, no one spoke to them to either encourage resumption or to determine why the individual had stopped using the system. The lack of concern reinforces the conclusion that there was a complete absence of a strategic plan for the incorporation of the HCIS into the care delivery model.

VI. DISCUSSION

The last of the pilot studies ended December 31, 2012 and currently the HCIS is not being used within any care organization. Thus, it is fair to conclude that the HCIS is a failure, even though it was used for over eight years to create and share information on the care of hundreds of at-risk older people. In the end there was a lack of acceptance on the part of the very end-users who designed the system. This is an irrefutable fact, but the reasons for the lack of acceptance are not so clear cut. Two different explanations have been offered for this result: one based on conclusions drawn exclusively from the evidence generated by questionnaires and interviews; and the second, although based on the same evidence, using the tenets of change management. Neither is wrong, per se; rather it is the difference in the ability to generalize the findings beyond these particular pilot studies that differentiates them. The first explanation is idiosyncratic and can only be compared with other studies using a pilot study model, while the second can be compared with any number of examples of how the process of change management can be used to understand the introduction of new technologies in a variety of industries.

It is clear that the seven pilot studies, to one degree or another violated each of the ten principles outlined in the previous section. It now appears to have been completely foolhardy to believe that at the end of the pilot study the HCIS would be adopted and incorporated into the existing care delivery model, simply because the HCIS required a new care delivery model to have been created. This new technology, if adopted, required a new way of organizing the delivery of care, necessitating a dramatic change in the way carers, supervisors and upper management did their jobs. In order to bring about such dramatic change, the leaders in the care organizations needed to develop a coherent strategic plan that addressed the challenges that this disruptive process put in front of them. In retrospect, the administrators at the organizations either needed to apply the tenets of change management themselves to this process or to have hired a consulting firm to direct the process. However, these options were never seriously considered simply because the pilots were viewed as "tests of a new technology". The belief was that if the technology worked, then it would just be seamlessly incorporated into the care delivery model. There was no recognition that this incorporation either could change the way people did their jobs or that people would just stop using the technology.

As it turned out, the fact that the end-users played a major role in the design of the HCIS was of little consequence to the final outcome. This inclusion of the end-users was just a furthering of the belief that the pilots were a test of the technology. The belief was that if the carers designed the technology, they would be more likely to use the technology in their work. Thus, the carers in the various pilot studies were encouraged to design the system, but once the technology changed how they did their jobs and no inclusive strategic plan that explained the long term goal of the introduction was produced, they stopped using the system. The original idiosyncratic explanations for the lack of acceptance of the HCIS were not wrong, but instead they were symptoms of the much larger issues captured by applying the tenets of change management to the process. The carers' worry about how the information contained in the HCIS could be used to evaluate their job performance was a real concern and the lack of response to these concerns illustrates that it was impossible for the pilot studies to have led to adoption of the HCIS: the complexity of the process was just underestimated by everyone involved.

VII. CONCLUSION

Several findings are apparent from the material presented above, some of which are specific to the pilot studies, while others are of a more general and important nature. Of the former, two are the most obvious: 1) it is possible to create an ehealth data system for home care and that such a system can be used effectively to coordinate care and services and contribute to the maintenance of independent living; and 2) the success of such a system is dependent on issues that do not concern design and functionality, but instead on its acceptance by the people employing the system. Even when much of the system's design was driven by these users, there was a lack of acceptance on the part of these very same people. Three reasons were put forward as the main factors for this lack of acceptance: the potential use of information in the system for the evaluation of job performance; the pilot study model; and the unwillingness of the carers to alter their normal routine. However, these reasons were unsatisfactory because of their lack of generalizability. In place of these idiosyncratic reasons, an analysis based upon the tenets of change management was offered which led to a second set of outcomes which prove to be much more important to understanding, not only the lack of acceptance of the HCIS, but of more fundamental challenges facing the introduction of new technologies.

The most important finding from this analysis was that there is no such thing as a test of technology in the real world. Tests of technology take place in the laboratory where conditions can be controlled, not in the world in which people have to provide care on a daily basis to at-risk elderly. In the real world what is being tested is the organization's willingness to change the way things are done so that the new technology will be used. The fact that the HCIS worked as designed and, thanks to the improvements made at the behest of the care providers, worked much better at the end of the eight years during which it was used, supports the view that tests of technology can only occur in the lab. It was not the technology that resulted in the lack of acceptance, but the fact that the carers did not want to change the way they did their jobs and there was no plan to alter the status quo. There is no reason to review the mistakes that were made from a change management perspective to conclude that there was no chance that the HCIS would be adopted while everyone involved believed that the only thing being tested was the technology.

At an even more general level, two additional conclusions can be derived from the findings. First, it can be argued that new ehealth technologies, such as the HCIS, will only become widely adopted when the traditional care models are unable to meet the needs of the burgeoning elderly population. Basically it comes down to the argument—if it's not broken why change it. The slow pace of the adoption of electronic medical records in hospitals and physician practices in the United States, even with government mandates and financial inducements from insurance companies, supports this simplistic argument. There will be a point at which the demographics overwhelm the ability of the current care models to provide for the huge number of elderly individuals who will need an ever increasing amount of services in the home and it will only be at that point that the new technologies will be adopted.

Second, when this time comes, not all organizations will be able to make the adjustments necessary to effectively adopt the new technologies. The internal inertia at many organizations will be so great that changing the basic work routines will just not be possible and thus, the only way the necessary changes will occur is for new organizations to emerge. This has been the history of technological change; new companies that embrace new technologies emerge and leap frog over the companies that refuse to undergo the necessary transformations. To think that this same process will not occur in the field of health care, in general, and home health care, in particular, is naïve.

REFERENCES

[1] A. Glascock, "If they designed it, why don't they want it? The lack of acceptance of an ehealth data records system," Proceedings of eTELEMED 2014, L. Van Gemert-Pijnen, H. C. Ossebaard, Eds, 2014, pp. 116-122.

[2] B. Shilling, "The federal government has put billions into promoting electronic health record use: How is it going?," www.commonwealtfund.org, retrieved October, 2014.

[3] B. Kaplan and K. Harris-Salamone, "Health IT success and failure: recommendations for literature and an AMIA workshop," JAMIA, vol 16, no. 3, pp. 291-299, 2009.

[4] K. Ashish, C. DesRoches, E. Campbell, K. Doneian, S. Rao, and T. Ferris, "Use of electronic health records in U.S. hospitals," NEJM, vol. 360, pp. 1628-1638, 2009.

[5] D. Ludwick and J. Doucette, "Adopting electronic medical records in primary care: Lessons learned from health information systems implementation experience in seven countries," International Journal of Medical Informatics, vol. 78, pp. 22-31, 2009.

[6] National Association for Home Care & Hospice, "Basic statistics about home care updated 2010," http://www.nahc.org, retrieved October, 2014.

[7] P. Esmaeilzadeh, M. Sambasivan and N. Kumar, "The challenges and issues regarding e-health and health information technology trends in the healthcare sector," in E-business Technology and Strategy, M. Zaman, Y. Liang, S. Siddiqui, T. Wang , V. Liu and C. Lu, Eds, Berlin: Springer-Verlag, 2010, pp. 23-37.

[8] N. Zakaria, S. Affendi and N. Zakaria, "Managing ICT in healthcare organization: culture, challenges, and issues of technology adoption and implementation," in Handbook of Research on Advances in Health Informatics and Electronic Healthcare Applications: Global Adoption and Impact of Information Communication Technologies, Y. Dwivedi, K. Khoumbati, B. Lal and A. Srivastava, Eds., Hershey, PA: IGI Publishing, 2010, pp. 153-168.

[9] Robert Wood Johnson Foundation, "Health information technology in the United States: Progress and challenges ahead, 2014," http://www.rwjf.org/en/research-publications/find-rwjf-research/2014/08/health-information-technology-in-the-united-states.html, retrieved November, 2014.

[10] Altarum Institute, "Overcoming challenges to health IT adoption in small rural hospitals," http://www.healthit.gov/sites/default/files/pdf/OvercomingChallenges_in_SmallRuralHospitals.pdf, retrieved November, 2014.

[11] A. Glascock, "A conflicted view of telehomecare after a 20 year journey," Amsterdam: IOS Press, 2012.

[12] A. Glascock and D. Kutzik, "A care informatics approach to telehomecare applications," in The Handbook of Research on IT Management and Clinical Data Administration in Healthcare, A. Dwivedi, Ed., Hershey, PA: IGI Publishing, 2009, pp. 368-382.

[13] A. Glascock and D. Kutzik, "The impact of behavioral monitoring technology on the provision of health care in the home," JUCS, vol. 12, no. 1, pp. 59-79, 2006.

[14] D. Kutzik and A. Glascock, "An Evidentiary Report of QuietCare®: Selfhelp Community Services," unpublished.

[15] C. Willems, M. Spreeuwenberg, L. van der Heide, A. Glascock, D. Kutzik, L. de Witte, and J. Rietman, "Activity monitoring to support independent living in Dutch homecare support," in Everyday Technology for Independence and Care, G. Gelderblom, M. Soede, L. Adriaens, and K. Miesenberger, Eds., Amsterdam: IOS Press, 2011, pp. 145-151.

[16] M. Schapranow, "Personal communication," during the Fifth International Conference on eHealth, Telemedicine, and Social Medicine, Nice, France, February 2013.

[17] E. Rogers, "Diffusion of innovations," New York: Free Press, 2003.

[18] J. Jones, D. Aguirre, and M. Calderone, "10 principles of change management," Strategy + Business, Booz Allen Hamilton Inc., pp. 1-4, 2004.

Wireless Health: Making Your Devices Talk

A Review, Solution, and Outlook for Wireless Health Connectivity

Stephan Hengstler

MeshEye Consulting
Campbell, California, United States
hengstler@alumni.stanford.edu

Abstract—**With the technological revolution in digital communications and connectivity over the past two decades, the healthcare sector is at the beginning of a dramatic overhaul. These technologies have already made their way into our everyday lives and thus changing the way we do things. The healthcare industry with its resistance to change has started considering, evaluating, and embracing the way connectivity can change medical treatment and personal health. In this paper, we review the state-of-the-art in medical device connectivity with a focus on wireless solutions. Throughout the paper, the discussion primarily applies to the United States and it separately studies its three major care delivery settings: clinical, office, and home. Based on the challenges and requirements that each of these settings present, we discuss the key aspects needed for medical device connectivity to succeed from both a technological and financial perspective. Cellular connectivity can satisfy many of these key aspects. Therefore, we have proposed and operated a testbed for cellular-enabled upload into Electronic Health Record (EHR) systems, which we present here and outline its implementation in detail. The paper concludes with a longer term outlook on the adoption of digital communications and connectivity in the healthcare sector.**

Keywords-cellular; connectivity; devices; health; wireless

I. Introduction

There is much excitement in the electronic health (eHealth) and mobile health (mHealth) industry about the promise that wireless technologies can bring to healthcare. Many grassroots efforts are underway promising everything from vital sign monitoring to aging in-place. Naturally, one may ask which technologies and solutions truly create value, which will survive in the end and ultimately benefit us humans?

The business environment feels similar to the beginnings of cellular technology in the mid to late 1990s. Many companies offer complementary, overlapping, or competing product solutions for improving healthcare through the use of wireless connectivity—the same kind of wireless connectivity we already use on a daily basis in our laptops, tablets, and cell phones. Although they share the same base technology, the rules of engagement differ for the healthcare sector in many aspects from consumer markets. It is us, as the end–user, driving market success in consumer markets and hence deciding the fate of a product solution or technology. A consumer product succeeds if it meets the needs of the consumer in terms of cost, features, and usability. Not so in the healthcare industry: the success of a medical product is driven by multiple other factors such as distributor alliances, marketing efforts, insurance reimbursement, quality reputation and regulatory track record. With all the parties involved in the chain of treatment, who have a stake in deciding the means of treatment, it is us, as the patient, who has the least say in the medical devices that facilitate our diagnosis and treatment.

The research presented here expands upon our seminal paper [1] on wireless health and our presentation [2] given at the mHealth Summit 2013. In this paper, we will cover and discuss the deployment and usefulness of wireless connectivity technology in a variety of medical instruments primarily in the United States. The paper starts out in Section II with a survey of existing connectivity solutions used in medical devices today. In Section III, we introduce several key aspects that are necessary for a connectivity solution to succeed in the healthcare market. Section IV applies these keys to cellular connectivity exclusively and presents our technology solution for connecting medical devices equipped with cellular modems to Electronic Health Record (EHR) systems. In Section V, we discuss the direction that we see the market taking and our view of what the future holds for connectivity solutions in healthcare. Section VI concludes the paper with a summary of the insights gained and final remarks.

II. Existing Connectivity Solutions

The deployment of wireless technology in healthcare delivery settings today is widespread. Many solutions already exist or are under development aiming to streamline the healthcare system [3]. But, as varied as the patient groups are, so are the treatment options. Today, wireless solutions in healthcare are highly fragmented with little standardization beyond the medium access layer. While this fragmentation facilitates a high degree of targeted solutions, which address specific needs, it makes it difficult for medical instrument companies to capitalize on their R&D investments.

Two different ways of categorizing solutions in use today help to shed light on wireless deployment: (i) grouping by the intended healthcare setting (clinical, office, and home setting) and (ii) grouping by the target patient group (teenagers, baby boomers, and general population). Grouping solutions by target patient group will lead to a separation by health condition, whereas categorizing them by healthcare setting will result in a separation by treatment and usage environment. The latter grouping is more meaningful with respect to medical device functionality and connectivity. Therefore, let us take a closer look at which connectivity solutions have made their way into the three different care delivery settings.

A. Clinical Setting

In clinical settings, i.e., clinics and hospitals, the objective of connected devices lies in preventing medical errors and reducing the cost of treatment. Connected devices facilitate this through streamlining the flow of admission, diagnosis, billing, and release information.

Clinical healthcare providers still prefer wired solutions for most of their medical instruments. Table I lists the main advantages and disadvantages of wired versus wireless connectivity in medical devices. For one, wired solutions are more secure, reliable, and easier to maintain once installed and configured. Such wired instruments include for example vital sign monitors, surgical instrumentation, and hospital lab equipment. The use of mobile devices that doctors and nurses carry around is limited to smart phones, tablets, personal digital assistants, and most notably bedside monitors [4]. Both wired and wireless devices that are used in diagnosis and treatment typically integrate into the facility's Health Information System (HIS) and Laboratory Information System (LIS) through the use of instrument middleware.

With few exceptions, IEEE 802.11 Wi-Fi [5] is the preferred connectivity technology for such devices. Cellular technology [6] is only used for text message notifications to personnel involved in patient care activities. So far, wireless connections only make sense for instruments that doctors and nurses carry with them to perform routine tasks or for patient bedside monitors according to a clinical laboratorian at the Palo Alto Medical Foundation. A prime example of such a patient bedside monitor is Abbott's i-STAT 1 Wireless System [7]. It is a handheld blood analyzer that, for instance, allows quantitative point-of-care measurements of Cardiac Troponin I, which is used in the diagnosis and treatment of heart attacks. The handheld analyzer can transmit measurement results via Wi-Fi into the patients' electronic medical records. The primary motivators for connecting medical devices into electronic medical records lie in the reduction of the overall cost structure, prevention of transcription errors, and, in the United States, by federal mandate [8], in the reduction of the rate of readmission.

The deployment of cellular technology in clinical settings still faces strong opposition. A frequently encountered argument against cellular technology is weak reception or no coverage. Cellular coverage largely depends on the network carrier. One thing to keep in mind is that coverage can only improve over time—if paying customers reside at any given location, at least one cell phone carrier will be there to go after them. Some hospitals and clinics are installing indoor Distributed Antenna Systems (DAS) thus actively improving cell coverage to the benefit of caregivers and patients alike.

TABLE I. WIRED VERSUS WIRELESS CONNECTIVITY

	Advantages	Disadvantages
Wired	• Robust, stable and reliable • Access control on premise • Simple to monitor security	• Higher cost of installation • More complicated to scale • Upgrade can be expensive
Wireless	• Easy to install and deploy • Supports device mobility • Readily upgradable to latest wireless standard	• Access control challenging • Devices need to be configured individually • Requires coverage testing

B. Office Setting

Doctors' offices are currently undergoing a fundamental change. The federal incentives and mandate towards the adoption and meaningful use of electronic health records [9] causes smaller doctors' offices to switch from primarily paper-based record keeping to electronic health records for their patient base. With it, the use of instrumented testing becomes also more lucrative as test results can automatically find their way into a patient's digital medical record. However, very few of such devices are in use today; let alone advanced devices offering cellular connectivity.

Especially for smaller practices, the main hurdle is the affordability of diagnostic test instruments and their limited insurance reimbursement. Test labs service most diagnostic testing needs arising in doctors' offices with an established cost structure for reimbursement. This flow of patient testing is more cost efficient as long it remains below the cost of ownership of in-house instrumented testing. This is primarily dependent upon the volume of tests run in a doctor's office. Furthermore, if an instrumented test has not been waived through the Clinical Laboratory Improvement Amendments (CLIA), a staff member of the doctors' office needs to undergo training to be authorized to perform this particular test. Another hurdle lies in that doctors, who have run their practice paper-based for most of their career, are unlikely open to adopting new technology and change the way they have been practicing medicine. This hurdle will diminish over time as more and more younger doctors take over—especially "digital natives" [10] who are accustomed to computer, internet and smart phone usage. Note that the implied assumption here is that younger doctors bring with them a comparatively higher willingness to try new technologies in patient treatment.

The situation is very different in an adjacent point-of-care setting: minute clinics. They specialize in the rapid diagnosis and immediate treatment of only the most commonly occurring infections such as Influenza or Streptococcus, and diseases such as diabetes, high cholesterol, high blood pressure or asthma. Their volume of tests performed is large enough to justify the use of instrumented testing. Therefore, medical instruments have started to make their way into these point-of-care facilities. Instrument connectivity is of little value thus far unless it can relay the prescribed drug treatment through the patient's health record to the pharmacy or send reminders of dosage or refills to the patient's cell

Figure 1. The BD Veritor™ System.

phone [11]. Since several of these minute clinics operate under the same business as pharmacy services, forwarding the prescribed drug treatment immediately to the pharmacy does make business sense—it saves paper and avoids human transcription errors.

C. Home Setting

There is a plethora of solutions already available in the wireless health market today. The industry has come up with enticing catch phrases to market the products in this market segment: quantified self, patient-centric, personalized medicine, and aging in place. Products ranging from vital sign monitoring, such as body weight, body fat, heart rate, blood glucose, and oxygen saturation to dieting, fitness and sleep trackers are readily available. They generate massive amounts of data which, in most cases, are continuously uploaded via Bluetooth, WiFi, or USB to an associated smart phone app, which analyzes and visualizes the data. The ultimate objective has to be the improvement of one's individual personal health [12] through changes in behavior and lifestyle. A reduction in healthcare cost is often a desired side effect for the people using these devices on a regular basis.

There are two sizeable markets in the United States for these personal health products: the teenage population and the baby boomers. The two population groups have different health challenges and hence the solutions are tailored to their needs. Baby boomers are entering the retirement age and with it come the onset of several health concerns such as congestive heart failure, hypertension, and diabetes. Hence, baby boomers spend money on solutions that enable graceful "aging in place," i.e., solutions that detect, prevent, or manage such chronic conditions in the convenience of their homes [13].

In case of the teenage population, who are sometimes referred to as "the Fat Kids of America," the primary health concerns are obesity, diabetes, and asthma. The objective here is not only the management of these chronic conditions under the supervision of the teenager's parents, but to maintain or improve his or her overall health through enforcing medication adherence and ultimately creating a persistent change in behavior. A representative example of a personal health monitor with cellular connectivity is Telcare's cellular-enabled Blood Glucose Meter [14]. To our knowledge, this is the first product in the personal healthcare market that directly uses cellular connectivity to upload glucose measurements in real-time to the patient's diabetes record, which resides on a secure Telcare server. In case of a minor, the parents as well as authorized doctors are given access to the diabetes record to review glucose level charts. Moreover, Telcare's server provides instant feedback and coaching to patients via the smart phone style glucose meter.

III. Keys to Success

Table II summarizes our review of medical device connectivity in the three care delivery settings. With these opportunities and challenges in mind, let us take a closer look at the keys required for a solution to succeed in each of these setting. The overarching key for success of any new healthcare solution is overall cost reduction in the healthcare delivery process. And that is the premise of wirelessly connected medical devices: their attraction lies in cost reduction, measurement objectivity, and ease of use. While the above mentioned keys are common across all care delivery settings, each setting weighs them differently or has additional keys to success.

For illustration and consideration purposes, a good example of a medical device that exhibits measurement objectivity and ease of use is the BD Veritor™ System [15], which the United States Food and Drug Administration (FDA) approved in 2010 for the clinical as well as the point-of-care care (POC) care delivery setting. It is a rapid testing platform for the detection of infectious diseases such as Influenza Type A and B and Group A Streptococcus. The BD Veritor System [16], as shown in Fig. 1, consists of the device and the consumables, that is, the mobile reader and the sample extractor, test tube, and test cartridge (in the figure, the cartridge is shown inserted in the reader). The reader is priced at around $300 USD; however, in its current version it is lacking the option of connectivity into HIS or LIS installations.

To perform a test with BD Veritor System, the physician mixes the patient sample (nasal fluid for an Influenza test and saliva for a Strep test) that resides on the sample

TABLE II. Wireless Connectivity in Healthcare Settings

Care Setting	Opportunities	Challenges
Clinical	• Bedside monitoring during routine patient visits • Patient self-monitoring after hospital discharge	• Clinics are slow in adopting new technologies • Reduction in overall cost of care not yet proven
Office	• Facilitate adoption of electronic health records • Seamlessly relay treatment to pharmacy or insurance	• Insurance reimbursement limits return on investment • Smaller offices not setup for wireless connectivity
Home	• Detect, prevent, and manage chronic conditions • Self-tracking to create persistent lifestyle changes	• Monitoring products lack standard and aggregation • Gap between tracking and persistent behavior change

extractor with the reagents in the test tube. Three drops are then dispensed into the sample well of the test cartridge. After ten minutes, the physician inserts the cartridge into the reader. Finally, the BD Veritor System reader analyzes the test strip for ten seconds and displays the final test result.

A. Clinical Setting

Since the hospital's clinical lab along with external central labs cover most of the testing needs arising in patient treatment, there is not a great deal of potential for adding wireless medical devices in the hospital setting. As discussed in Section II, the exceptions are devices that doctors and nurses use in routine patient treatment or patient bedside monitors.

There is however another emerging class of devices that can greatly benefit from wireless connectivity: devices that track the state of health of a patient after his or her release from the hospital. To achieve this, the patient could be given a monitoring device that facilitates home testing and wireless data upload into the hospital's HIS or LIS. One advantage is that the patient could recover in the comfort of his or her own home while the critical parameters of his or her state of health are still being monitored by the hospital's medical staff. The other benefit is that this would lower the readmission rate—in line with the United States' Affordable Care Act [8]—while reducing the cost of care at the same time.

The key to making this a reality is to combine a test approved for home usage with an easy-to-use device that is able to wirelessly transmit the patient's health parameters reliably and securely into the hospital's HIS or LIS.

B. Office Setting

To successfully place wireless medical devices in the point-of-care setting, minute clinics or physician offices, requires foremost that the solution makes financial sense. In this setting, a patient testing service has a fixed reimbursement amount no matter how the test is performed, i.e., visually read, instrument read, or by a central lab. Hence, doctors' offices will have a difficult time financially justifying the expense of instrumented testing if the per annum test volume for that particular test is low. In other words, wireless medical instruments can only succeed in this market if they prove to be less expensive to purchase, install, and operate than the already existing solutions in place. Although the federal mandate towards the use of medical health records may aid in deploying more wired and wirelessly connected instruments, most instruments are just too expensive to be financially viable testing solutions for most doctors' offices.

Nevertheless, rapid tests that occur frequently such as for infectious diseases (Influenza, Streptococcus, sexually transmitted diseases, etc.) may justify usage of wireless medical instruments. The keys here are that such instruments are cleared for the point-of-care setting, i.e., Clinical Laboratory Improvement Amendments (CLIA) waived, and that their cost of ownership lies approximately below $500 per year. A BD Veritor System that features connectivity into EHR systems would meet these requirements satisfactorily.

C. Home Setting

While each of the solutions offered for home deployment may address a particular health issue quite adequately, there are many challenges facing the wireless health home market today. For one, there is little to no standardization. Each solution works on its own independent of other health products in use. Each solution also requires frequent interaction and manual data entry by its user—something a society governed by convenience strongly shuns. For this reason, the average duration of regular usage does not exceed 30 days for the majority of these health improvement apps: just 5% of all apps (including health apps) are still in use 30 days after download [17]. In short, they are too intrusive to many people's already hectic and packed life.

Decentralized storage of data collected through different personal health solutions creates another significant challenge. How is one to get a comprehensive picture of one's health if the data resides in several different, unique applications? There are of course a few solutions like Google Health (discontinued as of January 2013) and Microsoft HealthVault [18] attempting to address the need of centralized data storage through offering a single landing page service. But, most personal health products do not interface with them and hence data would have to be entered manually. Therefore, a major key to succeed in this market is easy and seamless integration of the medical sensing devices, that is, the ones that provide personal health metrics, into personal health record systems. This can only be achieved effectively through standardization of the health data interfaces. The Continua Health Alliance [19] and the Institute of Electrical and Electronics Engineers [20] for instance are actively pushing this standardization and have been issuing design guidelines and standards for interoperability in personal healthcare [21].

Another fundamental issue of personal health tracking is that it is not sufficient to create persistent and lasting lifestyle changes. Living in a society of instant gratification, we expect solutions with this promise of success; but behavior change is a process of perseverance—very much in contradiction to instant gratification. In fact, Joseph Kvedar [22] has found "that only a small portion of the population, around 10 percent, will change their behavior based on tracker information alone." Knowing the right thing and doing the right thing are worlds apart. Even if personal health trackers provide us with vital information of what foods to avoid for example, we are still subjected to the marketing exposure of unhealthy eating habits. In the United States, good examples are the Carl's Jr. TV commercials for its selection of big and juicy burgers [23]. How can one not watch one of these commercials without leaving with the thought that relishing one of these irresistibly delicious burgers results in tremendous pleasure? Knowing that they are an unhealthy diet will likely not kill that thought! It is like running a marathon with a rock tied to one's ankle.

In essence, our lifestyle choices are not only impacted by reading our personal health statistics, but also by what we expose ourselves to in the form of billboards, commercials, and magazines. And to extract oneself from this omnipresent

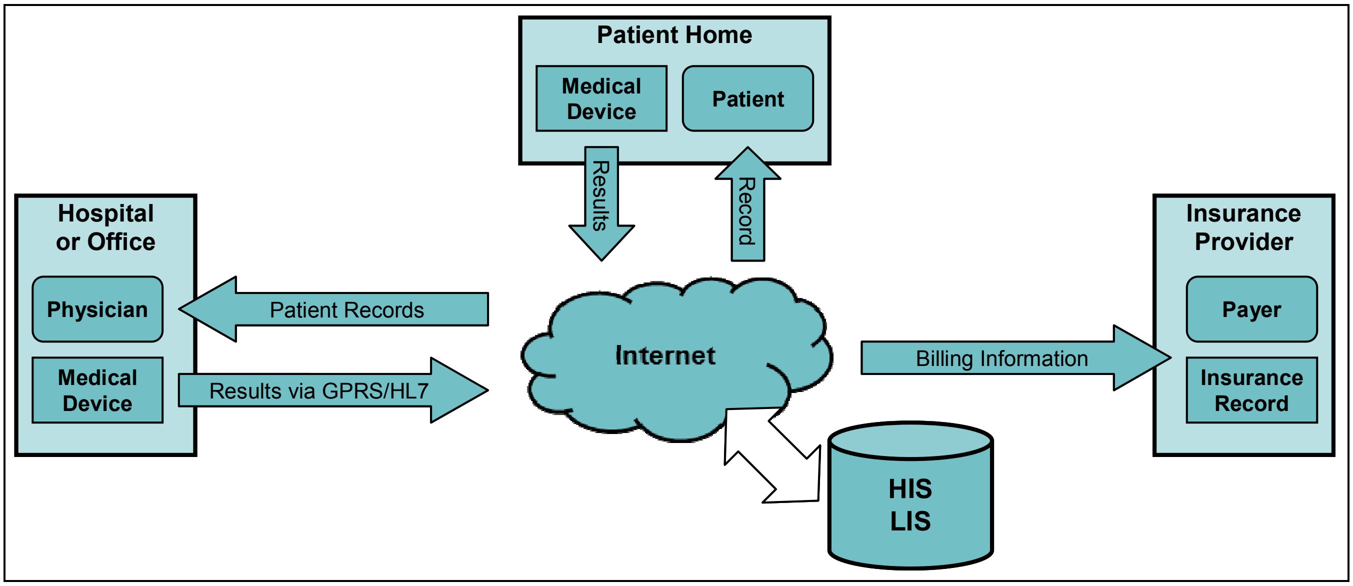


Figure 2. Healthcare information flow with cellular connectivity of medical devices.

exposure in the United States is a deliberate effort that has to be made daily. To assist us in this effort, our personal health systems would also have to tie into our flat panel TVs and web browsers and block out commercials and banner ads that are inappropriate for our current health condition. But, in reality, the opposite is happening. The marketing industry exploits our internet browsing behavior for the most part to entice us into buying and consuming more and more.

IV. The Case for Cellular

At this point, it should have become clear that there is no one-size-fits-all solution. The three care delivery settings considered have overlapping but also diverging requirements, which cannot be met by one solution all at once. Therefore, there are many product offerings from small to large companies, which focus on one or a few aspects in the healthcare delivery process. In short, the market is highly fragmented and proprietary solutions are prevalent.

But for solutions to be cost effective and scalable demands standardization and interoperability that in turn can proliferate integrated solutions [24]. Therefore, in the near-term, healthcare solutions will have to target seamless integration into the flow of care from patient over provider to payer [13]. Clearly, this is a good idea in theory but not enough to succeed in the healthcare market. The present reality is that the adoption of mHealth connectivity standards has been inconsistent [25].

We are convinced that the adoption of cellular connectivity in medical devices is the starting point to enabling higher levels of standardization and interoperability—at least at the front-end, where patient health data needs to make it into the digital medical record. It is crucial for subsequent treatment to consistently store this data digitally in a secure and reliable manner. But, if the interface method is lacking any of these attributes, the patient data will not be stored consistently leading to patchy health records. While there are several connection technologies and dataflow models conceivable, cellular technology is already dominating the personal consumer space and, as a result, has been widely adopted, is standardized, and continuously increases in data throughput and geographical coverage. Moreover, cellular hardware cost is held down by the large scale consumer market and service providers continue to drive down data transmission cost. Therefore, medical devices equipped with cellular modems can meet several of the keys for success discussed in Section III.

Let us discuss this cellular connectivity solution in more detail. Fig. 2 illustrates the flow of healthcare information when medical devices are equipped with a cellular GSM modem. This enables them to directly communicate with the HIS/LIS, or, more generally, the EHR system, through a General Packet Radio Service (GPRS) Internet connection. Test results can then readily be uploaded into the patient health record via the HL7 protocol [26]. Note that this direct connection eliminates the need for and expense of middleware software, a “middle man”, which, for the most part, reformats the device’s proprietary data output to the standardized EHR data format. Even more importantly, this dataflow model does not depend on another database server or cloud service operated, for instance, by the device manufacturer. Once the patient results have been uploaded to the EHR, which can either occur from a hospital, physician office, or the patient’s home, other need-to-know parties can readily access or be notified of the results. Such parties are the primary care physician, the insurance payer, as well as the patient itself.

A. *Testbed Implementation*

To explore and validate the feasibility of this cellular connectivity solution, MeshEye Consulting has been operating an Electronic Medical Record (EMR) connectivity testbed with an HL7 portal for test record upload since November 2010. The testbed deploys the open-source EMR software FreeMED [27] in lieu of HIS/LIS software. The

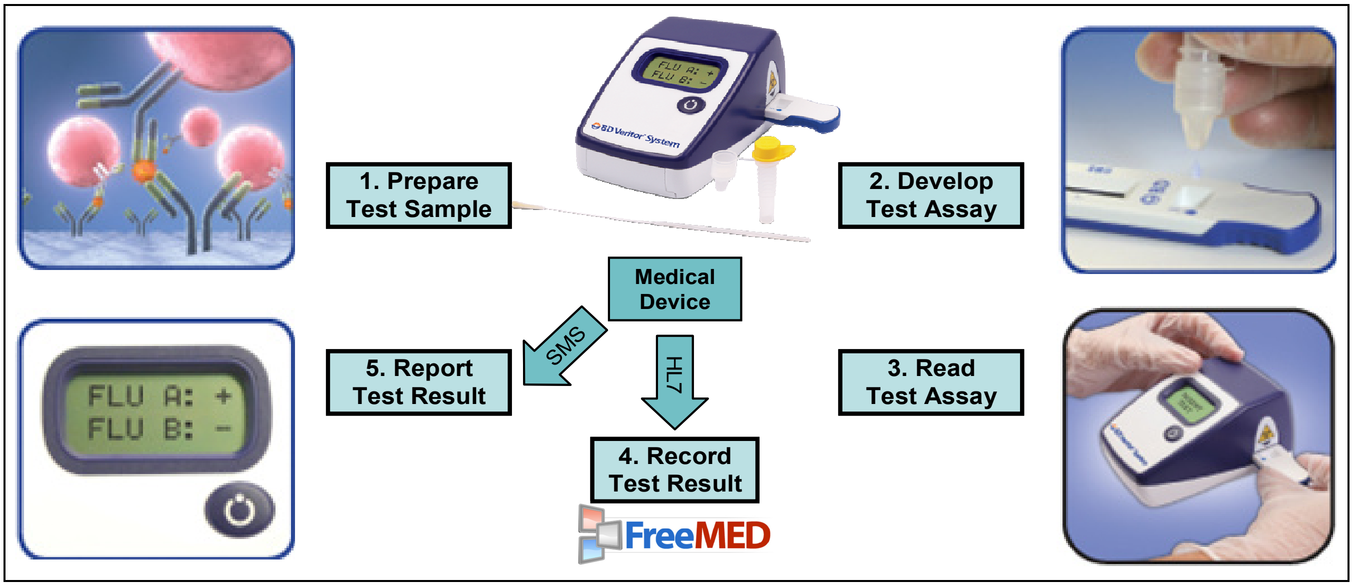


Figure 3. Process steps from patient sample to test result in the MeshEye EMR Connectivity Testbed.

FreeMED installation has been modified to accept test records from medical devices in the form of HL7 Version 2.3 messages encapsulated in XML-RPC requests. A medical device prototype equipped with a cellular GSM modem was designed to upload its test records to this EMR system via GPRS. The testbed has proven that this approach is feasible and easy to implement.

To generate the HL7 messages, the test result fields first needed to be mapped to fields in an HL7 ORU-R01 observation result message. The results of this field mapping are shown in Table III for Influenza test results that the BD Veritor System reports. To import the HL7 ORU-R01 messages into FreeMED v0.8.1.1, we used Java to invoke the FreeMED XML-RPC call

TABLE III. FIELD MAPPING OF VERITOR RESULT TO HL7 MESSAGE

HL7 Field Identifier	Veritor Result Field(s)	HL7 Message Field
MSH-10	Index of test result	Message Control ID
PID-4	Barcoded patient ID	Alternate Patient ID – PID
ORC-12	Reader serial number	Ordering Provider
OBR-2	Reader serial number and index of test result	Placer Order Number
OBR-4	“UPPER RESPIRATORY SAMPLE” and “INFLUENZA”	Universal Service ID
OBX-3	“UPPER RESPIRATORY SAMPLE”	Observation Identifier
OBX-5	Display of test result: “FLU A: [+/–] FLU B: [+/–]”	Observation Value
OBX-6	None	Units
OBX-7	[+/–]	Reference Range
OBX-8	[+]	Abnormal Flags
OBX-11	“F” for final result	Observ Result Status

```
FreeMED.Transport.parse('HL7v2', message);
```

where *message* is the HL7 message in string format.

To notify the physician of completed tests, the EMR connectivity testbed has been configured to send out text messages with the test results. The end-to-end delay commonly encountered is in the order of 10 to 20 seconds. Considering that rapid diagnostic tests typically take at least 10 minutes to complete, such quality of service (QoS) would be acceptable. But cellular network carriers do not make any guarantees of end-to-end delay for text messaging, and hence it is only a solution good enough for demonstration purposes but not for professional field deployment. Moreover, text messaging does not lend itself to encryption, which brings us to another area of frequent concern: compliance with the Health Insurance Portability and Accountability Act (HIPAA).

HIPAA compliance requires the implementation of reasonable safeguards for the protection of patient-identifiable information. Although the EMR connectivity testbed does not transmit any information that would allow identification of a patient by name, only an assigned patient identifier, it makes sense to encrypt the entire payload. This usually diffuses any concerns around patient privacy but adds the burden of encryption key management.

The process steps necessary to turn a patient sample into a test result notification in the MeshEye EMR connectivity testbed are shown in Fig. 3. In the first step, the patient sample is mixed with the test reagent in a test tube, which takes no more than five minutes. Next, the sample mixture is dispensed onto the test cartridge and the test assay develops within 10 minutes. In step 3, the cartridge is then inserted into the reader which reads the test assay and determines the test result with a read time of 10 seconds. Finally, the reader uploads the test result via HL7 transparently to FreeMED, the MeshEye EMR system, and reports the test result in the

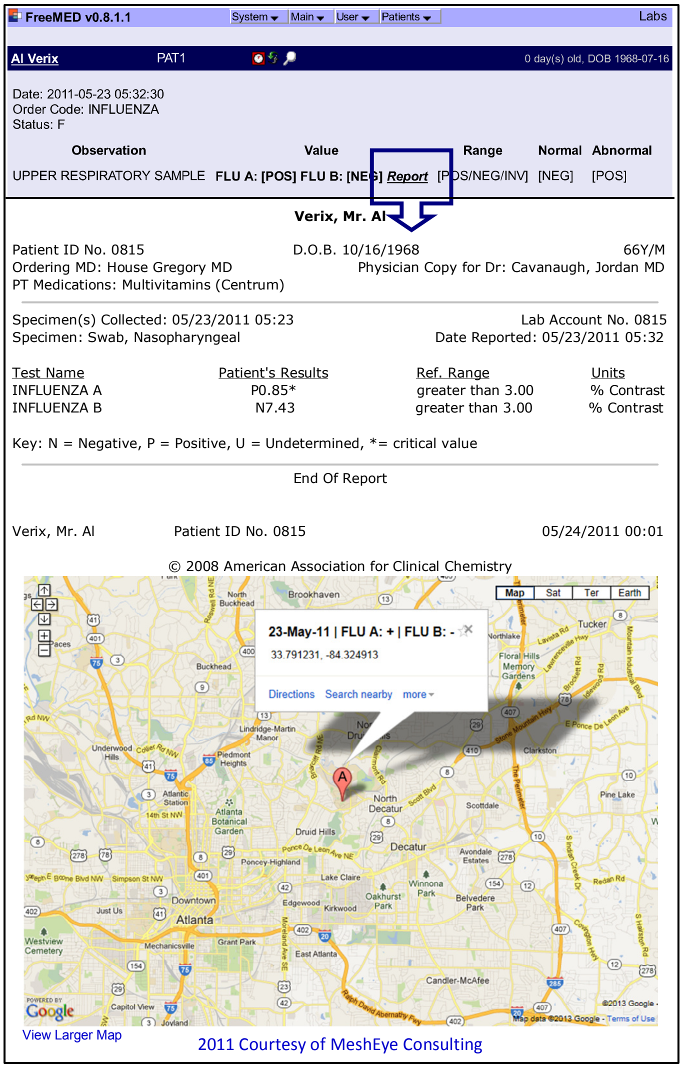


Figure 4. Patient test record (top) and instrument test report (bottom) of the MeshEye EMR Connectivity Testbed.

form of a text message notification to one or more cell phones. Steps 4 and 5 typically take just under a minute when cellular coverage is available. Thus, the entire process from test sample preparation to test result reporting approaches 15 minutes. Such test process duration is well suited for diseases and infections that benefit from being detected and treated during a single patient visit.

B. Testbed Demonstration

The EMR connectivity testbed was demonstrated to several hospitals in California as well as to the Centers for Disease Control and Prevention (CDC) in Atlanta, GA, in May 2011. Fig. 4 shows the patient test record and instrument test report that the testbed generated during the demonstration. The top of the figure shows the view of the patient's test result entry while the bottom of the figure shows the automatically generated instrument test report. The report contains all the fields expected of a lab test report. In addition, it maps the rough location of the testing site, which is derived from the cell tower identifiers within communication range of the modem. Most importantly, the test result upload completes in real-time, i.e., it usually takes less than a minute. This solution would allow the CDC to publish their "Influenza Surveillance Report" in real-time rather than with data lagging by two weeks. Especially CDC's recently launched influenza app [28] could benefit greatly from real-time reporting of infectious disease testing.

Further advantages of uploading and storing test results that have been patient de-identified but tagged with geo-location information, are their value for data mining purposes. In case of Influenza testing, this would allow monitoring real-time progression of the infection and spotting of pandemics early on. In case of diseases that are not seasonal, but chronic in nature, availability of such data could help to establish correlations between geographical area and the likelihood of developing such diseases. Maps that predict the likelihood of developing and progressing asthma for instance have already been generated and serve as an example of the value of such maps. Taking it even one step further, these likelihood maps could help shedding light on prevalent factors causing chronic diseases, that is, whether they are primarily hereditary or predominantly location dependent.

C. Testbed Limitations

While deployment of our EMR connectivity testbed was straightforward, we found that its deployment in the real world is far from this simple. More specifically, we came across three major challenges to real-world deployment: (i) HIS/LIS installations sit behind a tight firewall, (ii) HL7 is not really a universal standard, and (iii) healthcare providers may not accept unsolicited observation results. The following paragraphs will give more details on these challenges.

Since healthcare providers have to protect patient information from unauthorized access, HIS/LIS installations are only accessible from within the organization's intranet or at least sit behind a tightly configured firewall. That means that it is impossible for our cellular-enabled reader to directly upload test records via HL7 over Transmission Control Protocol (TCP)/Internet Protocol Suite (IP). Virtual Private Network (VPN) tunnels to the organization's intranet are typically established to overcome this obstacle. Nonetheless, running a VPN client inside the reader is not feasible given its limited computational resources. An alternative would be to open up a TCP/IP port in the organization's firewall although this would increase the vulnerability to unauthorized intrusion.

Anyone dealing with HL7 messaging quickly realizes that HL7 is not really a universal standard. The saying among experts goes "If you have seen one HL7 interface, you've seen one HL7 interface." One reason behind this is that every HIS/LIS installation configures the mapping of HL7 message fields to HIS/LIS database fields differently.

The third challenge comes from the reader sending an HL7 ORU-R01 observation result message without the order number issued for the diagnostic test. Therefore, this is considered an unsolicited result message which not all HIS/LIS installations accept. This challenge could be addressed by scanning the order identifier instead of or in addition to the patient identifier.

We are currently working on finding solutions to the above mentioned limitations of our EMR connectivity solution. The success depends to a large degree on the flexibility of healthcare providers to make changes to their IT infrastructure. How much value the set of tests performed by a particular instrument creates will primarily drive the willingness for such changes.

V. LONGER TERM OUTLOOK

There is no doubt that interoperability through standardization will continue to increase in healthcare solutions. From a technology perspective, that is what is required to make any medical device talk to any EHR system [29]. It also makes sense from a business perspective since interoperability is an essential component for a scalable connected health market [25]. In short, interoperability through standardization will likely pave the way for widespread use of connected medical devices. Not surprisingly, the history of digital technology already features numerous examples of standards that triggered widespread use: Ethernet, GSM, Wi-Fi, and UMTS to name just a few.

But, knowing the right thing does not necessarily translate into doing the right thing. In fact, the healthcare industry in the United States is known for its resistance to change and slow rate of technology adoption. For instance, Thompson states that "I feel frustrated that physicians don't quite seem to be practicing in the 2012 world of technology I see on the exhibit floor [at the annual AACC Clinical Lab Expo 2012]" [30]. Healthcare investor G. Kurtzman puts it this way [31]: "Unless there is a "pull" from customers, patients, providers, or payers, an entrepreneur in healthcare IT won't be able to capitalize on just a good idea." Along these lines, the two parties that still need to drive the idea of connected health with more conviction are the payers and the regulators. The roadblock, that regulatory agencies and healthcare payers pose, and the outlook, that we anticipate for them, are summarized in Table IV.

The regulatory agencies' mandate includes issuing regulations for marketability of medical devices and enforcing them in the marketplace. There still remains a lot of uncertainty concerning the regulation of mobile health applications and related connected health devices. Therefore, the regulatory agencies have to clarify the approval process of these emerging technologies. In the United States, the FDA already took a big leap forward towards more clarity in 2013 when it issued its final guidance on mobile medical apps [32]. The next step is to speed up their approval process. This will also make the pursuit of connected health solutions more attractive to the investment community.

With respect to regulatory approval of wireless medical devices, there is an important distinction between unidirectional and bidirectional communication of the device to a web server or a cloud service. When the device only transmits data to the web server but does not receive any data back, regulatory approval is generally only required for the medical device as an autonomous, stand-alone device. But once the wirelessly connected device receives any information back from the web server or cloud service that could alter its functionality, the approval process applies to the system encompassing both, the medical device and the web server or cloud service.

TABLE IV. THE REGULATORY HURDLE AND THE INSURANCE BATTLE

	Regulatory Agencies	Healthcare Payers
Road-block	• Regulations are uncertain and approval process is slow	• Payers not convinced about overall reduction in cost
Outlook	• Issue clear guidance around connected medical devices • Simplify submission and speed up approval process	• Conclusive case studies and clinical trials needed • Account for entire chain of healthcare services in studies

The healthcare payers, that is, the insurance providers, have to be persuaded that connected healthcare solutions not only make sense but also reduce the overall cost of treatment. This is especially important in the United States, which has the highest cost structure in healthcare. It will require several more case studies and clinical trials to make a convincing case for the overall reduction in healthcare cost. Such studies and trials are however intricate and costly since the entire chain of healthcare services involved in patient treatment has to be accounted for.

Finally, a strong push for wireless connectivity in healthcare is coming from several players at the bottom of the food chain of healthcare reimbursement: medical device manufacturers and cellular network providers. Device manufacturers have an increasing interest in equipping their products with connectivity. This would provide them with instrument quality control (QC) data as well as access to test results, which may allow them to move up in the food chain. Network providers see the opportunity in high-volume data contracts in machine-to-machine (M2M) communication, which is viewed as their next big market after the cell phone market volume has started to level off. This alliance between device manufacturers and network providers is however not without challenges. Device manufacturers will have to negotiate pricing for cellular data subscriptions that meet their revenue models without directly transferring increased cost to healthcare providers. In other words, network providers prefer per monthly billing whereas healthcare providers purchase medical devices in one time transactions. This is the cash flow gap that device manufacturers will have to bridge.

With respect to cellular connectivity in medical devices, the outlook is the same as for connectivity in general. Nevertheless, it has to bear the additional burden of subscription fees paid to cellular network service providers. But, there is hope in sight [33]: "[…] The number of devices with integrated cellular connectivity increased from 0.73 million in 2011 to about 1.03 million in 2012, and is projected to grow at a CAGR rate of 46.3 percent to 7.1 million in 2017." And by the laws of supply and demand, increased deployment will result in lower cost of cellular connectivity in medical devices. Certainly, in-home monitoring devices, which we mentioned in Section III for the clinical setting, will contribute to this market growth [33]: "For example, in the U.S., readmission penalties established by the Centers for Medicare & Medicaid Services will drive hospitals to adopt telehealth solutions for monitoring post-discharge patients." However, most likely countries other than the United States will lead the way—countries, in which cellular subscription fees adapt more rapidly to market supply and demand, as is the case in most countries across Europe and Asia.

VI. Conclusion and Future Work

We reviewed the current state of connectivity technology for medical devices in the healthcare sector giving special attention to wireless connectivity. The review highlighted the diversity and fragmentation of existing solutions to address the demands in the clinical, office, and home care setting in the United States. Therefore, the one key aspect to increase adoption of connected medical devices is interoperability through standardization. Cellular connectivity can enable standardized, seamless, and ubiquitous integration of medical devices into EHR systems. For this reason, we proposed and presented a cellular connectivity testbed that confirms and demonstrates the validity of this approach. Our EMR connectivity testbed indicates that medical devices can be seamlessly integrated into the flow of patient treatment across all three healthcare delivery settings. However, it remains to be seen whether wireless connectivity can actually lead to an overall reduction in the cost of care and change towards healthy lifestyle choices. Moreover, regulators and payers still have a long way to go before wireless connectivity becomes the norm in everyday patient diagnosis and treatment.

Acknowledgment

The author gratefully acknowledges the discussions, inspiration, and involvement of his colleagues at Alverix Inc. in San Jose, California, and the Becton, Dickinson and Company (BD) Diagnostics Group located in Baltimore, Maryland, and San Diego, California.

References

[1] S. Hengstler, "Wireless Health: Making Your Devices Talk," Proceedings of the Fifth International Conference on Emerging Network Intelligence (EMERGING 2013), Sept. 2013, pp. 19–25.

[2] S. Hengstler, "Making Your Devices Talk: The Case for Cellular Connectivity," Fifth Annual mHealth Summit (mHealth '13), Dec. 2013.

[3] "mHealth in an mWorld: How mobile technology is transforming health care," Deloitte Center for Health Solutions, Deloitte, 2012, pp. 1–21.

[4] P. B. Luppa, C. Müller, A. Schlichtiger, and H. Schlebusch (2011), "Point-of-care testing (POCT): Current techniques and future perspectives," TrAC Trends in Analytical Chemistry, vol. 30, no. 6, Jun. 2011, pp. 887–898.

[5] Wi-Fi® in Healthcare: Security Solutions for Hospital Wi-Fi Networks," White Paper, Wi-Fi Alliance, Feb. 2012, pp. 1–11.

[6] S. Saw, T. P. Loh, S. B. L. Ang, J. W. L. Yip, and S. K. Sethi (2011), "Meeting Regulatory Requirements by the Use of Cell Phone Text Message Notification With Autoescalation and Loop Closure for Reporting of Critical Laboratory Results," American Journal of Clinical Pathology, vol. 136, no. 1, Jul. 2011, pp. 30–34.

[7] Abbott Point of Care i-STAT 1 Wireless System, Product Website, [retrieved: December, 2014] Available: http://www.abbottpointofcare.com/Products-and-Services/iSTAT-Wireless.aspx

[8] United States Department of Labor, Affordable Care Act, [retrieved: December, 2014] Available: http://www.dol.gov/ebsa/healthreform/

[9] Government Health IT, Policymaking, Regulation, & Strategy: Meaningful Use, [retrieved: December, 2014] Available: http://www.healthit.gov/policy-researchers-implementers/meaningful-use

[10] M. Prensky (2001), "Digital Natives, Digital Immigrants Part 1", On the Horizon, vol. 9, no. 5, Oct. 2001, pp.1–6.

[11] I. Sachpazidis, S. Fragou, G. Sakas, "Medication adherence system using SMS technology," Proc. Intelligent Sensors, Sensor Networks and Information Processing Conference (ISSNIP 2004), IEEE Press, Dec. 2004, pp. 571–575.

[12] P. Sonnier, POLL RESULTS: What is the single greatest health benefit afforded by Digital Health solutions? [retrieved: December, 2014] Available: http://popperandco.com/2012/07/

[13] L. C. Baker, S. J. Johnson, D. Macaulay, and H. Birnbaum (2011), "Integrated Telehealth and Care Management Program for Medicare Beneficiaries with Chronic Disease Linked to Savings," Health Affairs, vol. 30, no. 9, Sept. 2011, pp. 1689–1697.

[14] Telcare Cellular-Enabled Blood Glucose Meter, Product Website, [retrieved: December, 2014] Available: http://www.telcare.com

[15] A. Paxton, "Rx for optimizing rapid flu test performance," CAP Today, vol. 27, no. 1, Jan. 2013, pp. 1–5.

[16] BD Veritor System, Product Website, [retrieved: December, 2014] Available: http://www.bd.com/ds/veritorsystem/

[17] R. Gary, "Why 95% of Mobile Apps are Abandoned—and Tips to Keep Your Apps from Becoming Part of that Statistic," Nuance Communications Inc., iss. 20, Nov. 2011, pp. 1–2.

[18] Microsoft HealthVault, Product Website, [retrieved: December, 2014] Available: https://www.healthvault.com

[19] Continua Health Alliance, Organization Website, [retrieved: December, 2014] Available: http://www.continuaalliance.org

[20] IEEE Standards Association, IEEE Standard 11073-10103-2012 – Health informatics—Point-of-care medical device communication, [retrieved: December, 2014] Available: http://standards.ieee.org/findstds/standard/11073-10103-2012.html

[21] D. Bowman, "Continua Health Alliance Releases 2011 Design Guidelines," FierceHealthIT, iss. 9, Sept. 2011, pp. 1–2.

[22] E. Waltz, "How I Quantified Myself: Can self-measurement gadgets help us live healthier and better lives?," IEEE Spectrum Magazine, vol. 49, no. 9, Sept. 2012, pp. 43–47.

[23] Carl Karcher Enterprises, Inc., Company Website, [retrieved: December, 2014] Available: http://www.carlsjr.com

[24] J. Hatcliff et al., "Rationale and Architecture Principles for Medical Application Platforms," Proceedings of the 2012 IEEE/ACM Third International Conference on Cyber-Physical Systems (ICCPS '12), Apr. 2012, pp. 3–12.

[25] "Interoperability: An essential component for scalable mHealth," mHealthInsights, PwC, Mar. 2013, pp. 1–3.

[26] Health Level Seven International, Introduction to HL7 Standards, [retrieved: December, 2014] Available: http://www.hl7.org/implement/standards/

[27] FreeMED Electronic Medical Record Software, Product Website, [retrieved: February, 2014] Available: http://freemedsoftware.org

[28] G. Slabodkin, "CDC's influenza app should be more reliable," Editor's Corner in FierceMobile Healthcare, iss. 15, Jan. 2013, pp. 1–2.

[29] B. Malone, "The Cutting Edge of Lab Connectivity: Will New Standard Deliver Plug-and-Play Solutions?" Clinical Laboratory News, vol. 38, no. 9, Sept. 2012, pp. 1–5.

[30] B. Malone, "Spotlight on Point-of-Care Testing: Innovation, Expansion Evident at AACC Clinical Lab Expo in Los Angeles," Clinical Laboratory News, vol. 38, no. 10, Oct. 2012, pp. 1–8.

[31] G. Kurtzman, GUEST POST: The question every healthcare IT startup must answer, [retrieved: December, 2014] Available: http://venturebeat.com/2012/08/25/healthcare-it-startup-tips/

[32] U.S. Department of Health and Human Services Food and Drug Administration, "Mobile Medical Applications," Guidance for Industry and Food and Drug Administration Staff, Sept. 2013, pp. 1–43.

[33] G. Slabodkin, "Research predicts strong growth of patient home monitoring systems by 2017," FierceMobile Healthcare, iss. 18, Jan. 2013, pp. 1–2.

Social Networking, the Catalyst for Identity Thefts in the Digital Society

Eric Holm
Federation University Australia
Mount Helen, Australia
PhD student, Bond University,
e.holm@federation.edu.au

***Abstract* - This paper explores the vulnerability of social network users to identity theft when they share personal identification information online. The sharing of details like age, sex, address and other personal information like photographs can assist in establishing an identity. Identity criminals exploit social network users and the weaknesses of social networking sites to gather the information needed to commit identity theft and identity fraud using this identification information. While there are mechanisms that can reduce the incidence of this crime, information sharing on social networks is voluntary, which, makes its control difficult. This paper presents an exploration of existing literature from Australia, the United States and United Kingdom and highlights the importance of the relationship between social networking and identity crime. The drivers to sharing information on these platforms are considered. The paper provides opportunities to improve the understanding of the relationship between personal information and the crime. A difficulty in having preventative mechanisms in place is that social networking sites have a vested interest in promoting rather than preventing the sharing of information. Further, identity crime is pervasive which, makes the amelioration of risks difficult. In conclusion, efforts have been made in this paper to outline arguments that will assist in resolving the crime given vulnerability of social network users to identity theft.**

Keywords- social networking, privacy, identity theft, identity fraud.

I. INTRODUCTION

Social networking has inspired computer users to share information online. Social networking sites bring together people with common interests and they enable mass social interaction [1]. This mechanism of communication overcomes geographical constraints and can bring together disparate groups [2]. Social networking is attractive due to its social inclusiveness [3] as well as its interactive nature [4]. For example, over 500 million people have used Facebook to create profiles to express themselves across this social networking platform [5]. Facebook is the most popular social networking site, followed by Myspace according to college students in the United States [6]. The social linkages these platforms create are particularly attractive to these users [7]. Many of the new innovative forms of communication are accompanied by new ways of accepting and exploiting this interaction. Acceptance of these new ways of interaction expands social connections between people, but this communication can be exploited, leading to something far more sinister: identity crimes.

This paper considers why identity crime is serious in the context of the strong uptake of social networking. The basis of the discussion is around literature obtained from the United States, United Kingdom and Australia. The paper then discusses the responses to identity crime in social networking including the suitability of criminal law and privacy responses to this crime. Thereafter the paper discusses the international dimensions to dealing with identity crime in social networking and provides some recommendations and foreshadows future work.

II. RELATED WORK

The extent that individuals share information on a social networking site is determined by the decisions they make and are influenced by many behavioral drivers. The control mechanism used on social networking sites is typically the user privacy settings, which, allows an individual to determine the visibility of their profile to others. Most users tend to leave these at the default setting established by the social networking provider, which, may be less than optimal to the end user of these services in respect of privacy [8]. A social networking profile is how the social networking users represent themselves online and it facilitates their presence and disseminates information about them and this is at the heart of social networking [9]. The opportunity to share information is attractive to users who aspire in particular to share their emotions, expressions and experiences online [10]. One of the attractions of social networking sites is the reciprocal nature of such information sharing [11], but social networking sites seek to balance the security needs of user with their ease of use [12]. While the sharing of information provides the foundation under which, many relationships are formed [13] it also provides the basis for rekindling relationships with old friends [14]. In addition, many social networking sites provide incentives for promoting the creation of these friendships, sharing general interests or religious beliefs, and numerous other activities [11]. In this regard there is a vested interest for social networking site providers to encourage the sharing of information and there are many positive outcomes that can be derived from social networking [15].

Social networks have become an alternative to communication in many traditional social contexts [15]. Increasingly communication takes place online and social networking has become a platform that functions in place of (or in conjunction with) existing social contexts. However, social networking is a relatively new phenomenon and many of the social conventions around it are still developing [8] and it may be for this reason that many users are complacent about the potential risks associated with sharing personal information online. For instance, accepting 'friend' requests may occur far more readily through a social networking site than it might off it [16]. Friend requests on a social networking site may appear innocuous but later become harmful particularly if granted to an identity criminal. Gender influences the preparedness of users to share information, with men being prepared to share information online more freely than women [16]. Furthermore, younger men are seemingly more prepared to share information than older men [17] and factors like peer pressure may play a role in this. Nonetheless, there seems to be complacency in relation to the risks associated with information sharing on social networking sites with many users sharing information about themselves including their full name, their location, date of birth and also photographs [13]. This information can be used by identity criminals to form an identity that they subsequently use to perpetrate crime.

The interest in social networking is profound, with the social networking site Facebook having an estimated 1.15 billion users. Platforms such as Twitter, Google plus, Myspace and LinkedIn have all attracted masses of members [18]. While user uptake in social networking sites is staggering, user engagement is equally astounding with 20 per cent of Facebook users checking their accounts numerous times per day [18], this is evidence suggesting that many users are also using different platforms to log into their chosen social networking site whether through a computer, mobile phone, tablet or an assortment of these [19]. Indeed, new technologies are facilitating an even more committed user base for social networking activity. This may contribute toward the complacency around the sharing of information.

In addition to individual users, many businesses interact on social networking sites to increase their business exposure to customers and clients [20]. They utilize the services of social networking providers to share information, advertise, promote and position themselves in the wider market and this makes them susceptible to identity crime on social networking sites in the same way as it does for individuals [20]. For corporations, the victimization from identity crime may be arising from corporate disclosure rules as this adds to the volume of information that is readily available to identity criminals [20].

In most instances, information acquired by an identity criminal is taken without the knowledge or consent of the victim [21]. The victim might not be aware that their information has been stolen until they find themselves exposed to financial liability. The usual motivation for the identity criminal is monetary gain but there may be other motivations for this crime [22]. The distinction between identity theft and fraud is important as identity theft is based on the theft of information, and identity fraud results from that theft [23]. In Australia, these crime types are distinguished through offenses that relate to the possession of identification information and offenses related to the dealing with it [24]. While both are crimes, offenses that result in financial loss are the more detrimental to the victim. Although in absence of financial loss it is still possible to suffer detriment from this crime.

Personal identification information is information that identifies a person, such as a passport, a driver's license or a bank statement [25]. However, there are other identifiers that could be regarded as personal identifying information and include demographic details including name, address or date of birth [26]. Past research has suggested that sensitive information also includes personal photographs, names and gender which, are prone to leakage on social networking sites [27]. Past research has also shown that Facebook users in particular are more prepared to reveal personal information (including their real name) on this site as well as including email addresses in their profiles [27]. While the documents needed to establish identity vary, most governments accept a range of identification documents [21] and by world standards, name, gender, date of birth and nationality are unique personal identifiers that are considered collectively to satisfy identity requirements [28]. The information stolen may be used by the identity criminal to achieve their desired outcomes of establishing identity. Once established they may use it for identity fraud or other to commit other crimes [21]. In the United States, for instance, identities have been stolen and used for perpetrating a range of criminal offences where the victims becomes wrongfully accused [21].

What is interesting in relation to the existing literature is that there is a body of literature exploring social networking as well as identity crime with little confluence between these topics. There is a real risk of identity crime through the disclosures of personal identification information on the on social networking sites. This paper aims to explore through relevant literature how social networking plays an important role as an enabler for identity crime. Identity crime is pervasive and will exploit emerging social interactions online, it is important to understand this vulnerability to better mitigate the crime.

III. METHODOLOGY

This paper is a selective literature review largely from countries including Australia, United Kingdom and the United States on issues of social networking disclosures and identity crime. The literature review is discussed from cognate areas with the view to exploring the confluence of social networking disclosure and identity crime and informing future research into this. As an emerging area of research little data presently exists on this topic and the outlook of this paper is to bring together existing work to point toward future opportunities for research work.

IV. How the crime is committed?

An estimated 16.6 million Americans were the victims of identity crime in the United States in 2012 [29]. Around 7% of households in the United States experienced identity theft victimization in 2010 [30] totaling about 8.6 million households [30]. The most quantifiable data relating to loss pertains to financial losses and this is expressed in the selected literature as follows. In 2010-2011 the estimated cost of personal fraud to Australians was $1.4 billion [31] with about 44,700 Australians being victims of identity crime [31]. Statistics from the United Kingdom suggest that identity crime is increasing prodigiously with the reported number of cases almost doubling from 77,500 to 123,600 between 2007 and 2012 [32]. These statistics suggest that identity crime is global and significant in terms of its impact and financial cost.

It is likely that the uptake of social networking has contributed to the volume of information exchanged and subsequently identity crime. Further, among the methods used by an identity criminal to obtain information is to utilize social engineering to gather information from other users [33]. A criminal may purport to be someone else like a friend or relative to gather the information they need to commit identity crime [34]. Hence caution should be exercised with friend requests in addition to the promulgation of information. In the context of corporate crime, a criminal might affiliate themselves with an organization or someone known to the organization [23]. A common rationale for this activity is that it is easier to obtain information through manipulation than by exploiting system security [35]. This approach seeks to exploit social interaction by playing on emotions [36].

While computer crime occurs through many highly technological means and is a highly sophisticated crime, paradoxically a basic understanding of computers is all that is needed to commit identity crime on the Internet [37]. Unlike other computer crime, this crime is pervasive as it is not restricted to those with specialist skills. A rudimentary understanding of information is what is needed along with an understanding of the crime and Internet access. In addition, identity crime is easy to commit and there is a low cost to committing it [37]. This crime is more readily accessible to criminals than many other Internet-related crimes because it is an instantaneous crime that is open to many prospective criminals [38]. In many respects the Internet has opened up numerous online communication mechanisms, as well many new ways of committing crime. Many of these aspects of identity crime make it attractive to criminals.

The dissemination of information has increased the risk of identity crime as well as establishing a separate industry based on the trade of personal identification information. Past research has suggested that purchasing information is the most common way of obtaining the information needed to perpetrate identity crime [39]. Beyond that, there are other techniques that can be used to obtain information [40]. However, the increased availability of information online has created an underground market for that information [41]. This presents numerous commercial imperatives for sharing that can also feed into identity crime and may devolve from social networking. The availability of personal identification information is the enabler for this crime and increasingly has a measurable monetary value [20]. Further, social networking feeds into the mass of information that has such a value and in many respects supports the current research around the cloud computers and the ways the Internet is emerging [42].

V. Why is this crime serious?

Identity crime has the potential to reach anyone. Research conducted at Carnegie Melon University suggests that children 15-18 years of age are those most likely to be victimized by identity criminals [35]. However, people of working age are at also at risk due to their levels of income as well as their relevant engagement with emerging technologies on the Internet [35]. Working age victims present ready-made targets to identity criminals and it is also probable that the risk of victimisation is linked to increased levels of engagement with technology [39]. Having said that, children have become victims of this crime for reasons that include the inadequate supervision of children’s Internet usage [43]. Children have a vulnerability to identity theft crime as they usually possess unblemished personal histories and remain relatively undefended as targets of this crime [35]. In addition, children often unknowingly share information about themselves that can place them at risk particularly if left unmonitored [35]. However, anyone using social networking can become a target of this crime and social networking: it seems reasonable to suggest that the more one reveals about oneself online, the more that can be used to perpetrate this crime.

Identity crime is serious because of the financial and emotional cost of the crime. The cost of identity crime comprises both direct and indirect costs. The most significant cost of identity crime is the financial cost [44], but the cost of identity crime extends beyond financial loss and incorporates additional costs referred to as soft costs [44]. The financial costs (the hard costs) are easily quantified whereas the non-financial costs (soft costs), such as those that relate to the cost of damage to reputation and the emotional cost of the crime, are more difficult to quantify and to prevent [44]. However, the cumulative losses can only be determined by considering both the hard and soft costs of this crime [44]. The banking sector for instance, is exposed to significant losses in relation to identity crime [44] but its spokespersons remain reluctant to disclose the losses arising from this crime [45]. Nonetheless, bank losses in the United States have been estimated to amount to over $2 billion per year [46]. However, due to the commercial sensitivities many banks are reluctant to share data [45]. This reluctance contributes towards the difficulty of establishing an accurate measurement of the true cost of identity crime [47]. Furthermore, there are issues with victims not reporting victimisation that also contributes toward the lack of accurate data [48]. The crime has a profound impact on an individual in terms of damaging their reputation and confidence as well as being financially reprehensible.

VI. Responding to the Crime

There are many practical difficulties in convicting identity criminals [47]. In the first place, in an international context, no central body is responsible for overseeing crime committed via the Internet. For this reason, controlling crime perpetrated through social networking sites is fraught with difficulties in the investigation and enforcement [48]. The Internet is a dispersed communication entity that permeates country boundaries, making regulatory responses difficult [48]. Further, different values influence the ways in which, crimes are viewed domestically and most international instruments continue to require attention through domestic laws. Success of responsive efforts will be dependent on the stance maintained by each country in question [49].

The European Cybercrime Convention has worked to harmonize the regulation of cyber-crimes internationally [50] and it provides domestic criminal law authorities with cooperative mechanisms to investigate and prosecute computer crimes [50]. The term 'cybercrime' is a phrase the European Convention uses to describe crimes where the computer or computer network is the target. Computer crime is distinguishable from traditional crimes because a computer is used to commit the crime [51]. This therefore subsumes frauds where the computer is used as a tool to commit the crime [50]. Likewise, when identity crime takes place through the computer it is arguably captured within the scope of the European Convention. However, the European Convention fails to deal directly with identity crime [50]. Rather, it captures computer-related forgery (article 7) as well as computer-related fraud (article 8) and it would apply to related offenses including identity crime but this is not made explicit [50]. The significance of this convention is that it assists in the investigation and enforcement of identity crime despite not making reference to it [50]. Given its scope for computer crimes, it would arguably encompass identity crime. Unfortunately, there is nothing simple about applying criminal sanctions to international identity crime particularly when they fall outside globally acknowledged crimes and atrocities such as genocide. Even so, the effectiveness of such responses is reliant on the preparedness of countries to agree and cooperate on responses to crime.

VII. Privacy Perspective

International responses to privacy share comparable challenges with the international regulation of crimes on the Internet. There is a lack of centrality when it comes to the regulation of privacy internationally [52]. Domestic laws are often based on international agreements that are relied upon to regulate privacy [48]. International principles of privacy protection are provided for in international agreements like the Universal Declaration of Human Rights [53]. This international agreement recognizes the protection of the inalienable rights of all humans to privacy, highlighting the need for them to enjoy freedom of speech and belief [53]. Further, Article 12 suggests that no one should be subjected to interference with respect to their privacy [53]. This international agreement provides the foundation for the development of domestic laws in the same way as the European Convention does for cybercrimes [48]. However, despite the operation of this agreement, a limitation of the Australian privacy responses, for instance, is that they are not prescriptive. Further, privacy is constrained by the same jurisdictional boundaries that limit the extraterritorial reach of criminal sanctions explained above [54]. This means that there are challenges of dealing with identity crime and privacy in an international context.

VIII. Focus on the Victims

The perpetrators of identity crime are illusive and many victims will often not know that they have become victims until considerable time has passed [55]. The time between when an identity crime occurs and an investigation takes place makes it difficult to gather evidence about the crime and to locate and prosecute the offender [55]. During this time, the victim must withstand the frustration and emotional distress and the financial losses caused by the crime. The impact of this is worsened the longer it takes for the situation to be resolved [56]. The subversive nature of this crime adds to a victim's frustration as well as facilitating the criminal's opportunity to evade capture. The crime also places an emotional burden on the victim and once an identity crime is discovered it can also take considerable time to resolve [56]. This delay influences the ability of law enforcement to investigate the crime effectively and provides the criminal with a greater likelihood of avoiding capture. This crime is pervasive and the criminal's illusive natures makes it difficult to capture them. Identity criminals may also harvest personal information over a period of time [39]. Further, the victim of the crime will find it difficult to determine when and where the information was obtained and what corrective action they should take to avoid future victimisation.

IX. Difficulties of Responding to the Crime

A major challenge in responding to identity crime is the ability of law enforcement agencies to obtain evidence for the prosecution of perpetrators of this crime. The gathering of evidence involves obtaining digital evidence both on- and off-line [57]. As there are many new ways of using information, it is essential for investigative efforts to deal with the speed data transference takes place on the Internet, making the investigation of identity crimes difficult [58]. Furthermore, as identity crime is cross jurisdictional, cooperation between law enforcement authorities is essential [59]. This also makes the civil responses to identity crime difficult given the scarcity of resources the individual has to prosecute criminals. Similar issues around detecting and locating the offender also exist for these actions.

A key weakness in the integrity of data is the way individual users manage their own information. Social networking users need to be more accountable for the information they willfully share on social networking sites. Each activity we engage in on the Internet leaves traces and a commonsense response to dealing with the exploitation of social networking by identity criminals is for social networking users to improve their behavioral practises on the Internet [57]. An educational program is necessary to ensure that social networking users are aware of the risks and of the need to exercise caution with respect to the sharing of personal information [60]. Moreover, this should take into account the ways information might potentially be misused by criminals [59]. While education could have a direct impact on crime reduction there will typically be a proportion of the population not responsive to such efforts [59]. The role of education is not going to resolve the crime entirely but irrespective it should be regarded as a way of dealing with this crime. However, social networking sites also should accept some responsibility for the protection of the users as they are responsible for attracting and retaining them. This should be broader than the general technological security measures and needs to include the architecture underpinning the sites to give users better understanding [61]. This might involve reconsidering the architecture that facilitates information exchange. Ultimately, identity crime can be reduced through better understanding of and mitigation of these risks [62].

Social networking providers put forward mechanisms to assist in the protection of information [15]. However, many of the tools used by social networking sites are underutilized, which, may be due to their complexity or lack of integration with the interface [15]. Alternatively, it may be due to the lack of engagement with the technology. Many users tend to utilize default functions within their profile that could mitigate many of the benefits of these tools [63]. More could be done by social networking providers to apply enhanced measures through related tools to protect users.

A number of additional and general technical responses can be applied to prevent identity crimes. The responses include improved authentication and encryption measures and might also involve elementary information security measures [62]. The purpose of technological responses is to ensure data integrity is maintained while correspondingly preventing unwanted misuse of information or intrusion [64]. However, as with most responses, these efforts aim to improve information integrity [62], but the strength of the responses to identity crime are often balanced against the perceived costs of such preventative action. In this respect, the threat of identity crime and the need for technological protection is understated. Nonetheless, these important technological measures provide additional ways to deal with information security and identity crime.

X. The Profile of the Social Networking User at Risk

Research conducted by Fogel and Nehmad indicated that certain social networking users are more prepared to engage in risk taking behaviors than others. Further, Facebook users have a greater sense of trust in the service they use than Myspace users [9]. In this regard, it has been found that men are more prepared to accept requests for friendship on social networking sites than are women. Men are also more prepared to share details like phone numbers and addresses than women [9]. This is where there appears to be a dichotomy between those that share information on the Internet and victims. But interestingly and somewhat conversely, it is women that are more often victimized in identity crime than men [9]. However, anyone using social networking sites is at risk and it seems that the more information that is shared equates to greater risk.

As mentioned, the development of protocols for communication on social networking sites is still developing and this influences the ways that information is shared. An interesting example of this is how people accept friends on social networking sites. The likely behaviors online are expected to be quite different to those undertaken in person outside the Internet [17]. Permitting a friend to have access to a profile is viewed far differently than the parameters of friendship that exist beyond the Internet. However, in many respects there are some elements of these relationships that are likely to be similar and shared. Walther and Boyd refer to friendship as a relationship of support based around emotional support [2]. However, the characteristics of friendship on this basis offline are difficult to transpose to a social networking 'friend' online [65]. What is interesting about this interaction is that the characteristic behaviors offline are not transposed online, this is interesting to observe as the protocols continue to develop, and this changes the profile of the victim.

The issue of consent on social networking raises questions about the right to share information belonging to another. Given the lack of prevalence of privacy principles on social networking sites, it is difficult to assume that consent is freely given for the use of personal information on social networking sites [66]. The issue of consent extends to the timeframes that information is retained on social networking sites. On these sites, data are typically subject to retention periods but these are often not adhered to, as interesting information is used in profile histories to attract new users [67]. Much of the information on social networking sites is also used to derive commercial benefits, this may be contrary and likely to be different from users' expectations about how this information will be used [15]. It is important to note that the business model of social networking providers is based on the dissemination of information and these providers arguably challenge the legal boundaries of privacy through the way they exchange data [15]. Once information has been passed on, particularly to third parties, it is unclear as to what obligations will be adhered to and the responsibilities of these parties are not defined [15]. Despite this being beyond the scope of this paper, users need to understand the risks attached to third party applications and understand the specific consent they are providing to the use of their sensitive information [8]. There are also significant privacy-related issues with these providers and the providers of third party applications to social networking sites, but this falls beyond the scope of this discussion.

There is an obvious dichotomy between the stakeholders involved with the protection of personal identification information and end users. Despite the many attempts to warn of the risks of information disclosure, information is still shared. The motivations of social networking providers are at odds with that of users [15]. There is a need to find the middle ground to ensure that a shared understanding is formed around the sensitivity of information online [68]. An educative effort is needed to deal with crimes [69], like identity crime that should include the social networking providers but should not be administered by them, given their divergent interests. There are few motivations for social networking sites to change their approach as there would be commercial ramifications in doing so [69]. However, they are in the best position to understand the architecture behind the interface and to deal with the problem.

Some researchers suggest that the improvement to privacy must come from improvements in the technology that underpins the architecture used by social networking services [62]. It is the technology that encourages information sharing in the first place, so the same technology can mitigate the incidence of crime in the future. The commercial interests may need to put aside interest for the greater good to reduce identity crime [27]. While many ways of exploiting individuals still exist, social networking has brought about new ways to exploit individuals and the service itself can play a role in reducing it [65]. The accessibility of social networking is an enabler for identity crime and the low cost of identity crime plays a role perpetuating the crime [66]. This crime brings new ways of committing old offenses [66]. In short, it would seem right that the mechanism itself should play a role in the solution.

XI. Discussion

Ultimately, policymakers should consider a multi-faceted approach for dealing with identity crime [70]. A mixture of techniques is necessary to counteract the threats of identity crime as it requires a ubiquitous response [71]. The relationship between social networking and identity crime is unique and therefore requires unique and creative responses. A major obstacle to responding to social networking and identity crime is the availability of accurate data relating to the relationship between the two concepts. While not all approaches to dealing with this phenomenon have been canvassed in this paper, the ones that have provide insight into the many issues evident. In particular, there is a need for a greater understanding of the behavioral factors of individuals in interacting with technology [72]. In addition, steps need to be taken to deal with the dissemination of information to avoid victimisation as well as better mechanisms to deal with this crime after it occurs [73].

The motivation for this research has been to explore the relationship between identity crime and social networking which, has scarcely been explored in existing literature and to establish a basis for further research to take place. More empirical research is needed to probe the parameters of this relationship. It is hoped that more interest in this research will be generated by raising awareness of this relationship.

XII. Evaluation

The material discussed in this paper has largely been drawn from secondary sources to identify a relationship between social networking and identity crime. To develop the contention further, empirical research is needed to discover the scope of this relationship. This paper has explored a number of responses to this phenomenon but these are by no means exhaustive and further research into the relationship between social networking and identity crime would be likely to provide greater insights into the mechanisms that might better deal with this crime.

XIII. Recommendations

Information is a vehicle for identity crime and considerable information is stored on social networking sites. Legal and technological responses have limitations in relation to the extent they can mitigate this crime particularly given the voluntary nature of information dissemination and the issues around jurisdiction and cooperation discussed [74]. The individual vulnerability arises because of personal identification information that eventually means that behavioral factors are important in mitigating risk. Therefore, it is hoped that through the dissemination of research and information that individuals may become better informed of the risks inherent in the activities they engage in on the Internet involving information sharing, particularly social networking. Individual users of social networking need to take greater responsibility for the personal identification information shared on social networking sites to avoid victimisation. In this respect, if behavioral norms can be changed on social networking sites then the risk inherent with identity crime can be reduced.

At the same time, individuals remain ambivalent to the risks that come from information sharing. A difficulty with information on social networking sites is that once it is shared with another person, it becomes harder to control [68].

Likewise the privacy mechanisms to prevent this are not strong. Social networking makes the protection of information far more difficult than traditional means, as information can be transferred instantaneously. With social networking, a tension exists between the technical designers of social networking sites and users concerning the disclosure of information, as one requires it for survival and the other for the joy of sharing experiences [75]. For social network providers this involves striking a balance, as far as they are compelled to, between the interests of members sharing information and their self-interest in promoting information sharing and risk [12].

More regulatory development is needed around privacy in an international context to develop principles to reflect the changing ways that information sharing takes place on the Internet. The regulatory environment needs to be progressive in the way that it deals with the changing risks present on the Internet [71]. In this, there are behavioral factors that need to be the focus as these relate to the decisions made by users in sharing information. Similarly, a better understanding of the crime and the new ways it can be committed through using social networking need to be developed [76].

XIV. IDENTITY CRIME REGULATION GLOBALLY

A number of barriers exist to dealing with this and related internet crime from a regulatory standpoint [77]. Criminal law has limitations in particular in supporting the victims of the crime. There are multiple issues that need to be resolved for regulation to be effective such as the need for cooperation to deal with the jurisdictional barriers to facilitate the investigation and prosecution of identity crimes [78]. The barriers to criminal sanctions also have an impact on actions in privacy as well as civil causes of action for the victims of this crime. In the absence of clear legal pathways, victims are left with few options for obtaining reparation. The issues of state sovereignty also present obstacles for actions against criminals [71]. The varied responses to identity crime complicate this. These present limitations on the ability of states and individuals to bring the offenders to justice and for victims to obtain support [80]. Internationally, there is a need for mechanisms to be developed in this realm to deal with the unique characteristics of this crime [71]. This will then translate to better domestic responses to this crime and related crime in Australia, as well as elsewhere [81].

XV. CONCLUSION AND FUTURE WORK

Social networking has encouraged many users to share personal information online, and social network users frequently engage in the sharing of information about themselves [12]. The sharing of personal identification information is prominent on social networking sites in ways that promote profiles which, might include details such demographic details and others. This article has considered the many ways that social networking can potentially nourish the transference of personal information on the Internet, in turn providing identity criminals with the information needed to commit identity crime. While there are many ways to respond to this crime, a blend of techniques is likely to work best, given the pervasive nature of this crime and barriers presented by multiple jurisdictions. These issues pertain not only to the difficulty of applying criminal sanctions but also to those relating to privacy in a transnational context. Future research is needed to explore responses to this crime in detail. An important starting point for dealing with this crime is to increase awareness of the risks associated with information sharing around social networking. More research is also needed to develop further knowledge about this crime and to understand the data surrounding identity crime and the nexus of this to the responses to it. This research would aim to recognize identity crime as the pervasive and significant crime it is now and will continue to be into the future.

ACKNOWLEDGMENT

I would like to thank the support of The Business School at Federation University Australia (formerly University of Ballarat) which, provided the assistance needed to prepare this article and Dr Ian Dobson for his outstanding editorial support. I would also like to acknowledge the support of the institutions I have studied at Australia and my research supervisors at Bond University. I would like to thank the ICDS peer reviewers for their comments on the draft sent to the ICDS 2014 Conference in Spain.

REFERENCES

[1] E. Holm, “Social Networking and Identity Theft in the Digital Society” The Eight International Conference on Digiatal Society ICDS2014) Mar. 2014, pp. 169-175, ISBN: 978-1-61208-324-7.

[2] J. B. Walther and S. Boyd, “Attraction to computer-mediated social support,” in Communication Technology and Society: Audience Adoption and Uses, C. A. Lin and D. Atkin, Eds. Cresskill: Hampton Press, pp. 153-188, 2002.

[3] J. Bargh and K. McKenna. “The Internet and Social Life,” Annual Review of Psychology, vol. 55, pp. 573-590, Feb. 2004, doi:10.1089/cpb.2005.8.423.

[4] I. Berson, M. Berson, and J. Ferron, “Emerging risks of violence in the digital age: Lessons for educators from an online study of adolescent girls in the United States,” Journal of School Violence, vol. 1, pp. 51-71, Jan. 2002.

[5] P. Valkenburg and J. Peter, “Internet communication and its relation to well-being: Identifying some underlying mechanisms,” Media Psychology, vol. 9, pp.23-58, Dec. 2007, doi:10.1080/15213260709336802.

[6] F. Stutzman, “An evaluation of identity-sharing behavior in social network communities,” International Digital and Media Arts Journal, vol. 3, pp.10-18, Aug. 2006.

[7] M.Green and T. Brock, “Antecedents and civic consequences of choosing real versus ersatz social activities,” Media Psychology, vol. 11, pp. 566–592, Dec. 2008, doi:10.1080/15213260802491994.

[8] P. Van Eecke and M. Truyens, "Privacy and social networks," Computer Law & Security Review, vol. 26, pp. 535-546, Sep. 2010, doi: http://dx.doi.org/10.1016/j.clsr.2010.07.006.

[9] J. Fogel and E. Nehmad, "Internet social network communities: Risk taking, trust, and privacy concerns," Computers in Human Behavior, vol. 25, pp. 153–160, Jan. 2009, doi:10.1016/j.chb.2008.08.006.

[10] F. Stajano and P. Wilson, "Understanding Scam Victims: Seven Principles for Systems Security," Communications of the ACM, vol. 54, pp. 70-75, Mar. 2011, doi :10.1145/1897852.1897872.

[11] A. Ledbetter, J. Mazer, J. DeGroot, K. Meyer, Y. Mao, and B. Swafford, "Attitudes toward online social connection and self-disclosure as predictors of Facebook communication and relational closeness," Communication Research, vol. 38, pp. 27–53, Feb. 2011, doi: 10.1177/0093650210365537.

[12] M. Lucas, and N. Borrisov, "flyByNight: Mitigating the Privacy Risks of Social Networking," Proc. of the 7th ACM workshop on Privacy in the electronic society (WPES '08), ACM, Oct. 2008, pp. 1-8, doi10.1145/1456403.1456405

[13] S. Hindujaa and J. Patchin, "Personal information of adolescents on the Internet: A quantitative content analysis of MySpace," Journal of Adolescence, vol. 31, pp. 125-146, Jan. 2008, doi:10.1016/j.adolescence.2007.05.004.

[14] N. Ellison, C. Steinfield, C, and Lampe. C, "Benefits of Facebook "friends:" Social capital and college students' use of online social network sites," Journal of Computer-Mediated Communication, vol. 12, pp.1143-1168, Aug. 2007, doi: 10.1111/j.1083-6101.2007.00367.x.

[15] Y. Yum and K. Hara, "Computer-mediated relationship development: A cross-cultural comparison," Journal of Computer-Mediated Communication, vol. 11, pp. 133–152, Aug. 2006, doi: 10.1111/j.1083-6101.2006.tb00307.x.

[16] H. Jelicic, D. Bobek, E. Phelps, and R. Lerner, "Using positive youth development to predict contribution and risk behaviors in early adolescence: Findings from the first two waves of the 4-H Study of Positive Youth Development," International Journal of Behavioral Development, vol. 31, pp. 263–273, May. 2007, doi: 10.1177/0165025407076439.

[17] J. Huang, D. Jacobs, D. Derevensky, J. Gupta, R, and T. Paskus, "Gambling and health risk behaviors among US college student-athletes: Findings from a national study," Journal of Adolescent Health, vol .40, pp. 390-397, May. 2007, doi:10.1016/j.jadohealth.2006.11.146.

[18] SocialMediaToday. *Social Media in 2013: By the Numbers*. [Online]. Available from: http://www.socialmediatoday.com/content/social-media-2013-numbers

[19] S. Abraham and I. Chengalur-Smith, "An overview of social engineering malware: Trends, tactics, and implications," Technology in Society, vol. 32, pp. 183-196, Aug. 2010, doi: 10.1016/j.techsoc.2010.07.001.

[20] South Korea. Organisation for Economic Co-operation and Development. Scoping Paper on Online Identity Theft [Online]. Available: http://www.oecd.org/sti/40644196.pdf 2008.01.09

[21] K. Saunders and B. Zucker, "Counteracting Identity Fraud in the Information Age: The Identity Theft and Assumption Deterrence Act," Cornell Journal of Law and Public Policy, vol. 8, pp. 661, Spring 1999

[22] T. Hemphill, "Identity Theft: A Cost of Business?," Business and Society Review, vol. 106, pp. 51-63, Dec. 2001, doi:10.1111/0045-3609.00101.

[23] Australian Crime Commission. *Organised Crime in Australia 2011*. [Online]. Available: https://www.crimecommission.gov.au/sites/default/files/oca2011.pdf 2011.01.28

[24] Criminal Code Act 1995 (Cth) div 372 (1)(a).

[25] P. Powale and G. Bhutkar, "Overview of Privacy in Social Networking Sites (SNS)," International Journal of Computer Applications, vol. 74, pp 39-46, July 2013.

[26] C. Dwyer, S. Roxanne, and K. Passerini. "Trust and privacy concern within social networking sites: A comparison of Facebook and Myspace," Proc. 2007 Thirteenth Americas Conference on Information Systems (*AMCIS 2007)*, Keystone, Aug. 2007, pp. 1-13.

[27] United Kingdom. Cabinet Office. *Identity Fraud: A Study*. [Online]. Available from: http://www.statewatch.org/news/2004/may/id-fraud-report.pdf 2013.11.10

[28] International Civil Aviation Organization. *Towards Better Practice in National Identity Managemen*t. [Online]. Available from: http://www.icao.int/Meetings/TAG-MRTD/Documents/Tag-Mrtd-19/TagMrtd19-wp03.pdf 2013.10.09

[29] Bureau of Justice Statistics. Victims of Identity Theft, 2012 (2013) http://www.bjs.gov/index.cfm?ty=pbdetail&iid=4911 http://www.antifraudcentre-centreantifraude.ca/english/documents/Annual%202011%20CAFC.pdf 2013.12.12.

[30] National Crime Justice Reference Service. *Identity Theft – Facts and Figures.* [Online]. Available from: https://www.ncjrs.gov/spotlight/identity_theft/facts.html 2013.10.13.

[31] Australian Bureau of Statistics. *Personal fraud costs Australians $1.4 billion*. [Online]. Available from: http://www.abs.gov.au/ausstats/abs@.nsf/mediareleasesbytitle/B634CE9C7619C801CA25747400263E7E?OpenDocument 2012.04.11

[32] CIFAS. *Is Identity Fraud Serious?* [Online]. Available from: http://www.cifas.org.uk/is_identity_fraud_serious 2012.06.06

[33] Symantec. *Social Engineering Fundaments, Part 1: Hacker Tactics. [Online]. Available from:* http://www.symantec.com/connect/articles/social-engineering-fundamentals-part-i-hacker-tactics 2001.12.18.

[34] Australian Bureau of Statistics. *Personal Fraud 2010-2011*. [Online]. Available from: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/65767D57E11FC149CA2579E40012057F?opendocument 2013.11.15.

[35] Carnegie Mellon. *Child Identity Theft. 2011* [Online]. Available from: http://www.cylab.cmu.edu/files/pdfs/reports/2011/child-identity-theft.pdf 2011.03.11

[36] S. Willis and B. Tranter, "Beyond the "Digital Divide": Internet Diffusion and Inequality in Australia," Journal of Sociology, vol. 42, pp. 43-59, Mar. 2006, doi:10.1177/1440783306061352.

[37] CALPIRG Education Fund. *Policing Privacy: Law Enforcement's Response to Identity Theft.* [Online]. Available from http://www.calpirg.org/sites/pirg/files/reports/policingprivacy2003.pdf 2003.05.01.

[38] Ran Wei, "Lifestyles and New Media: Adoption and Use of Wireless Communication Technologies in China," New Media & Society, vol. 8, pp. 991-1008, Dec. 2006, doi: 10.1177/1461444806069879.

[39] K. Anderson, "Who Are the Victims of Identity Theft? The Effect of Demographics," Journal of Public Policy & Marketing, vol. 25, Fall, 2006.

[40] J. Paradiso, J. Heidemann, and J. Zimmerman, "Hacking involves the unauthoried interaction with computer systems," IEEE Pervasive Computing, vol. 7, pp. 13-15, July. 2008.

[41] L. Plowman, O. Stevenson, C. Stephen, and J. McPake, "Preschool children's learning with technology at home,"

Computers & Education, vol. 59, pp. 30-37, Aug. 2012, doi:http://dx.doi.org/10.1016/j.compedu.2011.11.014

[42] United States. State of New Jersey Commission of Investigation and Attorney General of New Jersey. *Computer Crime: a Joint Report; 2000*. [Online]. Available from: http://csrc.nist.gov/publications/secpubs/computer.pdf 2014.06.12

[43] S. Livingstone and E. Helsper, "Parental mediation and children's Internet use," Journal of Broadcasting and Electronic Media, vol. 52, pp. 581–599, Dec. 2008, DOI: 10.1080/08838150802437396.

[44] M. Perl, "It's Not Always About the Money: Why the State Identity Theft Laws Fail to Adequately Address Criminal Record Identity Theft," Journal of Criminal Law and Criminology, vol. 94, pp.169-208, Fall, 2003.

[45] D. Lacey and S. Cuganesan, "The Role of Organizations in Identity Theft Response: The Organization-Individual Victim Dynamic," The Journal of Consumer Affairs, vol. 38, pp. 244-261, Jul. 2004, doi:10.1111/j.1745-6606.2004.tb00867.x.

[46] Kroll Advisory Solutions. *Global Fraud Report: The Strategic Impact of Fraud*, Regulation, and Compliance [Online]. Available from: http://www.krollconsulting.com/media/pdfs/KRL_FraudReport2011.pdf 2011.12.05

[47] R. Smith, *Addressing Identity-Related Fraud*. [Online]. Available from: http://www.aic.gov.au/about_aic/research_programs/staff/~/media/conferences/other/smith_russell/2003-09-identity.ashx 2012.12.09

[48] B. Fitzgerald, A. Fitzgerald, T. Beale, Y. Lim, and G. Middleton, Internet and C-Commerce Law: Technology, Law and Policy. Pyrmont: Lawbook Co, 2007.

[49] S. Brenner, "Cybercrime Investigation and Prosecution: The Role of Penal and Procedural Law," Murdoch University Electronic Journal of Law, vol. 8, pp 1-40, June 2001, pp. 1.

[50] J. Clough, "The Council of Europe Convention on Cybercrime: Defining 'Crime' in a digital world," Criminal Law Forum, vol. 23, pp. 363-391, Dec. 2012, doi:10.1007/s10609-012-9183-3.

[51] Council of Europe. *Convention on Cybercrime: Member States of the Council of Europe – Article 12* [Online]. Available from: http://conventions.coe.int/Treaty/Commun/ChercheSig.asp?NT=185&CM=&DF=&CL=ENG 2011.07.15

[52] K. Grewlich, Governance in 'Cyberspace' Access and Public Interest in Global Communications, Boston, USA: Klewer Law International, 1999.

[53] Universal Declaration of Human Rights, GA Res 217A (III), UN GAOR, 3rd sess, 183rd plen mtg, UN Doc A/810 (10 December 1948).

[54] P. Argy, "Internet Content Regulation: an Australian Computer Society Perspective," University of New South Wales Law Journal, vol. 23, pp. 265-267, July. 2000.

[55] R. Mercuri, "Scoping identity theft,", Communications of the ACM, vol. 49, pp 17-21, May. 2006, Doi: 10.1145/1125944.1125961.

[56] W. Kim, O. Jeong, C. Kim, and J. So, "The dark site of the Internet: Attacks, costs and responses," Information Systems, vol 36, pp. 675-705, May. 2011, doi:10.1016/j.is.2010.11.003.

[57] Australian Government: Office of the Australian Information Commissioner. *Scanning "Proof of Identity" Document*s [Online]. Available from: http://www.privacy.gov.au/materials/types/infosheets/view/6553 2007.08.06

[58] C. Blakesley, "United States Jurisdiction over Extraterritorial Crime," The Journal of Criminal Law and Criminology, vol. 73, pp.1109-1163, January 1982.

[59] Federal Bureau of Prisons. *Recidivism among Federal Prisoners Released in 1987* [Online]. Available from: http://www.bop.gov/resources/research_projects/published_reports/recidivism/oreprrecid87.pdf 1994.08.22

[60] Organisation for Economic Co-operation and Development. *OECD Policy Guidance on Online Identity Theft* [Online]. Available from: http://www.oecd.org/dataoecd/49/39/40879136.pdf 2008.06.06

[61] B. Schneier, Secrets and Lies: Digital Security in a Networked World. New York, USA: John Wiley, 2000.

[62] J. Lynch, "Identity Theft in Cyberspace: Crime Control Methods and Their Effectiveness in Combating Phishing Attacks," Berkeley Technology Law Journal, vol. 20, pp. 266-67, January 2005.

[63] B. Krishnamurthy and C. Wills, "On the Leakage of Personally Identifiable Information Via Online Social Networks," Proc. Of the 2nd ACM workshop on Online Social netowrks (WOSN'09) ACM, Aug. 2009, pp. 7-12, doi:10.1145/1592665.1592668.

[64] R. Sullivan, "Can Smart Cards Reduce Payments Fraud and Identity Fraud," Economic Review, vol. 93, pp. 35-62,.Third Quarter, 2008.

[65] G. Weir, F. Toolan, and D. Smeed, "The Threats of Social Networking: Old Wine in New Bottles?," Information Security Technical Reports, vol. 16, pp.38-43, January 2011.

[66] Trustwave, Global Security Report 2011 Trustwave. Available: http://www.secretservice.gov/Trustwave_WP_Global_Security_Report_2011.pdf 2011.05.01

[67] D. Rosenblum, "What Anyone Can Know: The Privacy Risks of Social Networking Sites," Security & Privacy, vol 5, pp.40-49, May. 2007, doi: 10.1109/MSP.2007.75.

[68] A. Nosko, E. Wood, and S. Molema, "All about me: Disclosure in Online Social Networking profiles: The case of FACEBOOK," Computers in Human Behaviour, vol 26, pp. 406-418, May. 2010, doi:10.1016/j.chb.2009.11.012.

[69] G. Gackenbach and J. Karpen, J, "The Coevolution of Technology and Consciousness," in Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications, J. Gackenbach. Ed. California: Academic Press, 337-357, 1998.

[70] G. Newman, "Policy Thoughts on "Bounded Rationality of Identity Thieves,"' Criminology & Public Policy, vol 271, pp. 271-278, Jun. 2009, doi: 10.1111/j.1745-9133.2009.00562.x.

[71] Rand Europe, *Comparative Study on Legislative and Non Legislative Measures to Combat Identity Theft and Related Crime: Final Report*. [Online]. Available from: http://ec.europa.eu/dgs/home-affairs/e-library/documents/policies/organized-crime-and-human-trafficking/cybercrime/docs/rand_study_tr-982-ec_en.pdf 2011.0601

[72] Centre for Problem-Oriented Policing. *The Problem of Identity Theft*. [Online]. Available from: http://www.popcenter.org/problems/identity_theft/ 2014.06.10

[73] S. Livingstone, K. Ólafsson, and E. Staksrud, "Risky Social Networking Practices Among "Underage" Users: Lessons for Evidence-Based Policy," Journal of Computer-Mediated Communication, vol.18, pp.303-320, Apr. 2013, doi: 10.1111/jcc4.12012.

[74] A. Cassese, International Law. Kansas, USA: Oxford University Press, 2001.

[75] FirstMonday. *Friends, friendsters, and Top 8: Writing community into being on social network sites* First Monday. Available from: http://firstmonday.org/issues/issue11_12/boyd/index.html 2006.12.9

[76] Australian Government. *Identity Theft*. [Online]. Available from:

http://www.scamwatch.gov.au/content/index.phtml/tag/identitytheft 2011.03.11

[77] Australian Institute of Criminology. *Examining the Legislative and Regulatory Controls on Identity Fraud in Australia* [Online]. Available from: http://www.aic.gov.au/media_library/conferences/other/smith_russell/2002-07-fraud.pdf 2002.07.12

[78] Parliament of Australia. *House of Representatives Standing Committee on Communication Chapter 6: Criminal and Law Enforcement Framework* [Online]. Available from: http://parlinfo.aph.gov.au/parlInfo/search/summary/summary.w3p;adv=yes;orderBy=customrank;page=0;resCount=Default;query=Criminal+and+Law+Enforcement+Framework 2010.06.01

[79] Council of Europe. *Internet-Related Identity Theft Discussion Paper*. [Online]. Available from: http://www.coe.int/t/dghl/cooperation/economiccrime/cybercrime/documents/reports-presentations/567%20port%20id-d-identity%20theft%20paper%2022%20nov%2007.pdf 2007.11.17

[80] US Department of Justice. *Victims of Identity Theft, 2008* [Online]. Available from: http://bjs.ojp.usdoj.gov/content/pub/pdf/vit08.pdf 2008.08.01

[81] Australian Institute of Criminology. *Counting the Costs of Crime in Australia* [Online]. Available from: http://www.aic.gov.au/documents/A/A/3/%7BAA329573-5D62-46FB-9E6F-4D86A6DDD9BC%7Dti247.pdf 2003.04.10

Managing Language Diversity Across Cultures: the English-Mongolian Case Study

Amarsanaa Ganbold
School of Information Technology
National University of Mongolia
Ulaanbaatar, Mongolia
amarsanaag@disi.unitn.it

Feroz Farazi
DISI
University of Trento
Trento, Italy
farazi@disi.unitn.it

Moaz Reyad
DISI
University of Trento
Trento, Italy
reyad@disi.unitn.it

Oyundari Nyamdavaa
School of Information Technology
National University of Mongolia
Ulaanbaatar, Mongolia
oyundari.n@gmail.com

Fausto Giunchiglia
DISI
University of Trento
Trento, Italy
fausto@disi.unitn.it

***Abstract*—Developing ontologies from scratch appears to be very expensive in terms of cost and time required and often such efforts remain unfinished for decades. Ontology localization through translation seems to be a promising approach towards addressing this issue as it enables the greater reuse of the ontological (backbone) structure. However, during ontology localization, managing language diversity across cultures remains as a challenge that has to be taken into account and dealt with the right level of attention and expertise. Furthermore, reliability of the provided knowledge in the localized ontology is appearing as a non-trivial issue to be addressed. In this paper, we report the result of our experiment, performed on approximately 1000 concepts taken from the space ontology originally developed in English, consisted in providing their translation into Mongolian.**

Keywords: Ontology localization, space ontology, space domain, ontology, Semantic Web, knowledge, provenance

I. INTRODUCTION

This paper is a long version of [1], in which it is described that building a true, flourishing and successful Semantic Web [2] should involve the participation from all cultures and languages across the world. In the development of the traditional Web, this participation was spontaneous and has been made possible as the necessary tools and resources were available. In the Semantic Web one crucial feature is the capacity to assign precise meaning to words, for instance in order to diminish the impact of polysemy. Still for many languages, one example being Mongolian, such resources are not developed at all and for some others what is out there cannot be used effectively as they could not achieve critical mass. However, for English much progress has been made and the WordNet (http://www.princeton.edu) developed at Princeton is one of the well-known and most widely used resources in the field. Yet its coverage is often unsatisfactory when dealing with domain specific tasks [3].

Towards solving the issue of the lack of coverage and to gain a critical mass of concepts, some domain ontologies have already been developed. A prominent example is the *space ontology* [4] developed in English with comparatively very large coverage of geo-spatial features and entities around the globe. Domain ontologies can also deal with the specificity of an area of knowledge, for example, by providing relations and attributes specific to the domain. By reducing polysemy (the amount of words with same meaning), they can enable better semantic interoperability.

Ontologies that are developed to perform NLP tasks in one language can hardly be used with their full potential for another. Representing an existing ontology in a new language, taking into account cultural and linguistic diversities, is defined as ontology localization.

In this paper, we describe the development of the space ontology in Mongolian starting from its English counterpart as it is contained in the Universal Knowledge Core (UKC), as described below. Building an ontology without human-level accuracy is a potential obstacle in developing applications (e.g., word sense disambiguation and document classification). Synset base resources (linguistic representation of ontologies) such as WordNet and FinnWordNet [5] are built manually to obtain better quality.

Knowledge is created to be consumed by others in a multitude of activities including daily life, education, research and development for the advancement of our society. Through ontology development and ontology localization we create new knowledge. Trustworthiness and reliability of the produced knowledge are crucial measures that if comprehended, modelled and communicated properly would make consumers lives comparatively easier.

Being concerned about the quality and giving utmost importance to it, we followed a manual approach. The contributions of our paper include:

i) The development of an ontology localization methodology that is domain and language independent and seems to achieve very high quality
ii) The development of a methodology for dealing with diversity (e.g., lexical gaps) across cultures and languages
iii) The lessons learned from the execution of the whole process in the generation of the space ontology in Mongolian
iv) The development of the provenance model that manages information about various contributors to the ontology localization process for ensuring quality and credibility of the knowledge produced.

The paper is organized as follows. In Section 2, we provide detailed description of the UKC. Section 3 gives an overview of the space ontology. In Section 4, we describe the macro-steps of the translation process. In Section 5, we present the provenance model. In Section 6, we describe the diversity across English and Mongolian cultures in terms of space related features. Section 7 reports the experimental results. Section 8 discusses the lessons learned while Section 9 describes the related work. Finally, in Section 10, we provide the concluding remarks.

II. THE UNIVERSAL KNOWLEDGE CORE

The UKC [4] is a large-scale ontology, under development at the University of Trento, which includes hundreds of thousands of concepts (e.g., lake, mountain chain) of the real world entities (e.g., Lake Garda, Alps). It consists of three main components: *domain core*, *concept core* and *natural language core* (See Fig. 1).

Domain core: As described in [4], the domain core consists of various **domains,** where each of them represents an area of knowledge or field of study that we are interested in or that we are communicating about [6]. In other words, a domain can be a conventional subject of study (e.g., mathematics, physics), an application of pure disciplines (e.g., engineering, mining), the aggregation of such fields (e.g., physical science, social science) or a daily life topic (also called Internet domains, e.g., sport, music). Each domain is organized in **facets,** where a facet can be defined as a hierarchy of homogeneous concepts describing the different aspects of meaning [7]. According to our methodology [8], called DERA, where D stands for Domain, facets are classified into three categories: Entity class (E), Relation (R) and Attribute (A). For example, in the space ontology, country and continent are **entity classes**. **Relations** describe relations between entities; examples of spatial relations are near, above, far etc. An **attribute** is a property of an entity, e.g., depth of a lake.

Concept core: The concept core consists of concepts and semantic relations between them. The concepts in the concept core form a directed acyclic graph, which provides the terms and the structure from which facets are defined. Entity class, relations and attributes are all codified as concepts. A **concept** is a language independent representation of a set of words (synset), which are synonym of a given word in natural language. For example, country, city, etc. The concept *city* can be represented as *city* in English, *città* (chit'a) in Italian, *хот* (khot) in Mongolian.

A semantic relation is a relation that holds between two concepts. Some examples of semantic relation are is-a (or *hyponym-of*), part-of (part-meronym-of) and value-of. An instantiation of the is-a relation can be given as *city* is-a *populated place*.

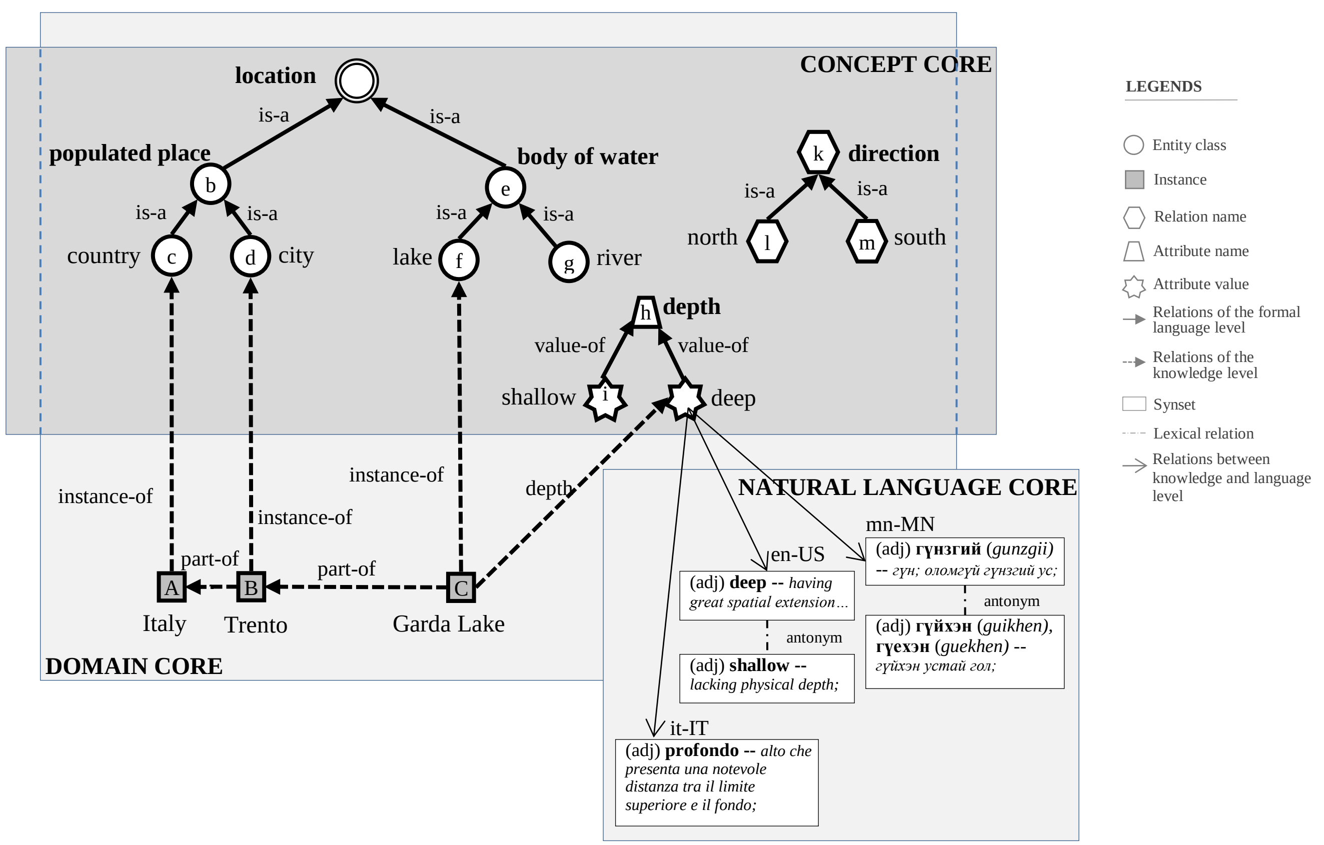


Figure 1. Knowledge Organization in the UKC

Natural language core: The natural language core consists of a set of languages, each representing a set of linguistic objects and relations between them. The objects of this core are words, senses, synsets and exceptional forms. A **word** is the basic lexical unit of the natural language core represented as a lemma. It can be multiword, phrase, collocation, etc. The words in the natural language core provide, for any given language, the translation of the concepts stored in the concept core.

Word senses are organized into four part-of-speeches -- noun, verb, adjective and adverb. One word may have more than one part-of-speech, and synonym word senses with the same part-of-speech are grouped into a synset. A **sense** is a possible meaning for a word. A word can have one or more senses each having a part-of-speech tag. Each sense corresponds to and belongs to only one synset. All senses of a given word are ranked according to most preferred usage. A **synset** is a set of words, which share the same meaning. In fact, words in a synset have semantically equivalent relations. Each synset might be accompanied by a gloss consisting of a definition and optionally example sentences.

Relations of the language core are of type lexical and semantic lexical. This kind of relations holds between the objects of the same language.

A lexical relation is a relation that holds between the words of different synsets. Antonym, derivationally-related-form and also-see are examples of such relation. An example of the antonym relation can be provided as *lowland* is an antonym of *highland*. Note that hereinafter we represent synsets with double hyphens distinguishing words (comma separated) from glosses, which are formatted in *Italics*.

(a) lowland -- *low level country*

(b) highland, upland -- *elevated (e.g., mountainous) land*

The word *highland* of the synset reported in (b) is in antonymy relation with the word *lowland* of the synset (a). Notice that the same relation does not hold between the other word *upland* of the synset (b) and *lowland*.

A semantic-lexical relation is a relation that holds between two synsets. Some examples of this kind of relation are similar-to, troponymy and verb-group. An example of the semantic-lexical relation can be *adjacent* is similar-to *near*.

(c) adjacent -- *near or close to but not necessarily touching*

(d) near, close, nigh -- *not far distant in time or space or degree or circumstance*

In this case the synset (c) that consists of only one word is in similar-to relation with the synset (d) that consists of three words. This means that the very same relation can be applied between any word of the synset (c) and any word of the synset (d).

The natural language core is built with the complete integration of hierarchically organized synset bases, as it is the case, for instance, for WordNet and the Italian part of MultiWordNet (http://multiwordnet.fbk.eu).

III. THE SPACE DOMAIN

The space domain [4], [6] is a large-scale geospatial ontology built using the faceted approach. It was developed as the result of the complete integration of GeoNames (http://www.geonames.org) and WordNet. It is also known as space ontology and in this paper, we refer to it with any of these names. It currently consists of nearly 17 facets, around 980 concepts and 8.5 million entities. The ontology (excluding entities) is integrated into the UKC. Some examples of facet are *land formation* (e.g., mountain, hill), *body of water* (e.g., sea, lake), *administration division* (e.g., state, province) and *facility* (e.g., university, industry).

In Fig. 2, we provide a partial bird's eye view of the whole set of facets. Note that facets are not connected to each other and they do not have concept overlap across or within them.

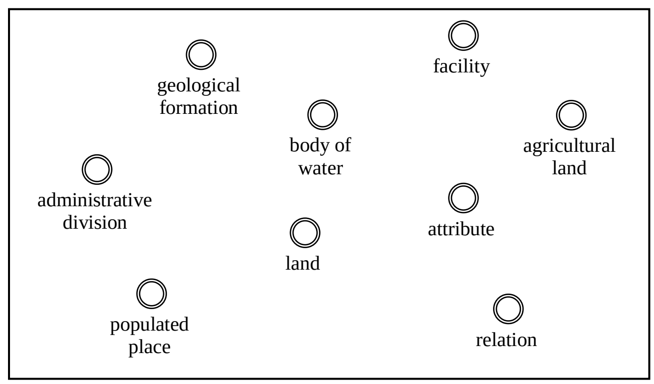


Figure 2. A subset of the facets of the *Space* domain

Fig. 3 shows a small portion of the facet *geological formation* in which the second level represents *natural elevation*, *natural depression* and the level below the *natural elevation* is organized into *oceanic* and *continental elevation*, and so forth.

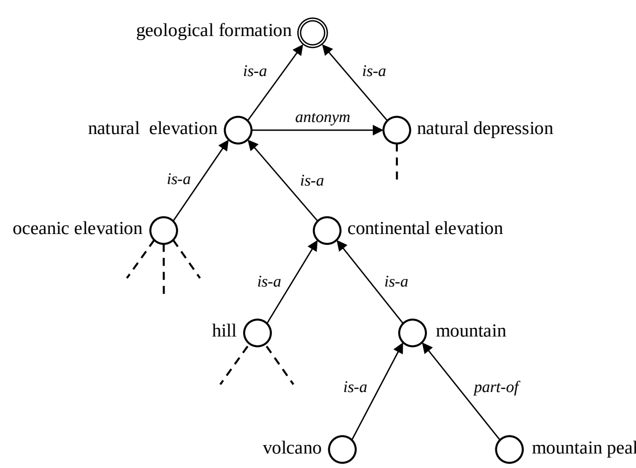


Figure 3. An entity class (E) category facet (partial view)

Note that, within a facet with double circled node we distinguish the root concept from the rest of the concepts that are represented with single circle.

In the Space domain, the *relation* category contains around 10 facets such as *spatial relation* and *primary outflow*. A partial representation of the *spatial relation* facet is shown in Fig. 4.

The *spatial relation* is the spatial property between geological physical objects or the way in which something is located. Leaf nodes of this facet represent relations between entities. For instance, Mongolia is *south* of Russia and *north* of China. The relation *primary outflow* connects two bodies of water.

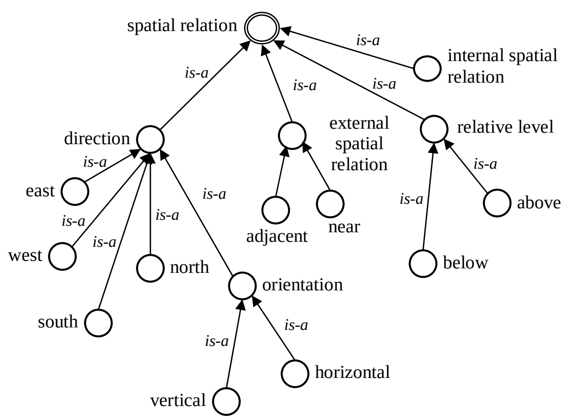


Figure 4. A relation (R) facet (partial view)

Within the domain the *attribute* category consists of around 20 facets such as *rain* and *temperature*.

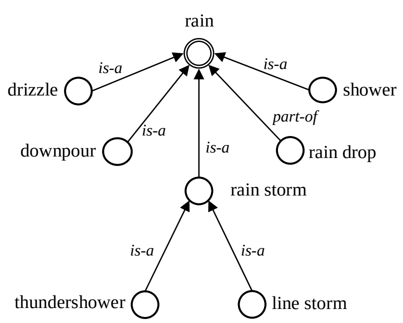


Figure 5. An attribute (A) facet (partial view)

As shown in Fig. 5 the facet *rain* includes among others *rainstorm*, *downpour*, *drizzle* and *shower*. With rain we mean *falling of water in drops from vapor condensed in the atmosphere*. The temperature indicates *the degree of hotness or coldness of an object or environment*.

IV. TRANSLATION APPROACH

In the following subsections we describe the general process for translation and its instantiations both for translating and creating a concept in the target language.

A. General process

The main idea of the translation process is to take the objects of the domain of interest from a source language, in this case English, and to produce the corresponding representation in a target language, e.g., Mongolian in order to extend the UKC with translations. The process includes the translation of the synset words and glosses. A direct translation of them is provided whenever possible. However, the world is full of diversity and people of a particular culture might not be aware of some concepts. For instance, Mongolia is a landlocked country, thus some terms (e.g., *dry dock*, *quay*, *pier*, etc.) related to seaport are not known to the community or are rarely used and are often a source of lexical gaps for Mongolian. Lexical gaps are those concepts that do not have a succinct representation in a given language. However, they can be expressed as a free combination of words [10].

We select English as a master source language for all localization activity since the language is the second most widespread language and will be a common language to use in scientific and research oriented discussions. For executing the translation process, English language representation of the concept core is copied to an LKC (Local Knowledge Core) repository, which contains translations in the target language.

In order to provide the most suitable translation for a synset, we follow the macro-steps described below and represented in Fig. 6.

1(a) A **language translator** takes a synset provided in the source language and gets a clear understanding of its meaning. In case of difficulty, he/she finds the corresponding images or videos of the synset word(s) on the Web to perceive the concept through visualization.

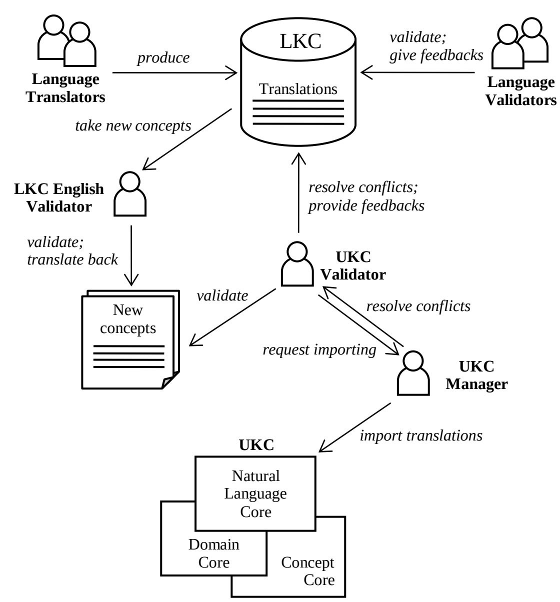


Figure 6. Translation phases of UKC

1(b) The **language translator** provides a suitable translation of the word(s) in the target language.

With suitable we mean word, multiword, co-occurrence and phrasal representation as we do not allow a free combination of words as translation of a word. In case of unavailability of the word(s) for the given meaning, the translator can mark it as a lexical gap. However, the translator always provides the translation of the gloss.

1(c) A **language validator** evaluates the translation of the word(s) and the gloss of the synset. In case the concept is marked as a gap, the validator either confirms the gap or suggests a translation for the word(s).

1(d) Upon receiving feedback on the synset, the **language translator** goes through the comments and updates the translation when necessary. In case of disagreement, the language translator provides comments including mostly the rationale about the disagreement.

1(e) The **language validator** reevaluates the updated translation. In case of disagreement, the validator generates further feedback and sends it back to the language translator (step 5). Even if after a few iterations a disagreement is not resolved, a second language validator is consulted. If agreed upon, the validation for the given synset is over.

2 In the cases where the **language translator** finds out a new concept that might be a lexical gap or a missing concept in the source language, she suggests a suitable synset for this concept in the target language and if possible also the corresponding synset in the source language. The **LKC English validator** evaluates the source language synset (if suggested) for the new concept coming from the target language. Otherwise, she translates back this target language synset into English.

3 A **UKC validator** takes the translations resulting from steps 1 and 2 to evaluate their correctness from both the language and UKC perspectives. The validator corrects the mistakes and resolves the issues (if any) communicating with the language validator and LKC English validator (if necessary), possibly in a few iterations. Finally, she asks a UKC manager for importing the translation to the UKC.

4 The **UKC manager** runs an automatic validation tool to evaluate if the provided input is compliant with the UKC. In case of errors are found, they are corrected with the help of the UKC validator (if needed) possibly iterating a few times. The manager also decides new concepts originating from the target language whether to accept or reject them. Once all the issues are resolved, the UKC manager imports the translations to the UKC.

Following these steps we translated the *space ontology* into Mongolian end-to-end, evaluated and finally imported the translations to the UKC.

To achieve optimal quality while executing the whole process depicted in Fig. 6, we set the criteria that translators and various validators must possess the competences necessary for the task. The language translator should be a native speaker from the country of origin of the target language with a good command of the source language. The language validator should be a linguist possessing the necessary language competences. The LKC English validator should be as close as possible to an English native speaker who should understand well the target language. The UKC validator is a native speaker of the target language with knowledge of the UKC. Both the UKC validator and LKC English validator are in charge of the language dependent tasks in the translation process. The UKC manager is an expert on the UKC with no specific competence on the language.

From a geographical point of view we expect that, in most cases, the language core will be developed in the countries where that language is spoken, while the UKC is and will be developed centrally. The UKC validator, whenever possible, should operate centrally where the UKC manager is. This spatial distribution of operations and operators has been designed as an attempt to preserve local diversity and, at the same time, to deal with the need for central coordination required because of existence of a unique, single UKC. The underlying model is that there is a single world, represented by the UKC, and many different views of the world, each represented by a different natural language. The diversity of the world is therefore captured, as it will be described in detail in the next section, in the mapping from the informal natural languages and the unique UKC formal concept language.

B. Translating a concept

Here we instantiate the general translation process described in section A by taking the concept subtree, rooted at “mountain”, shown in Fig. 3, Translation is performed from English to Mongolian. All three concepts of the subtree are provided as follows:

mountain, mount -- *a land mass that projects well above its surroundings; higher than a hill*

volcano -- *a mountain formed by volcanic material*

peak, crown, crest, top, tip, summit -- *the top point of a mountain or hill*

Note that the subtree should normally get translated according to the macro steps 1, 3 and 4, whose executions are marked with prime (') and described below.

1(a)' The **language translator** perceives the meaning of the concept *mountain* by reading its gloss. The translator could understand the concept as a massive land that is highly raised than surrounding geological formations and it is more elevated than the hills. She checks whether there is at least a term to refer to this notion in Mongolian culture. In case of dilemma in understanding its availability, she also visualizes the meaning by consulting resources (e.g., images, videos) on the web.

1(b)' The **language translator** represents *mountain* as *уул (*uul*)* in Mongolian. In this case, the Mongolian representation is a lexical unit, therefore the concept is not marked as a lexical gap. The gloss is translated as *эргэн тойрон буюу хүрээлэн буй* *орчноосоо дээш өргөгдөн гарсан өндөрлөг газар; толгодоос өндөр* (a high land raised above and elevated from its surroundings and all-around; higher than hills).

1(c)' The **language validator** agrees with the term provided as the name (or label) of the concept, but she suggests an improved translation of the gloss as *эргэн тойрны орчноосоо дээш өргөгдсөн өндөрлөг газар; дов толгодоос өндөр;* (a high land raised above from its surroundings; higher than hills). Here, the **language validator** removes the words *эргэн тойрон* (all-around) and *гарсан* (elevated) from the gloss developed by the **language translator** because from her point of view without these terms in the gloss, the concept can clearly be understood by the native speakers.

1(d)' The **language translator** receives the validation feedback on the translation of the concept *mountain*. She accepts the validated result and updates the translation according to the language validator's comments.

1(e)' As the **language translator** accepts the modifications proposed by the language validator, no conflict is left to be resolved and the validator proceeds with the next steps.

3' The UKC validator checks the translation of the terms and glosses of the concept. Since there are no disagreements from both the language and UKC perspectives, she asks a UKC Manager to import the translations into the UKC.

4' Finally, the UKC Manager runs an automatic import function to integrate the translation into the UKC.

Similarly, the other two concepts *volcano* and *peak* are translated and then integrated into the UKC.

C. Adding a new concept

New concepts can be added executing macro steps 2, 3 and 4 of the general process of translations. The instantiations of these steps are marked with prime ((') and described below.

2' The **language translator** realizes that in the given subtree a concept, which is part-of the concept *mountain*, is missing. Therefore she proposes to create a concept and develops a synset for this in Mongolain as follows:

гэзэг (gezeg) -- *уул толгодын ар шил*

The **LKC English validator** recognizes that this concept is a gap in English and translates its gloss back to English as follows:

GAP -- *northern ridge of a mountain*

3' The UKC validator verifies the translation of the gloss, confirms that it is a gap in English and evaluates the correctness of the added relation of the concept in the subtree. As she conceded with the produced knowledge from linguistic and UKC viewpoints, she proceeds through asking the UKC Manager to incorporate the translations into the UKC.

4' The UKC Manager takes the necessary steps for the integration of the new knowledge to the UKC.

Note that, the UKC infrastructure was developed taking into account the fact that new knowledge can come to the system at any point in time. Therefore, it supports integration of new objects (e.g., concepts and relations). UKC manages the lifecycle and evolution of the objects by exploiting the timestamp attached to them.

V. Handling Provenance

Provenance can be defined as the source of a piece of knowledge. In the context of this work, we assign fine grained provenance in terms of knowledge objects and their contributors.

For ensuring trustworthiness and reliability, we employ a provenance model that is defined and then instantiated with the use cases -- concept translation and concept addition -- in the following subsections.

A. Representing provenance

In this section, we present the data structure that we use to represent the provenance. Our provenance model is designed to maintain information about the elements that can be created in the knowledge base during the ontology localization process. These elements are concepts, lexical gaps, lexical relations, semantic lexical relations, synsets, senses, words and sense ranks.

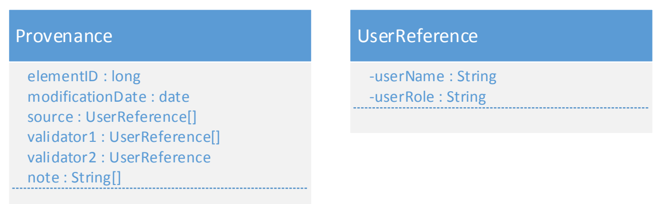


Figure 7. UML Diagram for provenance

Fig. 7 shows the UML diagram of our provenance model, which consists of two classes – *Provenance* and *UserReference*. *Provenance* represents information about the source of an element and *UserReference* represents a human user who is involved in the localization process.

Provenance is a 5-tuple <*elementID*, *modificationDate*, *source*, *validator1*, *validator2*, $note^{op}$>, where *elementID* is the unique identifier of the knowledge base element to which the provenance is applied, *modificationDate* is a timestamp specifying the latest date of modification of the provenance, *source* refers to a list of contributors who are language translators responsible mainly for translating an element and also for proposing a missing concept in the target language, *validator1* refers to a list of contributors who are language validators responsible for validating translations, *validator2* is the UKC validator and $note^{op}$ refers to the additional remarks that can optionally be provided by LKC developers/ validators or UKC validator/manager.

UserReference is a pair <*userName*, *userRole*>, where *userName* represents the name and email address of a human user and *userRole* indicates one of the following editorial roles: *LKC_Developer, LKC_Validator* and *UKC_Validator, UKC manager*.

Note that, in this provenance model, for the translation tasks, there are references to at least three human contributors – *source*, *validator1* and *validator2*. That means each manually translated element is validated by one or more user, who can also be communicated via email to discuss the rationale behind the translations and concepts developed by them. This we believe is the main strength of our provenance modelling and that helps increase the reliability of the ontologies localized in any target language following our approach.

B. Provenance in concept translation

In the concept translation macro-steps, we create and update provenance in the following cases.

a. If a language translator develops a word, a synset, a lexical gap or a concept in Steps 1(a)-1(b) or Step 2, for each element created, a new provenance will be generated with the *source* referring to the translator.
b. If a language validator confirms an already developed element in Step 1(c) and 1(e), the provenance of that element is updated by linking *validator1* to the instance of the UserReference, which corresponds to the name and email and the role of the language validator that is LKC_Validator. This marks the element as validated.
c. As soon as the UKC validator confirms an element in Step 3, the provenance of that element is updated with the instantiation of the attribute *validator2* referring to the name and email and role of the validator (i.e., UKC_Validator) of the given context. This marks the element as completely validated and accepted.

Note that the modification date changes whenever a new operation is performed on the provenance. In (a) it refers to the date of translation, in (b) it is updated with the date of the language validation and in (c) it is replaced with the date of the UKC validation.

C. Provenance in concept addition

While adding concept, we create and update provenance in the following possible scenarios.

a. In the concept addition phase in Step 2, the language translator may create a new concept and its related lexical components such as synset, word, etc. in the target language and optionally in the source language In this case the provenance *source* for each of the objects will be instantiated with the translator for her LKC developer role.
b. Again in Step 2, (i) if the LKC English validator evaluates the source language synset provided by the language translator, the provenance *validator1* is instantiated with her LKC Validator role (ii) if it happens that the LKC English validator translates back the new concept into the source language, in this case the *source* is filled in with her role as LKC Developer and the *validator1* is left un-instantiated.
c. Similarly to the concept translation, a UKC validator checks the correctness of the concept addition and she becomes *validator2* with the corresponding role.

VI. TYPES OF DIVERSITY

The translation or localization is the adaptation of a piece of knowledge to a particular language and culture [9]. This is nontrivial and linguistic experts might help in this task. Moreover, the localization should be based on the perception of the concepts and entities in the real world within the local communities and not on the literal translation.

A. Concepts

We assume concepts to be universal. However, their representation in the different natural languages changes. Within the same language a concept might be referred with multiple terms (known as synonymy) and multiple concepts might be referred with the same term (known as polysemy).

The concepts *valley*, *dale* and *hollow* are represented with the same term in Mongolian.

valley *– (a long depression in the surface of the land that usually contains a river)*

dale *– (an open river valley (in a hilly area))*

hollow *– (a small valley between mountains; "he built himself a cabin in a hollow high up in the Appalachians")*

Moreover, in the UKC *dale* and *hollow* are subordinate concepts of *valley*. In this case, translating them into the target language increases polysemy. However, we translate them because within the Mongolian culture people can classify their (real world) entities under the specific concept.

Moreover, a concept might not have a name in a target language the fact that it can be a lexical gap. For example, the concept *parish - (the local subdivision of a diocese committed to one pastor)* is a lexical gap in Mongolian. The variation in the concept lexicalization from the source language (S) to the target language (T) is depicted in Fig. 8(a).

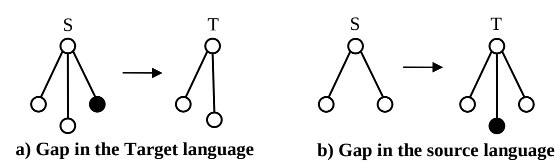


Figure 8. Variations of concept localization

As the lexical gap is a feature of the languages, it does happen with all of them. There can be a gap also from the target to source language. For instance, the Mongolian words **бууц** (buuts) and **буйр** (buir) are gaps in English. The word *buuts* can be represented in English as *an area of dried and accumulated manure where a nomadic family was living* and the word *buir* can be represented in English as *a round shaped spot where a nomadic yurt was built*. Note that these words lack a succinct representation in English. Therefore we consider them as gaps. This phenomenon is drawn in Fig. 8(b).

The nomadic lifestyle of Mongolians is the source of these concepts that are not used in the English speaking cultures across the globe.

Words pointing to lexical-gap concepts might appear also in the glosses. For instance, the term *piers* appearing in the gloss of *Romanesque architecture* is a lexical gap in Mongolian. In such cases, the translation is produced with a free combination of words.

> *Romanesque architecture – (...characterized by round arches and vaults and by the substitution of* ***piers*** *for columns and profuse ornament and arcades)*

B. Senses

In the space ontology, some words have multiple senses that have subtle difference in meaning. For instance, the word fissure has two senses:

> [S1]: *crack, cleft, crevice,* ***fissure****, scissure – (a long narrow opening)*
>
> [S2]: ***fissure*** *– (a crack associated with volcanism)*

The two concepts associated with the given word are hyponyms of *continental depression* and they can be represented with the same word(s) in the target language. This phenomenon is shown in Fig. 9(a).

Polysemous words in the source language might correspond to lexical gaps for a subset of senses. For instance, *gorge* has two senses within Space ontology and one of them is a gap as depicted in Fig. 9(b), where 'mn' and 'en' denote Mongolian and English, accordingly.

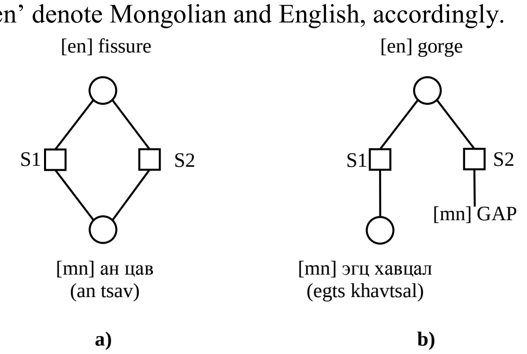


Figure 9. Word sense diversity

C. Synsets

Words in a synset can be directly translated into the target language. However, for some of them there might be a lack of translation. For example, the synset *mountain peak (the top point of a mountain or hill)* has 6 words of which 3 of them lack translation into Mongolian as shown below.

1 peak	→	оргил (ogril)
2 crown		
3 crest		
4 top	→	орой (oroi)
5 tip		
6 summit	→	дээд оргил (deed orgil)

In gloss paraphrasing, some parts of the glosses sometimes are obtained using words with a very close or similar meaning instead of exact translation. Though our first preference is to provide the exact translation, in many cases this could not be achieved. The following example shows a paraphrased translation where the phrase "near a shore" is eliminated from Mongolian version. In this situation, there is no difference between bank and shore in Mongolian language.

> [in English] ***oceanic sandbank*** *– a submerged bank of sand near a shore, can be exposed at low tide*
>
> [in Mongolian] ***далайн элсэн эрэг*** (gl. oceanic bank of sand) *– шунгаж орсон далайн элсэн эрэг, далайн давалгааны намхан хаялганд үзэгддэг (gl. a submerged sea bank of sand, visible at low tide)*

Example sentences in glosses were also paraphrased or added newly in order to provide a better explanation. For example, well-known place names are often substituted in the target language because famous names within a culture might give better understanding about a concept being translated. The highest mountain peak of the Alps ridge is Mont Blanc that is substituted with Everest as it is known to the most of the people in the East Asian region. Moreover, symbols are kept in their original forms, e.g., measurement unit symbol, pH.

Date and time format, measurement unit and currency were converted into the ones used regionally. For example, 5 inches is converted into 12.7 centimeters because of the pervasive use of MKS system in Mongolia. Note that these types of words appear only in glosses. However, using these types of word might not be suitable as fractions are less intuitive than whole numbers. For example, 3 feet is converted into 0.9144 meters. Such fractions cannot be mapped easily to the real world entities and most often become tedious to remember.

VII. Results

In this section, we report the results of our experiment. We could translate 91.88% of the concepts of the space ontology into Mongolian and the remaining 8.12% were identified as lexical gaps. In Table I, we report the detailed statistics of the translation task and the obtained results.

In Table I, the number of concepts per facet is shown separately, e.g., administrative division has 18 concepts, agricultural land has 19 concepts and so on. Note that for the sake of space, we group the statistics of all attribute facets as attribute and relational ones under relation.

Language Translators provided the Mongolian translation for 905 concepts *Language Validators* provided feedback on each of the produced synset words and glosses separately that help us achieving better quality. The validation procedure identified 188 disagreed words and 243 disagreed glosses. Cases such as disagreements and modifications for improvement were solved in iterations (as many as needed) between the translators and validators until they reached to an agreement. The highest number of iterations was recorded as 4.

TABLE I. LOCALIZATION RESULT OF THE SPACE DOMAIN

Facets	Concepts	Translated	Disagreed words	Disagreed glosses	Translator Identified Gaps	Finally accepted Gaps	Finally Localized Concepts
administrative division	18	18	2	4	0	0	18
agricultural land	19	19	2	1	0	0	19
attribute	85	73	1	23	12	10	75
barren land	7	7	1	0	0	0	7
facility	357	357	54	64	0	2	355
forest	5	5	5	4	0	0	5
geological formation	200	150	73	87	50	52	148
land	15	15	2	3	0	2	13
plain	12	12	0	0	0	3	9
rangeland	8	8	1	4	0	0	8
region	46	44	6	0	2	2	44
relation	54	54	8	32	0	0	54
wetland	8	8	3	1	0	0	8
abandoned facility	16	15	4	1	1	1	15
body of water	116	106	24	17	10	3	113
populated place	13	10	2	1	3	2	11
seat of government	6	4	0	1	2	2	4
Total number of objects	**985**	**905**	**188**	**243**	**80**	**79**	**906**

Language Validators' evaluation of the lexical gaps revealed that the translators proposed 10 false positives out of 80. We also identified that the translators produced 9 false positive translations of the concepts whereas they are gaps. In the end, we found that there are in total 79 gaps and 906 concept translations being accepted. The *UKC Language validator* and *UKC validator* reported a few (around 5) conflicts, which were then solved with little effort. It is worth mentioning that *Language Translators* proposed to add 7 new concepts to the space ontology. This is only initial work and we expect that a few more concepts will be added with the evolution of the space ontology.

VIII. LESSONS LEARNED

Assigning word sense rank appears as a difficult task to accomplish since the *Language Translators* provide their the results indipendently. In the translation work, they were aware of the fact that concepts translated by others might have the same word label. But it remained obscure until the whole translation task was finished. This ranking could be defined once all the concepts are translated. This is a non-trivial task to accomplish because deciding acceptable ranks might require local community agreement or the consultation of high quality linguistic resources that are often insufficient for domain specific tasks in many languages.

Synonymous words within the synsets were often increased after translations were evaluated by the Language Validators. This was the case since Language Translators concentrate in providing the target language correspondence representation of the knowledge objects taken from the source language within a reasonable amount of time. This often results in the postponement of the addition of synsets.

In the cases where an example sentence in a gloss contains a number that has to be converted according to some suitable measurement, we should freely change values and corresponding units since the numbers always give some extra information to provide glosses. For instance, 6000 meters can be changed to 6 km (while value remains same) and 3 kilograms to 3 pounds (while value modifies). Nevertheless, in case of sensitive information found in a gloss, we should exactly convert the number to relevant measurement unit in order to preserve the meaning of the gloss. For example, for understandable measuring unit of the target users 500 feet can be converted into 152.4 meters.

Parts of the glosses that follow the same syntactic pattern in the source language can be translated with little effort. For instance, the gloss part *a facility for [verb]+ing [object]* appeared in around one tenth of the concepts. We repeated the same translation for the part that matched completely. Moreover, we used the translation memory technique that provides a translation with recurrent structure in the same way as previous translations.

In order to introduce foreign cultures to the community, we can translate lexical gaps as free combination of words. However, this should not always be the case. A first reason is computational: the explicit marking of the lexical gaps could support the KB-based applications in reducing computation time by avoiding the management of (multi)words that will be very rarely or never used. A second, more important reason, is related to the actual existence of a free combination of words capable of capturing, in the mind of a native speaker with no knowledge of the original concept (as it exists in the foreign culture) what the concept actually means, in the real world.

IX. RELATED WORK

MultiWordNet [10] consists of several European language WordNets. It was developed under a model that reuses semantic relations from WordNet as follows: when there are two synsets and a relation holding between them, the same relation is assumed to hold between corresponding synsets in the new language. There is no literal translation in the case of developing Italian version of MultiWordNet of the synsets, words and exceptional forms but the contributors have produced the best possible Italian equivalents according to their skills and experiences in knowledge organization and linguistics. However, a limited number of glosses has been provided, e.g., around 2k in Italian over 33k.

The ontology localization activity described in [11] is an attempt to address the localization and diversity issues. They proposed guidelines and methodologies for enriching ontology with multilingual information. However, we differ from them with respect to the target language and the development approach.

Universal Multilingual Knowledge Base also known as UWN [12] was developed leveraging on the Wikipedia data and linking multilingual terms that are connected to the same page. However, automatically built KB resources often suffer from quality issues, e.g., around 10% of the terms in UWN are attached to the wrong senses, whereas we achieved human-level accuracy.

FinnWordNet [5] was produced from WordNet with the help of professional translators and the output is monitored by bulk validation. While producing the whole WordNet in Finish in 100 days, they traded off the quality for reducing the amount of translation time. Diversity in the languages such as lexical gaps is overlooked in this task.

Concerning provenance modelling and representation, the PROV-O ontology [15] was developed to be used to trace resources belonging to any domain. Despite its richness and well-coverage in terms of classes and relations, it could fulfill our need only partially.

X. Conclusion

In this paper, we proposed an experiment for generating ontologies through translation from one language into another. This experiment was developed to be applied independently of domain and language and to deal with the diversity across the languages. While translating the ontologies, we identified the various diversity features and their presence in a given target language by working together with the linguistic experts and/or native speakers living in the country where it is spoken. We evaluated the effectiveness of the methodology by performing a case study for translating the space ontology into Mongolian.

Thanks to the reuse of the ontological backbone structure, we achieved space ontology in Mongolian that is as high quality as the original one in English. Though manual approach is usually known to be time consuming, adopting this methodology in a crowdsourcing setting can help increase throughput and make this suitable for dealing with large ontologies. We also have presented a provenance model for ontology localization tasks to keep track of the translators and validators the fact that it helps increase the reliability of each single object of the knowledge base. The generated ontology can be exploited to improve the accuracy of *NLP tasks* [13] and *Concept Search* [14] in space domain. The Mongolian version of space domain is currently in use in the School of Information Technology of the National University of Mongolia as background knowledge in their NLP pipeline.

Acknowledgment

The research leading to these results has received funding (partially) from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n. 600854 Smart Society: hybrid and diversity-aware collective adaptive systems: where people meet machines to build smarter societies http://www.smart-society-project.eu/. The experiment described in this paper is part of a long term project, whose goal is to create the Universal Knowledge Core (UKC), a worldwide multilingual linguistic and ontological resource. The work on provenance is part of the PhD thesis of the third author. We are thankful to Vincenzo Maltese for his valuable feedback.

References

[1] A. Ganbold, F. Farazi, and F. Giunchiglia. "An Experiment in Managing Language Diversity Across Cultures," in eKNOW 2014, pp. 51–57.

[2] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific American, vol. 284, no. 5, 2001, pp. 34–43.

[3] F. Giunchiglia, V. Maltese, F. Farazi, and D. Biswanath, "GeoWordNet: a resource for geo-spatial applications," in ESWC'10, Volume Part I, 2010, no. December 2009, pp. 121–136.

[4] F. Giunchiglia, V. Maltese, and D. Biswanath, "Domains and context: first steps towards managing diversity in knowledge," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 12–13, 2012, pp. 53–63.

[5] K. Lindén and L. Carlson, "FinnWordNet – Finnish WordNet by Translation," LexicoNordica – Nordic Journal of Lexicography, vol. 17, 2010, pp. 119–140.

[6] F. Giunchiglia, B. Dutta, V. Maltese, and F. Farazi, "A facet-based methodology for the construction of a large-scale geospatial ontology," Journal on Data Semantics, vol. 1, no. 1, 2012, pp. 57–73.

[7] F. Giunchiglia, B. Dutta, and V. Maltese, "Faceted Lightweight Ontologies," in in Conceptual Modeling Foundations and Applications, vol. 5600, 2009, pp. 36–51.

[8] F. Giunchiglia, B. Dutta, and V. Maltese, "From Knowledge Organization to Knowledge Representation," in ISKO UK Conference, 2013, pp. 44–56.

[9] M. C. Suárez-Figueroa and A. Gómez-Pérez, "First Attempt towards a Standard Glossary of Ontology Engineering Terminology," in TKE08, 2008, pp. 1–15.

[10] L. Bentivogli and E. Pianta, "Looking for lexical gaps," in EURALEX International Congress, 2000, pp. 663-669.

[11] M. Espinoza, E. Montiel-Ponsoda, and A. Gómez-Pérez, "Ontology localization," in K-CAP '09, 2009, pp. 33-40.

[12] G. De Melo and G. Weikum, "Towards Universal Multilingual Knowledge Bases," in Principles, construction, and applications of multilingual wordnets : proceedings of the Fifth Global WordNet Conference, 2010, pp. 149–156.

[13] I. Zaihrayeu, L. Sun, F. Giunchiglia, W. Pan, Q. Ju, M. Chi, and X. Huang, "From Web Directories to Ontologies : Natural Language Processing Challenges," in ISWC'07/ASWC'07, 2007, no. 60673038, pp. 623–636.

[14] F. Giunchiglia, U. Kharkevich, and I. Zaihrayeu, "Concept search," The Semantic Web Research and Applications, vol. 5554/2009, 2009, pp. 429–444.

[15] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, and J. Zhao. PROV-O: The PROV Ontology, 2013, W3C Recommendation. URL: http://www.w3.org/TR/2013/REC-prov-o-20130430/.

The Diffusion of a Personal Health Record for Patients with Type 2 Diabetes Mellitus in Primary Care

Floor Sieverink[1], Louise M.A. Braakman-Jansen[1], Yvonne Roelofsen[2], Steven H. Hendriks[2], Robbert Sanderman[1], Henk J.G. Bilo[2,3], Julia E.W.C. van Gemert-Pijnen[1]

[1]Center for eHealth Research & Disease Management
University of Twente
Enschede, the Netherlands
{f.sieverink, l.m.a.braakman-jansen, r.sanderman, j.vangemert-pijnen}@utwente.nl

[2]Diabetes Centre
Isala
Zwolle, the Netherlands
{y.roelofsen, s.hendriks, h.j.g.bilo}@isala.nl

[3]Department of Internal Medicine
University of Groningen and University Medical Center Groningen
Groningen, the Netherlands

***Abstract*—A Personal Health Record is a promising technology for improving the quality of chronic disease management. Despite the efforts that have been made in a research project to develop a Personal Health Record for patients with type 2 diabetes mellitus in primary care (e-Vita), considerable differences have been reported between the number of registered users in the participating primary practices. Interviews among practice nurses indicated that a lack of infrastructure (integration in daily care processes), the ability to try the Personal Health Record with minimal investments and without commitment (trialability) and the relative advantage of using a Personal Health Record in comparison with other methods were important factors for the diffusion of the Personal Health Record in primary care.**

Keywords - Personal Health Record; type 2 diabetes mellitus; implementation; diffusion of innovation; interviews; contextual inquiry; value specification; summative evaluation.

I. Introduction

In this paper, we provide an extended version of our paper as presented at eTelemed 2014, the sixth International Conference on eHealth, Telemedicine and Social Medicine [1]. We present the results of an interview study to identify factors perceived as important in the diffusion of a personal health record (PHR) for patients with type 2 diabetes mellitus (T2DM) in primary care, from the perspective of health care workers.

A. *Personal Health Records*

The aging population and the increasing need for chronic care requires an integral approach to disease management that is well coordinated and consistent with (inter)national care standards in order to support a shift from institutionalized care to home care [2]-[4]. Disease management may be viewed as a set of interrelated services that spans from prevention and self-management to intramural care for patients with chronic diseases [5]-[7]. Information- and communication technology (eHealth) will play an important role in disease management, e.g., in providing online support for self-management, in improving information exchange among professionals and with patients, as well as in monitoring the performance of the disease management program [8][9].

The electronic PHR is a promising technology for improving the quality of chronic disease management [10] [11]. A PHR can be defined as “an electronic application through which individuals can access, manage, and share their health information and that of others for whom they are authorized, in a private, secure and confidential environment” [12], a definition that is adopted by many researchers over the years (e.g., [13]-[15]).

However, PHRs are becoming more complex and potential functions of current PHRs may not only include sharing clinical and personal data (e.g., history, test results, treatment, appointments), but may also include self-management support (patient-provider communication, information about the illness, peer support or monitoring health behavior data) [14].

Potential benefits of a PHR include empowering patients in managing their diseases and the reduction of geographical and communication barriers. This may, in turn, lead to a transition from episodic to continuous care, which has the potential to shorten the time to address disease-related complaints that may arise [13][14].

Despite these benefits, the use of such systems in diabetes care has only led to small improvements in diabetes quality measures that were of marginal clinical relevance [10], and up to now, evaluations have only provided little insight into why a particular outcome did occur [16][17]. Consequently, the added value of the existing evidence is often limited for decision making in relation to the strategic direction of implementation efforts [18]. To gain insight into factors that contribute to a successful implementation of eHealth technologies in daily health care processes, it is therefore necessary to look for methodological approaches that go beyond a baseline and follow-up measurement of health outcomes.

B. e-Vita

The PHR e-Vita is an initiative of the Dutch foundation Care Within Reach (in Dutch: Zorg Binnen Bereik), a partnership between Philips and Achmea, a Dutch health insurance company. Currently, e-Vita is deployed in in the Netherlands via three trials to study the effects of the PHR for patients with T2DM, chronic heart failure or COPD. In this paper, we will focus on results from the T2DM study. For patients with T2DM, the main content of e-Vita consists of insight into personal health data (e.g., lab values, blood pressure), self-monitoring health values (e.g., weight), education and a coach for reaching personal health-related goals.

The T2DM research project consists of two parts. First, a prospective observational cohort study (a benchmarking study) is being conducted to assess clinical parameters and, on the long term, quality of life and disease-related complications. Within this study, questionnaires among participants are administered periodically and health data and blood samples are collected.

When patients agreed to participate in this benchmarking study, they are invited to participate in the PHR trial. Main goal of this trial is to study the effects of using a PHR in primary care for patients with T2DM (ClinicalTrials.gov number NCT01570140) [19]. In total, 44 primary care practices participate in this trial.

C. Diffusion of Innovations

Despite the efforts that have been made to develop a technology that has added value in the treatment of patients with T2DM in primary care, differences in the diffusion of the PHR are signaled between the primary practices that participate in the research project.

The pace at which new innovations in health care diffuse through the system, depends on several factors. In Table I, an overview of critical factors for the diffusion of innovations according to Cain & Mittman [20] is given. To gain insight into the factors that influence the diffusion of a PHR in a primary care practices, an evaluation via interviews has been conducted. The main research question was:

What factors influenced the diffusion of a PHR for patients with type 2 diabetes (T2DM) in primary health care, according to primary health care workers?

The sub-questions were:

1. What were the reasons and incentives of primary health care workers to participate in research project regarding the use of a PHR?
2. What training did the primary health care workers receive at the start of the research project?

TABLE I. CRITICAL DYNAMICS FOR THE DIFFUSION OF INNOVATIONS [19]

Critical Dynamic	Explanation
1. Relative advantage	The higher the potential of the technology in comparison to current practice, the more rapidly it will diffuse.
2. Trialability	The ability to try out an innovation without total commitment and with minimal investment.
3. Observability	The extent to which potential users follow the adoption of an innovation by others.
4. Communication channels	The communication channels through which others communicate about the innovation.
5. Homophilous groups	Innovations diffuse faster among groups with similar characteristics.
6. Pace of innovation	The extent to which innovations evolve and are being altered by its users.
7. Norms, roles and social networks	Innovations are shaped by the rules, formal hierarchies, and informal mechanisms of communication operative in the social systems in which they diffuse.
8. Opinion leaders	Individuals whose opinions are respected by others in a population affect the pace of diffusion.
9. Compatibility	The ability of an innovation to coexist with technology and social patterns already in place.
10. Infrastructure	The presence of some form of infrastructure that cluster with the innovation.

3. How did the primary health care workers embed the PHR in their daily care routines?
4. What were the perceived and expected barriers and facilitators for embedding a PHR in daily care routines, according to primary health care workers?
5. What are the expectations of primary health care workers regarding the use of PHRs in the future?

The outcome of the interviews provides critical factors for the improvement of the diffusion and implementation process of a PHR in primary care.

This paper is organized as follows: in Section II, we will describe how the interviews were conducted. In Section III, the results of the interviews are described. In Section IV, the results and directions for future research are discussed. Finally, in Section V, the conclusions of this paper are given.

II. METHODS

In this section, we will present the framework for the interview study, the participants, the design and procedure of the interviews and how the data was analyzed.

A. The Center for eHealth Research (CeHRes) Roadmap

The CeHRes roadmap [21] is a framework that is used to develop new and to evaluate and improve existing eHealth technologies. The roadmap states that eHealth development is a participatory process and that development is intertwined with implementation into daily health care processes. The roadmap consists of five different phases (Fig. 1):

1. *Contextual inquiry*: First, the needs and problems of the stakeholders (e.g., patients, caregivers, health insurers) are described, in order to gain insight into the context and whether or how technology can contribute to minimizing problems.
2. *Value specification*: Second, information about the added values that key-stakeholders attribute to the eHealth intervention is gathered. Together with the contextual inquiry, the value specification provides the functional requirements for the design of the technology.
3. *Design*: On the basis of these requirements, prototypes of the technology are developed and tested.
4. *Operationalization*: The final version of the eHealth technology is launched.
5. *Summative evaluation*: Finally, the uptake, effects and impact of the eHealth intervention is evaluated.

According to the roadmap, the development of eHealth technology also requires continuous evaluation cycles after every step, in order to create eHealth technologies that have the potential to diffuse among its end-users.

For this study, the interviews serve as both a forward evaluation (contextual inquiry and value specification) and a backward (summative) evaluation to gain insight into the uptake and impact of e-Vita as well as into the possibilities to improve the content of e-Vita according to health care providers.

B. Participants

For the interview study, primary care nurses (PNs) of general practices in Drenthe, a province in the north of the Netherlands, were invited to participate in the interview study. In the Netherlands, PNs are the main responsible caregivers for educating patients about their (chronic) disease, advising patients regarding medication use and lifestyle changes and performing health checks. In the

current trial, all selected PNs are responsible for explaining the purpose of the PHR to the participants in the study and administering questionnaires regarding the effects of the PHR. No guidelines for intended use of the PHR are defined.

To reveal the differences between the diffusion processes of practices with high and low numbers of participants, potential practices for the interview study were selected by the means of an inclusion percentage (high, middle, low). The inclusion percentage was calculated as follows:

Inclusion percentage =
*(number of included patients for the benchmarking study / total number of patients with T2DM in the practice)*100.*

Our aim was to conduct five interviews in every group, 15 interviews in total. We therefore invited the five primary practices with the highest and lowest percentages. Also, five primary practices with an average inclusion percentage were invited to participate. When primary practices had indicated on beforehand that the inclusion of participants was postponed due to explainable circumstances (e.g., long-term diseases among the staff), practices were not contacted to participate in the interview study.

All PNs who met the criteria for the interview study received an e-mail with information about the purpose and the topics of the interview. When PNs agreed to participate, they were contacted to make an appointment for the interview.

C. Design and Procedure

Semi-structured in-depth interviews were conducted. During the interviews, questions were asked regarding the following themes: reasons and incentives to use and implement a PHR in their primary practice, the use and users of the PHR so far, bottlenecks and barriers that are encountered or expected, the (positive) results so far and the expected changes a PHR will make in the primary health care for patients with T2DM. The duration of the interviews was 45-60 minutes (non-stop). All participating PNs received a gift voucher of 50 euros.

Ethical approval for this interview study was obtained by the ethics committee of the University of Twente.

D. Data Analysis

All interviews were transcribed verbatim and themes and categories were subsequently coded via open coding, axial coding and selective coding [22]. In this way, recurring

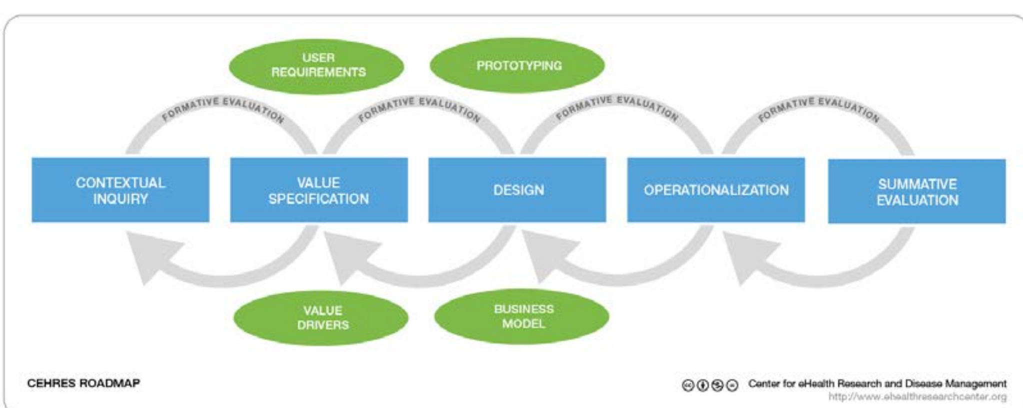


Figure 1. The Center for eHealth Research and Disease Management Roadmap [21]

themes and items of interest regarding the implementation and use of eHealth technologies in primary health care practice were identified. Occurring themes were categorized using the critical dynamics for the diffusion of innovations, according to Cain & Mittman [20]

III. RESULTS

In this section, we will present the results of the interview study.

A. General Results

After receiving the invitation, 11 PNs agreed to participate in the interview study. An overview of the number of potential and included participants for the primary care practice of every PN at the start of the interviews (August 2013) is given in Table II.

TABLE II. OVERVIEW OF PARTICIPATING PRIMARY PRACTICES IN THE INTERVIEW STUDY

PN number	Inclusion group	Patients	
		# T2DM patients	*# participants included (%)*
1	High	204	126 (62)
2	High	56	33 (59)
3	High	61	37 (61)
4	High	88	45 (51)
5	High	146	63 (66)
6	Middle	98	22 (22)
7	Middle	122	25 (20)
8	Middle	182	45 (25)
9	Middle	94	18 (19)
10	Low	163	4 (2)
11	Low	235	7 (3)

In total, 5 PNs of practices with high inclusion percentages, 4 PNs of practices with average inclusion percentages and 2 PNs of practices with low inclusion percentages participated in the interview study.

B. Reasons and Incentives to Participate

In total, 5 PNs (interview numbers 2, 4, 7, 8 and 11) indicated that they participate in the trial because they find it important to stimulate the development of self-management skills of their patients:

> *"If we can offer patients tools to learn about their own disease and to take their own responsibilities, we must not miss this opportunity."*

Two PNs indicated that they want to keep abreast of the times (interview numbers 1 and 2), that they are curious about the added value of a PHR in primary practice (interview numbers 1 and 6), and that they want to offer their patients something extra during their treatment, for example information about T2DM (interview numbers 7 and 8).

Also, one PN indicated that the primary care practice wanted to participate because of their (relatively young) patient population (interview number 3) or to participate in a research project besides the daily working routines (interview number 9).

Two PNs (interview numbers 6 and 11) indicated that they were curious about the results of the benchmarking study: how do their practices perform in comparison with other practices and how satisfied are their patients about the care they receive?

C. Training

Regarding the training the PNs received before the start of the project, 6 PNs indicated that they attended a plenary information meeting (interview numbers 1, 2, 3, 4, 8 and 9). During this meeting, information was given about the purpose and the course of the research project and instructions were given regarding the inclusion of participants and administering the questionnaires for the benchmarking study. One PN (interview number 4) indicated that there was a short explanation about the purpose and the functions of the PHR during that meeting.

Eight PNs indicated that they were not trained in using the PHR and how to integrate the PHR in daily care routines (interview numbers 1, 3, 5, 7, 8, 9, 10, 11). One PN received some instructions for using the PHR (interview number 6) and another PN logged in once with one of the researchers (interview number 2).

D. Integrating the PHR in Daily Care Routines

In total, 8 PNs indicated that they did not integrate the PHR with consultations with their patients. The remaining 3 PNs occasionally ask their patients if they visited the PHR and if they have questions regarding the information on the PHR.

E. Perceived Barriers for Using a PHR

All PNs indicated that there is a lack of time to use the PHR in the treatment of their patients. According to 7 PNs (interview numbers 1, 2, 3, 5, 6, 7 and 8), this is due to a lack of integration of the PHR in daily work routines:

> *"For me, it [the PHR, red.] is quite difficult to fit in the 20 minutes that I have for every patient. I need that time for the health checks. How do you start a dialogue about the PHR, then?"*

Also, 9 PNs (interview numbers 1, 2, 3, 5, 6, 7, 9, 10 and 11) indicated that they did not have access to the PHR. They were not able to see what patients see, and find it therefore difficult to promote and explain the PHR among their patients:

> *"It is quite difficult. You don't know e-Vita, and you have to explain it to the patients. That doesn't work."*

> *"I was not able to see what patients see for a long time. I have to recommend something I don't know."*

When they were able to log in to the PHR via their patients, 4 PNs (interview numbers 1, 5, 6 and 9) experienced usability problems:

> *"When you want to visit e-Vita, you have to take the hurdle of logging in first."*

Also, 4 PNs (interview numbers 2, 8, 9 and 11) indicated that they have easy accessible alternatives:

> *"One patient was quite motivated, so I printed the health data and gave it to him. That is a much shorter way."*

Regarding their patients, 5 PNs indicated that many of their patients do not have Internet access (interview numbers 1, 6, 7, 8 and 10) or experience usability problems (interview numbers 1, 2, 3, 4 and 6).

Furthermore, according to 5 PNs, the PHR does not meet the needs of their patients (interview numbers 1, 4, 5, 6 and 7), and patients are afraid that the usual care will become less personal (n=4, interview numbers 1, 6, 8 and 10).

F. Perceived Facilitators for Using a PHR

Although the PHR is not yet used in daily care routines, some potential factors for the successful implementation of a PHR were mentioned during the interviews.

All PNs indicated that they could easily contact a help desk when they experienced technical problems with the PHR or when they have questions regarding the benchmarking study:

> *"Nothing is too much. It is really important that it is never inconvenient to have contact."*

> *"A strong point is that they are easy to reach. When I have a problem, I send an email and the next day, I have an answer."*

Second, 5 PNs (interview numbers 1, 3, 4, 7 and 10) indicated that they were able to fit the activities necessary for the benchmarking study into their own workflow:

> *"I made a Word file with a list of participants, this gives me an overview of the procedure."*

> *"Every time I realize that I have to establish a system to create a routine."*

Third, 4 PNs (interview numbers 4, 5, 7 and 9) indicated that triggers for using the PHR are probably important, for example, via news letters or via (financial) incentives:

> *"I think you should give a financial incentive, for example, a bonus from the health insurance company."*

> *"Maybe a newsletter, a stimulus to let patients think, 'let's visit e-Vita again'."*

G. The Future of PHRs in Primary Care

Finally, PNs were asked about their opinion regarding the future of PHRs in primary health care. In total, 5 PNs (interview numbers 1, 2, 4, 5 and 11) indicated that they find it hard to predict whether there is a future for PHRs in health care, and how PHRs will be used in the future.

On the other hand, 5 PNs (interview numbers 2, 4, 6, 9 and 11) expect better-informed patients during consultations, but they also believe that their own role is not likely to change.

Three PNs (interview numbers 3, 5 and 6) believe that a PHR in primary care will mainly be used for communication purposes between patients and health care providers.

IV. Discussion

The interviews indicated that, despite the participants' enthusiasm and understanding of the importance of stimulating self-management skills of patients with chronic conditions, the diffusion of the PHR for patients with T2DM in primary care is still rather low. The goal of this study was to identify the factors that influenced the diffusion. We signaled differences in the inclusion percentages between the participating primary practices in the research project, and therefore we wanted to gain insight into the factors that influenced the diffusion from the view of primary health care workers.

Because we believe that the development of eHealth technologies is an ongoing process that requires continuous evaluation cycles [21], we conducted both a forward and a backward evaluation. With this evaluation, we tried not only to gain insight into the factors that influenced the uptake and impact of a PHR, but also to identify possibilities for improving the PHR in the future.

Although we aimed to identify differences between the factors that influenced the diffusion of the PHR experienced by PNs from primary practices with high, medium and low inclusion percentages, the experiences of all PNs were fairly similar, which indicates that a high inclusion of participants in the study does not necessarily lead to using the system in daily practice. This finding made it difficult to identify factors that contributed to the use of the PHR. However, we did find some important factors that influenced the diffusion of the PHR.

First, a lack of *infrastructure* that is necessary for the implementation of an innovation in health care [20][23] played an important role. Most PNs indicated that at the start of the research project, little attention has gone towards education and guidance regarding the integration of the PHR with daily practice, and thus, with national guidelines for the treatment of chronic diseases in primary care. Most PNs indicated that they were mainly trained to administer the questionnaires for the benchmarking study, and during that training, only little attention has gone to the content and the functions of the PHR and the integration of the system with daily practice.

As a result, the awareness of PNs regarding the added value of the PHR is low, which can reduce its diffusion and subsequent use [24]. PNs indicated that they do not use the PHR in the treatment of their patients. They have a certain amount of time for every patient during the consultations, and in this time, PNs have to finish the health checkups (blood pressure, control of the feet, et cetera), talk with the patient about how they are doing and administer the questionnaires for the benchmarking study. Because no guidelines were given regarding the use of the PHR, asking patients about their experiences is not on top of the minds of the PNs. Also, PNs indicated that there is often no time to ask their patients about the experiences and the use of the PHR.

A first study of log data of the PHR showed that the use by patients is suboptimal. After an invitation to visit the renewed PHR, 28% of all registered users visited the PHR at least once, with a mean of 1.5 visits in the first six weeks [25]. It is well possible that creating a routine in using a PHR during consultations will lead to an increased use of the PHR by patients at home. After all, it is likely that patients will use a system that is being promoted by their health care providers, who are often in a relationship of trust with their patients and are therefore seen as *opinion leaders* [20].

An another finding of the interview study is that the *trialability* [20] of the PHR played an important role in the diffusion of the technology. Due to technical problems, PNs were not able to log on to the PHR with test accounts and could therefore not see what their patients see. Subsequently, PNs reported that they find it difficult to use and promote a technology they hardly know.

At the time of the interviews, test accounts were available for PNs, which enhanced the *trialability* of the system [20]. However, because of the reported work pressure and established working routines for administering the questionnaires regarding the benchmarking study, it often had no priority anymore to visit the PHR.

Also, 4 PNs indicated that they have easy accessible alternative tools and resources available for the PHR, for example by providing patients with lab values on paper, instead of viewing them on the PHR. In other words, the *relative advantage* [20][23] of the PHR is rather low, which is probably another important factor for the slow diffusion of the PHR [23].

To increase the diffusion of the PHR in the future, it is therefore useful to guide health care providers in integrating the system in daily routines. By creating an infrastructure for the use of the PHR, new working routines are being established. Also, it might be of added value to appoint ambassadors: health care providers who already successfully use the PHR in daily practice. By increasing the *observability* of the PHR, other health care providers are able to see how others use the PHR and can acknowledge that the use of the PHR is safe and beneficial [20].

However, the results also showed that the PHR is mostly illness-driven instead of user-centered (with little attention for the needs of the end-users), indicating that involving the end-users (via a contextual inquiry and value specification [21]) is valuable in the development of new technologies [24]. By involving the end-users and having an eye for their needs, the added value of the new technology is already evident in the first stages of the developmental cycle [26], which may in turn lead to a better diffusion of the technology.

The interviews not only served as a backward evaluation, but also as an forward evaluation to gain insight into the possibilities to improve the PHR in the future. However, because of the lack of insight into the system, PNs found it rather hard to give directions for the improvement of the PHR. Also, they find it hard to indicate how PHRs will be used in the future. In general, PNs expect better informed patients during consultations, but they also believe that their own role in the treatment of patients is not likely to change.

In the future, we are planning to conduct a further process analysis of the implementation of the PHR using a mixed-methods approach via interviews and usability tests among both patients and health care providers. Also, we are planning to conduct advanced log file analyses, containing real-time data about the actual use of the PHR, collected by the web server. With this data, we will analyze who the actual users of the PHR are, how they use the system and how the PHR supports the users in reaching their health related goals in order to gain insight in how a PHR can be of added value in primary care.

V. Conclusion and Future Work

The results of the interviews indicate that PNs understand the importance of stimulating self-management skills of patients with chronic diseases via a PHR. However, the diffusion of the PHR is still rather low, mostly due to a lack of training in using the PHR and a lack of guidance in integrating the system in daily care routines. Also, the *trialability* and the *relative advantage* of the PHR played an important role in the uptake and impact of the new system.

In the future, we will involve end-users (both patients and health care providers) in our research, in order to create a PHR that is of added value for patients with chronic diseases in primary care.

References

[1] F. Sieverink, S.M. Kelders, L.M.A. Braakman-Jansen, and J.E.W.C. Van Gemert-Pijnen, "The Uptake and Impact of a Personal Health Record for Patients with Type 2 Diabetes Mellitus in Primary Care: A Research Protocol for a Backward and Forward Evaluation," The Sixth International Conference on eHealth, Telemedicine, and Social Medicine (eTelemed 2014) IARIA, March 2014, pp. 170-173, ISBN: 978-1-61208-327-8.

[2] JE. Shaw, R.A. Sicree, and P.Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030," Diabetes

Research and Clinical Practice, vol. 87(1), pp. 4-14, 2010, doi:10.1016/j.diabres.2009.10.007.

[3] D.C. Klonoff, "Using telemedicine to improve outcomes in diabetes—an emerging technology," Journal of Diabetes Science and Technology, vol. 3(4), pp. 624-628, 2009.

[4] R. Bellazzi, "Telemedicine and diabetes management: current challenges and future research directions," Journal of Diabetes Science and Technology, vol. 2(1), pp. 98-104, 2008, doi: 10.1177/193229680800200114.

[5] T. Bodenheimer, K. Lorig, H. Holman, and K. Grumbach, "Patient self-management of chronic disease in primary care," JAMA, vol. 288(19), pp. 2469-2475, 2002, doi:10.1001/jama.288.19.2469.

[6] Task Force on Community Preventive Services, "Recommendations for healthcare system and self-management education interventions to reduce morbidity and mortality from diabetes," American Journal of Preventive Medicine, vol. 22(4 SUPPL. 1), pp. 10-14, 2002.

[7] E.H. Wagner, R.E. Glasgow, C. Davis, A.E. Bonomi, L. Provost, D. McCulloch, P, Carver, and C. Sixta, "Quality improvement in chronic illness care: a collaborative approach," The Joint Commission Journal on Quality Improvement, vol. 27(2), pp. 63-80, 2001.

[8] F. Verhoeven, K. Tanja-Dijkstra, N. Nijland, G. Eysenbach, and L. van Gemert-Pijnen, "Asynchronous and Synchronous Teleconsultation for Diabetes Care: A Systematic Literature Review," Journal of Diabetes Science and Technology, vol. 4(3), pp. 666-684, 2010, doi: 10.1177/193229681000400323.

[9] H. Goldberg, J. Ralston, I. Hirsh, J. Hoath, and K. Ahmed, "Using an Internet comanagement module to improve the quality of chronic disease care," Joint Commission Journal on Quality and Patient Safety, vol. 29(9), pp. 443-451, 2003.

[10] M. Tenforde, A. Nowacki, A. Jain, and J. Hickner, "The Association Between Personal Health Record Use and Diabetes Quality Measures," Journal of General Internal Medicine, vol. 27(4), pp. 420-424, 2012, doi: 10.1007/s11606-011-1889-0.

[11] C. Osborn, L. Satterwhite Mayberry, S. Mulvaney, and R. Hess, "Patient web portals to improve diabetes outcomes: A systematic review," Current Diabetes Reports, vol. **10**(6), pp. 422-435, 2010, pp. 422-435, doi: 10.1007/s11892-010-0151-1.

[12] "Connecting for health. The Personal Health Working Group final report," Markle Foundation, 2003, last retrieved: November 19, 2014 from http://www.providersedge.com/ehdocs/ehr_articles/The_Personal_Health_Working_Group_Final_Report.pdf.

[13] P. Tang, J. Ash, D. Bates, J. Overhage, and D. Sands, "Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption," vol. 13(2), pp. 121-126, 2006, doi: 10.1197/jamia.M2025.

[14] C. Pagliari, D. Detmer, and P. Singleton, "Potential of electronic personal health records," BMJ, vol. 335(7615), pp. 330-333, 2007, doi: 10.1136/bmj.39279.482963.AD.

[15] M. Tenforde, A. Jain, and J. Hickner, "The value of personal health records for chronic disease management: what do we know?," Family Medicine, vol. 43(5), pp. 351-354, 2011.

[16] A. Black, J. Car, C. Pagliari, C. Anandan, K. Cresswell, T. Bokun, B. McKinstry, R. Procter, A. Majeed, and A. Sheikh, "The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview," PLoS Med, vol. 8(1), pp. e1000387, 2011, doi: 10.1371/journal.pmed.1000387.

[17] B. Kaplan, "Evaluating informatics applications—some alternative approaches: theory, social interactionism, and call for methodological pluralism," International Journal of Medical Informatics, vol. 64(1), pp. 39-56, 2001, doi: 10.1016/S1386-5056(01)00184-8.

[18] S. Clamp and J. Keen, "Electronic health records: Is the evidence base any use?," Informatics for Health and Social Care, vol. 32(1), pp. 5-10, 2007, doi: doi:10.1080/14639230601097903.

[19] Y. Roelofsen, S. Hendriks, F. Sieverink, M. van Vugt, K. van Hateren, F. Snoek, et al. "Design of the e-Vita diabetes mellitus study: Effects and use of an interactive online care platform in patients with type 2 diabetes (e-VitaDM-1/ZODIAC-40)," BMC Endocrinde Disorders, vol. 14(22), 2014, doi:10.1186/1472-6823-14-22.

[20] M. Cain and R. Mittman, "Diffusion of innovation in health care," California Healthcare Foundation Oakland CA, 2002.

[21] J. van Gemert-Pijnen, N. Nijland, M. van Limburg, H. Ossebaard, S. Kelders, G. Eysenbach et al. "A holistic framework to improve the uptake and impact of eHealth technologies," Journal of Medical Internet Research, vol. 13(4), pp. e111, 2011, doi:10.2196/jmir.1672.

[22] J. Corbin and A. Strauss, "Basics of qualitative research: Techniques and procedures for developing grounded theory," 2008: Sage.

[23] Nazi, K.M., "The personal health record paradox: Health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings," Journal of Medical Internet Research, vol. 15(4), pp. e70, 2013, doi:10.2196/jmir.2443.

[24] N. Archer, U. Fevrier-Thomas, C. Lokker, K. McKibbon, S. Straus, "Personal health records: A scoping review," Journal of the American Medical Informatics Association, vol. 18(4), pp. 515-522, 2011, doi: 10.1136/amiajnl-2011-000105.

[25] F. Sieverink, S. Kelders, L. Braakman-Jansen, J. van Gemert-Pijnen, "The Added Value of Log File Analyses of the Use of a Personal Health Record for Patients With Type 2 Diabetes Mellitus: Preliminary Results," Journal of Diabetes Science and Technology, vol. 8(2), pp. 247-255, 2014, doi: 10.1177/1932296814525696.

[26] J. van Gemert-Pijnen, O. Peters, and H. Ossebaard, "Improving eHealth," The Hague, the Netherlands: Eleven International Publishing, 2013.

Remote Synchronous Usability Testing as a Strategy to Integrate Usability Evaluations in the Software Development Process: A Field Study

Fulvio Lizano
Informatics School
National University
Heredia, Costa Rica
flizano@una.cr

Jan Stage
Department of Computer Science
Aalborg University
Aalborg, Denmark
jans@cs.aau.dk

Abstract*—Although Human-Computer Interaction techniques, as usability evaluations, are considered strategic in software development, there are diverse economic and practical constraints in their application. The integration of these tests into software projects must consider practical and cost-effective methods such as, the remote synchronous testing method. This paper presents results from a field study designed to compare this method with the classic laboratory-based think-aloud method in a realistic software development context. Our interest in this study was to explore the performance of the remote synchronous testing method in a realistic context and its effectiveness to provide an integration method of usability evaluations into the software development process. The results demonstrate that the remote synchronous testing method allows the identification of a similar number of usability problems achieved by conventional methods at a usability lab. Additionally, the time spent using remote synchronous testing is significantly less. Results obtained in this study also allowed us to infer that when using the remote synchronous testing method, it is possible to handle some practical constraints that limit the integration of usability evaluations into software development projects. In this sense, the relevance of the paper is based on the positive impact that remote synchronous testing could have in the digital accessibility of the software, by allowing the extensive use of usability evaluation practices in software development projects.

Keywords - usability evaluations, remote synchronous testing method, integration of usability evaluation in software development projects, field study.

I. INTRODUCTION

In this extended paper, we improve the work presented at the eighth International Conference on Digital Society (ICDS 2014) in Barcelona, Spain, in March 2014 [1].

In this extended paper, we improve the introduction section with more elements that reflect on the relevance of usability evaluations. We have also better enriched and structured, the related work section. Special mention is deserved regarding additional references related to economic limitations (i.e., the cost obstacle) and the conceptual background on the remote synchronous test. In the method section, we included two new sub-sections. First, there is a new subsection called “The context of the study”, which presents the context where the study was made. Second, we included a subsection called “Overcoming the limitations of the field study”, where we present the main theoretical concepts of these kinds of studies. In addition, we also explain how we intended to handle such limitations. In this section we also included several figures. The result and discussion sections were also structured more clearly, with some additions. We have also enhanced the conclusion section and separated the limitations of the study into a new section. Our intention is to bring a final reflection on the main limitation presented in our study, which is related to the type of participants. Finally, the bibliographic references were extended.

Usability has a significant impact on software development projects [2]. Common usability activities, such as usability evaluations, are relevant and strategic in diverse contexts, such as users, software developers, development organisations, and software development projects.

In the case of the user, who requires a high level of usability in the software [3], usability evaluations are important because they assess whether the software under evaluation considers users’ skills, experiences, and expectations [4]. A high level of usability in a software system enables users to perform their work while saving time and resources, and this allows them to be more effective and efficient.

In the case of software developers, usability evaluations provide them with clear details about usability problems in a software system. This information becomes valuable feedback [5], which allows them to produce better results in their work. Furthermore, improved usability in software increases the developers' confidence levels regarding their technical ability and creates a personal identification with a software product; these are strong motivators for developers [6][7].

For development organisations, usability evaluations are important because they provide benefits, such as cost savings, increased sales, increased productivity, lower training costs, and decreased technical support requirements for users [8]. More usable software implies less user support and training, which increases the development organisational efficiency and productivity.

Finally, in software development projects, Human-computer Interaction (HCI) techniques have a high valuation [2]. In fact, Abran et al. [9] considered usability evaluations to be relevant and strategic activities within software projects. One of the main reasons for this high valuation is due to the application of usability methods (e.g., usability inspection methods, usability testing with users, etc.), and it is possible to improve the quality of the software by providing useful feedback about usability.

The importance of usability in the above-cited cases has motivated the integration efforts of usability evaluations into software projects [10][11][12].

However, economic and practical issues limit integration of usability evaluations into software projects, where limited schedules and high expectations of stakeholders to obtain effective/efficient results faster, are common. Productivity has been a recurrent concern in the industry [13][14] and is something that makes it very difficult to justify certain HCI activities [15].

Considering this, any effort to integrate usability evaluations into software projects must necessarily consider practical and cost-effective methods. Many of the studies conducted to explore this integration have been made on limited realistic contexts (e.g., literature reviews [15][16][17][18][19], surveys [2][13][20][21][22], experiments in labs [23][24] and case studies [5][25]). Other papers cited above present proposals of projects or methods [12][26][27]. There are only three studies with a more empirical base in more realistic contexts [5][28][29]. Confidence in the results of these studies should be improved by other studies made in a realistic developmental context.

This is the reason we present the results of a field study that aimed to compare the remote synchronous test method against the classic laboratory-based think-aloud method in a realistic software development context.

In the following section, we offer an overview of related works. The next section presents the method used in our research. Following this, we present the results of our study. After the results are summarised, the paper presents the analysis before concluding with suggestions for future work.

II. Related works

This section presents the economic and practical constrains to integrate usability evaluations into software development projects. The literature on practical constrains included here considers only studies focused on the perspectives between practitioners, methods, and user participation. On the other hand, in this section we present the main concepts of remote synchronous usability testing. Our purpose in the literature review is to provide a basic framework to analyse the results of our study.

A. *Economic constrains: the cost obstacle*

High consumption of resources in usability evaluations, also known as the cost obstacle, is a recurrent perception in diverse contexts [17][20][22][23][30]. This fact could explain why usability has a lower valuation for an organisation's top management [12], becoming manifest by the lack of respect and support for usability and HCI practitioners [21]. Therefore, cost-justification of usability may be difficult for many companies, as it is perceived as an extra cost or feature [15].

It is possible to define the cost obstacle as the constraints of applying usability evaluations due to the high consumption of resources required by this kind of testing. The cost of usability evaluations is a measurable obstacle presented in both of the following cases: development organisations and software projects.

In the case of development organisations, this obstacle is presented in the form of a 'perceived cost obstacle'. In this case, the perception can be understood as the perspective that development organisations have regarding the cost of the usability evaluations. This perspective is normally based on the value judgment presented within development organisations. Some examples of this modality of the cost obstacle were reported in [20] and [30], where it is possible to see that in development organisations there exists the idea that usability testing is expensive, and this limits its application– even though such evaluations have not been conducted. In addition, Nielsen [23] argues that the perception of the cost of usability engineering techniques is the reason why such techniques are not used extensively in development organisations. Coincidentally, Bellotti [31] reports that software developers view usability methods as too time-consuming and intimidating in their complexity.

Within software projects, the cost obstacle appears in the form of an 'actual cost obstacle'. Considering the dynamic presented in the software development project, based on a specific product (i.e., the software), the cost obstacle is more tangible and is related to the real cost presented in such a project. Nielsen [23] offers some examples of actual costs. Ehrlich and Rohn [32] referred to the actual costs in terms of 'initial costs' and 'sustaining costs'. The initial costs include the settings and laboratory or similar facility equipment that are required for usability evaluations. The sustaining costs correspond to those costs related to the conduction of the usability evaluation process and include the staff, recruitment of participants, transportation, allowances, special equipment, software, and etc.

The cost obstacle can be quantified by defining and collecting information about diverse usability metrics. Time consumption in usability evaluations is one of the most commonly used measures to assess cost [23][33][34][35]. Time consumption relates some ideas about the consumption of resources in usability evaluations. For example, based on this measure, some studies concluded that classical protocols, such as 'thinking aloud', have a high consumption of time [36][37]. In addition, Kjeldskov and Graham [38] found that the analysis of the data collected during the usability evaluations normally demands a high time consumption, especially in the video data analysis process. Finally, Borgholm and Madsen [39] argue that usability reports could be impractical due the extensive time used in their preparation.

It is possible to identify the following two main strategies for reducing the cost of usability evaluations: 1) use of alternative usability evaluation methods, and 2) the improvement of the usability evaluation process.

Alternative usability evaluation methods aim to reduce costs in classical usability evaluations with users. One example of these methods is the heuristic evaluation, which is a method where the software interfaces are evaluated based on usability heuristics in order to generate an opinion about the usability of the software [40]. The process starts with individual reviews by three, four, or five expert evaluators of the software. During this process, each evaluator checks whether the usability principles, used as a reference for good practices (i.e., heuristics), are included in the software. Next, evaluators compare their results and produce an integral usability report [37].

The approach of improving the usability evaluation process has been widely discussed; for example, the time consumption issue within analysis activities was addressed by Kjeldskov and Graham [38]. They proposed an analysis technique called Instant Data Analysis (IDA) that is used in the analysis process of the results of the sessions with users. The aim of IDA is to conduct usability evaluations in one day, obtaining similar results to traditional video data analysis methods. Alternatively, Borgholm and Madsen [39] suggest focusing on the report of the results of the usability evaluations. These researchers found that some HCI practitioners prepared two kinds of reports with different formats and contents. The first report was oriented to the developers and provided an executive summary with information useful for their work. The second report, which is more extensive, was delivered several days after the evaluation for documentation purposes. Supplementary meetings, at which the developers and HCI practitioners discussed the usability findings, and posters, which described the main usability problems, were used to mitigate this problem.

B. *Other practical constrains*

On the other hand, regarding practical constrains, three of the most cited are related to the difference of perspectives between HCI and Software Engineering (SE) practitioners, the absence or diversity of methods, and user participation.

The first constraint related to the difference of perspectives between HCI and SE practitioners is contextualised in the difference of opinions regarding what is important in software development [27]. This diversity of perspectives results in contradictory points of view regarding how usability testing should be conducted and may result in a certain lack of collaboration between HCI and SE practitioners. It is possible to find the origin of this discrepancy between these two perspectives in the foundations of the HCI and SE fields. Usability is focused on how the user will work with the software, whereas the development of that software is centred on how the software should be developed in a practical and economical way [19]. These conflicting perspectives result in tensions between software developers and HCI practitioners [19][25].

The second constraint relates to the absence or diversity of methods and has two opposing views. First, some researchers report a lack of appropriate methods for usability evaluation [20][22] or a lack of formal application of HCI and SE methods [2]. This situation may explain why the User-Centred Design UCD community has expressed criticism about the real application of some software development principles [18]. Second, it is reported that the existence of numerous and varied techniques and methodologies in the HCI and SE fields could hamper the integration [25].

Finally, the participation of customers and users has become another relevant limitation for the integration of usability evaluations into software projects [20][30][22]. This matter is a permanent challenge to the dynamic of the software development process. Users and customers have their own problems and time limitations, and these normally limit their participation in software development activities, such as usability evaluations.

The literature reported different proposals for handling the aforementioned three practical constraints. First, in the case of the difference of perspectives between HCI and SE practitioners, some studies have suggested that increased participation by developers in usability testing could positively impact the valuation of usability [5]. This improvement in developer perspectives could make them more conscious of the relevance of HCI techniques.

Second, with respect to the absence or diversity of methods, an integration approach based on international standards is proposed [16] in order to enable consistency, repeatability of process, independence of organisations, quality, etc. A similar approach suggests the integration of HCI activities into software projects using SE terminology for HCI activities [26].

Finally, regarding the constraint related to the participation of customers and users, some researchers have suggested several practical actions (e.g., smaller tests in iterative software development processes, testing only some parts of the software, and using smaller groups of 1–2 users in each usability evaluation) [29].

C. *Remote Synchronous Usability Testing*

The aforementioned obstacles can be handled by using remote synchronous usability testing, which is a method that allows software developers to conduct/participate in usability evaluations with users in a practical and economical way.

Remote Usability Testing (RUT) was defined as a usability evaluation technique in which the evaluator remains separated in space and/or time from the users while performing observation and analysis of the process [41]. The RUT techniques can be synchronous or asynchronous. The synchronous format allows the evaluators to receive and conduct the evaluation in real time with users who are located elsewhere. In contrast, in the asynchronous format, the evaluators do not access the data nor conduct the evaluation in real time [42]. The RUT method allows usability testing without the constraint of geographical limitations, and therefore requires fewer resources.

The main uses of RUT are:

- to evaluate the usability of web applications [43],
- to reduce the costs of the usability evaluation process [43][44],
- to collect a high volume of data [42], and
- to make usability evaluations by considering an international context [42].

The practicality of logistic considerations and the resource-saving advantage make RUT a promising alternative for reducing the aforementioned limitations.

III. Method

We have conducted an empirical study aimed at comparing the remote synchronous testing method (condition

R) with the classic laboratory-based think-aloud method (condition L).

Using remote synchronous testing, the test is conducted in real time, but the evaluators are separated spatially from the users [33]. The interaction between the evaluators and users is similar to those at a usability lab. There are many studies that confirm the feasibility of RUT methods [33][45][46]. Actually, there is a clear consensus regarding the benefits obtained using this method (e.g., no geographical constraints, cost efficiency, access to a more diverse pool of users, and similar results as a conventional usability test in a lab) [33][47]. The main disadvantages are related to problems generating enough trust between the test monitor and users, longer setup time, and difficulties in re-establishing the test environment if there is a problem with the hardware or software [33].

Three usability evaluations were made by three teams using a classic usability lab. In addition, another three usability evaluations were conducted by another three teams using a remote synchronous testing method.

Final-year students of SE who had 18 months of practical experience working in software development formed all of these teams. This experience is the result of an academic project created by the students to develop a software system in a real organisation.

In the next subsections, we present in detail the main elements of the context of the study, participants, training, and advice received for the participants, procedure followed in the study, settings, data collection and analysis, and the actions taken in order to overcome the limitations of the field study.

A. The context of the study

The study considered usability evaluations conducted on software systems made in the context of the internship/academic project for which the students put into practice what they learned in the courses of the System Engineering Bachelor degree. The organisation in charge of these academic projects is the Department of System Engineering (DSI), School of Informatics, National University (UNA), located in Heredia, Costa Rica. The UNA is one of the five public universities of Costa Rica. Funded in 1973, this university has five faculties that have an enrolment of around 18000 students in 65 undergraduate and postgraduate programs. Informatics School is the second school of the university with an approximate enrolment of 1300 students, 250 new students every year.

Starting in the third year, over three semesters, around 30 software projects are developed by student teams formed by three to four students who were also learning regular topics related to system engineering theory. In this process, the student teams receive supervisor feedback (a DSI professor) and also interact with stakeholders/users. The students design, develop, and implement a software system in a real organisation (private, public, Small and Medium Enterprise (SME), or Non-Profit Organisations (NPO)). This organisation provides regular assessments of the students'

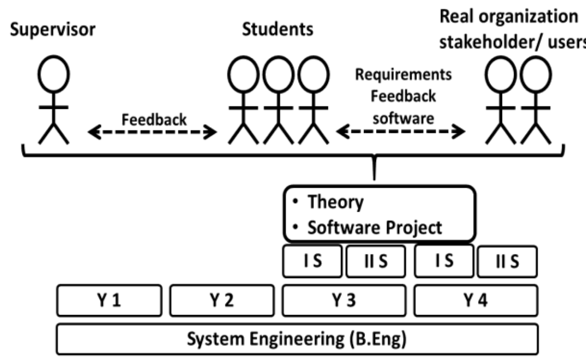


Figure 1. The context of the study: the academic project.

work. In addition, at the end of the process, the organisation should formally accept or reject the software product. There is a main user assigned by the organisation. This user normally plays a role as main contact between the students and the organisation. Usually, this user is also one of the main stakeholders. It is also possible that other regular users can be considered in the process. Fig. 1 presents the context around these projects.

B. Participants

In order to be considered for our research, the software projects must meet our requirements regarding users being available for the tests. Considering these criteria, 16 of 30 teams and their software projects were preselected as potential participants in the experiment. Finally, we randomly selected six teams who were randomly distributed throughout the R and L conditions.

Final-year students who were finishing their last course in System Engineering formed the teams. These participants were organised into six teams consisting of three members each. A total of 18 people participated in our study. The average age was 22 (SD=2.13), and 17% were female. In addition to the courses taken previously, the participants had amassed nearly 18 months of real experience in practical academic activity by developing a software system in a real organisation that sponsored the project. These organisations provided regular assessments and formal acceptance (or rejection) of the software. Several users and stakeholders were also involved in the process. The scope of the software projects was carefully controlled in order to guarantee a similar level of effort from all of the participants. The average of the final assessment of the project was 9.67 on a scale of one to ten (SD=0.33). As an incentive for participation, the participants received extra credits. The conditions, code, members, and software are presented in Table I.

C. Training and advice

All participants received training and advice during the experiments (remotely for R condition). In the training, we presented and explained several forms and guidelines based on commonly used theories [38][47]. In addition, a workshop was created in order to put into practice the contents of the

TABLE I. TEAMS, MEMBERS, AND STAFF FOR THE TESTS

Cond.	Code	Members	Software
L	L1	3 males	Students' records in a college.
	L2	1 female, 2 males	Internal postal management system in a financial department of a public university.
	L3	1 female, 2 males	Laboratory equipment management in a biological research centre belonging to a public university.
R	R1	1 female, 2 males	Criminal record in a small municipal police station.
	R2	3 males	Management of documents related to general procurement contracts in an official national emergency office.
	R3	3 males	Students' records in a public school.

training materials. The participants received specific instructions in order to consider the following three categories of usability problems: critical, serious, and cosmetic [33]. The number of hours spent in training was ten (four hours in lectures and six hours in practice). Furthermore, the advice provided to the participants included practical issues concerning how to plan and conduct usability evaluations.

D. Procedure

The design of the experiment increased the confidence in the results and the objectivity of the development teams during the evaluation process. Under the two conditions, each team had to test the software system made by another team, who also tested another software system made by a third team.

Each test had two main parts. The first part, under the responsibility of the team who made the software, corresponded to the planning of the complete process (e.g., planning, checklists, forms, coordination with users, general logistics, etc.). The planning included a session script with ten potential tasks of the software.

In the second part of the tests, another team conducted the sessions with the users. The test monitor of this team had to select, for each user, five tasks from those previously defined. We hypothesised that this measure would increase the impartiality of the process; the developers of the software could not interfere in the selection of the task, and the users had to work with different tasks in each session. Next, the test monitor guided the users in the development of the task while the logger and the observers took notes. The test ended with a final analysis session conducted by a facilitator [38]. Table II shows the assignment of the planning and conducting responsibilities in each test. In Fig. 2 and Fig. 3 it is possible to see examples of sessions conducted in condition L and R, respectively.

TABLE II. PLANNING AND CONDUCTION OF TESTS.

Software made by	Planning made by	Conduction made by
L1	L1	L3
L2	L2	L1
L3	L3	L2
R1	R1	R3
R2	R2	R1
R3	R3	R2

E. Settings

The test conducted under the L condition used a state-of-the-art usability lab and think-aloud protocol. Using this technique it is possible to collect information regarding what the test participants are thinking while they perform the usability tasks. Participants are guided along the test in order to express commentaries of their thoughts by thinking aloud [8][47]. Each test included three sessions where the users were in front of the computer and the test monitor was next to the users. The logger and observers were present in the same room. In the case of the R condition, the tests were based on remote synchronous testing [33]. All participants were spatially separated. Users were in the sponsors' facilities. Each test included three sessions with the users.

F. Data collection and analysis

Each user session was video recorded. The video included the software session recorder (video capture of the screen) and a small video image of the user. Under R conditions, the video also recorded the image of the test monitor (see Fig. 2). We also used a test log to register the main data of each activity (i.e., date, participant, role, activity, and time consumed) and the usability problem reports.

Figure 2. Example of a test's session in L condition.

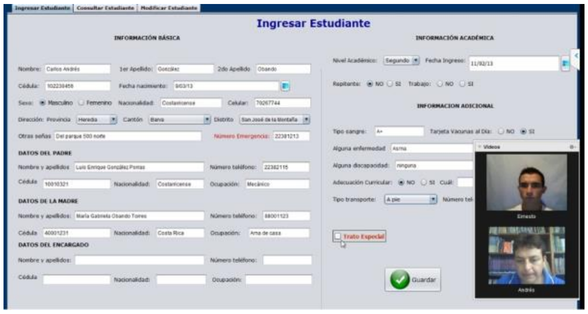


Figure 3. Example of a test's session conducted in R condition.

The data analysis was conducted by the authors of this paper based on all data collected during the tests. The tests produced six sets of data for analysis (i.e., six usability problem reports, six test logs, and six videos).

The consistency of the classification of the usability problems by participants was one of the main concerns in this study. Consequently, our analysis included an assessment of such classifications. Our intention was to be sure that this classification was done consistently according to the instructions given to all participants during the training. We assessed the problem categorisation by checking the software directly in order to confirm the categorisation given by participants to a usability problem. The videos were thoroughly walked through in order to confirm this categorisation.

The tests were conducted on different software systems. There is no joint list of usability problems. In our analyses this is the reason we compared the differences between both conditions using average and standard deviations calculated separately for each condition.

Using the test logs, we analysed the time spent on all the tests. We considered individual and group time consumption. We calculated totals, averages, and percentages to facilitate the analysis. We included in this process all the activities made by all members of the teams in the preparation for the test (e.g., usability plan, usability tasks, etc.) and conducting the test itself. In the analysis, we also considered other participants, such as the users and observers, in order to consider a more realistic context.

Finally, in order to identify significant differences in the data collected, we used independent-sample t tests.

G. Overcoming the limitations of the field study

Wynekoop and Conger [48] classified the field study as a natural setting method normally used for studying current practice and for evaluating new practices. As with any research method, the field study has strengths and weaknesses. The main strengths are that they are practical and in realistic settings.

Braa and Vidgen [49] argue that the field study method is an extension of lab experiments conducted in the particular context of an organisation, and this is something that implies less methodological rigor but conduction in a more realistic environment. The realistic settings used in a field study are useful in terms of exploring a specific phenomenon in conditions close to reality. For example, observation of users' natural behaviours in their own environments was highlighted by Nielsen [50] as an important method used in HCI research.

The main weaknesses of the field study are the difficulties in finding an adequate setting and the control and management of the study.

Considering that having an adequate environment is a key aspect of the design of the field study, the difficulty in finding such an environment becomes a relevant weakness [9]. Real environments may limit the research process (e.g., time restrictions, resource limitations, motivation of participants, etc.). Another weakness is the complexity of the control and management process [51]. The particular characteristics of the field study (i.e., made outside of controlled conditions existing in a lab) make the process complex. For example, a variety of logistics must be considered in the experimental design, as well as the particular conditions presented in the place where the study will be conducted. The management process, among other things, of the data collection is also complex and demands additional efforts considering that the study setting is not necessarily preconditioned to allow the conduction of regular experiments.

Our field study aimed to compare several usability evaluation methods in order to explore how practical and cost effective the methods were.

Our study had the following cited problems: finding adequate environments and controlling and managing the process.

To overcome the difficulty of finding an adequate environment, we did the following. First, we defined a set of conditions that potential organisations and participants had to meet. Second, once we identified potential actors, we randomly selected the number of organisations and participants needed for the study. Finally, using a random distribution, we grouped the actors into different conditions.

To overcome the problem of control and management, we did the following. First, we defined several guidelines to orient the conduction of the study. Second, we provided formal training to the study participants. Third, we provided personalised advice to the participants using alternative channels (i.e., in person, email, chat, and phone). Finally, all data collections were backed up using different alternatives (e.g., CD-ROM copies, public file hosting services, public video-sharing websites, etc.). Although all these measures were taken, it is a fact that the public nature of some tools used to back up the data collection (i.e., the hosting services and video-sharing websites), involves a certain level of risk for such data collections.

IV. Results

The results section is organised into problems identified by type, task completion time, and time spent on the tests.

A. Problems identified by type

Table III shows an overview of the usability problems identified under the two conditions. The problems are classified by type. The largest number of problems was critical. The lowest number of problems identified was in the category of cosmetic problems. The distribution of all types of problems between the two conditions was relatively uniform.

TABLE III. PROBLEMS IDENTIFIED PER TYPE OF PROBLEM. (%)= PERCENTAGE PER CONDITION.

Cond.-> Problems	L	R
Critical	36 (52%)	33 (56%)
Serious	29 (42%)	22 (37%)
Cosmetic	4 (6%)	4 (7%)
Total	69	59

An independent-sample t test for the number of usability problems identified for the three categories, under both conditions, showed no significant difference (p=0.404). The fact that there are no significant differences between the L and R conditions is a reflection of the similarity of the effectiveness of these methods in terms of the number of problems identified.

B. Task completion time

The task completion time was less in the tests completed under the L condition. In these tests, the users spent a total of 87.6 minutes completing the five tasks assigned to each one. The average time per user/task was 1.94 (SD=0.5). The average task completion time per usability problem identified under the L condition was 1.26. In the tests completed under the R condition, task completion time was 137.4, the average time per user/task was 3.10 (SD=1.3), and the average task completion time per problem was 2.32. In Table IV, we present these results.

An independent-sample t test for the task completion time of the nine users considered under the two conditions showed a significant difference (p=0.018).

The analysis of the videos recorded during the tests completed under the R condition showed delays due to technical problems–mainly in the communication between the actors (i.e., users, test monitor, technician, etc.). In addition, in general, the users in their normal jobs were more distracted. On the contrary, in the case of the tests completed at the laboratory, the users were more focused, and the guidance of the test monitors was more effective.

C. Time spent on the tests

The time spent to complete the tests presents an entirely different perspective to that shown in the previous section. Here, the tests conducted under the R condition consumed less time than that conducted under the L condition.

In Table V, we presented an overview of the time spent in the tests conducted under the two conditions. This table includes the average number of minutes spent on test activities. The standard deviation is shown between parentheses. At the end, the table also shows the average of time per problem in minutes.

These results included all the actors involved in the tests (i.e., users, test monitor, logger, observers, etc.). In this sense, it is possible to consider these results to be more realistic; here, all of the elements/persons required to perform the tests are included. An independent-sample t test, for the average time spent in the tests, for both conditions, showed an extremely significant difference (p<0.001).

TABLE IV. USERS' TASKS COMPLETION TIME AND TIME PER PROBLEM. UP= TOTAL NUMBER OF USABILITY PROBLEMS IDENTIFED PER CONDITION.

Condition-> Test–User	L (UP 69)		R (UP 59)	
	Tot. Minutes	Avg. per task (SD)	Tot. Minutes	Avg. per task (SD)
T1–U1	10.8	2.2 (1.9)	30.0	6.0 (1.3)
T1–U2	9.7	1.9 (1.0)	18.3	3.7 (1.6)
T1–U3	12.8	2.6 (2.5)	18.7	3.7 (1.6)
T2–U1	6.1	1.2 (0.4)	17.6	3.5 (1.8)
T2–U2	14.3	2.9 (0.8)	13.3	2.7 (1.3)
T2–U3	8.4	1.7 (0.7)	8.9	1.8 (0.7)
T3–U1	7.4	1.5 (1.0)	11.2	2.2 (2.4)
T3–U2	6.9	1.4 (0.9)	9.0	1.8 (1.4)
T3–U3	11.1	2.2 (1.1)	10.5	2.1 (2.1)
Total Avg. por task (SD)	87.6 1.94 (0.5)		137.4 3.10 (1.3)	
Avg. task completion time per problem, in minutes	1.26		2.32	

The time spent on each activity during the tests confirms these extremely significant differences for all of the activities–except in the analysis. In preparation, conducting the tests, and moving staff, the independent-sample t tests for the time spent in the three tests conducted under each condition, showed extremely significant differences (p<0.001 for all of the cases). In the case of the analysis, the difference was significant (P=0.045).

V. DISCUSSION

The discussion section is organised into two parts. First, we will reflect on the effectiveness of RUT to overcome the cost obstacle. Next, we reflect on how RUT helps to handle the practical constraints previously presented in the related works section.

TABLE V. TIME SPENT IN THE TESTS. UP= TOTAL NUMBER OF USABILITY PROBLEMS IDENTIFIED PER CONDITION .

Condition-> Activity	L (UP 69)	R (UP 59)
Preparation	2500 (102)	1580 (123)
Conducting test	1320 (73)	840 (42)
Analysis	980 (157)	710 (71)
Moving staff/users	1110 (107)	160 (57)
Tot.time spent per test	5910 (220.5)	3290 (102)
Avg. time per problem in minutes	85.7	55.8

A. *Overcoming the cost obstacle*

Usability evaluations made using the remote synchronous testing method are a cost-effective alternative to integrating usability evaluations into software projects. The number of usability problems identified by this method is similar to that obtained by conventional tests made in a usability laboratory. Additionally, there is a significant difference between the time spent on the remote synchronous test method and that spent on the tests made in the lab.

We confirmed the feasibility of conducting usability evaluations by software developers using diverse methods, including the remote synchronous testing method [24][28][41]. In parallel to this practical feasibility, our study also proved the economic feasibility of the remote synchronous testing technique by taking economic advantage of consideration of the developers' conduction of usability tests as was suggested by Bruun and Stage [28], and Skov and Stage [24]. Using developers to conduct usability evaluations, it was not necessary to hire external independent usability experts, thus reducing the cost of the process as suggested by Bruun [52].

In addition, we also confirmed the similarity to the number of problems identified by the conventional lab method [33]. However, in the case of the time spent, our results differ from those of others [33] who argue that the time spent to conduct tests using lab and remote synchronous tests was quite similar. In our case, the difference in time consumption for both methods was significantly favourable in the remote synchronous testing method. A detailed analysis of the test logs showed us that, in the tests made under the L condition, the logistic matters consumed much more time than in the tests under the R condition. Considering our aim of confirming previous findings in a realistic development context, logistic matters must be considered as factual components of any usability test.

The analysis of the procedures followed the conducting of the tests (reported in the usability problem reports) and the test logs showed that when using the remote synchronous testing method, it is possible to achieve several practical advantages that save time in the tests.

It is possible to contextualise these advantages in the results of the time spent on test activities shown in Table V. First, in the case of the preparation activities, the virtualisation of the complete coordination process saved time and effort. The coordination between teams and other actors was easier and more efficient using email, chat, videoconferences, etc.

Second, in the activities of conducting the tests, it was easy and efficient to use all the software tools used during the tests. Even when considering that the task completion time was shown to be better in the tests made under the L condition (see Table IV), differences in the overall process were evident due to this task completion time only being related to the time spent by users to complete the tasks. On the contrary, in the conducting activities of the tests, all of the elements and actors required to conduct the whole test are included (i.e., users, test monitor, logger, observers, etc.)

Third, the difference in the analysis was also significant due to the technological tools that facilitated the conducting of the analysis sessions by the facilitator. In a certain way, the videos also showed that the virtualisation of the process seems to produce a shared feeling about the relevance of productivity during the virtual sessions.

Finally, the results in the moving activities explain themselves. In the realistic development context used in this study, it is clear that avoiding the movement of the usability evaluation staff is one of the most relevant advantages in terms of time consumption.

In general, all of the advantages of the remote synchronous test cited in literature were confirmed in the realistic contexts considered in our study [33][47]. In the case of the disadvantages, we could only identify–in the analysis of the test logs–some problems in the setting of the hardware and software tools used in the process [33].

At this point in the discussion, the economic advantages of the remote synchronous testing method had become evident. Furthermore, this method also helps to handle other practical problems of the integration of usability evaluations into software projects.

B. *Overcoming practical obstacles*

In our study, we have also confirmed the feasibility of the active participation of software developers in usability evaluations [5][24][28]. The participants played several roles in the usability evaluation teams (e.g., test monitor, logger, observer, and technician). This confirmation is relevant when considering the context used in our study (i.e., lab and remote synchronous tests under more realistic conditions). The design of our experiment proved to be very useful because all of the teams actively participated in all of the processes (i.e., planning and conducting the test) and with impartiality. It is a fact that these levels of participation of developers in usability evaluations may positively impact their perspective regarding usability and the HCI practitioners [27] and will reduce the tensions between SE and HCI practitioners [19][25].

Furthermore, in the case of the problem related to the lack of a formal application of HCI techniques, our experiment found that using guidelines and basic training, it is possible to prepare developers for conducting usability evaluations. In a certain way, the theory used to inspire the guidelines used in the tests has followed the suggested approach [16] of using standards to help the integration of usability evaluation in software projects. The analysis of the dynamic of the tests registered in the videos did not show any particular significant problems.

In the case of the tests made using the remote synchronous testing method, the guidelines were fundamental in conducting the remote process. Considering the similarity of the results in the remote synchronous tests and those obtained in the lab, it is clear that the guidelines served their purpose.

Considering these facts, we can conclude that using guidelines based on standards, it is possible to improve the perception of the lack of appropriate methods for usability evaluation [20][22].

Finally, our study also found that the reported problem [20][22][30] relating to the participation of customers and users could be handled well using the remote synchronous testing method. The users do not need to drastically change their activities. Certainly, the task completion time was higher in the remote synchronous testing method however, putting this element in perspective for the entire process; it is always possible to see the strengths of the remote synchronous testing method. Furthermore, other actors did not have to go to the lab.

VI. Conclussion and Future Work

In this paper, we presented the results of a study aimed at comparing the remote synchronous test method against the classical laboratory-based think-aloud method in a realistic software development context. Final-year students who had 18 months of practical experience conducted several tests. Although, the tests were made on software systems for different organisations and purposes, the scope of these software systems was carefully controlled in order to provide similar settings for the study.

Our study confirmed that remote synchronous testing is a practical and cost-effective alternative for integrating usability evaluations into software projects. The study has shown that there is no statistical significant difference in the number of problems identified using the remote synchronous test method when compared to the usability lab. Even considering that the task completion time in the lab was 37% quicker, the time spent to complete all of the remote synchronous tests was 44% quicker. The statistical analysis has shown that the difference was extremely significant. These results included all the actors involved in the tests (i.e., users, test monitor, logger, observers, etc.), which implies a more real context in terms of the whole testing process. In this case, the field study has shown that with the remote synchronous test it is possible to reduce the 'actual cost obstacle' in order to allow economical conduction of usability evaluations.

The identification of a similar number of usability problems and lower time consumption make remote synchronous test method a good alternative. Using this method, it is possible to involve more software developers in the conduction of usability testing. Such an aim only requires basic training, guidelines, and essential advice. Basic guidelines and training allow handling the problems related to the methods. Finally, one of the most relevant advantages of this method is to facilitate the participation of users, developers, and other potential actors in the tests. Avoiding unnecessary movements allows their participation to be easily justified.

In our study, we were focused on the problems identified and the time consumption metrics in a realistic development context. For future work, it is suggested that for the same context a deeper analysis of other metrics, such as the improvement of the perspective of software developers regarding usability, which is another expected result of close participation of developers in usability evaluations, should be conducted.

VII. Limitations of the Study

Our study has two limitations. First, the participants in the study were final-year undergraduate students. Nevertheless, the real conditions present in our study have allowed for a control of this bias. In addition, we think that it is possible to consider these advanced students as novice software developers because they share similar characteristics. We base such a statement on three main facts. First, Bruun and Stage [28] defined novice developers as persons with limited job experience related to usability engineering and no formal training in usability engineering methods. In this sense, advanced students share similar characteristics to the novice developers. Second, the students mainly conducted usability evaluations in the PhD project. To perform these activities, students had to use several soft skills (e.g., defining user tasks, documenting results, following a method, working with real users, working in teams, etc.). According to Begel and Simon [53], novice developers (as well as the students who participated in my research), usually have serious constraints when it comes to these soft skills because these issues are normally less well supported in university pedagogy. Finally, as with novice developers, the students who participated in the PhD project were not preconditioned with extensive previous work experience.

Second, we used only two usability evaluation techniques. However, our selection considered an ideal benchmark of high interaction with users (lab) and the alternative option, which was the focus of our study.

Acknowledgment

The research behind this paper was partly financed by UNA, MICIT, CONICIT (Costa Rica), and the Danish Research Councils (grant number 09-065143). We are very grateful to the participants, observers, facilitators, organisations, and users that helped us in this research.

References

[1] F. Lizano and J. Stage, "Usability Evaluations for Everybody, Everywhere: A field study on Remote Synchronous Testing in Realistic Development Contexts," Proc. ICDS 2014, pp. 74-79. 2014.

[2] Y. Jia, "Examining Usability Activities in Scrum Projects–A Survey Study," Doctoral dissertation, Uppsala Univ., 2012.

[3] G. Lindgaard and J. Chattratichart, "Usability testing: what have we overlooked?," Proc. SIGCHI, ACM Press, pp. 1415-1424, 2007.

[4] P. Bourque and R. Fairley, "SWEBOK : Guide to the Software Engineering Body of Knowledge Version 3.0.," IEEE Computer Society, 2014.

[5] R.T. Hoegh, C.M. Nielsen, M. Overgaard, M.B. Pedersen, and J. Stage, "The impact of usability reports and user test observations on developers' understanding of usability data: An exploratory study," in International journal of Human-Computer Interaction, 21(2), pp. 173-196, 2006.

[6] R.H. Rasch and H.L. Tosi, "Factors Affecting Software Developers' Performance: An Integrated Approach," MIS quarterly, vol.16(3), 1992.

[7] G. Hertel, S. Niedner, and S. Herrmann, "Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel," Research policy, vol. 32(7), pp. 1159-1177, 2003.

[8] J. Nielsen, Usability engineering, Morgan Kaufmann Publishers, 1993.

[9] A. Abran, J.W. Moore, P. Bourque, R. Dupuis, and L.L. Tripp, "Guide to the Software Engineering Body of Knowledge: 2004 Edition-SWEBOK," IEEE Computer Society, 2004.

[10] N. Juristo and X. Ferre, "How to integrate usability into the software development process," Proc. the 28th international conference on Software engineering, ACM Press, pp. 1079-1080, 2006.

[11] K. Radle and S. Young, "Partnering usability with development: How three organizations succeeded," IEEE Software, vol.18(1), pp.38-45, 2001.

[12] T. Granollers, J. Lorés, and F. Perdrix, "Usability engineering process model. Integration with software engineering," Proc. HCI International, pp 965-969, 2003.

[13] P.F. Drucker, "Knowledge-Worker Productivity: The Biggest Challenge," in California management review,vol.41(2), pp.79-94, 1999.

[14] A. Hernandez-Lopez, R. Colomo-Palacios, and A. Garcia-Crespo, "Productivity in software engineering: A study of its meanings for practitioners: Understanding the concept under their standpoint," Proc. Information Systems and Technologies (CISTI), IEEE Press, pp. 1-6, 2012.

[15] G.H. Meiselwitz, B. Wentz, and J. Lazar, Universal Usability: Past, Present, and Future, Now Publishers Inc., 2010.

[16] H. Fischer, "Integrating usability engineering in the software development lifecycle based on international standards," Proc. SIGCHI symposium on Engineering interactive computing systems, ACM Press, pp. 321-324, 2012.

[17] D. Nichols and M. Twidale. *The usability of open source software*. [Online]. Available from: http://firstmonday.org/ojs/index.php/fm/article/view/1018/939 2014.11.14

[18] A. Seffah, M.C. Desmarais, and E. Metzker, "HCI, Usability and Software Engineering Integration: Present and Future," In Human-Centered Software Engineering—Integrating Usability in the Software Development Lifecycle, Springer: Berlin, Germany, pp. 37-57, 2005.

[19] O. Sohaib and K. Khan, "Integrating usability engineering and agile software development: A literature review," Proc. ICCDA, IEEE Press, vol. 2, pp. V2-32, 2010.

[20] C. Ardito, P. Buono, D. Caivano, M.F. Costabile, R. Lanzilotti, A. Bruun, and J. Stage, "Usability evaluation: a survey of software development organizations," Proc. SEKE, pp. 282-287, 2011.

[21] J. Gulliksen, I. Boivie, J. Persson, A. Hektor, and L. Herulf, "Making a difference: a survey of the usability profession in Sweden," Proc. NordiCHI, ACM press, pp. 207-215, 2004.

[22] F. Lizano, M.M. Sandoval, A. Bruun, and J. Stage, "Usability Evaluation in a Digitally Emerging Country: A Survey Study," Proc. INTERACT, Springer Berlin Heidelberg, pp. 298-305, 2013.

[23] J. Nielsen, "Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier," in Cost-justifying usability, pp. 245-272, 1994.

[24] M.B. Skov and J. Stage, "Training software developers and designers to conduct usability evaluations," in Behaviour & Information Technology, 31(4), pp. 425-435, 2012.

[25] J.C. Lee and D.S. McCrickard, "Towards extreme (ly) usable software: Exploring tensions between usability and agile software development," in Proc. Agile Conference (AGILE), IEEE Press, pp. 59-71, 2007.

[26] X. Ferré, N. Juristo, and A. Moreno, "Which, When and How Usability Techniques and Activities Should be Integrated," in Human-Centered Software Engineering - Integrating Usability in the Software Development Lifecycle, Springer Netherlands, pp. 173-200, 2005.

[27] J.C. Lee, "Embracing agile development of usable software systems," In Proc.CHI'06 extended abstracts, ACM Press, pp. 1767-1770, 2006.

[28] A. Bruun and J. Stage, "Training software development practitioners in usability testing: an assessment acceptance and prioritization," Proc. OzCHI, ACM Press, pp.52-60, 2012.

[29] Z. Hussain, M. Lechner, H. Milchrahm, S. Shahzad, W. Slany, M. Umgeher, P. Wolkerstorfer, "Practical Usability in XP Software Development Processes," in Proc. ACHI, pp. 208-217, 2012.

[30] J.O. Bak, K. Nguten, P. Risgaard, and J. Stage, "Obstacles to Usability Evaluation in Practice: A Survey of Software Development Organizations," Proc. NordiCHI, ACM Press, pp.23-32, 2008.

[31] V. Bellotti, "Implications of current design practice for the use of HCI techniques," Proc. the Fourth Conference of the British Computer Society on People and computers, Cambridge University Press, pp. 13-34, 1988.

[32] K. Ehrlich and J. Rohn, "Cost justification of usability engineering: A vendor's perspective," in Cost-justifying usability, 1994.

[33] M.S. Andreasen, H.V. Nielsen, S.O. Schrøder, and J. Stage, "What happened to remote usability testing?: an empirical study of three methods," Proc. SIGCHI, ACM Press, pp. 1405-1414, 2007.

[34] A. Bruun, P. Gull, L. Hofmeister, and J. Stage, "Let your users do the testing: a comparison of three remote asynchronous usability testing methods," Proc. SIGCHI, ACM Press, pp. 1619-1628, 2009.

[35] R. Jeffries, J.R. Miller, C. Wharton, and K. Uyeda, "User interface evaluation in the real world: a comparison of four techniques," Proc. SIGCHI, ACM Press, pp. 119-124, 1991.

[36] S. Alshaali, Human-computer interaction: lessons from theory and practice. Doctoral dissertation, University of Southampton, 2011.

[37] A. Holzinger, "Usability engineering methods for software developers," Communications of the ACM, vol.48(1), pp.71-74, 2005.

[38] J. Kjeldskov, M.B. Skov, and J. Stage, "Instant data analysis: conducting usability evaluations in a day," Proc. NordiCHI, ACM Press, pp. 233-240, 2004.

[39] T. Borgholm and K.H. Madsen, "Cooperative usability practices," Communications of the ACM, vol.42(5), pp. 91-97, 1999.

[40] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," Proc. SIGCHI, ACM Press, pp. 249-256, 1990.

[41] H.R. Hartson, J.C. Castillo, J. Kelso, and W.C. Neale, "Remote evaluation: The network as an extension of the usability laboratory," Proc. CHI, ACM Press, pp. 228-235, 1996.

[42] S. Dray and D. Siegel, "Remote possibilities?: international usability testing at a distance,", Interactions, vol.11(2), pp. 10-17, 2004.

[43] F. Menghini. *Remote usability testing*. [Online]. Available from: http://internotredici.com/article/remoteusabilitytesting/ 2014.11.14

[44] F. Paternò, "Models for universal usability," Proc. the 15th French-speaking conference on human-computer interaction on 15eme Conference Francophone sur l'Interaction Homme-Machine, ACM Press, pp. 9-16, 2003.

[45] M. Hammontree, P. Weiler, and N. Nayak, "Remote usability testing," in Interactions, 1, 3, pp. 21-25, 1994.

[46] K.E. Thompson, E.P. Rozanski, and A.R. Haake, "Here, there, anywhere: Remote usability testing that works," Proc. Conference on Information Technology Education, ACM Press, pp. 132–137, 2004.

[47] J. Rubin and D. Chisnell, Handbook of usability testing: how to plan, design and conduct effective tests, John Wiley & Sons, 2008.

[48] J.L. Wynekoop, S.A. Conger, "A review of computer aided software engineering research methods," Department of Statistics and Computer Information Systems, School of Business and Public Administration, Bernard M. Baruch College of the City University of New York, 1992.

[49] K. Braa and R., Vidgen, "Interpretation, intervention, and reduction in the organizational laboratory: a framework for in-context information system research,", Accounting, Management and Information Technologies, vol.9(1), pp.25-47, 1999.

[50] J. Nielsen. *Best Application Designs.* [Online]. Available from: http://www.nngroup.com/articles/best-application-designs/ 2014.11.14

[51] J. Kjeldskov and C. Graham, "A review of mobile HCI research methods," Human-computer interaction with mobile devices and services, Springer Berlin Heidelberg, pp. 317-335, 2003.

[52] A. Bruun, Developer Driven and User Driven Usability Evaluations. Doctoral dissertation, Videnbasen for Aalborg Universitet, Det Teknisk-Naturvidenskabelige Fakultet, Aalborg University, The Faculty of Engineering and Science, 2013.

[53] A. Begel and B. Simon, "Novice software developers, all over again," Proc the Fourth international Workshop on Computing Education Research, ACM Press, pp. 3-14, 2008.

A Pseudo-HDR Method Implementation for Medical Images Enhancement

New post-processing method for X-ray chest image quality enhancement

Desislava Valentinova Georgieva
New Bulgarian University, NBU
Sofia, Bulgaria
e-mail: dvelcheva@nbu.bg

Vesselin Evgueniev Gueorguiev
Ivan Evgeniev Ivanov
Technical University Sofia, TUS
Sofia, Bulgaria
e-mail: veg@tu-sofia.bg, iei@tu-sofia.bg

***Abstract*— Medical images are an important part of the diagnostic process. In many cases the accuracy of diagnosis depends on the quality of the image. The investigation and development of new medical image processing methods and systems has received great attention over the last two decades. This is due to its wide range of applications in computer-assisted methods and computer-aided methods. The proposed paper presents a new image enhancement method for X-ray images. The method uses an HDR-image creation as a technique to increase the image dynamic range. This allows after mapping an HDR-image to an LDR-image (low dynamic range image) to get a better distribution of the intensity over all pixels in the image. The result is enhancing brightness, contrast and/or sharpness of images without the appearance of visible medical artefacts.**

Keywords - X-ray medical imaging; HDR-images; image quality enhancement.

I. Introduction

Medical imaging is the process of creating images (visual representations) of the human body organs and tissues for medical purposes to diagnose and treat disease - it is intended to reveal internal human body structures hidden by the skin and bones. The process is defined as non-invasive as it does not disturb the physical integrity of the skin and tissues.

The main method of medical images interpretation is the visual one, makes the process depend on the observer [1]. It has been found that some doctors have systematic underestimation or overestimation of the information in the image - they reject some information as insignificant or overestimate the importance of structures in the image. Additional influence on the final image is exerted by overall readability of the image because the conclusions "no changes" or "no noticeable changes" are not the same thing. Some other factors for the proper medical image interpretation are time for investigation, environment, image lighting, type of media of the image, and the physiological/functional status of the physician [2][3]. All this leads to the need to improve the readability of images and to reduce the factors influencing the perception of information in images.

The techniques for obtaining medical images are not harmless to patients: different amounts of radiation are absorbed by the various body tissues (bones, muscles, soft tissues, etc.), leading to resulting physical and chemical reactions, as well as to a variety of physiological changes in the human body. Clinical practice mainly uses two types of X-ray images: the standard radiological images obtained by projection, and fluoroscopic images representing X-ray images in real time, used as an aid in medical procedures. Notwithstanding the advantages of CT and PET imaging, these two radiographic techniques are widespread due to their low price, relatively high quality and relatively low doses of patient radiation.

To obtain better quality X-ray images with greater contrast and clearer details it is necessary to increase the amount of radiation that a patient is subject to. To avoid this harmful effect on humans, a variety of technologies have been developed by using additional grids to the X-ray detectors for absorbing the scattered radiation and reduction of the size of the X-ray beam. Another approach, which is widely used, is the use of methods for computer-assisted diagnosis, to improve the quality of medical images without increasing the radiation load on the patient.

Radiographic studies are one of the most commonly used techniques of the diagnostic process in pulmology (the X-rays and CT scans are the most frequently used image acquisition techniques). For the proper interpretation of X-ray images the quality of the image has the greatest importance. While the image quality of modern digital X-ray apparatus is good, with conventional X-ray apparatus in many cases the quality is unsatisfactory. This is one of the main sources for the wrong interpretation of the image, resulting in hypo- or hyper- diagnosis. An example of a disease that may be "masked" is the peripheral lung carcinoma. In the initial stages of this carcinoma the X-ray density is usually not big and when the X-ray image is overexposed it may not be visible. Detection of this carcinoma in this early stage is very important because it rarely manifests any symptoms at this stage, but in an advanced stage radical treatment is impossible. So the X-ray image reading is the only way for early diagnosis.

Other examples of lung diseases that require high quality images are secondary tuberculosis (especially in the case of circular fireplace), viral pneumonia and pulmonary embolism.

This paper presents a new image enhancement method for X-ray images. The method uses HDR-image creation as a technique to increase the image dynamic range. This allows after mapping HDR-image to LDR-image (low dynamic range image) to get a better distribution of the intensity over all pixels in the image. The result is enhancing brightness, contrast and/or sharpness of images without the appearance of visible medical artefacts.

This present paper is structured as follows: Section II looks into the set of methods for HRD imaging; Section III presents the proposed new enhancement method; Section IV presents the implementation and analyses of the presented method; Section IV presents the conclusion.

II. WHY MEDICAL IMAGE ENHANCEMENT METHODS ?

The investigation and development of new medical image processing methods and systems has received great attention over the last two decades. This is due to its wide range of applications in computer-assisted methods and computer-aided methods.

Regardless of the quality of X-rays, there are many reasons that necessitate the creation of new methods and techniques for improving quality and readability of digital images used for medical diagnosis. Our research showed that the most frequent by encountered reasons for using methods and techniques for medical images post-processing can be grouped in the following way:

- The image isn't captured with the optimal X-ray apparatus settings. The most important for the X-ray picture quality is the choice of exposure, because it can get overexposed or underexposed images (for digital the X-ray apparatus the selected type of linearization of X-ray sensor affects also the choice of exposure values). Often the resulting image is acceptable, but the contrast is not the best for a correct interpretation of the image (Fig. 1a). This has a significant impact on reading images containing small size structures or in need of early diagnosis of diseases, where the change in X-ray density of the abnormality is less pronounced, i.e. the change in gray tones for an examined area is still small and hardly noticeable.
- The characteristics of the display or output media do not allow you to see all the special features of the examined image. This most often occurs when the image has a greater depth than the possibilities for visualisation. Since, in this case, the computer system performs compression of the image depth; the result is loss of contrast and dynamic range. A very common manifestation of this problem is recording the medical image in a format different from the DICOM one (Fig. 1c). Most often this is used to reduce the volume of archieved images (the digital patient archive), or if necessary to make a copy for the patient.
- It is physicians who diagnose using X-ray picrutres. It has been established that the physician's functional, physiological and mental status has direct impact on the ability to correctly read the X-rays [2][3][4]. In this case, the image post-processing allows altering the digital image characteristics so that they become readable for the physician.

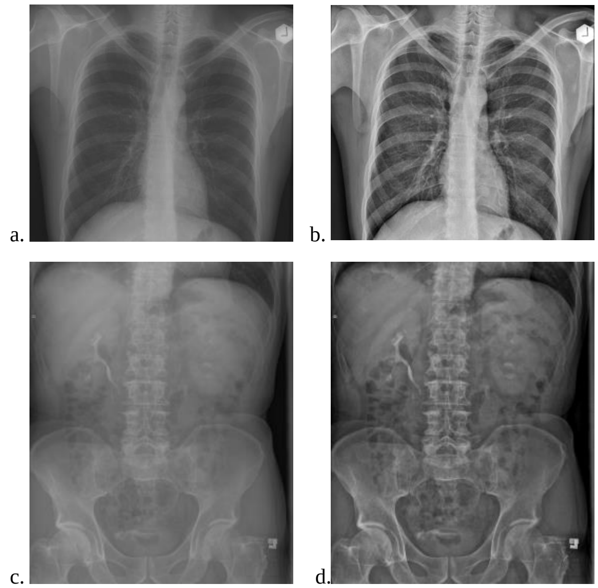


Figure 1. Examples of the necessity to use a post-processing of the images before they are used for diagnostic: a) image after capturing by incorrect apparatus settings; b) improvement of quality of image (1.a) with our method; c) image from digital X-ray apparatus, recorded in a file with 8 bits depth (JPG without compression); d) improvement of quality of image (1.c) with our method.

Among the many types of image processing tools, image enhancement is one of the vital processes in medical imaging systems – it is one of the preparatory steps and it is applied before starting the image analyses. Image enhancement refers to any technique that improves or modifies digital images, so the resulting image is better suited than the original for a particular application. Essential image enhancement includes but is not limited to intensity and contrast manipulation, noise reduction, background removal, sharpening and filtering edges. In this context, 'image enhancement' means any method or technique which change digital images, so the resulting image is better suited than the original to a particular application. Due to this the basic types of image enhancement include manipulation of intensity, changing the local or global contrast, noise reduction, filtering and sharpening edges. During the image enhancement process one or more attributes of the image are modified. The choice of attributes and the way they are modified is specific to a given task. Moreover observer-specific factors such as the human visual system and the observer's experience will introduce a great deal of subjectivity into the choice of image enhancement methods [1].

X-ray images are grayscale images with 12-14 bits depth and their visual perception depends on the three most common image characteristics: brightness, contrast (local and global) and sharpness. Apart from these saturation and image dynamic range have a significant influence on the human perception of the images but they are not directly relevant to X-ray images, because images are grayscale (no saturation), and the dynamic range of the visualization systems (computer displays) is less than human vision dynamic range. Therefore, all quality enhancement methods change the intensity of pixels so as to provide optimal brightness, contrast and sharpness values. While brightness, contrast and sharpness may appear to be the simplest of image controls on the surface and may appear to be mutually exclusive controls, they are related and intertwined in such a way that changing any one of them can create quite complex effects in post-processed images. This specifies a wide variety of methods that have been proposed and are being created now – each of these methods seeks to solve the task of determining the image characteristics optimal values.

A sample classification of medical image enhancement methods is shown in the Fig. 2. [8][9][10][11]

In the image processing literature there is a wide variety of image processing operators, but only a part of them are used for the processing of medical images. The main reason is the need for preservation of the medical authenticity of the image after applying the selected operator. Our analyses show that medical systems most often use the following functional groups of operators.

- Arithmetic and Logic operators: these are different forms of unification/fusion of the pixels of the two images to obtain the resultant pixel.
- LUT operators (Look-up-table operators): these operators change pixels' grayscale levels by functional transformation on a single pixel.

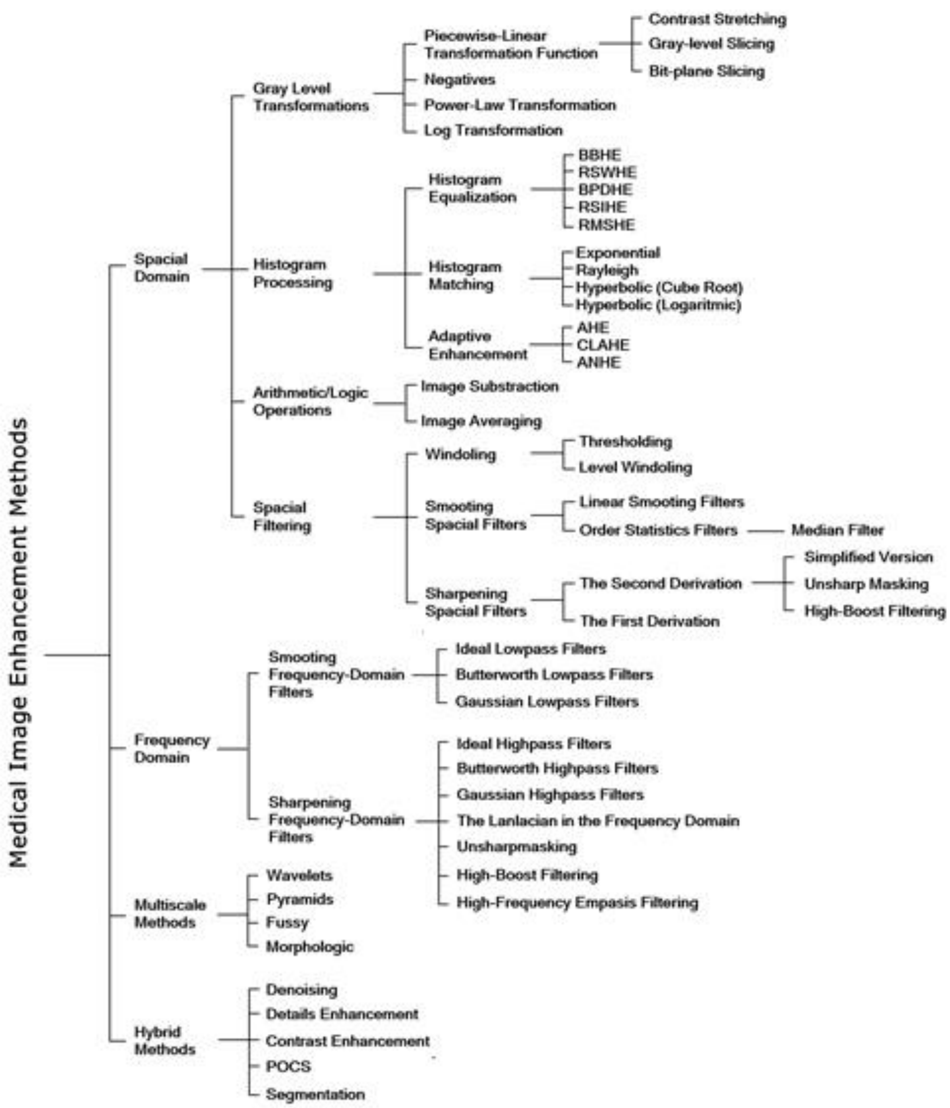


Figure 2. Medical image enhancement methods classification.

Geometric operators: the most common purpose of these operators is to remove the geometric distortions and shifting of the image, due to the image acquisition. Operators for image translation, rotation, and scaling are among those most often used.

- Image analysis operators.
- Morphologic operators: these are operators for analyzing pixels groups using mathematical morphology.
- Digital filters
- Attribute operators: these are operators for detection of basic properties or characteristics; this group includes operators, which are used to detect edges, lines or other specific structures.
- Image transformation operators: they are used to transform the image into another presentation in order to more easily process it for certain needs.
- Synthesis operators: most often these are noise generation methods.

Fig. 3 shows a more extended classification of the most commonly used operators for image processing in medical imaging systems.

III. X-RAYS IMAGE ACQUISITION AND QUALITY CHARACTERISTICS

During the X-ray image creation process the number of X-rays which interact with the human body tissues and organs depends on the thickness and distribution of the anatomical structures in the body. The ‘diagnostic’ X-rays are primarily those that occur in photoelectric and Compton processes:

- Photoelectric interactions are the most important ones in the formation of images, because of the strong dependence of the photoelectric process on the location of the absorbent tissue atoms and the absence of secondary radiation.
- Compton-type interactions occur in soft tissues, depending mainly on their density. During these interactions multiple scattered photons occur, which leave the patient's body and reach the detector. Their direction is not associated with the focal point of the X-ray tube, which primarily causea a decrease of the image contrast.

There are three main options for the interaction of photons with the human body organs and tissues:

- All the X-ray photons pass through the patient and reach the detector (the X-ray plate or the digital X-ray sensor). Such is the case when the X-ray photons pass through areas with air in the lungs. In this case the result is an image with dark (almost black) areas in the lungs.
- The X-ray photons are absorbed completely, for example in interaction with metal elements. In this case the metal objects are displayed in white on the X-ray.
- The most common case of interaction of radiation with human tissues is when the X-ray photons reach tissues and there is photons scattering and attenuation. So only part of the photons reach the detector and this creates areas with different levels of gray in the final X-ray image.

If the X-ray photons will be absorbed or not by the patient’s tissues depends on several factors, the most important of which are the energy of the X-ray beam and the density of the human tissue. Therefore, one of the main tasks of the radiologist is the correct selection of the energy of the X-ray beam according to the patient tissues, which will be explored. This directly affects the X-ray image quality.

In medical X-rays five basic grey levels (from black to white) can be defined. This classification is based on the main basic X-ray densities of tissues and organs, as the tissues with low density are plotted with darker tones, compared to those with greater density:

- black – air in the lungs;
- dark gray – subcutaneous tissue (fat);
- light gray – the heart and blood vessels (soft tissues, water, blood, etc.)
- near to white – clavicles, ribs and bones;
- white - metal items (jewelry, orthopedic items, hospital equipment, marker left side).

The principle of reading the X-rays is based on the fact that the structures or organs can be recognized in the image only based on the differences in X-ray density of adjancent structures, such as light gray areas of the heart to black areas of air in lungs. In this way the interpretation of X-ray images requirs the areas with normal and pathological coloring in gray tones to be defined [12][13][14][15].

According to modern concepts the medical image quality can be described by three basic physical characteristics: contrast, sharpness and noise [16][21][22]. Between these characteristics there is a mutual interdependence, the improvement of one can lead to the deterioration of another [17][19]. Three levels, characterizing the quality of the image [1], are used in practice:

- Visualization – the object is detected without its details being reproduced;
- Display - details of anatomical structures are outlined, but out necessarily distinct;
- Imaging - details of the anatomical structures are clearly outlined.

Contrast is one of the main characteristics of the X-ray image, directly affecting the general readability of the image. Ray contrast is defined as the relationship between the difference of power transmission at two different points, and its average (background) value [20]. It depends on the tissues attenuation coefficients and the tissues thickness. Radiographic contrast is determined by the resulting visible image: it represents the difference between the optical densities (or between brightness) in two adjacent areas [21]. At the same time, the effect of the scattering of X-rays is expressed in almost constant intensity of the background I_S, imposed on the real image. This results in the decreasing of the image contrast which can be expressed by the formula:

$$C = \log_{10}\left(\frac{I_1 + I_S}{I_2 + I_S}\right) \tag{1}$$

where I_1 is the intensity of X-rays passed through the soft tissue and I_2 is the intensity of the X-rays passed through the bones.

The type of transition in optical density (or brightness of the image) characterizes the sharpness of the X-ray image.

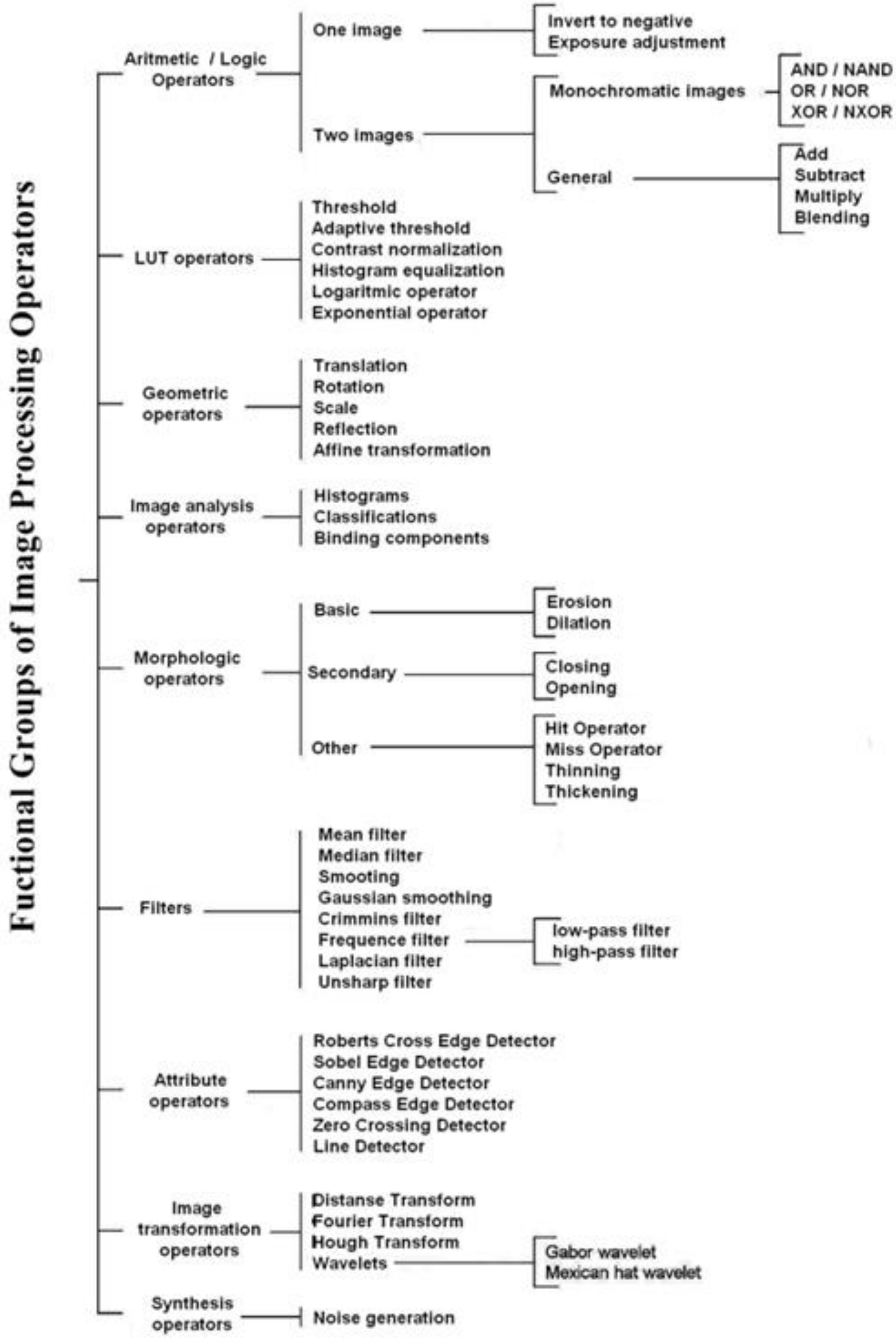


Figure 3. The most commonly used operators for the image processing in the medical imaging systems.

In medical imaging the main reasons for the decrease in sharpness are the following:

- Loss of geometric sharpness is a consequence of the size of the optical focus of the X-ray tube. Apart from reduction of its size, it can be limited through the use of as little as possible zoom for less distance between the object and the detector.
- Loss of dynamic sharpness is due to patient movement during the X-ray image capturing and can be reduced by shorter exposure times.
- Loss of contour sharpness is caused by rounded shapes of the organs and structures in the patient body.

Total loss of sharpness of the image U_{TOT} depends on several factors and the influence of each of them can be expressed by the expression:

$$U_{TOT} \approx \sqrt{\left(U_G^2 + U_M^2 + U_D^2 + U_I^2\right)} \qquad (2)$$

where:

- U_G characterizes the position of the patient relative to the X-ray tube and the detector and can be determined by the formula (FP_{size} is the size of the focal point; $D_{obj\text{-}sensor}$ is the distance between the object and the detector; and $D_{focus\text{-}obj}$ is the distance between the focus point and the object)

$$U_G = \frac{FP_{size} * D_{obj-sensor}}{D_{focus-obj}} \qquad (3)$$

- U_M expresses the effect of the patient shift during image capturing.
- U_D is called a detector blur and characterizes the effect of the detector type;
- U_I is the so-called ‘inherent blur’, which is determined by the shape of the captured object.

In this way, the total loss of sharpness characterizes the change in the size of the smallest visible detail in the image.

The X-ray image noise appears as random variations in the signal of the converter in a homogeneous structure of objects in digital radiographic systems. There are several noise sources and the main source of noise is quantum noise – it is the result of statistical fluctuations in the number of X-rays reaching the detector [19].

For grayscale images, as most medical images are, the image histogram can be used as an initial characteristic of image quality – it may show the areas that need digital enhancement. It is particularly important to take into account the fact that grayscale levels and pixels values are not the same thing, because the change of the image can be done by changing the grayscale levels, without changing the pixels values [18]. This is realized through the change of the Look up Table (LUT).

Another important characteristic and criterion for the image quality evaluation is the dynamic range – it describes the difference between the maximum (a_{max}) and minimum (a_{min}) value for a pixel in the image:

$$D = 20 * \log_{10} (a_{max} - a_{min}) \qquad (4)$$

The correlation between the dynamic range and the contrast of the image has been a long-known fact. For the grayscale medical images this relationship is even more pronounced, but that does not mean that contrast and dynamic range are synonymous – contrast depends not only on the dynamic range, but also on the average distribution of the values of the pixels in the image, as well as on the number of peaks in the image histogram.

Theoretically, for medical X-ray image capturing the dynamic range of the detector should be similar to the range of the emitter. In cases in which the emitter has a greater dynamic range than the detector, a cut of received values outside the dynamic range of the detector is obtained – it can easily be seen in the histogram of the image (the histogram has peaks in the black part and/or in the white part). In practice the reverse also occurs: the dynamic range of the detector is sagnificantly larger than the dynamic range of the transmitter – in this case, the values are grouped in a limited part of the histogram, and in the black and/or in the white part of the histogram there are no values.

IV. The Most Common Problems with the X-ray Image Quality

Digital X-ray images can be obtained in two ways – after digitization of X-ray plates/films, or from digital X-ray apparatus. These are two completely different sources of images, so there are different reasons for change in the quality:

- With digital X-rays the main problem is related to the image format in a long-term registers – the DICOM format is very unconvinient for those purposes due to the large volume of images. Our studies showed that a hospital with 50 000 patients needs 5-8 GBytes per year new external space only for storage of digital X-rays. So increasingly, hospitals after a certain period transfer images from external memory on new generation X-ray plates or films.
- The main problems of quality of X-rays captured by the classical X-ray apparatus are related to the calibration of the X-ray apparatus and with changes in X-ray plate during a prolonged storage. The digitalization process can also lead to changes in the quality of the final digital X-rays, but these kinds of problems of are not part of the discussion in this article.

The two most common problems resulting from the of X-ray apparatus calibration (other than the noise) are obtaining overexposed or underexposed X-rays. This leads to a change in the image dynamic range and/or the image contrast, as well as to the possibility to improve X-image quality by post-processing techniques:

- Underexposed X-rays - these are very bright images and the main problem is related to the lack of sufficient visual information in the image. Such images have small dynamic range and a unimodal histogram with a clearly expressed peak in the white end of the histogram (Fig. 4). Additionally, the histogram may have several distinct local peaks (usually one or two). In this case, the classical methods for histogram manipulation do not create medical trustworthy images, because a post-processed image has characteristics of pronounced pathologies.

- Overexposed X-rays – these images are too dark areas, due to the large amount of radiation passed through the patient. They have a bimodal histogram with peaks near the white and black ends, i.e. they have a good dynamic range (Fig. 5). This is the result of overexposure in the soft tissues (the tones close to the black) and the regions with higher radiographic density (e.g., thicker bone) that have colors near to white. This pronounced bimodal histogram with large and very distant peaks greatly decreases the applicability of classical methods for image contrast improvement.

Problems, which are the result of improper storage of X-ray plates, are due to the change of the physical and/or photo-optical characteristics of the plates. These changes have led to a change in the quality of the recorded image:

- The whitening of the X-ray image – this leads to loss of the image dynamic range (Fig. 6).

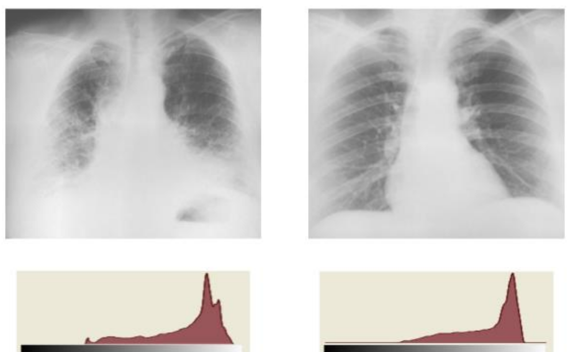


Figure 4. Examples of underexposed X-rays and their histograms.

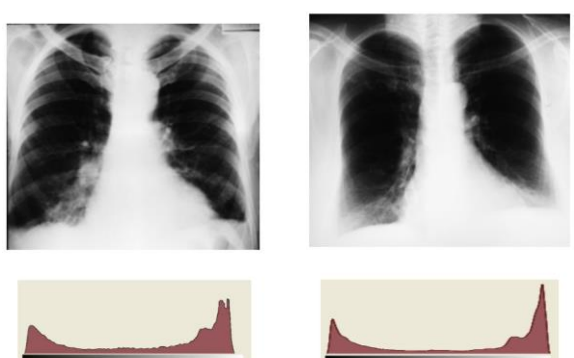


Figure 5. Examples of overexposed X-rays and their histograms.

- The X-rays cease to be grayscale images and turn into colored images (Fig. 7).
- The appearance of characteristic and clearly distinct linear or point changes in the X-ray plate surface.

In terms of quality enhancement methods the X-rays whitening is the most interesting problem. The reason for its occurrence is exposing plates to constant external light (e.g., sunlight). This leads to change of photo-optical characteristics of the plate, i.e. light levels in adjacent dark and light areas converge. The result is a reduction of the image dynamic range and/or decreasing of the image contrast. In this case the histogram change resembles the change, which is the result of gamma-correction.

The coloring of the X-ray plates is also a significant problem, because quality enhancement methods of digital X-ray images are developed for grayscale images. The problem is very interesting, but its discussion is out of the scope of this article.

V. HDR Imaging

High Dynamic Range Imaging is a set of methods in photography/imaging, supposed to capture/create a greater dynamic range between the darkest and lightest image areas than current standard digital imaging methods [5][6]. The information in the HDR-image file isn‘t based on based on the chosen color model – the stored values are logarithmically encoded or floating-point linear values, or gamma compressed values of luminance [5].

Figure 6. The result of whitening of X-rays: a) clean whitening; b) whitening+coloring.

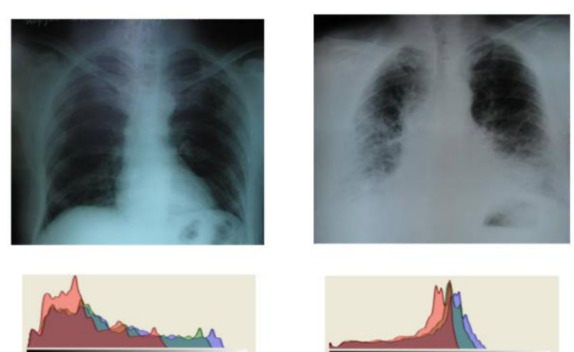


Figure 7. Examples of X-ray plates coloring resulting of aging of the plate material.

At the same time the HDR-images don’t use fixed presentation of individual color channels – to represent the increased amount of colors they use floatin-point values (the LDR-images use integer values to represent the color channels of selected color model). That is why HDR-images provide much more correctly the luminance levels in the captured scene compared to classic images with low dynamic range (LDR-images).

The human eye covers the dynamic range of about 10^5:1 at one time and this is bigger than the top dynamic range of most real-world scenes. For comparison, computer displays have dynamic range of 10^3:1 and digital cameras have dynamic range of 10^4:1. In the last two years HDR cameras with dynamic range just over normal human vision dynamic range and displays with near to human vision dynamic range began to appear on the market.

The human vision can be accommodated to a dynamic range of 10^{14}:1 but the iris is simply not as flexible and the human perception of intensity changes is logarithmic (the Weber law). This is much more than the capabilities of modern devices for image creation and visualization. Therefore, a non-HDR image device takes pictures at one exposure level with a limited contrast range. This leads to the loss of details in dark or bright image areas, depending on the camera exposure setting. HDR methods compensate detail loss by taking multiple pictures at different exposure levels and stitching them together to create an image which presents the greatest number of details in both dark and bright areas. Data stored in HDR-images typically corresponds to the physical values of luminance/radiance that can be observed in the real world and this presents a great difference from classical digital images: classical digital images represent intensities and colours that should appear on an output device (display, printer, plotter, etc.). Therefore, HDR image formats are called scene-referred while classical digital images are called device-referred.

In photography dynamic range is measured in EV (Exposure Values) differences between the darkest and brightest parts of the image that show detail: an increase of 1 EV is a doubling of the amount of light. Using EVs not very strict categories of images are [7]:

- High Dynamic Range (HDR) images: These have a dynamic range of about 14EV and these images (they use 32-bit float values without limitation for channels bits depth) are usually produced by merging multiple 12-14 bits images of different exposures (most often these are raw data files).
- Medium Dynamic Range (MDR) images: These have a dynamic range of [9 EV, 12 EV] and can originate from a file with 16-bit depth, or by merging 3 or more 8-bits images with different exposures.
- Low Dynamic Range (LDR) images: These have a dynamic range of lover than 8 EV. This means one 8-bits image.

Historically, the first versions of obtaining images with greater dynamic range began in the 19th century when films having several layers with different sensitivities were developed. By analogy, the computer-based generation of HDR-images require several images captured at different exposure, and then these LDR-images are merged. So each separate exposure captures various illuminations in the natural scene, i.e. each LDR-image stores information for a different part of the real dynamic range of the scene. Therefore, the information about the luminance from the individual parts of the scene dynamic range can be merged to generate the final HDR image [23]. The main problem in this case is the correct selection of the number of LDR-images and the choice of exposure values suitable to the characteristics of the scene.

VI. Our method for chest X-ray image quality enhancement

X-ray images are 12-14 bit grayscale images and their visual perception depends on the three most common image characteristics: brightness, contrast (local and global) and sharpness. Thus, when the image has no sufficient quality, this is the result of some incorrect values. As stored information in the grayscale images is the values for intensity of the image pixels, then all methods for quality enhancement are aimed at changing the pixel intensity as a way to change the basics characteristics of the image. This limits the opportunities for selection of optimal values, because a limited amount of information about the luminosity/radiance power stored as pixel intensity is used.

The most common feature of the existing medical image quality enhancement methods is the fact that they cannot substantially increase the dynamic range of the image, which is why the improvement of the characteristics of the image (contrast, sharpness, brightness) is achieved through redistribution of the values of the intensity of the pixels in the image. Therefore, these methods improve the basic image characteristics (contrast, sharpness, brightness) by redistribution of intensity values to the pixels in the image. The main problem of this approach is the possibility of occurrence of medical artifacts – the classical criteria for evaluation of contrast enhancement (eg, increasing the dynamic range and maximizing the contrast) are inapplicable, since the resulting grayscale levels of pixels can be an indicator of pathological changes in tissues and organs (Fig. 8). Since there is no established computer metrics for medical authenticity, it cannot rely on automated techniques for evaluation of image quality of post-processed medical images. It is necessary to seek new approaches to solve this task which enable a qualified physicion to control the image quality enhancement process manually.

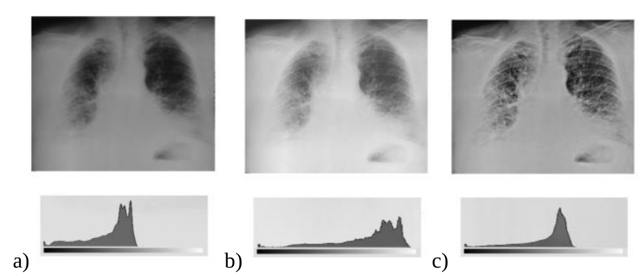


Figure 8. The most important factor of the medical image post-processing is the preservation of the authenticity of the image: a) the generic image; b) the image after classical enhancement of the dynamic range and the contrast; c) the most improved image without the occurrence of medical artifacts.

Our method uses a different approach to solve the issue of the optimal intensity distribution over image pixels. Following this approach a model of the luminosity distribution is created instead, which has led to the current image. This is achieved by creating a HDR-image because it represents the description of the luminosity/radiance in the nature scene. After a HDR-image is created the method allows determining the optimal mapping from a HDR-image to a LDR-image.

The use of the HDR-image to describe the scene luminosity model is justified, because the photosensitive material of the X-ray films and the X-ray-plates doesn't represent a monolayer structure – it is a multilayer structure. Thus, the energy of the photons passing through the material determines the level of the change in depth. Therefore, the photo-sensor catches different illumination information from different levels of photosensitive material under different exposure levels. This approach can be used directly for digitalization of X-ray plates [24] but if you want to use it as an image quality enhancement method you needed to create a technique for generating LDR-images with pixels intensity values corresponding to the relevant exposure values of the scene. This is one of the main tasks of the proposed method.

The proposed new quality enhancement method of for digital X-rays has the following three consecutive stages:

- Creating the LDR-images simulating capturing at different exposure;
- Creating an HDR-image as a luminosity model.
- Obtaining the final image by controlling the process of transformation from the HDR-image to the LDR-image.

A. *Stage 1: Simulation capturing at different exposure*

To achieve the correct results, it is necessary to establish a correct luminosity model of the simulated scene. For the HDR-image this is achieved by correctly selected additional images with different exposure. In photography this is achieved through capturing a new image with a selected exposure. Here this is not applicable and the main problem is to obtain an image that is accurate enough to simulate changes in the original image after changing the exposure.

From the image processing point of view increasing or decreasing the exposure changes the values of brightness, contrast and sharpness. Therefore, if the change of image pixels intensity resulting from the exposure change can be imitated, it can be used to simulate the image exposure change when a HDR-image is created. Our tests and analyses of results showed that for simulation a change in intensity a few different techniques can be used: using the brightness and the contrast control; using the gamma-correction; using the brightness and the contrast control followed by a gamma-correction; using the gamma-correction followed by a brightness and contrast correction

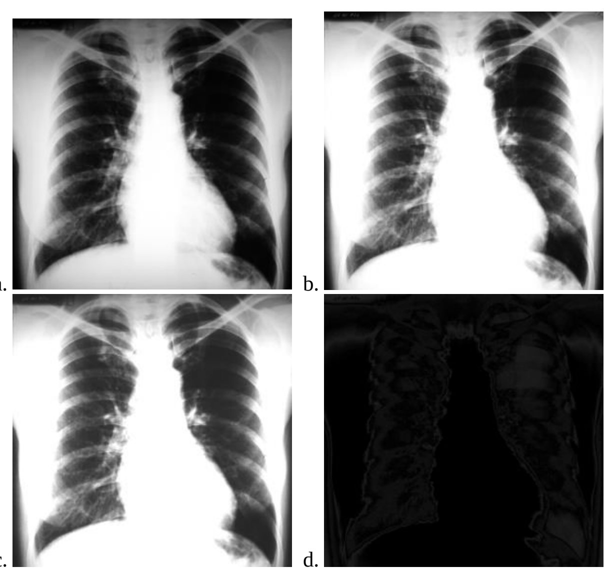
1) *Using the brightness and the contrast control*

One approach to solve the problem is based on the understanding that exposure change by 1 EV means doubling the amount of light. As the visual result is increasing of the pixels intensity for the entire image, the imitation of intensity shift requires calculation of brightness shift. Unfairness of this approach is that doubling the amount of light does not lead to doubling pixels intensity, because graphic devices and the characteristics of the created images reflect the human vision characteristics (logarithmic law for change of the intensity sensibility). Therefore, besides brightness there is also a considerable change in contrast.

Tests to determine brightness and contrast values were conducted: X-rays are captured with different exposures (from -3 EV to +3 EV by a 0.5 EV step) and the difference between the real image and the simulated image is evaluated to select values for brightness and contrast – Table I shows the results obtained for brightness and contrast (values of brightness and contrast are between -100 and +100). An example of -1.5 EV exposure simulations is shown in Fig. 9.

TABLE I. EXPOSURE SIMULATION: GAMMA-CORRECTION VALUES

	Exposure (EV steps)					
		-2.5	***-2***	***-0.5***	***-1***	***-0.5***
Brightness		-81	-70	-55	-38	-21
Contrast		-35	-27	-20	-11	-5
	Exposure (EV steps)					
	0.5	***1***	***1.5***	***2***	***2.5***	
Brightness	22	40	56	71	83	
Contrast	6	16	26	34	46	



a. b. c. d.

Figure 9. Using brightness and contrast control: a) the original image; b) the image with +1.5 EV; c) the simulated image with -1.5 EV; d) the difference between images (b) and (c) - the histogram is stretched twice in order to see the difference.

Our experiments show that simulation of exposures above 2.5 EV and below -2.5 EV is unrealistic and cannot be used for HDR-like image generation – when mapping to a LDR-image the result always contains medical artefacts.

However, for bone X-rays, this approach gives very good simulations.

2) Using the gamma-correction control

Another way to simulate changing the intensity of pixels is by changing the gamma-correction.

The difference between brightness and gamma-correction control is that increasing the value of gamma-correction can make the image to look brighter, but it is a non-linear change and it only increases brightness of the shadows and mid-tones in the image without affecting the highlights. Our experiments showed that this is particularly useful for simulating the overexposed images or the lung X-rays.

Another significant difference is the ability to simulate exposure values in the range [-5 EV, +5 EV]. Fig. 3 shows an example from Fig. 10, and Table II shows calculated values for gamma-correction.

TABLE II. EXPOSURE SIMULATION: GAMMA-CORRECTION VALUES

	Exposure (EV steps)					
	-3	***-2.5***	***-2***	***-0.5***	***-1***	***-0.5***
gamma-correction	6.0	4.9	3.7	2.8	1.9	1.3
	Exposure (EV steps)					
	0.5	***1***	***1.5***	***2***	***2.5***	***3***
gamma-correction	0.81	0.71	0.6	0.52	0.45	0.4

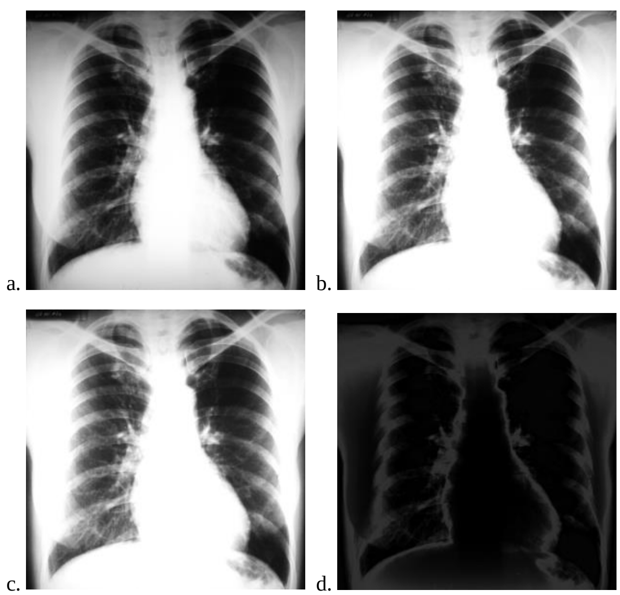


Figure 10. Using gamma-correction control: a) the original image; b) the image with +1.5 EV; c) the simulated image with -1.5 EV; d) the difference between images (b) and (c) - the histogram is stretched twice.

3) Using the brightness and the contrast control followed by gamma-correction

The main disadvantage of using brightness and contrast control is the incorrect change of local contrast between lung structures and ribs. That is why we tested additional image correction – the gamma correction. The result is a significant improvement of the simulation - Fig. 11 shows the example from Fig. 9, but now with the new way of correction. Table III shows calculated values for simulation of an exposure change.

4) Using the gamma-correction control followed by brightness and the contrast correction

The last approach to create an exposure simulation is gamma-correction control followed by brightness and contrast correction. This approach differs from the previous one, because the operations are not commutative. When comparing the result with the second approach, it appears that in this case the lighter areas are correctly changed.

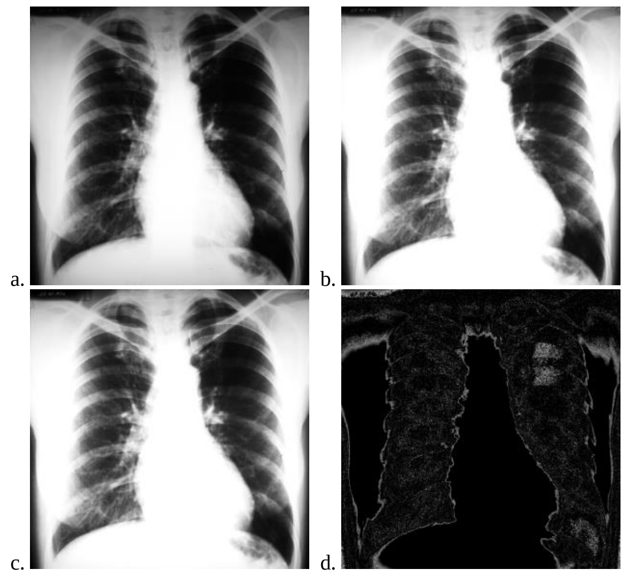


Figure 11. Using brightness and contrast control: a) the original image; b) the image with +1.5 EV; c) the simulated image with -1.5 EV; d) the difference between images (b) and (c) - the histogram is stretched 32 times.

TABLE III. EXPOSURE SIMULATION: BRIGHTNESS AND CONTRAST FOLLOWED BY GAMMA-CORRECTION

	Exposure (EV steps)					
		-2.5	***-2***	***-0.5***	***-1***	***-0.5***
brightness		-81	-70	-55	-38	-21
contrast		-35	-27	-20	-11	-5
gamma-correction		1.55	1.34	1.21	1.12	1.05
	Exposure (EV steps)					
	0.5	***1***	***1.5***	***2***	***2.5***	
brightness	22	40	56	71	83	
contrast	6	16	26	34	46	
gamma-correction	0.95	0.87	0.78	0.66	0.53	

The result is the best simulation of exposure change of an image - Fig. 12 shows the example from Fig. 10. Table IV shows calculated values for simulation of an exposure change.

Another advantage of the third approach is the possibility to simulate a much larger range of exposure values.

B. Stage 2: The HDR-image generation

To implement this step of the method, it is necessary to use a product for HDR-image generation based on a set of LDR-images with known exposure values. Analysis of the reqirements for digital X-ray images was conducted. The possibilities for creating an image with different exposures have been determined.

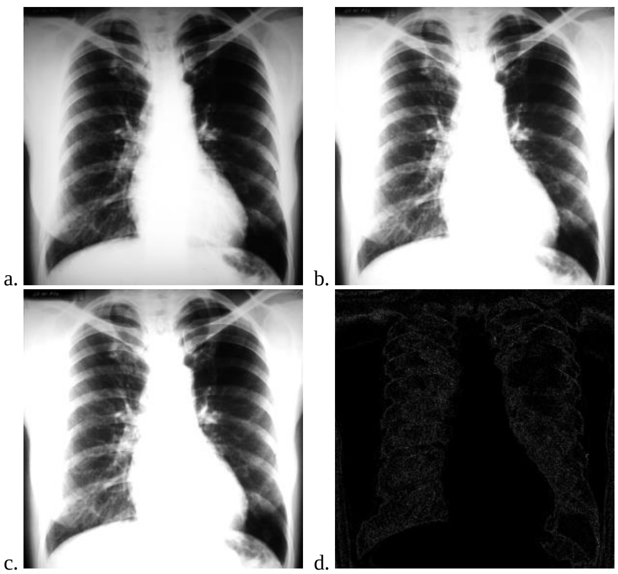


Figure 12. Using brightness and contrast control: a) an original image; b) an image with +1.5 EV; c) a simulated image with -1.5 EV; d) the difference between images (b) and (c) - the histogram is stretched 32 times.

TABLE IV. EXPOSURE SIMULATION: GAMMA-CORRECTION CONTROL FOLLOWED BY BRIGHTNESS AND CONTRAST CORRECTION

	Exposure (EV steps)					
	-3	***-2.5***	***-2***	***-0.5***	***-1***	***-0.5***
brightness	-52	-41	-26	-13	-4	-2
contrast	-55	-48	-34	-23	-17	-9
gamma-correction	6.0	4.9	3.7	2.8	1.9	1.3
	Exposure (EV steps)					
	0.5	***1***	***1.5***	***2***	***2.5***	***3***
brightness	6	14	18	22	25	27
contrast	7	17	24	32	38	42
gamma-correction	0.81	0.71	0.6	0.52	0.45	0.4

Using the above analysis and the deretmined possibilities the following criteria for usability of the existing sowtwate sysytems for HDR-generation have been defined:

- Criterion 1: The level of realism of the resulting images after mapping an HDR-image to an LDR-image – this is the main criterion for the product selection because it is characterized by the appearance of artifacts in the final image, which is unacceptable for medical diagnosis. The evaluation of products by this criterion was done by a pulmonary diseases specialist.
- Criterion 2: The available set of input file formats – mainlywe evaluate the ability to work with 16-bits or 32- bits formats for input LDR images (8-bits images strongly restrict the possibility of achieving a large dynamic range)
- Criterion 3: Feasibility of alignment control by geometrical criteria or by image objects features - in the theory of HDR-images for calculation of the scene dynamic range, it is necessary to specify the matching objects in images with different exposures, because changes in their illumination is used for determination of illumination of the real scene, i.e. to calculate the real dynamic range of the scene. As for the proposed method the scene is static, the change of exposure greatly changes the forms and structures of the objects, but not their location relative to the boundary of the image. That is why the use of geometric criteria (available to pivot) gives much better quality results and there are much less artifacts in the final image. In real scenes (shooting nature) this is very rare and therefore most products use object alignment only.
- Criterion 4: the ability to edit the HDR-image – This is very significant, because the generation of an HDR-image can lead to the occurrence of noise or chromatic aberration in the ldr-image. At the same time, the ability to edit the scene luminance model allows generation of an LDR-image with much higher quality.
- Criterion 5: Ability to manually control the process of mapping an HDR-image to an LDR-image – the process of selecting the optimum dynamic range of an LDR-image directly affects the quality of the final image, because changing the contrast, the brightness, and the dynamic range affect the distribution of intensities levels over tissues with different radiographic densities.
- Criterion 6: Ability to generate a 16 bits output ldr-image - this is a mandatory requirement for X-ray images (digital X-rays are 12- or 14-bits images).

The results of the survey on the individual criteria are shown in Table V – the marks are between 1 star (the lowest) and 5 stars (the highest).

Based on these analyses, the Photomatix Pro system was selected, since it best meets the needs of the proposed method – a medically inauthentic image was not generated in any of the tests.

C. Stage 3: The mapping of HDR-image to LDR-image

When setting the tasks for this method stage, analogy with methods for image pre-processing in digital X-ray apparatus has been used – medical X-rays apparatus have a group of methods for the digital sensor calibration and linearization (the goal is the best possible distribution of intensities levels over captured image). Since in the developed method the HDR-image acts as a digital sensor, the task at this stage is to use a mapping procedure between the HDR-image and the LDR-image to select optimal LDR-image characteristics (dynamic range, contrast, ryaskost, noise, brightness, etc.) without generating medical artifacts.

TABLE V. EVALUATION OF PRODUCTS FOR HDR IMAGE GENERATION.

	1	2	3	4	5	6
Artizen HDR	****	*****	***	**	*****	****
Dynamic Photo HDR	***	***	***	****	****	****
easyHDR	***	*****	**	***	****	****
Essential HDR	**	***	**	***	***	***
HDR Darkroom	****	****	***	**	***	***
HDR Photo Studio	****	*****	***	***	****	*****
Luminance HDR	***	***	***	****	****	**
Photomatix Pro	*****	****	*****	****	*****	*****
Photoshop CS5 HDR Pro	*****	*****	****	****	*****	*****
Picturenaut	****	****	***	**	***	****
HDR Efex Pro	****	****	***	****	****	****
HDR Expose	****	***	**	***	***	***
Oleono HDREngine	*****	****	***	*****	****	*****
SNS-HDR	***	***	***	***	*****	**
Fhotoroom HDR	***	*****	**	***	****	*****

The tests and the analysis of the results showed that the chosen approach allows physicians to make the following kinds of image quality improvements:

- Control of the global characteristics of the final LDR-image as a dynamic range, global contrast, global sharpness, and global brightness.
- Control of the level of details for the LDR-image - for medical images the size of the structures that will be visible in the final image is of gtear importance.
- Local contrast of small structures and details – in medical imaging, it is important to increase the image readability.
- Adjustment of the simulated light conditions - in digital X-ray machines the final image has illumination that simulates the standard for background lighting on X-ray plates. This type of control in our method provides the same opportunity because the medical information perception in the X-rays is reduced without the background lighting.
- Halo-effect adjustment – this effect occurs around concentrated structures with low permeability when strong lighting is applied (the central part becomes darker, and the periphery becomes lighter). Very strong halo-effect may cause medical artifacts, but full suppression can lead to unrealistic contrast between structures.
- Extreme areas managing - these are image areas that should not be involved in determining the mapping between an HDR-image and an LDR-image (these areas aren't parts of the human body). This affects the quality of the contrast adjustment between organs and tissues with different radiographic density.
- Noise adjustment –all X-ray images have low-frequency noise, because the sensor elements are not fully homogeneous. Since this noise is part of the feel of background lighting, the complete removal of its alleged influence does not always give reliable results (the low-pass filter removes not only the results of the noise, but also the effects of different radiographic density of human organs and tissues). At the same time the low-pass filter creates a sense of non-uniformity in the backlight, which further reduces the readability of the image.

Fig. 13 shows the changes in the standard image when some discussed techniques are used. According to medical consultants results obtained by using the Oleono HDREngine and the Photomatix Pro have the highest medical reliability.

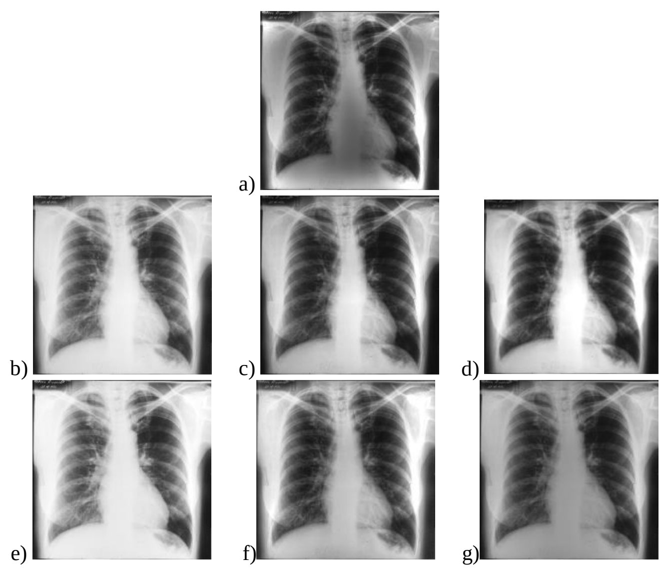


Figure 13. Changes in image, resulting from the use of stage 3techniques: a)image without corrections; b) the detail-level contrast reduction; c) the detail-level contrast increase; d) extermely high global smoothing; e)level of detail adjustment; f) global brightness manipulation; g) extreme areas manipulation.

VII. METHOD IMPLEMENTATION AND ANALYSIS OF RESULTS

Using this method for enhancing X-ray quality gives a significant change even in the exposure values [0 EV, -0.5 EV, +0.5 EV] but in general this is not the best combination of values. The correct implementation of the proposed method and the maximum use of its capabilities needs answers to three main questions:

- How many LDR-images with different exposure values should be used?
- Symmetrical or asymmetrical exposure values should be used?
- What should be the maximum and the minimum value of the simulated exposure, so that there is a significant image quality improvement, but without medical artifacts?

The experiments and the analysis of chest X-rays have led to the following conclusions:

- The difference in the final image when 3 or 5 LDR-images have been used is irrelevant for the diagnosis of lung diseases, since the changes are primarily in the field of the trachea and the heart, ie the space outside the lung (Fig. 14).
- The use of asimetric exposure values for LDR-images with positive and negative exposure most often does not lead to noticeably better results in the area of the lung (Fig. 15).

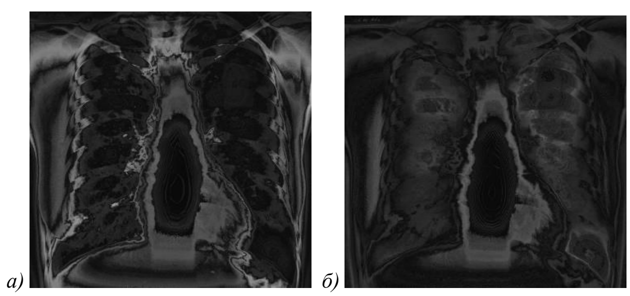


Figure 14. Comparing the results using 3 or 5 LDR-images - images show the subtraction of both final images and the image histograms are stretched 10 times to show the differences: a) the different exposures are simulated by gamma-correction; b) the different exposures are simulated by brightness and contrast adjustment.

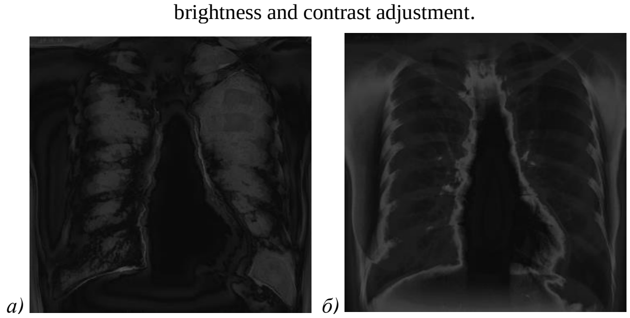


Figure 15. Comparing the results using asymmetrical exposure values - the image shows the difference of the final images of two exposure values sets (the image histograms are stretched 10 times to be able to show the differences): a) the differens between [0 EV,-1.5 EV,+1.5 EV] and [0 EV, -1.5 EV,+1 EV]; b) the differens between [0 EV,-1.5 EV,+1.5 EV] and [0 EV,-1 EV,+1.5 EV].

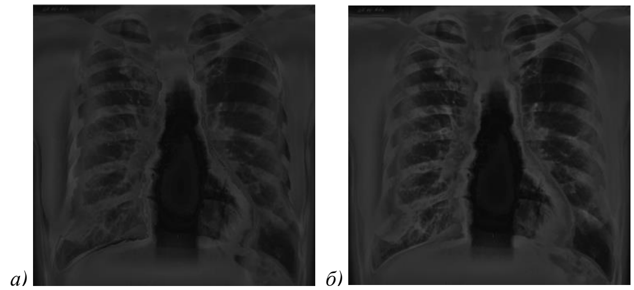


Figure 16. Comparing the results using asymmetrical exposure values - the image shows the difference of final images of two exposure values sets (the image histograms are stretched 10 times to be able to show the differences): a) the differens between [0 EV,-1.5 EV,+1.5 EV] and [0 EV, -2 EV,+2 EV]; b) the differens between [0 EV,-1.5 EV,+1.5 EV] and [0 EV, 2.5 EV,+2.5 EV].

- The use of exposure values distant more than 1.5 EV from the base image doesn't lead to noticeable contrast change, but increases the probability of medical artifacts occurrence (Fig. 16).

So, as a standard set of exposure values, we used [0 EV, -1.5 EV, +1.5 EV]. This set of exposure values can be used in most cases, but for some specific purposes there are other parameters:

- In case of overexposed images, the best results are achieved with a set of 5 images with exposure values [0 EV, -1.5 EV, +1.5 EV, -2 EV, +2 EV].
- In case of underexposed images, the best results are obtained when using the set of exposure values [0 EV, -1 EV, - 2.5 EV].
- In case of X-rays of bones, good results are obtained with asymmetric values for the minimum and maximum exposure – for example [0 EV, -2 EV, +1 EV]. This set increases details in lighter areas (like bone structures).
- In case of lung or soft tissues X-rays, good results are obtained with opposite asymmetric values for the minimum and maximum exposure – for example [0 EV, -1 EV, +2 EV]. This set increases details in darker areas.
- In case of an image with a small dynamic range, a set of 5 images has to be used. This increases the details for all structures with different radiographic densities.

Another major advantage of the proposed method is the ability to manage the transformation from a HDR-image to the final LDR-image. This allows an optimal image quality to be obtained without the occurrence of medical artefacts.

The comparison of the results of the proposed method with other techniques showed that this method can help to obtain a major improvement in quality without the occurrence of medical artefacts. Especially important is the opportunity to use the same characteristics in all cases and always to get good quality – for example 5 images with exposure values [0 EV, -1.5 EV, +1.5 EV, -2 EV, +2 EV].

A few examples of the method implementation and comparison with Laplacian pyramids filter and CLAHE are shown in Fig. 17.

VIII. Conclusion

Image quality enhancement is very important because it increases readability and understandability of the analysed images, their details and structure. When the exploited for image generation model is known this increases possibilities to correct the image without generation of medical artefacts. The presented method for pseudo HRD enhancements of medical images enables increasing quality of understanding and information gathering.

The next steps of this research are oriented to X-ray images of other body parts like bones, abdominal cavity, and other soft tissues as well analyses of images from CT and other medical image sources.

Acknowledgment

This work is funded partially by the Bulgarian NSF under DO02/113 and DRNF02/3 projects.

References

[1] V. Gueorguiev, I .E. Ivanov and D. Georgieva, “Medical Images Enhancement with Pseudo-HDR Method,” eTELEMED 2014 - The Sixth International Conference on eHealth, Telemedicine, and Social Medicine, Barcelona, Spain, ISBN: 978-1-61208-327-8, pp. 134-139, 2014

[2] K. Arulmozhi, S.A. Perumal, K. Kannan, and S. Bharathi, “Contrast Improvement of Radiographic Images in Spatial Domain by Edge Preserving Filters,” IJCSNS International Journal of Computer Science and Network Security, Vol.10 No.2, pp. 233-240, February 2010.

[3] E. Reinhard, W. Heidrich, P. Debevec, and S. Pattanaik, “High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting,” 2nd Edition, Morgan Kaufmann, ISBN 978-0123749147, 2010.

[4] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, “Advanced High Dynamic Range Imaging: Theory and Practice,” A K Peters/CRC Press, ISBN 978-1568817194, 2011.

[5] D. Nightingale, “Practical HDR: A complete guide to creating High Dynamic Range images,” 2nd edition, ISBN 978-0240821221, 2011.

[6] R. Maini and H. Aggarwal, “A Comprehensive Review of Image Enhancement Techniques,” Journal of Computing, Vol. 2, issue 3, pp.8-13, March 2010.

[7] A. M. Trifas, Medical Image Enhancement, PhD Thesis, Louisiana State University and Agricultural and Mechanical College, Dec. 2005.

[8] Y. Abu Sadah, N. Al-Najdawi, and S. Tedmori, “Exploiting Hybrid Methods for Enhancing Digital X-Ray Images,” The International Arab Journal of Information Technology, Vol. 10, No. 1,pp.28-32, January 2013.

[9] S.S. Bedi and R. Khandelwal, “Various Image Enhancement Techniques - A Critical Review,” International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 3, pp.1605-1609, March 2013

[10] C. Clarke and A. Dux, “Chest X-rays for Medical Students,” John Wiley & Sons, ISBN: 0470656190, 2011.

[11] E. E. Coche et al. (eds.), “Comparative interpretation of CT and standard radiography of the chest,” Medical Radiology, Springer-Verlag Berlin Heidelberg, 2011

[12] E. J. Hall and A. J. Giaccia, “Radiobiology for the Radiologist,” 7th edition, Lippincott Williams & Wilkins, ISBN-10: 1608311937, 2011.

[13] E. Svensson, J. E. Starmark, S. Ekholm, C. von Essen, and A. Johansson, “Analysis of interobserver disagreement in the assessment of subarachnoid blood and acute hydrocephalus on CT scans,” Neurol Res., vol. 18, pp. 487–494, 1996.

[14] E. Trägårdh, P. Höglund, M. Ohlsson3, M. Wieloch, and L. Edenbrandt, “Referring physicians underestimate the extent of abnormalities in final reports from myocardial perfusion imaging,” EJNMMI Research, vol. 2, no. 27, 2012

[15] F. Chen, J. Pan, and Y. Han, “An Effective Image Quality Evaluation Method of X-Ray Imaging System,” Journal of Computational Information Systems, vol. 7, no. 4, pp. 1278-1285, 2011

[16] G. De Lacey and S. Morley, “The Chest X-Ray: A Survival Guide,” Elsevier Health Sciences, ISBN: 070203046, 2008

[17] G. Dougherty, “Digital image processing for medical applications,” Cambridge University Press, ISBN-13 978-0-511-53343-3, 2009

[18] H. Aichinger, J. Dierker, S. Joite-Barfuß, and M. Säbel, “Radiation Exposure and Image Quality in X-Ray Diagnostic Radiology,” 2nd ed., Springer, ISBN 978-3-642-11241-6, 2012

[19] H. McAdams, E. Samei, J. Dobbins III, G. Tourassi, and C. Ravin, “Recent Advances in Chest Radiography,” Journal of Radiology, vol. 241, no. 3, Dec. 2006

[20] International Commission on Radiation Units and Measurements, “Modulation transfer function of screen-film systems,” ICRU Report 41, Bethesda, ICRU, 1986

[21] International Commission on Radiation Units and Measurements, “Medical imaging - The assessment of image quality,” ICRU Report 54, Bethesda, ICRU, 1996

[22] J. Bushberg, J. Seibert, E. Leidholdt Jr., and J. Boone, “The Essential Physics of Medical Imaging,” 3rd ed., Wolters Kluwer, ISBN-10: 0781780578, 2011

[23] S. Mann and R. Picard, “On Being ‘Undigital’ With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures,” In proceeding of: IS&T's 48th Annual Conference, Cambridge, Massachusetts, 1995

[24] “Radiation-Emitting Products: Medical X-ray Imaging,” FDA U.S. food and Drug Administration, last update: 07.03.2013 <http://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandprocedures/medicalimaging/medicalx-rays/default.htm> [last accessed: 10.11.2014].

[25] I.E. Ivanov, V. Gueorguiev, and D. Georgieva, “New Approach to Information Extraction from X-ray Plates,” The Third International Conference on Global Health Challenges (GLOBAL HEALTH 2014), Rome, Italy, August 2014.

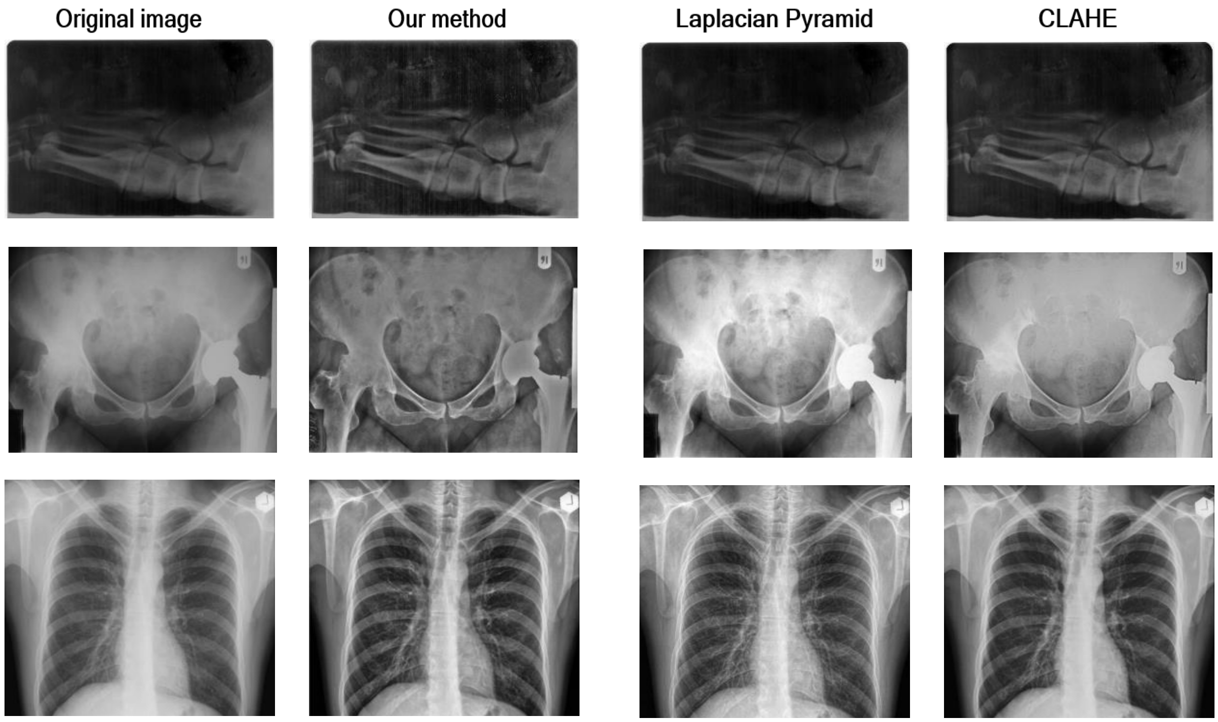


Figure 17. Comparison between our method, Laplacian pyramids and CLAHE: the proposed method improves contrast and details.

International Journal on Advances in Life Sciences, vol 6 no 3 & 4, year 2014, http://www.iariajournals.org/life_sciences/

Understanding eHealth use from a Persuasive System Design Perspective

An Antibiotic Information Application for Nurses

Jobke Wentzel, Nienke de Jong, Julia E.W.C. van Gemert-Pijnen
Dept. of PHT, University of Twente
Enschede, The Netherlands
m.j.wentzel@utwente.nl, n.dejong@utwente.nl, j.e.w.c.vangemert-pijnen@utwente.nl

Lars Nijdam, Regine van Drie-Pierik
Hospital Pharmacy, Medisch Spectrum Twente
Enschede, The Netherlands
l.nijdam@MST.nl, r.vandrie-pierik@mst.nl

***Abstract* - To ensure optimal treatment and patient safety, nurses need easily accessible information. An application was developed to support nurses in their antibiotic-related tasks. With log data analysis and user interviews we evaluated the application. We aimed to interpret log data by applying the Persuasive Systems Design model. The results show steady, continuing use that corresponds to nurse tasks regarding timing and content of use. Task support seems a top relevant motivation for using the system, given positive user comments and most popular content of instruction pages. The results show that a combination of log data and user interviews help to understand use and uptake of an application in practice.**

Keywords-eHealth; antibiotic stewardship; nurse; task-support; log data; persuasive technology

I. INTRODUCTION

In this paper, we provide a more in-depth analysis of actual use (log data), elaborating on the effects of an antibiotic information application for nurses that was presented at eTelemed 2014, the sixth International Conference on eHealth, Telemedicine and Social Medicine [1]. Managing constant information needs is among one of the many tasks of Health Care Workers (HCWs). Increasingly, integration of information and support via online or mobile applications can facilitate HCWs. This is especially the case when working in complex medical settings, or when dealing with situations that require good interpretation of information from multiple sources [2]. For good support, applications should be integrated with daily work practice. Therefore, besides studies on behavioral or clinical outcomes, understanding of application use and use context is pivotal to the development and improvement of eHealth applications for clinical settings. This paper focuses on understanding of use data as a prerequisite for interpretation of effectiveness.

A. Ehealth to support tasks

Information and communication technology holds the promise of facilitating information transfer and offering support in a variety of health care settings [3][4]. Applications aimed to support health care workers (HCWs) are abundant, both in mobile applications (for smartphone or tablet), in websites, and web apps [2]. However, sometimes applications and websites that are supposed to facilitate and support HCW's jobs are perceived as distracting, user-unfriendly and ill-fitted to work practice, thereby possibly compromising patient safety [5][6][7]. Taking a more user- or practice- based focus throughout the design phases may help to overcome these difficulties [6][8]. By taking this focus, user needs and implications of use context are balanced with stakeholder values and needs, to reach an optimal fit between user, organization and technology [8]. However, in professional clinical settings, conducting such user-driven formative studies can be a challenge, due to practical (time, access) and ethical considerations. Therefore, requirement elicitation strategies that take into account the specific boundaries of medical settings have been proposed [9].

B. Persuasive system Design

Another development approach that may facilitate implementation, uptake, and effectiveness of eHealth applications includes incorporating Persuasive Systems Design (PSD) into technology. In this model, the targeted behavior is reinforced, shaped or changed by the technology, thereby 'persuading' its users [10]. This persuasion can occur via different strategies, which depend on an analysis of the context. The system qualities should be adapted to this context, using one or more strategies: primary task support, dialogue support (facilitating interaction with the system), credibility support (credibility of the system and its content), and social support (strategies that include others, integrate the system in interpersonal interaction or comparison). The strategies are supported by design principles that exemplify how to operationalize a certain persuasive strategy. For example, including tokens of expertise (regarding content) in the system contributes to credibility support [10].

Designing technology in such way that it is unobtrusive, useful, and easy to use helps to blend the technology seamless into users' work (or life) and to quickly reach their goals. The model is applied to the design and evaluation of health interventions aimed at (changing health behavior of) patients or non-professional [11], but the principles apply to professional-aimed eHealth tools as well [12].

C. Information support for antibiotic stewardship

Information support can be especially useful in areas where complex medical situations arise [13]. Clinical antibiotic use is such a situation. Mis- or over-use of antibiotics contributes to the problem of antibiotic resistance. Due to their resistance to antibiotics, infections caused by resistant pathogens are difficult to treat. Therefore, prudent use of antibiotics is needed to stop resistance forming and preserve the effectiveness of these drugs [14]. Antibiotic Stewardship Programs (ASPs) aim to optimize antibiotic use in clinical settings. Formulary restrictions, antibiotic cycling, multidisciplinary bed-side consultation and improved diagnostics are some strategies that are used to this end. Infectious disease specialists, clinical microbiologists, pharmacists and physicians work together in ASP to improve antibiotic use. Optimal antibiotic therapy relies on timely adjustment of therapy, based on lab results (diagnostics) as well as careful monitoring of patient progress and vital signs. Correct execution of antibiotic therapy (including preparation and administration of the medications) as prescribed by the physician stands at the base of successful ASPs. In addition to expert input in ASPs, nurses contribute a great deal [15]. Nurses spend a lot of time caring for and observing patients; they are the 'eyes and ears' of the physician and notice changes in patient status that call for action (e.g., antibiotic therapy adjustment).

In earlier research, we identified nurses as an important stakeholder for ASPs [16]. However, in literature and practice, physicians and clinical experts are regarded as the main actors in ASP, and nurses' roles are not made explicit [15][17]. As earlier research indicated that nurses have high information needs regarding the antibiotic care process, we aimed to optimally support nurses in their antibiotic-related tasks. We applied human centered design to develop an information application, that takes user needs into account and provides bed-side task support. With regard to nurse support in antibiotic stewardship, no dedicated applications were identified in literature, as nurses are often overlooked as stakeholder in ASPs. Nurse tasks in ASPs are demanding and ask for good information integration and decision making: nurses gather information from different information systems, integrate it, and decide whether further action is required. Previous research indicates that especially task support and system credibility strategies from the Persuasive Systems Design model are influential; formative evaluations revealed a lack of task support and easy accessible information [16], which can be a barrier for nurse empowerment in ASP. Based on this understanding of the work context, we developed an application, the Antibiotic Information Application (AI Application).

D. Study aim

In this study, we aim to assess if and how the AI Application fits into nurses' workflow and nurses are motivated to use it. To this end, we analyze actual use patterns (time of use and content viewed) and combine this log data with data on user satisfaction as resulted from our qualitative evaluations.

II. THE ANTIBIOTIC INFORMATION APPLICATION

Based on formative evaluations, the Antibiotic Information Application (AI Application) was developed [17]. In focus groups, interviews, scenario tests and usability tests nurses in our pilot research expressed a need for an easily accessible, centralized information application, where they can find all information on antibiotics they need during their work. This includes instructions on preparation and administration, as well as background information and information that is needed on specific occasions only (e.g., when side effects occur). On the basis of the previous results, a prototype was created in WordPress [18], with some alterations to fit our specific content management demands (e.g., flexibility of button-order and features for import and export of content). Based on its evaluations (user tests, user interviews), a final release was launched. This application is web-based, and can be run from the hospital's own server. However, for the pilot study the system was made available via the internet, and was taken up in the nurses' personal hospital start page and medication administration system so that it can be accessed easily. Besides for use on a PC, the AI Application is optimized for mobile (tablet, smartphone) use. For the application's content, the information sources that were already used in this hospital served as input. This included general reference works and national guidelines as well as the hospital's own protocols as created by the hospital pharmacy. All sources were 'chopped up', where needed, selecting only the content that is necessary for nurses to execute their tasks (the original sources often contain much information aimed at physicians or medical experts, which can be irrelevant and confusing to nurses). This process was done based on formative research outcomes regarding information needs, and expert views (clinical microbiologist, pharmacists) on what types of information should be available to nurses to be able to contribute to ASPs. The different types of content are ordered according to the mental models of the nurses as they resulted from the card-sort study that was conducted during formative research

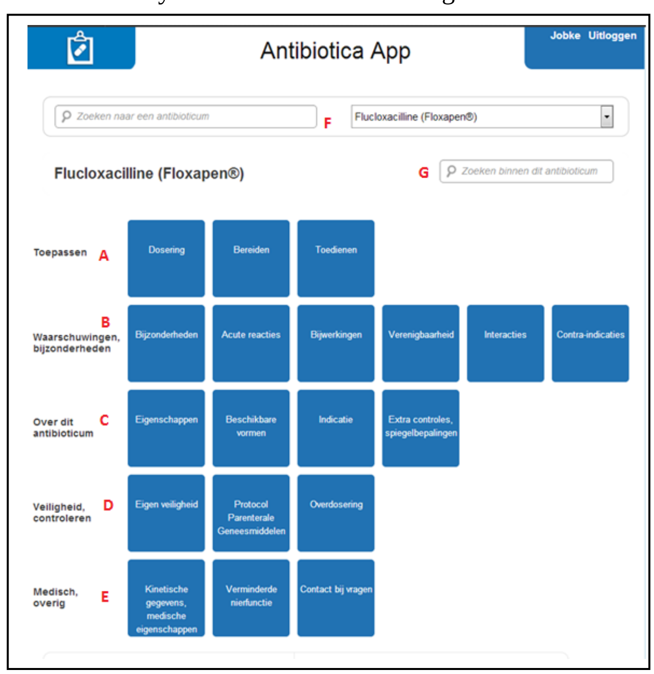


Figure 1. Screenshot of antibiotic overview page. A: task instructions, B: important information and warnings, C: background information, D: Safety checks, E: medical background, F: search field for antibiotic or select from list, G: search all pages (per antibiotic).

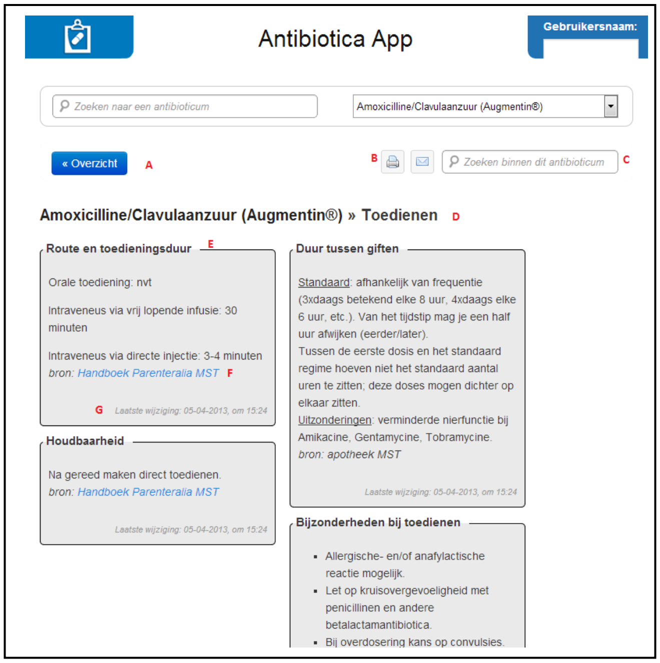


Figure 2. Screenshot of antibiotic overview page. A:back-button B: print and email buttons, C: search field, D: breadcrumb trail, E: instructions for administration, F: information source with link, G: information on updates

phases [17]. Users select an antibiotic from a drop-down menu or type the name in a search field to access an overview page (see Fig. 1). On these overview pages, information is grouped by the following categories: information needed to perform the primary tasks, warnings, general or background information, extra checks and safety information, and information for specialists and medical background (see Fig. 1). Fig. 2 shows an instruction page of the AI Application, showing information on drug administration, including what to pay extra attention to. A demo version (in Dutch) of the AI Application can be accessed online via [19].

By providing easily accessible information, nurses may be better equipped to perform antibiotic related tasks. For example, it helps them to recognize and address instances of suboptimal antibiotic use. This supports improved knowledge and recognition of instances to optimize antibiotic use and nurse empowerment to discuss patient therapy and alert physicians in case of suboptimal antibiotic use.

As nurses are generally aware the importance of looking up information, it is not a lack of awareness that hinders information use. Rather, inaccessibility and incomprehensibility of information is the problem: nurses lack time an knowledge to localize and interpret the various information sources to be able to apply them quickly during their work. Therefore, in this case, persuasive strategies that focus on creating unobtrusive, easy to use information, that supports tasks are most likely to contribute to a positive user experience and effectiveness. In the AI Application, reduction and tunneling (among others) realize these strategies: the amount of actions to reach information are decreased by integrating all information in one application (reduction) and users are guided directly to the desired information via the overview grid (tunneling), as can be seen in Fig. 1.

III. Methods

A. Participants

Two lung wards of a local 1000-bed teaching hospital participated in this research. The wards have a total of 57 beds. During the pilot phase, 62 nurses (45 FTE) worked at the two wards. Approximately 15 of them were informed about the AI Application and the importance of antibiotic stewardship in presentations and were encouraged to share this information with their colleagues. In addition, all nurses received an email with instructions and fact-sheets on antibiotic stewardship and the AI Application were distributed repeatedly throughout the wards.

A convenience sample (based on availability during the weeks of interviewing) of thirty-four nurses participated in the AI Application evaluation interviews. These nurses have an average age of 32 years, and 29 of them are female, 5 are male. On average, they have 9 years work experience as a nurse, and 6 years work experience on their current ward.

B. Log file analysis

To get more insight into the working mechanisms of the technology that is implemented, actual use of the technology can be studied. Log data analysis provide a means to study technology use in an ecological setting. The AI Application was introduced at the two pilot wards where users could access it directly on a pc. Login was performed based on IP recognition. To ensure whether the AI Application fulfills information needs, at what moment these needs arise and what information is viewed, AI Application use is logged using Google Analytics [20], as well as a Wordpress plug-in because of additional information the two systems provide; e.g., the plug in identified users (including admin) whereas Google was able to provide overviews of most viewed content. Log information of interest includes number of visits per day, time of day with most visits, most frequently viewed content and visit duration. Also, preferred search methods, entry and exit pages are assessed. Only log data from the hospital's IP address is considered, to filter out application use by third parties (e.g., maintenance) not included in the pilot.

C. Intended use

To make sense of the log data for answering the research questions, the intended use of the AI Application must be defined. With regard to intended use, some remarks from our earlier research can be made. Antibiotics are administered several times a day in the wards, but nurses indicated during development phases that they especially look for instructions when dealing with antibiotics that are unfamiliar to them or when a patient reacts to the medication in an unexpected way (side effect or allergy). Also, inexperienced nurses may need to look up information more often than experienced nurses [17]. The frequency of critical moments, when an unknown antibiotic is prescribed or when an unexpected event related to an antibiotic occurs, is unknown. Because nurses share information, not every individual nurse may need to consult the AI Application every time he/she cares for a patient who receives antibiotics. Therefore, a precise estimation of intended use is hard to give but multiple visits per day (overall use) are expected.

D. User interviews

Semi-structured interviews with nurses (n=34) were held to assess their satisfaction with the AI Application, eight months after implementation. First, participants were asked to perform a few information seeking tasks. Half of them was asked to use the AI Application (n=17) for this, while the other half of the participants (n=17) was asked to use the regular information sources that were already available and used prior to this pilot study (pre-implementation information sources), such as traditional guidelines. The participants were asked to 'think aloud' [21] to assess overall positive or negative user experience of the AI Application as compared to the pre-implementation information sources. In addition, during these sessions short interviews were held to ask participants about their opinion of the AI Application, how satisfied they are with it, and if they would like to change anything about the application. Audio recordings were made during these individual sessions. The transcripts of the 'think aloud' data were transcribed and categorized into positive, neutral, or negative remarks. The transcripts of the interview part of the sessions were coded using the PSD model, as previous work on a sub-sample of our dataset showed that this model [22] helps to clarify the cause of (dis)satisfaction with the application. The data was analyzed using the constructs of the Perceived Persuasiveness Questionnaire [23] as a coding scheme. This questionnaire measures perceived persuasiveness; the extent to which persuasive strategies are recognized (present) in the system by measuring to what extent users experience primary task support, dialogue support, perceived credibility, social support, unobtrusiveness, perceived persuasiveness, perceived effort, perceived effectiveness and use continuance. As not all persuasive strategies are present in the AI Application, because of the strong focus on primary task support, some constructs were adapted in our coding scheme: perceived effectiveness and primary task support were merged. Previous research showed that these constructs overlap in measuring the AI Application's perceived persuasiveness [22]. The data were analyzed by two independent coders [JW, NdJ], first to identify relevant transcript excerpts, and second to apply a certain code.

The transcript excerpts coded according to the PSD-based coding scheme are not the same excerpts that were used to establish satisfaction during use; the latter comments are the verbalizations of thoughts made while executing tasks with or without the application, while the PSD-coded comments constitute interviews transcripts.

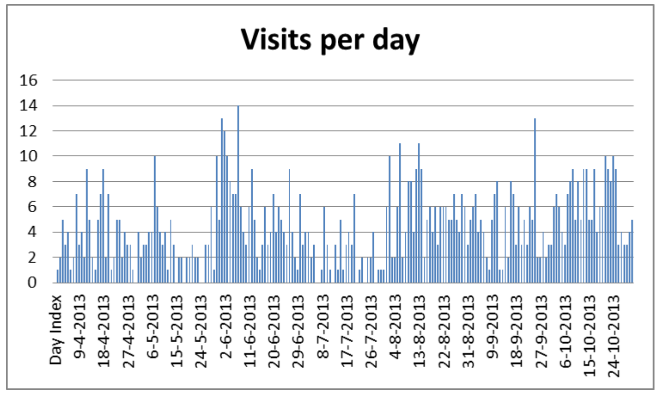


Figure 3. Daily AI Application visits throughout the pilot period.

IV. RESULTS

The eight month pilot period provides prolonged insight into use and user satisfaction with the AI Application.

A. Log file analysis

The AI Application was launched at the end of March 2013, and has been used 5.21 (SD 3.09) times per day on average during the pilot period (April 2013 – Nov 2013). The bounce percentage is 11%; in these cases users left directly after entering. On average, visits lasted 2.5 minutes. The log files show a variable, but steady use of the AI Application over time (see Fig. 3). This use pattern continues after the pilot phase (as the application continues to be available). During week days (Monday to Friday) the AI Application is used more often (14-16% of visits per week day are on one of these days) than during weekends (11% of all visits occur on a Sunday, 13% on a Saturday), as can be seen in Fig. 4. When data are arranged by visits per hour of the day, peaks around 08:00h, 12:00h, 17:00h and 21:00h are visible (see Fig. 5).

To access the antibiotic pages, users prefer the search field over the drop-down menu; 74% vs. 26% out of 3,060 search instances. Pages that are viewed most often, aside from the welcome page and antibiotic overview pages (see Fig. 1 for an example), are pages of the category task instructions (Fig. 1, section A): information on dosage, 189 page views (8.5%), instructions on preparation, 1,025 page

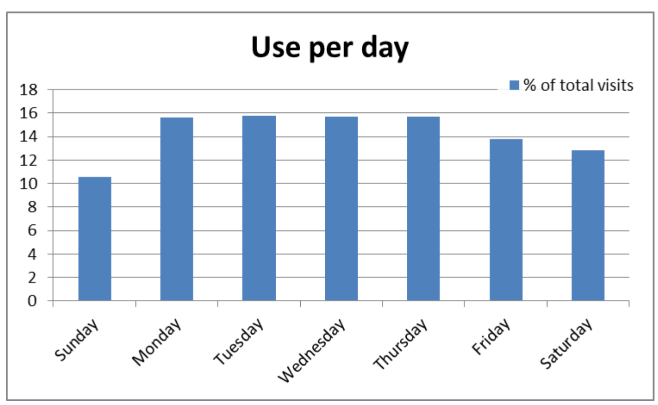


Figure 4. AI Application use per day of the week.

Figure 5. Hourly AI Application visits throughout the pilot period

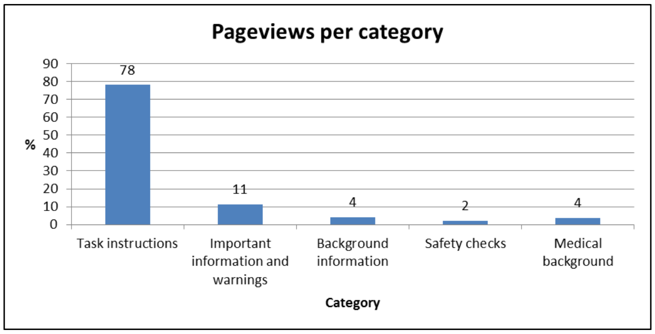


Figure 6. Page views in percentage per section.

views (46%), and instructions on administration, 516 page views (23%). The other categories (see Fig. 1, section B-E) generate less page views, ranging from 2-84 (0-4%) for individual pages and 49-247 (2-11%) per category (see Fig. 6). Information on the antibiotics Amoxicillin-Clavulanicacid, Ceftriaxon, and Ceftazidim was viewed most often with page views (including overview pages) of respectively 717 (16%), 609 (13%), and 537 (12%).

B. *User interviews*

While using the AI Application during the interview, users commented on the ease and support they experienced when working with it. Table I show the types of remarks participants made while searching for information with the application or without it (pre-implementation sources).

Overall, more negative than positive remarks about the pre-implementation information sources were made (see Table I). The negative comments regarding the pre-implementation information sources (negatively) addressed the ease of finding information and clarity of the sources:

"*I want to remark that I do not think this [table of drug-interactions, ed.] is good, because it is very, very small. And when you zoom in, you cannot see it precisely and that is when you make mistakes".*

TABLE I. USER COMMENTS DURING SEARCH

	Pre-implementation sources		AI Application	
	Comments	%	Comments	%
Positive	18	14	97	56
Neutral	10	8	12	7
Negative	104	79	63	37

However, positive remarks on the pre-implementation sources include that information is nonetheless available and there is always a way to find it:

> *"Well, there is always the drug leaflet, and you can find this type of information in there. And I think it is just as fast to search for it in the leaflet […] Takes precisely as much time [as other search methods, ed.]".*

The AI Application generated more positive (56%) than negative (37%) comments (see Table I). Positive comments included remarks on functionality, looks and usefulness:

> *"You see all these nice blocks"* and

> *"It is instantly clear where you can find the information".*

Negative comments addressed the available content and problems finding information:

> *"That's stupid, I cannot find it right now"* and

> *"Look, it does not show any results. I would want it to do that".*

To be able to deduct why the AI Application generates positive and negative qualifications, interview transcripts were analysed using PSD constructs for coding. Table II depicts the results of this analysis. Overall, for nearly all categories the amount of positive comments outweigh the negative comments, except for unobtrusiveness (11 positive vs. 12 negative) and credibility (equal amount of 4 comments). By far, most remarks concerned primary task support (42,45%), indicating that this is a relevant concept in the AI Application's setup. Participants typically commented on how the information and setup of the AI Application is specifically useful when performing their tasks. This quote illustrates it:

> "[…] *that it clearly shows: dose, preparation and administration. That is what I want to know. That's why I use the AI Application".*

Some negative remarks concerning primary task support were given (see Table II) and address information that lacks or ways in which the AI Application cannot be used while performing tasks when this is wanted:

> "*You miss [information, ed.] in the AI Application sometimes. One time I had something; one gram that should be given per two grams. It doesn't say if the speed of the drip should be adjusted yes or no. That are sort of the puzzles you have to [solve, ed.]".*

Comments concerning perceived persuasiveness were quite prevalent as well (see Table II), mostly indicating that using the AI Application is easy, nice or convenient. This seems a more pleasurable method of searching than former information sources. These quotes underline that:

> "*[…] And it is very convenient that it is so easy to search. That it's much like our good old 'yellow booklet' [paper-based antibiotic information, ed.]" and "It works quite easily, yes".*

However, some users encountered difficulties in using the AI Application that made them less motivated to use it:

> "*That it is difficult to read, so it is less interesting, because you'll soon feel like you don't understand and I would then just leave it to the physician".*

Table II shows that many remarks addressed the concept of perceived effort: how easy it is to use the AI Application (e.g., time, number of actions required). Especially saving nurses the trouble of scanning large texts and extensive searching within the system typify these comments:

> *"Clear. Especially these distinct blocks and preparation, administration and the like. That it is clear where you can find it because especially with all the other sites they give you a long list and you have to search through the entire page".*

However, some effort still has to be put into searching for information, as the negative comments indicate:

> *"Sometimes you have to search a bit further, but often you'll figure it out".*

TABLE II. OVERVIEW OF USER SATISFACTION ANALYSIS

PSD construct	Comment load	Number of remarks (*)	% of total remarks per construct
Primary task support	Positive	63 (27)	42,45
	Negative	27 (15)	
Perceived persuasiveness	Positive	20 (15)	11,79
	Negative	5 (4)	
Perceived effort	Positive	20 (14)	13,21
	Negative	8 (5)	
Use Continuance	Positive	18 (14)	12,26
	Negative	8 (7)	
Unobtrusiveness	Positive	11 (6)	10,85
	Negative	12 (9)	
Credibility	Positive	4 (3)	3,77
	Negative	4 (2)	
Social Support	Positive	7 (7)	3,30
	Negative	0 (0)	
Dialogue support	Positive	5 (4)	2,36
	Negative	0 (0)	
total		212 (34)	100

*number of unique participants making remarks in this category

Use continuance indicated if nurses were inclined to (keep) using the AI Application, which was mostly positive for as far as nurses made a remark on this theme:

> *"I actually always use the AI Application"* and

> *"When I search something about antibiotics or something I always start with the AI Application, [to see, ed.] if it is in there".*

A few users had little experience with using the AI Application, or mentioned they just did not use it as often as they could. This quote illustrates how the AI Application is not used in standard settings (where information still can be of value):

> *"Look, the standard stuff, we give a lot of Augmentin [an antibiotic, ed], you really don't look up in the AI Application anymore".*

Having a system right at hand that is easily accessible while performing tasks was mentioned as a positive aspect. Especially easy access to the AI Application via the medication registration system contributed to this unobtrusiveness:

> *"You want your information instantly. And this is useful for that. It is just nice that I can access it directly from this system [medication registration system, ed.]".*

However, some down-time of the system and the desire for even more and faster ways to access the AI Application accounted for a few negative remarks on unobtrusiveness:

> *"[…] then it would become a bit more striking. […] If you would just see it [the link, ed] on top here, that would be clearer for the people that do not use it very often".*

Credibility is not commented on very frequently (see Table II). Given the AI Application's location and support by the hospital, its credibility is valued positively, but some concerns were expressed with regard to a warning message that appeared when updates were overdue. These quotes illustrate it:

> *"[…] Information that you find on the internet is not specifically written for our hospital. This is".*

Concerning the updates, users mentioned:

> *"So how reliable is the AI Application then? If it says […] that it is not fully up to date?!"*

The AI Application may play a role in social interactions, via social support; it can offer nurses a reference source when discussing or communicating a problem to physicians:

> *"Or just for your own information. […] Because you want to be as well informed as possible when you call the physician".*

The AI Application does not require a great deal of interaction, but some positive comments on the way search queries are handled by the system (a form of dialogue) are given:

> *"As soon as you type in 'am', that* Amoxicillin and Augmentin *are already suggested to you. I personally find that really convenient".*

V. Discussion

A. AI Application use

The log file analysis and user satisfaction results show that the AI Application does support nurses during their work. AI Application use remains steady over time, and the bounce percentage is low (11%), indicating that most users do not leave after entering the home page but actually navigate through the site to content pages. Therefore, we conclude that the uptake is good. As the pre-implementation information sources remained available to the nurses, they had the option of not changing their primary source of information, but our data shows continuing use of the application, indicating a preference among the nurses of the wards to use it over conventional materials, months after its release.

A first possible contributor to the AI Application's uptake, is its successful introduction. To introduce the AI Application, relatively few activities were undertaken; some nurses attended an instructional meeting and fact sheets on antibiotic stewardship and the AI Application were distributed. This minimal introduction may have sufficed because the intended users initially proclaimed they liked the concept and look and feel of the application, which has possibly led to increased willingness to start using it, and spread the word. In addition, the AI Application probably fits in with nurses' work routines as it is available via the system that nurses are already using when dealing with antibiotics, and it is easier to find than the scattered information sources they previously used. A third reason for AI Application uptake may lie in the human centered design process, as involvement in the development can generate user commitment. It can be more difficult to reach this type of involvement on a large scale however.

Log data contributes to eHealth evaluation by giving insight into actual use (and uptake) [24]. However, to make sense of the log data, in terms of how the AI Application is used in daily care practice, having some concept of what intended use (how users should use the technology) should be present. For eHealth interventions in general, this usually relies on the program's setup (e.g., weekly modules require at least on log-in per week) [25]. However, for information applications that meet occasional information needs, it is difficult to determine whether users use it as intended. In the current study, we found that overall, AI Application use peaked on four moments per day; 08.00h, 12.00h, 17.00h, and 21.00h. These hours correspond roughly with medication rounds; typically nursing activities where information on preparation and administration of antibiotics is needed. In addition, in the (late) afternoon and evening there is no support from the hospital pharmacy available, which might motivate nurses to check the application in case of questions. AI Application use was slightly higher during week days than on Saturdays and Sundays. An explanation may be that usually few patients are admitted during weekends, so few new medication regimes are started, which causes a lower information need on these days. These results suggest that the AI Application is integrated in the care process as the most popular moments of use correspond to the timing of care activities involving antibiotics.

The most popular pages of the AI Application are instructions on preparation and administration of antibiotics and the antibiotics that are viewed most often are used frequently on these wards. As nurses are required to follow the instructions for preparation closely, these findings are not surprising: preparation and administration of the commonly used antibiotics are tasks that occur often. Based on the log data, the AI Application seems to primarily provide support via the procedural information [26]: instructions on how to perform certain tasks. In contrast, side effects or unexpected progress or deterioration in patient status occur less frequent, and the information that offers the background knowledge and explanations (e.g., working mechanisms, side effects), is not used as often as the concrete instructions. However, to optimally benefit from nurses' monitoring and alerting roles in ASP, having sufficient (background) knowledge that enables recognition of critical moments in antibiotic care, is needed too. The log data suggest that a more interactive learning method for information that supports this behavior is needed, as background information is used less often. In this sense, primary task support is a useful persuasive strategy to fine-tune information to support workflow. For learning new behaviors or creating awareness, it needs to be complemented with interactive learning strategies, as are applied in other persuasive behavior change systems, for example [11].

B. User satisfaction

The user interviews indicated that overall, nurses are satisfied with the AI Application. In contrast, the pre-implementation sources generated more negative comments. One of the possible explanations for this finding lies in the fact that the AI Application indeed greatly facilitates information-finding as compared to the sources that were available pre-implementation. This is underpinned by other preliminary research findings of our study that are currently analyzed more in-depth: a pre-post implementation questionnaire reveals that nurses judge the pre-implementation information situation worse than post-implementation situation [1].

The AI Application's clear setup that enables nurses to find information fast contributes to the positive evaluations. As was to be expected, especially primary task support as a persuasive strategy seemed relevant. For nurses, using an information system is not so much a matter of wanting to change some type of behavior, or reaching a long term goal (e.g., lose weight, increase physical activity, become less depressed), but concerns solving an instant information need.

In this context, task support, for example via tunneling, contributes most. This finding is supported by the log data, that show that the AI Application is used most often when concrete tasks of preparing and administrating antibiotics are performed. However, good task support possibly does not explain application use and positive appreciation entirely: we found that perceived persuasiveness, operationalized as pleasantness of using the application, was also present in users' comments on why they use the AI Application. Thus, an application can be efficient and clear, but positive user experience contributes to use and implementation as well. In addition, possibly related to this finding, unobtrusiveness and effort also contributed to, and may overlap with task support. Surely, if it costs little effort, and is not very obtrusive to use a system, users may use it more frequently thereby feeling more supported. The interrelatedness of these Perceived Persuasiveness constructs seems logical and is being studied [11][12]. The negative remarks that were made stipulate the importance of managing information applications properly: credibility is hampered when information is not up to date and when the application cannot be accessed easily, as was the case temporarily during the pilot, users may use it less often. Users also commented on information that they wished to access via the AI Application, but was not present. This included certain types of instructions that would make their tasks easier. Also, requests for an expansion of the scope of the application to all types of medications instead of antibiotics only. This also asks for application management: either comply with the users' wishes and provide the extra information and broaden the scope, or manage users' expectations. However, besides these requests for improvement, users seem to experience more advantages than disadvantages in using the AI Application, given the steady use and greater amount of positive than negative remarks.

Combining the quantitative log data analysis with qualitative user satisfaction interviews strengthens our understanding of the application's use: task supporting content is used most often, and is also most prevalent in our user comments. Also on other levels, log data support user comments and vice versa, thereby strengthening our view on application use and appreciation. Combining log data analysis with other methods is therefore recommendable to gain insight into why an (online) intervention is used (or not), as stated by others [24].

C. Considerations for implementation

Nurses spend much time and effort gathering information, so surely an application that centralizes information has much to offer. However, this implies that somehow the information from different sources needs to be centralized and the ability to do this automatically depends heavily on the type of information sources that need to be centralized. In our case, 'filling' the AI Application and managing its content was done manually because automatizing this process would be too complex and costly in this pilot phase. Individual hospitals and sometimes even individual wards use different information systems (thus, many different types of information sources to centralize), so a one fits all solution may be difficult to achieve. For long term implementation and sustainment this is an issue that must be resolved, for example by assigning quality staff or specialized nurses to keep information up to date. Expanding the application's scope to all types of medication should be taken into account too, as users explicitly requested this and problems with antibiotic information finding during formative research applied to these other types of medications as well.

With regard to implementation throughout the hospital, nurses outside our pilot ward found the AI Application on the intranet of the hospital and wanted to use it. In this sense, the application implemented itself just by being available and easily accessible. In addition, physicians to whom the AI Application was demonstrated expressed interest in a physician-aimed version. These findings show that implementation success depends on the operationalization of early design phases when user needs and context are studied, and meeting these needs within the boundaries of clinical settings. In this case, the mismatch of expert-driven content and user needs that was resolved with the application facilitated initial implementation. We tailored the various expert-based sources to fit clinical practice, something that is not done often because it costs time, effort, and multidisciplinary cooperation and understanding. In this sense, human centered design can help to meet end-user needs [27]. However, when these needs require highly tailored information via applications with dynamic content that requires frequent quality checks and updates, this poses some challenges. Design teams should then find a balance between available sources to manage the information, and generalizability and up-scaling possibilities.

D. Antibiotic Stewardship

The motive for creating the AI Application was to empower nurses so that they can make a contribution to antibiotic stewardship programs (aimed at more prudent use of antibiotics). Even though this study's results suggest the AI Application supports nurses' antibiotic related tasks, the extent to which it contributes to antibiotic stewardship remains unclear. To detect the effects of the AI Application on work processes regarding antibiotic stewardship is difficult. As nurses are not the actual prescribers of antibiotics, the effect of more knowledge and empowerment can indirectly influence actual antibiotic use. Surely, executing preparation and administrating tasks correctly contributes to prudent antibiotic use. This study shows that task support is a relevant persuasive strategy to reach this goal. Nonetheless, the log data indicate that information on side effects, acute responses, interactions and appropriate dosage is not used often. To benefit from the monitoring, recognizing and alerting role nurses can fulfill as 'eyes and ears' of the physician, we expected these types of information to be important for nurses. However, this type of information is not accessed often, possibly because critical moments occur scarcely. Another explanation can be that nurses are unaware that they can play an important role in recognizing the critical moments and/or lack time or motivation to pay attention to this. We found some evidence for this hypothesis in our formative study and preliminary analyses of app efficiency [1][17]. Thus, it appears the AI Application offers more information that nurses actually wish to utilize during their work, but there are no indications that this overload hampers usability and applicability. In addition, instead of information support, awareness and (e)learning may be more a more relevant approach to activate knowledge in the currently less popular information categories of the application. This view is supported by a study that showed that more active learning methods did result in increased awareness and knowledge of nurses regarding ASP [28]. In addition, the ASP-related nurse tasks possibly ask for a change in safety, cooperation and communication culture, which likely is influenced by communication tools or socially-aimed strategies [29]. Possibly, persuasive strategies such as social support might be more appropriate in these situations [10].

E. Limitations

The outcomes of this research must be interpreted with appropriate caution because of several possible limitations. First of all, the AI Application was developed and tested with the help of nurses of two wards in one teaching hospital. Generalizing the results greatly depends on the specific information sources in place and the AI Application's effectiveness may differ in other settings. This is to some extent a consequence of the design approach of zooming in in local needs and local contexts. Agile methods can be used for re-design in other wards or institutes. In this case, identification of local information sources and integrating them in the AI Application are among such up-scaling activities.

With regard to the log data it must be remarked that in the current use and log setup, no individual user can be identified. Therefore, some ambiguity remains with regard to the actual amount of users that use the AI Application. Possibly, a few nurses of the pilot wards use the AI Application often, or, a lot of nurses use it occasionally. Our interview data tells us that in our sample, the majority of respondents claim to use the AI Application whenever they need to look up information on antibiotics. How often this occurred, was very hard to assess by asking respondents. Using a personal login helps to overcome this user identity issue. However, we applied an IP-based login because this does enable easy (unobtrusive) access, which is of paramount importance for information accessibility. Furthermore, sharing logins can pose validity threats regarding correct identification of user identity, especially when multiple users use the same computer. The consequence of our approach is that outcomes can be interpreted on a population (ward) base, but not on an individual level. Still, the data are meaningful, because of the communicative nature of nursing: asking a colleague who is standing next to you is often the fastest way of resolving an information need, as other studies also demonstrate [30]. Therefore, information availability improves the information situation of an entire ward, even if not all individual nurses access the AI Application.

Another important consideration when interpreting the results lies in the pilot-ward setup of this research. Even though not all nurses were fully aware of ongoing evaluation of AI Application use during the pilot period, satisfaction measurements via user comments could have been slightly more positive because of nurses' positive attitude towards (participating in) the pilot and research instead of the AI Application. To identify and limit these types of biases, combining multiple methods to measure and correctly interpret use are crucial [24]. As we found that our user remarks correspond with prolonged AI Application use, we think this possible bias is negligible.

F. Future Work

Additional measurements are currently being analyzed and include scenario-based tests during which nurses are asked to resolve scenarios by searching for information with or without the app, to measure actual information seeking efficiency. Outcomes of these tests include time needed to search, number of encountered problems during search, and number of times a scenarios was resolved correctly. In addition, questionnaires to measure satisfaction pre- and post-implementation change regarding information seeking and applying information will be analyzed [1].

The AI Application will be introduced on different wards and in other hospitals, following a re-design or adaptation of the application and its content to fit local work methods. This broad implementation will be accompanied by evaluations regarding stakeholder needs, use context and user satisfaction, following the CeHRes roadmap [8].

ACKNOWLEDGMENT

This project was funded by the European Interreg IV "EurSafety Health-net" project. We thank the various participants in the pilot for their cooperation in this research: the nurses, and the chest physicians of the A4/C4 department of the Medisch Spectrum Twente hospital in Enschede, The Netherlands. We further acknowledge Marjolein Hartgerink for her assistance in analyzing data.

REFERENCES

[1] J. Wentzel and J. van Gemert-Pijnen, "Antibiotic Information App for Nurses," The Sixth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2014) IARIA, March 2014, pp. 174-179, ISSN: 2308-4359, ISBN: 978-1-61208-327-8.

[2] S. H. Koch, C. Weir, M. Haar, N. Staggers, J. Agutter, M. Görges, et al., "Intensive care unit nurses' information needs and recommendations for integrated displays to improve nurses' situation awareness," J Am Med Inform Assn, 2012, 19(4), pp. 583-590.

[3] M. Prgomet, A. Georgiou, and J. I. Westbrook, "The impact of mobile handheld technology on hospital physicians' work practices and patient care: a systematic review," J Am Med Inform Assn, 2009, 16(6), pp. 792-801.

[4] C. A. Caligtan, D. L. Carroll, A. C. Hurley, R. Gersh-Zaremski, and P. C. Dykes, "Bedside information technology to support patient-centered care," Int J Med Inform, 2012, 81(7), pp. 442-451.

[5] A. Holzinger, M. Baernthaler, W. Pammer, H. Katz, V. Bjelic-Radisic, and M. Ziefle, "Investigating paper vs. screen in real-life hospital workflows: Performance contradicts perceived superiority of paper in the user experience," Int J Hum-Comput St, 2011, 69(9), pp, 563-570.

[6] B. Middleton, M. Bloomrosen, M.A. Dente, B. Hashmat, R. Koppel, J.M. Overhage, et al., "Enhancing patient safety and quality of care by improving the usability of electronic health record systems: Recommendations from AMIA," J Amer Med Inform Assn, 2013, 20(e1), pp. e2-e8.

[7] R. J. Katz-Sidlow, A. Ludwig, S. Miller, and R. Sidlow, "Smartphone use during inpatient attending rounds: Prevalence, patterns and potential for distraction," J Hosp Med, 2012, 7(8), pp. 595-599.

[8] J. E. W. C. van Gemert-Pijnen, N. Nijland, M. van Limburg, H. C. Ossebaard, S. M. Kelders, G. Eysenbach, et al., "A holistic framework to improve the uptake and impact of eHealth technologies," J Med Internet Res, 2011, 13(4), pp. e111.

[9] L. van Velsen, J. Wentzel, and J.E.W.C. van Gemert-Pijnen, "Designing eHealth that matters via a multidisciplinary requirements development approach," JMIR research protocols, 2013, 2(1), pp. e21.

[10] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive systems design: Key issues, process model, and system features," Communications of the Association for Information Systems, 2009, 24(1), pp. 485-500.

[11] F. Drozd, T. Lehto, and H. Oinas-Kukkonen, Exploring Perceived Persuasiveness of a Behavior Change Support System: A Structural Model, in Persuasive Technology. Design for Health and Safety, M. Bang and E. Ragnemalm, Editors, 2012, Springer Berlin Heidelberg, pp. 157-168.

[12] N. de Jong, A. Eikelenboom-Boskamp, A. Voss, R. Sanderman, and L. van Gemert-Pijnen, "Combining user-centered design with the persuasive systems design model; the development process of a web-based registration and monitoring system for healthcare-associated infections in nursing homes," International Journal on Advances in Life Sciences, in press.

[13] F. Verhoeven, When staff handle staph: user-driven versus expert-driven communication of infection control guidelines, 2009, University of Twente: Enschede.

[14] C. MacDougall and R. Polk. "Antimicrobial stewardship programs in health care systems," Clin Microbiol Rev, 2005, 18(4), pp. 638.

[15] R. Edwards, L.N. Drumright, M. Kiernan, and A. Holmes, "Covering more territory to fight resistance: considering nurses' role in antimicrobial stewardship," Journal of Infection Prevention, 2011, 12(1), pp. 6-10.

[16] J. Wentzel, M. van Limburg, J. Karreman, R. Hendrix, and L. van Gemert-Pijnen, "Co-creation with stakeholders: a web 2.0 antibiotic stewardship program," The Fourth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2012) IARIA, Jan-Feb 2012, pp. 196-202, ISSN: 2308-4359, ISBN: 978-1-61208-179-3.

[17] J. Wentzel, L. van Velsen, M. van Limburg, N. de Jong, J. Karreman, M.G.R. Hendrix, et al., "Participatory ehealth development to support nurses in antimicrobial stewardship," BMC Med Inform Decis, 2014, 14(1), pp. 45.

[18] Wordpress, accessed on 31-01-2014 via: http://wordpress.org

[19] Demo Information Application for nurses. Accessed on 31-01-2014 via http://abnurseapp.infectionmanager.com

[20] Google Analytics, accessed on 2-9-2014 via: http://www.google.com/analytics/

[21] K.A. Ericsson and A. Herbert, Protocol analysis: verbal reports as data, Cambridge, MA: MIT Press, 1993.

[22] N. de Jong, J. Wentzel, S. Kelders, H. Oinas-Kukkonen, and J. van Gemert-Pijnen, "Evaluation of perceived persuasiveness constructs by combining user tests and expert assessments," The Second International Workshop on Behavior Change Support Systems (BCSS2014), May 2014, pp. 7-15.

[23] T. Lehto, H. Oinas-Kukkonen, and F. Drozd. "Factors affecting perceived persuasiveness of a behavior change support system," The 33th International Conference on Information Systems (ICIS 2012), Dec. 2012.

[24] J.Y. Han, "Transaction logfile analysis in health communication research: challenges and opportunities," Patient Educ Couns, 2011, 82(3), pp. 307-312.

[25] S. M. Kelders, E. T. Bohlmeijer, and J.E.W.C. Van Gemert-Pijnen, "Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial," J Med Internet Res, 2013, 15(8), pp. e172.

[26] J. Karreman, N. Ummelen, and M. Steehouder, "Procedural and declarative information in user instructions: what we do and don't know about these information types," Professional Communication Conference (IPCC 2005). IEEE, July 2005, pp. 328-333, ISBN: 0-7803-9027-X.

[27] S. Kujala, "User involvement: A review of the benefits and challenges," Behav Inform Technol, 2003, 22(1), pp. 1-16

[28] E. Gillespie, A. Rodrigues, L. Wright, N. Williams, and R.L. Stuart, "Improving antibiotic stewardship by involving nurses," Am J Inf Cont, 2013, 41(4), pp. 365-367.

[29] M. Leonard, S. Graham, and D. Bonacum, "The human factor: the critical importance of effective teamwork and communication in providing safe care," Qual Saf Health Care, 2004, 13(1), pp. i85-i90.

[30] M. Ndosi and R. Newell, "Medicine information sources used by nurses at the point of care," J. Clin Nurs, 2010, 19(17-18), pp. 2659-2661

Refining *Advanced* Professional Development for Online Teaching and Course Building: An Evaluation From the Faculty Perspective

Vanessa Slinger-Friedman, Joe Terantino, Christopher K. Randall, Philip J. Aust, and Tamara Powell
Kennesaw State University
Kennesaw, United States of America
{vslinger, jteranti, crandal2, paust, tpowel25{@kennesaw.edu

Abstract—**This paper reports the findings of a study conducted to assess an innovative online course building and instructor faculty development program offered at a major regional university. The training program, entitled "The Project," was designed to enhance the knowledge and skills of university-level instructors with previous online course development and instructional experience. At the conclusion of the five-month program, "Project" participants completed a 20-item questionnaire to identify best practices of the program and to refine its content for future offerings. Main findings include (a) faculty learn most when faculty development materials model best practices of online teaching, (b) faculty perceive pedagogy and technology-related modules as the most important content of online training, and (c) online training programs must be updated regularly to keep pace with the rapidly changing landscape of online teaching and learning. Future research needs to examine the impact that similar programs have on student learning, evaluate the transfer of developmental training programs to faculty use of technology, and provide a broader sense of how advanced training impacts faculty development in a university setting.**

Keywords-e-learning, faculty perceptions, online course building, online teaching, professional development.

I. Introduction

Building on previous research conducted with a pilot implementation of "The Project," which highlighted the need for an emphasis on pedagogy to further develop online teaching competencies within an online course development workshop [1], the main purpose of this article is to document the effectiveness of the refined program based on the perceptions of the faculty participants as determined by their responses to an end-of-program survey. This research comes as a response to the rapid increase in demand for more faculty training related to building and teaching online courses [2][3], which correlates directly to the considerable growth experienced in the number of online courses offered and their student enrollment [4][5]. In addition, the trend to offer more faculty training, such as the program described here, serves as a response to the increasing pressure from administration in institutions of higher education for faculty to provide more online course offerings [6].

The distinguishing factor of this study is that it evaluates the implementation of an advanced faculty professional development program for online teaching and course building. A*dvanced* is used to refer to the faculty members who have been trained in online course development and teaching, are experienced online course builders, and who have previously taught online courses. The paper begins with a review of the literature associated with advanced professional development for online teaching and course building and its role in developing online courses and programs. Then the design of the advanced faculty professional development program is detailed, including changes made to "The Project" since its initial implementation, along with the research method that was used for this study. Next, the results of the program evaluation and a discussion of the faculty perceptions are provided. Finally, the article concludes with broader implications of this research through a discussion of how the results from the program evaluation will inform future faculty development workshops.

II. Literature Review

Studies confirm a substantial increase in the availability of online courses and programs in recent years [4][5] and research related to online course development continues to be consistent with the creation, implementation, and facilitation of training for faculty new to online teaching. The motivation behind this line of research is the increasing requests for online course offerings and programs from students and, therefore, the increasing pressure for institutions to provide more online course offerings. Consequently, research about the effectiveness of online training models is also more in line with the needs of students [4][7][8][9] and the concerns of faculty new to online teaching. In the history of faculty development for online teaching, first, faculty development consisted of limited and specific opportunities to learn a task such as how to better use a learning management system. According to Meyer, the second stage consisted of assisting faculty with online-specific pedagogy/andragogy and instructional design [10]. There is little or no mention of advanced faculty who develop online courses [11][12][13][14]. However, it can be reasoned that advanced faculty are much more likely to be able to take advantage of self-directed learning in their faculty development experiences. Quinney, Smith, and Gailbraith found that a self-directed program to assist librarians in gaining/updating technological skills increased their motivation to continue updating their skills and increased their likelihood of using their new skills on the job [15].

Furthermore, there is a considerable amount of research that examines the effectiveness of the transfer of learning (TAM) models and the ease of use among faculty when training for online teaching [12][16]. For example, Agarwal and Prasad [16] describe how training affects the participants' perceptions of usefulness for the technology, and that people more highly educated or trained with the use of technology are more likely to adopt technology for teaching. More directly related to this study, Gegenfurtner, Veermans, Festner, and Gruber [17] found that the way the person perceives training may impact the decision to apply knowledge gained from the training. In an analysis of current and effective training strategies for preparing faculty to teach online Lackey concluded that online "preparation strategies should include both technical and pedagogical training" [18].

Meyer lists examples of 12 faculty development programs offered across the US. The programs come in a wide range of time spans and delivery modes. Meyer notes that some of the programs were offered with little or no in-depth evaluations from participants. As Meyer advances, and the writers of this paper agree, the evaluation of a faculty development program is a great way to learn what is working in the program [10]. Meyer found that often faculty development programs are created using Adult Learning Theory, or andragogy, which originated with Malcolm Knowles [10].

Knowles delineated several characteristics of adult learners: adult learners like to take charge of their own learning experiences, and they like to be respected for the knowledge they bring to the learning environment. In addition, they like to know why they are learning something and what they will do with what they have learned [19]. While adult learners like to take charge of their own learning experiences, institutions seeking to provide professional development opportunities must grapple with the fact that "individual faculty members have different learning styles, needs, knowledge, abilities, and attitudes with respect to developing online courses and teaching online." The challenge with regard to faculty development is to provide "a spectrum of support and professional development opportunities for individual faculty members" [20, p. 21]. Also in relation to adult learning, Mezirow describes transformative learning, which he explains assists adult learners in "reach[ing] their objectives in such as way that they will function as more autonomous, responsible thinkers" [21, p. 8].

Professional development for online instructors, in both online and face-to-face formats, can create effective informal learning whereby participants in the training collaborate, share, discuss, and reflect on different technologies, pedagogies and practices [22]. In this way, participants construct knowledge and transfer learning with each other. It is anticipated that this experience would be heightened and even more beneficial in an advanced faculty online professional development program where the participants already have online training and online teaching experience. Orr contends that as a faculty member gains more experience teaching online, his or her needs with regard to professional development change. Orr notes that faculty surveyed expressed a need for advanced faculty development, a need that institutions had not yet responded to [23, p. 265]. Additionally, researchers have not yet investigated such perceptions of implementing an advanced faculty professional development program, as addressed by this study. The specific research questions addressed in this article are the following:

1. How do faculty perceive the effectiveness of an advanced professional development workshop designed for online course building?
2. What changes do faculty recommend making to the advanced professional development workshop designed for online course building?

III. METHODS

To address the increasing demand for online and hybrid university courses and provide university-level faculty with the incentive, knowledge, and skills to develop and deliver well-constructed online and hybrid courses, a professional training program known as "The Project" was created in January 2013. "The Project" was created to provide faculty within a college who are already teaching online with advanced tools and pedagogy to develop future online courses. For the first iteration of "The Project" online modules were created, run, and analyzed by online coordinators, designated faculty representing nine academic departments in the College of Humanities and Social Sciences who act as a liaison between department faculty and the Office of Distance Education (ODE) in the college. Since faculty serving as online coordinators completed the alpha testing for "The Project" there was an element of 'creators bias' in the study [1].

In their role as program developers, the coordinators initially generated and tested 11 learning modules on various topics addressing online course building and instruction. The 11 modules created for alpha testing over the course of five months, January to May 2013, are offered in Table I.

TABLE I. OVERVIEW OF LEARNING MODULES INCLUDED IN THE'ALPHA TEST' ITERATION OF "THE PROJECT"

Learning Modules
Module 1: Latest Research Into Successful Online Learning
Module 2: Best Practices in Mobile Learning
Module 3: Faculty Presence in Online Courses
Module 4: Get Your Students' Heads INTO the Clouds: Cloud Computing
Module 5: Strong and Effective Types of Feedback
Module 6: Taking the Long View: How Online Learning Has Changed at the University
Module 7: Lessons Learned: Five Tips I Would Share with New Online Coordinators
Module 8: Creative Assignments in the Online Classroom: The Virtual Classroom
Module 9: Learner-Content Interaction in Online Courses
Module 10: Real Online Programs at the University
Module 11: Social Media in Online Teaching

Based on the findings from the alpha test of "The Project" [1], a team comprised of two online coordinators, the CHSS ODE director, and a student worker revised "The

Project" in August 2013 by eliminating extraneous content, adding modules that were found valuable to those with online course building and instruction experience, and reorganizing the program's content for other faculty who desired further training in online course building and online teaching. This version of "The Project" would run as the pilot test in fall 2013 with 12 faculty. The program's finalized content is offered in Table II.

TABLE II. OVERVIEW OF LEARNING MODULES INCLUDED IN THE PILOT VERSION OF "THE PROJECT"

Category of Module	Name of module
Pedagogy/ Online Teaching	Latest Research Into Successful Online Learning; "Faculty Presence" in Online Courses; Strong and Effective Types of Feedback; Learner -Content Interaction in Online Courses; Work Smarter, Not Harder*; Improving Retention in Online Courses*
Trends	Best Practices in Mobile Learning; Get Your Students' Heads INTO the Clouds: Cloud Computing; Creative Assignments in the Online Classroom: The Virtual Museum; The Use of Social Media in Online Teaching
Technology	Panopto*; SoftChalk*; VoiceThread*; Tiki Toki*; Doceri*

*New content added for the pilot version of "The Project"

Faculty members were accepted into the program by an application process. Only faculty who had successfully completed the college's "Build a Web Course" workshop and previously taught online or hybrid courses were eligible to apply. There were four expectations of the advanced faculty development training program: (a) Professional development modules must be completed on schedule, including participant or activity assignments; (b) An online or hybrid course must be built by the participant and delivered in one calendar year; (c) All of a participant's newly developed online or hybrid course must be submitted for college *Quality Matters* [24] pre-review nine months after starting the program; and (d) a participant's online or hybrid course must pass a university Quality Matters Peer Review within one year of beginning "The Project". Participants earned $1000 for completing all four tasks.

"The Project" was hosted on the learning management system Desire2Learn from September 2013 to February 2014. During the first three months of the program, participants had to complete 9 of 12 skills-update modules in one of three general topic areas: (a) pedagogy and online teaching, (b) trends in online teaching, and (c) technology for online teaching. By timeline, three modules had to be completed each month beginning in September and ending at the end of November. In terms of specific content areas, participants went through advanced online instructor training by covering such topics as successful student learning, faculty presence in online learning, improving retention in online courses, and best practices in mobile learning (Table II).

During this skills training, participants were asked to relate their own experiences and methods of achieving a particular teaching goal when teaching an online class. For instance, one module addressed establishing instructor presence in an online course. At times participation required that faculty members share samples from their own courses. In other cases, self-assessment quizzes were used to verify content comprehension and retention. The aim was to have participants share information while learning from others at the same time. In this way, participants were not just presented with content but were offered the opportunity to engage in transformative learning. After learning about a tool or theory or pedagogical method, participants were asked to use the tool and share that work or to reflect upon the theory or pedagogical practice in their own teaching. These final exercises were shared with others on an asynchronous discussion board where fellow participants could view, comment on, and discuss results and observations. In this way, "the [n]ew information [was] only a resource in the adult learning process" [21, p. 10]. The participants engaged in transformative learning and were moved to consider and share whether various presented content could be incorporated into their current teaching practices, and if so, how. While the initial "Build a Web Course Workshop," which was a pre-requisite for "The Project," was crafted according to adult learning principles, "The Project" was created with an additional emphasis on transformative learning principles [21].

Once the skills-update training was completed, participants next had to build an online or hybrid course. The course had to be fully functional, reflect the latest research in online teaching pedagogy, and be taught in the university within a year. In terms of particulars, participants were encouraged to incorporate such technologies as voiceover PowerPoint presentations, wikis, Panopto, SoftChalk, VoiceThread, Tiki Toki, Doceri, or social media in their course designs. Every module started with learning objectives. Modules routinely made use of discussion boards where participants were asked to reflect and interact by answering one or more directed questions relating to the material covered.

The twelve faculty members participating in "The Project's" first year rollout from Fall 2013 to Spring 2014 consisted of seven females and five males. In terms of faculty rank, participants included two full professors, five associate professors, three assistant professors, one senior lecturer, and one part-time instructor. Collectively, the group had an average of 15.5 years of university-level teaching experience, 9.2 years of technology-enhanced teaching, and 4.2 years of online teaching experience. At the conclusion of "The Project," participants completed a survey to gauge perceptions of the professional development program, distinguish what they learned, and clarify what changes needed be made for future iterations of the program. Of the 12 participating faculty, 9 responded to the survey. It should be noted that the relatively small sample size of this research (N=9) limits the generalizability of the findings. Therefore, the presentation of faculty perceptions regarding the advanced faculty workshop for online course building should be considered in context and only applied to other contexts of comparable nature. To bolster the generalizability of these findings, larger samples should be studied over longer periods of time**.**

This program afforded a unique opportunity to answer two emerging research questions in online instruction literature. These questions include (a) How do faculty perceive the effectiveness of an advanced professional development workshop for online course building? (b) And what changes do training program participants recommend for future advanced online course-building workshops? While "The Project" was principally designed to improve instructor online course-building and teaching abilities, it also served to gather information to refine an advanced course-building training program for future participants in the university it took place in as well as comparable settings.

This inquiry-based research focused directly on faculty perceptions of advanced online teacher training. To answer the research questions advanced here, participants were administered a 20-item (Appendix A) questionnaire at the end of "The Project". The questionnaire was electronically administered to the program's 12 faculty members. A general analysis of the open- and closed-ended survey responses was undertaken by focusing on the participants' satisfaction levels with "The Project" as well as by examining specific items addressing the program's components. The survey was designed with the Technology Acceptance Model in mind to ascertain whether the participants planned to use information and tools presented in the project. It was created and delivered via Qualtrics, a cloud-based application for surveys, which made the questionnaire more accessible. Nine of 12 participants completed the questionnaire, which was designed primarily to gauge the participants' perceptions of having participated in and completed the professional development course. The analysis of the survey responses targeted the participants' satisfaction or dissatisfaction with the course and its specific components. This focus was adopted to gather information in an effort to answer the research questions and refine the course for future use. The results of participant feedback follow.

IV. RESULTS

"The Project" participants were generally positive about the ability of "The Project" as an advanced faculty development workshop to effectively serve faculty who have completed the beginning "Build a Web Course" workshop (Fig. 1). A majority (73%) of the respondents were very satisfied or satisfied with the workshop (Fig. 2), and 27% were somewhat satisfied with the workshop, and 91% of the respondents stated that the training materials provided during the workshop assisted them in creating their online course content. All of the respondents liked the fact that "The Project" was completely online. This was the feature that most participants liked best about "The Project". One participant stated, "I like that it could be done completely on your own time." Ninety-one percent of the respondents stated that after completing "The Project" they were better equipped to teach an online or hybrid course (Fig. 3).

Related to this question, participants felt that the updated information about online pedagogy, along with technology expertise that they obtained through the workshop specifically helped better equip them to teach online. Pedagogy, first, and then technology were the two categories of modules from the workshop that were considered the most helpful for the participants. The modules on trends were not selected as helpful to the participants.

Some of the feedback indicated aspects of the workshop that the participants liked the least or felt that could be improved for future iterations of "The Project". These included the need for more timely feedback and instructor presence in the modules and the perceived drawback of being limited to completing three of the five modules in each of the three sections of the workshop. Some participants mentioned that they did not like the way that the workshop was designed where after the deadline to complete a section of the workshop passed, access to the modules in that section were closed. Some participants suggested having all the sections of the workshop open the entire time of the workshop, thereby allowing participants to move through the modules at their own pace and perhaps do more than the three required modules if desired. One participant even recommended permitting "binge" learning, where participants could go through the entire workshop in a very short space of time and then spend the rest of the time developing their own online courses.

All of the workshop participants stated that they would recommend "The Project" to other faculty. With the exception of one participant, the respondents to the survey about "The Project" felt that this workshop was designed at an appropriate level and that they were fairly compensated for the time and effort requirements of the workshop. While only one participant was not interested in participating again in any potential version of "The Project", the rest of the participants expressed an interest in participating in future versions of "The Project". Respondents were asked to what, if any, future version of "The Project" they would most like to see. The most popular potential version of the workshop was one in which "The Project" is completely updated every two years and opened to all faculty.

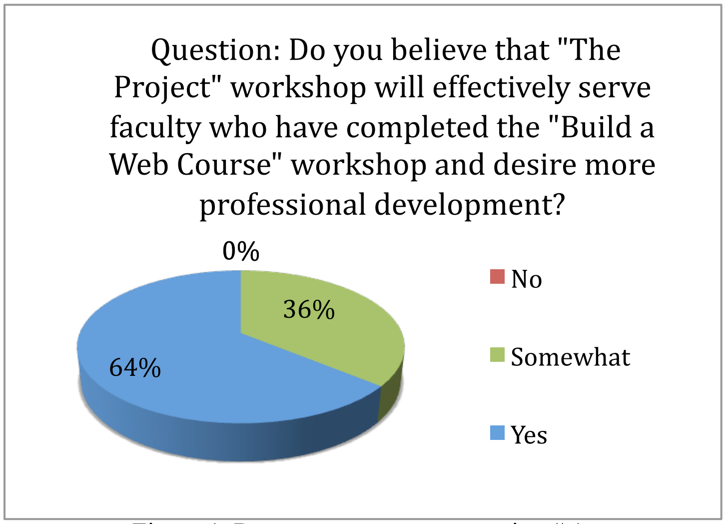


Figure 1. Responses to survey question # 1.

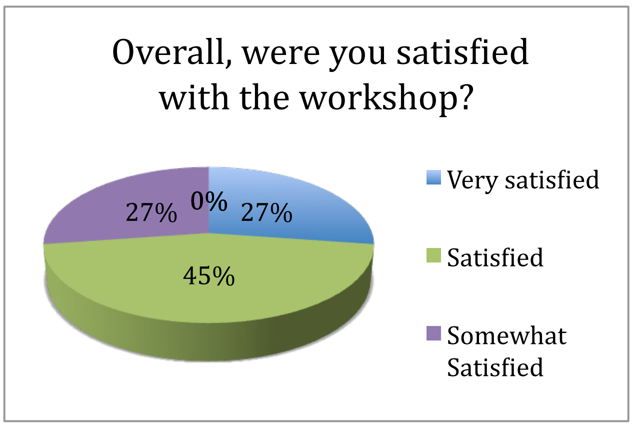


Figure 2. Responses to survey question # 20.

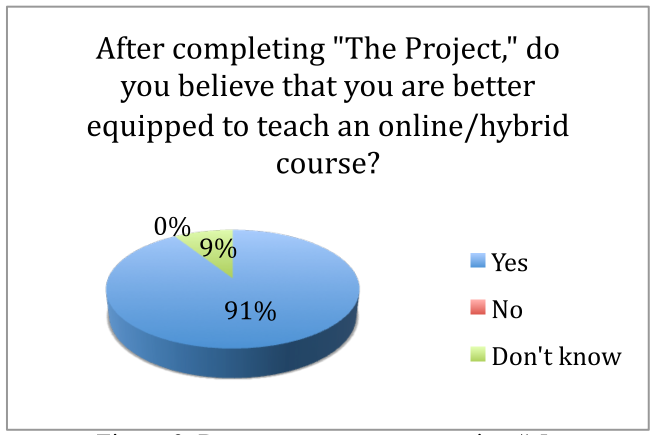


Figure 3. Responses to survey question # 5.

V. DISCUSSION

One of the concerns expressed about the feedback generated from our alpha test of "The Project" was the potential for biased evaluations from participants, given that these individuals also served as content developers for several of the project modules. While acknowledging that structure, content, and post-workshop survey questions were modified somewhat from the initial implementation, the feedback received from the pilot test participants closely mirrored that collected from the alpha test participants. Additionally, the post-workshop survey responses focus on three main findings: First, faculty in online workshops place similar value on the organization of content and immediacy of feedback. Second, participants perceive pedagogy and technology-related modules to be the most important content presented. Third, the revisions implemented were largely successful, but continual improvement is a necessary characteristic of any professional development programs in the rapidly changing landscape of online teaching and learning. A more detailed explanation of each of the findings follows.

A. Characteristics of a Successful Online Professional Development Program

Responses on the post-workshop surveys support the notion that modeling effective practices for online teaching and learning is arguably the preferred method for encouraging faculty to adopt those practices. Conversely, a poorly organized training environment with infrequent or uneven instructor presence does not appear to encourage faculty to reflect on and improve their own course design and online facilitation skills. Specifically, participants overwhelmingly reported that they liked the fact that this workshop was delivered entirely online. Qualitative responses to open-ended questions bemoaned the fact that in a few modules, or at various times during the workshop, instructor feedback was not as interactive or as prompt as participants would have liked. Similarly, several participants reported a desire for immediate acknowledgement of assignment submission and activity completion, which reinforces the suggestion that prompt feedback is important [25], and that may be especially true in the asynchronous online environment [26]. Finally, faculty as students did not like the fact that the module availability was paced across the workshop period, thereby enforcing distributed practice and content mastery. Instead, survey responses reflected the participants' desire for immediate access to all course materials so they could "binge' learn (actual survey response) and focus only on their preferred learning modules from among those developed for The Project. This expressed preference to direct their own learning reinforces Knowles' [19] earlier characterization of adult learners. Despite any real or perceived shortcomings of this second iteration/pilot version of this advanced professional development workshop, however, "The Project" continued to be quite positively received; all respondents reported that they would recommend it to other colleagues and reported that they would complete "The Project" again once the content is updated to reflect new research and technologies.

B. Content Focus for an Advanced Professional Development Program

A second finding that emerged from participant feedback is that they distinguished both pedagogy and technology modules as the most important content areas in the workshop for furthering one's online teaching abilities. Indeed, 64% of participants noted that of the 12 modules in "The Project," covering three online education topic areas (i.e., pedagogy and online teaching, trends in online teaching, and technology for online teaching), the topic deemed most valuable to the program was pedagogy for online teaching. Ninety-one percent of respondents believed they were improved online teachers based on their "The Project" experience, due especially to content that emphasized pedagogy. These comments substantiated this finding: "I really liked the theory part of the course, because it allowed us to step back and get a look at the context of what we are doing." And "The course made me think again about online pedagogy and how to teach an effective online class." These responses are in line with Orr's findings that experienced online faculty seek to create and deliver high quality courses and want feedback regarding how to improve their courses [23]. These responses also support Mezirow's theory of transformative learning [21]. By being asked to revisit their online practices and discuss these practices online with other online teaching veterans—often also providing screenshots of their online courses with explanations of various practices—faculty were able to reflect on their goals and motivations for various activities. As a result, they often "arrive[d] at a transformative insight" [27, p. 20] that allowed them to transfer concepts presented in "The Project" into improved learning experiences for students. Additionally, a clear preference for the technology-focused modules and their relevance to pedagogy emerged via both the quantitative ratings of those modules and the qualitative comments, such as: "Use of better technology to mark online presence," "I appreciated learning about SoftChalk and immediately incorporated into my online course," and "SoftChalk and Panopto are two very useful softwares that I am using in creating my classes." This second finding fits with Stephenson's [28] perennial work on online education that emphasizes the importance of pedagogy to online instruction. This finding also supports the significance of the growing body of pedagogical theories being taught in training programs for online instruction and being referenced and tested in social scientific research (e.g., Technology Acceptance Model [16] and Theory of Action [29]). Specifically, the Technology Acceptance Model suggests that the perceived usefulness and ease of use of the technology will increase the users' acceptance of the tool. Overall, these results lead us to concur with Lackey's [18] previously stated view about the importance of including both techniques and pedagogy in online training strategies.

C. Continual Improvement is Critical for a Successful Online Professional Development Program

Assessment has been and will continue to be an essential component of "The Project". The feedback generated during the first iteration of "The Project" underscored a need for and guided the revision process that resulted in the structural and content changes in the current version of the workshop. As evidenced by the qualitative feedback provided by the current participant group, the revisions were largely successful as "The Project" was well-received by a group of experienced faculty who were not involved in the development of the workshop content. Nevertheless, the most recent participants identified several areas for improvement and future enhancements. Additionally, participants continued to acknowledge the importance of continued skill development as part of their experiences in "The Project". As Burke and Hutchins [12] as well as Gegaenfurtner, Veermans, Festner, and Gruber [17] note, emphasis on skill development is an important component of any effective instructor development program, especially for one focused on online teaching. Several responses support attention to online skills training: (a) "Learning about and practicing new technologies for online teaching," (b) "I learned a bit more about online pedagogy (the importance of timely feedback and instructor presence, etc.) and a few techniques and tools to add to classes (like Tiki-Toki)," and (c) "It helps to keep "fresh" with this stuff! It's absolutely necessary actually!" In summary, "The Project" appears to serve as a meaningful experience and, with consistent assessment focused on continuous improvement and technological advances, this workshop model should remain a viable alternative for continuing professional development in the years to come.

VI. CONCLUSION AND FUTURE WORK

The favorable results of this study make plain that "The Project" provides a formal structure for colleges and universities who have yet to establish a similar advanced faculty development program. Moreover, the study offers added insights to developers of teacher training programs. These insights include: (a) based on participant feedback, emphasis on pedagogy must be the cornerstone of any faculty-development online teaching program; (b) colleges and universities are justified in offering skill-development training to faculty with minimum and extensive online teaching experience, because new teaching technologies continue to emerge; and (c) developers are encouraged to seek regular feedback from program participants in order to assess and revise program components and, as a consequence, maximize faculty development. It is important to seek faculty input on faculty development opportunities both to gain feedback for real improvement and to encourage faculty buy-in [20]. In this respect, this research goes far in addressing an area of deficiency in faculty development programs – the lack of participant analysis and evaluation previously noted by Meyer [10].

As can be seen in Table II, five modules were added to The Project after it was alpha tested. Moreover, two modules on pedagogy were added: “Work Smarter, Not Harder” with timesaving tips for online teachers; and “Improving Retention in Online, Hybrid, and F2F Courses” with research-based strategies and templates for improving course retention. The ODE will continue to revise and modify modules so that faculty members with advanced online teaching experience will benefit from The Project.

As noted in the introduction, this study provides a unique opportunity to examine faculty perceptions of an *advanced* development program for online course building. Although the findings of this research yield positive results, several areas merit future research. First, similar research should draw from a larger sample size. Second, researchers should evaluate the transference of technology skills as a result of completing the workshop. Third and last, future research should examine the impact such workshops may have on student learning. Such future research will provide a fuller picture of how an advanced faculty development workshop impacts faculty teaching and students’ learning.

ACKNOWLEDGMENT

We thank Kennesaw State University’s College of Humanities and Social Sciences Office of the Dean and its CHSS Office of Distance Education for its support of the college’s faculty development opportunities. We would also like to thank the faculty participants in “The Project” for taking the time to provide us with valuable feedback about the workshop.

APPENDIX A: SURVEY ITEMS

Question 1: Do you believe that "The Project" workshop will effectively serve faculty who have completed the "Build a Web Course" workshop and desire more professional development?

- No
- Somewhat
- Can’t say/don’t know
- Yes

Question 2: Do you like the fact that "The Project" was all online?

- No
- Somewhat
- Can’t say/don’t know
- Yes

Question 3: What category of modules was most helpful to you?

- None
- Pedagogy/online teaching
- Trends
- Technology
- Other

Question 4: What category of modules was least helpful to you?

- None
- Pedagogy/online teaching
- Trends
- Technology
- Other

Question 5: After completing "The Project," do you believe that you are better equipped to teach an online/hybrid course?

- Yes
- No
- Don’t know

Question 5a. If you answered "yes," above, would you please share an example of how you feel you are better equipped to teach an online/hybrid course

Question 6: After completing "The Project," do you feel you are better equipped to incorporate technology into your face to face classes?

- Yes
- No
- Don’t know
- N/A

Question 6a. If you answered "yes," above, would you please share an example of how you feel you are better equipped to incorporate technology into your face to face course

Question 7: What did you like least about "The Project"?

Question 8: What did you like most about "The Project"?

Question 9: What changes would you make to better serve faculty who enroll in "The Project" workshop?

Question 10: For each module below, if you did not participate in the module, click "not applicable." If you did participate in the module, please rank its value to you (usefulness, helpfulness) on a scale of 1-10, with 1 being "I felt like I wasted my time" and 10 being "It changed my life."

Quality Matters
Tiki Toki
Soft Chalk
Social Media
Hot Potatoes
MicroPoll
The Schedule Assignment
Grade Mark
The Virtual Owl
Brain Shark
Cloud Computing
The Virtual Museum/Blogging
VoiceThread
Faculty Presence
Mobile Learning
Doceri
Latest Research into Effective Online Learning
Strong and Effective Feedback
Panopto

Question 11: What would you add to "The Project"?

Question 12: Do you believe that “The Project” was an appropriate level/time requirement of faculty development for you,you were asked to contribute an appropriate amount of time and work both on your own course and the professional development, you were compensated adequately, and you gained skills needed?

Question 13: We have had suggestions for various ways to use “The Project” for advanced faculty development. Option 1: Run two versions: one focused on pedagogy and latest research, and one focused on tools and trends. Faculty can participate in each one once. Option 2: Run three versions: one focused on pedagogy, one focused on latest research, and one focused on tools. Faculty can participate in each one once. Option 3: "The Project" will be completely updated every two years, and all faculty will be eligible to participate in it again. Do any of these ideas interest you, or do you have other suggestions?

Question 14: Would you recommend "The Project" to other faculty?

- Yes
- No
- Don't know

Question 15: Did the training materials provided to you during the workshop assist you in creating course content?

- Yes
- No
- Don't know

Question 16: Was the design of the workshop logical and clear?

- Yes
- No
- Don't know

Question 17: Did the facilitator provide adequate support as you created your online or hybrid course?

- Yes
- No
- Don't know

Question 18: During the QM evaluation portion of the course (as the QM rubric was applied to your course), were you and your course treated with care and respect?

- Yes
- No
- Don't know

Question 19: During the workshop were your questions related to designing and delivering online and hybrid courses answered adequately and respectfully?

- Yes
- No
- Don't know

Question 20: Overall, were you satisfied with the workshop?

REFERENCES

[1] J. Terantino, V. Slinger-Friedman, G. Thomas, C. Randall, P. Aust, and T. Powell, "Faculty perceptions of "The Project": An advanced faculty professional development for online course building," The Sixth International Conference on Mobile, Hybrid, and Online Learning (eLmL 2014). *IARIA* March 23 – 27, 2014, pp. 69-74. [Online] Available from: http://www.thinkmind.org/index.php?view=article&articleid=elml_2014_3_40_50059.2014.12.01.

[2] L.M. Lane, "An open, online class to prepare faculty to teach online," Journal of Educators Online, vol. 10(1), 2013, [Online] Available from: http://www.thejeo.com/Archives/Volume10Number1/Lane.pdf. 2014.12.01

[3] M. Northcote, D. Reynaud, and P. Beamish, "Teaching the lecturers: Academic staff learning about online teaching," US-China Education Review, vol. A(4), pp. 384-393, 2012.

[4] I.E. Allen, J. Seaman, D. Lederman, and S. Jaschik, "Conflicted: Faculty and online education," A joint project of the Babson Survey Research Group and Inside Higher Ed, pp.1-48, 2012. [Online] Available from: http://files.eric.ed.gov/fulltext/ED535214.pdf 2014.12.01

[5] I.E. Allen and J. Seaman, "Class differences: Online education in the United States, 2010," The Sloan Consortium, pp. 1-30, 2010. [Online] Available from: http://sloanconsortium.org/publications/survey/class_differences. 2014.12.01

[6] C. Carlson, P. Aust, J. McNeill, T. Powell, and L. Witt, "Which technology do I use to teach online: New media and communication course instruction," The Journal of Online Learning and Teaching, vol. 8(4), 2012, pp. 334-347, [Online] Available from: http://jolt.merlot.org/vol8no4/carlson_1212.htm. 2014.12.01

[7] E.T. Jones, J.R. Lindner, T.H. Murphy, and K.E. Dooley, "Faculty philosophical position toward distance education: Competency, value, and educational technology support," The Journal of Distance Learning Administration, vol. 5(1), 2002, [Online] Available from: http://www.westga.edu/~distance/ojdla/spring51/jones51.htm. 2014.12.01

[8] B. Means, Y. Toyama, R. Murphy, M. Bakia, and K. Jones, Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. (Contract No. ED-04-CO-0040) Washington, D.C: U.S. Department of Education, pp. 1-66, 2010.

[9] M. Fetzner, "What Do Unsuccessful Online Students Want Us to Know?" Journal of Asynchronous Learning Networks, vol 17(1), 2013, [Online] Available from: http://onlinelearningconsortium.org/sites/default/files/2_What_Do_Unsuccessful_Online_Students_Want_Us_to_Know_0.pdf 2014.12.01

[10] K.A. Meyer, "An Analysis of the Research on Faculty Development for Online Teaching and Identification of New Directions," Journal of Asynchronous Learning Networks, vol 17 (4), 2014, [Online] Available from: http://onlinelearningconsortium.org/sites/default/files/8-meyer.pdf 2014.12.01

[11] A. Barker, "Faculty development for teaching online: Educational and technological issues," Faculty Development for Teaching Online, vol. 34(6), pp. 273-278, 2003.

[12] L.A. Burke and H.M. Hutchins, "Training transfer: An integrative literature review," Human Resource Development Review, pp. 264-297, 2007.

[13] S. Gold, "A constructivist approach to online training for online teachers," Journal of Asynchronous Learning Networks, vol. 5(1), pp. 35- 57, 2001.

[14] J. Terantino and E. Agbehonou, "Comparing faculty perceptions of an online development course: Addressing faculty needs for online teaching," Online Journal of Distance Learning Administration, vol. 15(2), 2012, available from http://www.westga.edu/~distance/ojdla/summer152/terantino_agbehonou152.html 2014.12.01

[15] K.L. Quinney, S.D. Smith, and Q. Gailbraith, "Bridging the gap: Self-directed staff technology training," Information Technology and Libraries, 29 (4), 2010. Available from http://ejournals.bc.edu/ojs/index.php/ital/article/viewFile/3131/2746 2014.12.01

[16] R. Agarwal and J. Prasad, "Are individual differences germane to the acceptance of new information technologies?," Decision Sciences, vol. 30(2), pp. 361-391, 1999. [Online] Available from: http://dx.doi.org/10.1111/j.1540-5915.1999.tb01614.x 2014.12.01

[17] A. Gegenfurtner, K. Veermans, D. Festner, and H. Gruber, "Integrative literature review: Motivation to transfer training: An integrative literature review," Human Resource Development Review, vol. 8, pp. 403-423, 2009. [Online] Available from: http://dx.doi.org/10.1177/1534484309335970 2014.12.01

[18] K. Lackey, "Faculty development: An analysis of current and effective training strategies for preparing faculty to teach online," Online Journal of Distance Learning Administration, vol. 16(2), 2013, [Online] Available from: http://www.westga.edu/~distance/ojdla/winter144/lackey144.html 2014.12.01

[19] M. Knowles, The Adult Learner: A Neglected Species. Gulf Publishing Company, Houston, Texas. 1973

[20] M. Britto, C. Ford, and J. Wise, "Three Institutions, Three Approaches, One Goal: Addressing Quality Assurance in Online Learning," Journal of Asynchronous Learning Networks, 17 (4), pp. 11-24, 2014. [Online] Available from:

http://jaln.sloanconsortium.org/index.php/jaln/article/viewFile/402/62. 2014.12.01.

[21] J. Mezirow, "Transformative learning: theory to practice," New Directions for Adult and Continuing Education, vol. 74, 1997, [Online] Available from: http://www.esludwig.com/uploads/2/6/1/0/26105457/transformative-learning-mezirow-1997.pdf. 2014.12.01

[22] S. Bolt, "Professional development: Then and now," International Association for Development of the Information Society, pp. 287-290, 2012. [Online] Available from: http://files.eric.ed.gov/fulltext/ED542831.pdf. 2014.12.01

[23] R. Orr, M. William, K. Pennington, "Institutional efforts to support faculty in online teaching," Innovative Higher Education, vol 34, pp. 257 – 268, 2009.

[24] Quality Matters: Higher Education > Course Reviews. (n.d.). [Online] Available from: https://www.qualitymatters.org/reviews 2014.12.01

[25] A. W. Chickering and Z. Gamson (Eds.), Applying the seven principles for good practice in undergraduate education. (47th ed.). San Francisco: Jossey-Bass, 1991.

[26] A.W. Chickering and S.C. Ehrmann, "Implementing the seven principles: Technology as lever," American Association of Higher Education Bulletin, October 1996, [Online] Available from: http://www.tltgroup.org/programs/seven.html 2014.12.01

[27] J. Mezirow, "Learning to Think Like an Adult," in J. Mezirow (Ed.) Learning As Transformation: Critical perspectives on a theory in progress. San Franscisco, CA: Jossey Bass, pp. 3-34, 2000.

[28] J. Stephenson, Teaching & learning online: Pedagogies for new technologies. Sterling, VA: Stylus Publishing, Inc., 2001.

[29] F. Davis, R. Bagozzi, and P.R. Warshaw, "User acceptance of computer technology: A comparison of two theoretical models," Management Science, vol. 35(8), pp. 982-1003, 1989. [Online] Available from: http://dx.doi.org/10.1287/mnsc.35.8.982 2014.12.01

Developing a Data Infrastructure for Patient-Centered Telemedicine

Juha Puustjärvi
University of Helsinki
Helsinki, Finland
juha.puustjarvi@cs.helsinki.fi

Leena Puustjärvi
The Pharmacy of Kaivopuisto
Helsinki, Finland
leena.puustjarvi@kolumbus.fi

*Abstract***— Patient-centered care is an emerging healthcare model that is changing how people think about health and about the patients themselves. It emphasizes the coordination and integration of care, and the use of appropriate information, communication, and education technologies in connecting patients, caregivers, physicians, nurses, and others into a healthcare team where the health system supports and encourages cooperation among team members. However, in spite of the widespread adoption of telemedicine, existing telemedicine applications neither support patient centered-care nor the interoperation of relevant e-health tools. In this paper we present a cloud-based telemedicine consultation server, which manages telemedicine consultation request and their allocation to consulting physicians. It also aids physicians to access patients' health documentation either through a Personal Health Record (PHR) or Personal Health Information System (PHIS). Although both ways are suitable for supporting patient centered telemedicine the PHIS is superior in that it integrates the functionalities of a variety of e-health tools, and thus provides a holistic approach for personal healthcare. However, the integration of e-health tools requires the introduction of a shared ontology and the transformation of patients' health records in the format that is compliant with the shared ontology.**

Keywords - telemedicine; patient-centered care; personal health records; CCD standard, cloud computing; ontologies; eHealth ecosystem

I. Introduction

Telemedicine is the use of telecommunication and information technologies in order to provide clinical health care at a distance [1]. It aims to eliminate distance barriers and can improve access to medical services that would often not be consistently available in distant rural communities [2] [3]. In particular, telemedicine is viewed as a cost-effective alternative to the more traditional face-to-face way of providing medical care [4].

At the same time, the introduction of new emerging healthcare trends, such as patient-centered care [5], pharmaceutical care [6], chronic care model, and personal health records (PHRs) [7], are changing how people think about health and about the patients themselves. In addition, many studies have demonstrated that the provision of these healthcare models can increase compliance with treatment regimens, satisfaction with the health care provider and medical facility, and improve the ultimate health outcome for the individual [8].

It is also true that patients who do not understand their treatment instructions, disease management, or prescription requirements are more likely to mishandle their health, be hospitalized more frequently, and have much higher medical costs than their more involved counterparts [9].

Unfortunately, none of the categories of telemedicine support these emerging healthcare trends: *Store-and-forward telemedicine* involves acquiring medical data and then transmitting this data to the system that is accessible to patient's physician. *Interactive services* provide real-time interactions between patient and physician. It includes phone conversations, online communication and home visits. *Remote monitoring* enables medical professionals to monitor a patient remotely using various technological devices.

In this paper, we describe our work on developing a data infrastructure that supports patient-centered telemedicine. In particular, we describe our work on developing a cloud-based telemedicine system that manages consultation requests by a software module called Consultation Server. It also assists physicians in accessing patients' health information that is stored either in Personal Health Records (PHR) [10] or Personal Health Information System (PHIS) [11, 12].

In our vision PHRs, and especially PHISs, can significantly contribute to the introduction of patient-centered care. Further, they do not only improve the quality of care but also significantly increase the efficiency of consulting physicians as they can be authorized to access patient's health documentation.

Our approach changes the traditional telemedicine paradigm: its main goal is not only to provide a cost-effective alternative to the more traditional face-to-face way of providing medical care but rather to provide a data infrastructure for information sharing among patient's healthcare team.

In the design of the Consultation Server, as well in the design of the PHIS, we have followed the idea of knowledge oriented organizations [13]. Its key idea is to revolve all applications around a shared ontology. The main gain of such architecture is that the applications can interoperate by accessing the shared ontology.

In our introduced ecosystem the participants that agree to work together for practicing telemedicine is called as a *telemedicine affinity domain*. Examples of possible telemedicine affinity domains include nationwide and regional affinity domains, regional federations made up of several local hospitals, healthcare providers, and insurance provider supported communities.

A useful feature of a telemedicine affinity domain is that it is global: its components can locate anywhere in the Internet, and it can be exploited by patients and consulting physicians as far as they have an Internet connection.

The notion of telemedicine affinity domain has similarities with the *clinical affinity domain* of the IHE XDS [14], which studies the problem of patient's scattered clinical documentation. Its key idea is that patient's clinical documents are dynamically retrieved from a clinical affinity domain by exploiting relevant registries. This model assumes that patients' clinical documentation follow them as they move from one clinical affinity domain to another. Another difference between IHE XDS and PHRs is that in the former patient's clinical documentation is managed by healthcare authorities while in the latter they are managed by a patient.

The rest of the paper is organized as follows. First, in Section II, we give a motivating scenario of our ideas of patient-centered telemedicine in practice. Then, in Section III, we give an overview of the emerging healthcare models that have inspired our work. Further, in Section IV, we give an overview of various approaches for storing patients' healthcare documentation. However, as in our case PHRs and PHISs have a key role in storing patients' health documentation, they are considered more detailed in Sections V and VI, respectively. With respect to PHRs we focus on considering the expression power of the HL7 CCD standard while with PHISs we introduce our designed system, which extends traditional PHRs by the functionalities of various e-health tools. The key point here is that these tools interoperate by sharing an ontology, which is an integration of the data sources of these e-health tools.

In Section VII, we consider operational aspects of telemedicine. In particular, we present the Consultation Server that manages telemedicine consultation request and their allocation to consulting physicians. It also aids physicians to access patients' health documentation either through a CCD standard-based PHR or through our designed PHIS. In Section VIII, we consider the Consultation Server as an ecosystem having many interconnected parts. Finally, Section IX concludes the paper by analyzing the potential risks that may jeopardize the deployment of our designed telemedicine system. We also shortly consider our future work on integrating the Consultation Server with other e-health tools used by the consulting physicians.

II. MOTIVATING SCENARIO

Assume that a patient, named Nancy Taylor, has an online Web-based free PHR, where her health data and information related to the care given to her is stored. She can access her PHR from any place having an Internet connection.

One day Nancy discovered a rash on her waist, and so she decided to visit the nearest general practitioner having a contract with a telemedicine affinity domain. The practitioner examines Nancy's rush but the practitioner does not know what kind of treatment or medication should be prescribed for Nancy, and so the practitioner decides to request medical consultation by his Web browser.

To carry out the consultation the practitioner first takes a photo from Nancy's waist. Then the practitioner fills the request document by describing the symptoms of the rash and attaches the photo to the document. The document also provides a hierarchical classification (i.e., a taxonomy) of diseases. The practitioner marks the node "skin disease". In addition, authorized by Nancy the practitioner adds a link to Nancy's PHR and authorization for its use (including required access rights). Finally, the practitioner clicks the submit button, and so the document is delivered to the Consultation Server, which maintains a queue of the pending requests.

Each request includes a set of metadata items such as disease(s), source of request, language and possible priority of the request. The function of the metadata items is to enable automatic matching of the requests and the consulting physicians. So the Consultation Server shows for a consulting physician only the requests that match with his or her profile. Therefore, each consulting physician of the affinity domain also has a profile, which has values for the metadata items and is stored in the Consultation Server.

Assume that a physician, named John Smith, is a specialist in a hospital of a telemedicine affinity domain. In addition his profile matches with the request concerning Nancy, and so the request is in the consultation request queue shown for him.

After a few minutes John Smith picks up the consultation request document and examines the symptoms described in the document as well as the attached photo. Immediately he recognizes Nancy's rash as shingles (herpes zoster), which can be treated by a medicinal product named Aciclovir. Then he checks from Nancy's PHR whether Nancy has some allergies that would prevent the use of the drug or whether she has some ongoing medical treatment that could cause mutual negative effects. As there are no such findings the physician updates Nancy's PHR by the prescribed medication and by the diagnosis he made. Finally John constructs and signs the prescription electronically, which is then electronically delivered to the general practitioner visited by Nancy.

Further, assuming that instead of PHR Nancy is using a PHIS that integrates the functionalities of various e-health tools including information therapy server, then the PHIS would automatically send by e-mail relevant medical information (or their links) dealing with the prescribed medication (i.e., Aciclovir) and the diagnosis (i.e., herpes zoster).

III. EMERGING HEALTHCARE MODELS

Patient-centered care emphasizes the coordination and integration of care, and the use of appropriate information, communication, and education technologies in connecting patients, caregivers, physicians, nurses, and other into a healthcare team where health the system supports and encourages cooperation among team members [15]. It is based on the assumption that physicians, patients and their families have the ability to obtain and understand health information and services, and make appropriate health decisions.

Pharmaceutical care emphasizes the movement of pharmacy practice away from its original role on drug supply towards a more inclusive focus on patient care **[**16**].** It emphasizes the responsible provision of drug therapy for the purpose of achieving definite outcomes that improve patient's quality of life [17].

Chronic care model [18, 19] emphasizes patients' long-term healthcare needs as a counterweight to the attention typically paid to acute short-term, and emergency care. In this sense, the traditional care models are not appropriate as the patients with chronic illness do not receive enough information about their condition, and they are not supported in caring themselves after they leave the doctor's office or hospital.

Information therapy is a type of healthcare information service that has emerged in the past decade. The goal behind information therapy is to prescribe the right information to right people at right time [20]. Information therapy is also described as "the prescription of specific evidence based medical information to specific patients at just the right time to help them make specific health decisions or behavior changes" [21].

Information therapy applies to a wide range of situations and context. For example, information therapy may be a physician-written prescription telling a patient what to read, or it may use to help a patient to make treatment decision such as whether to continue medication.

Information therapy can be compared to similar concepts in medicine such as drug therapy, physiotherapy or bibliotherapy [22]. However, information therapy differs from these in the sense that by exploiting information technology information therapy aims at providing personalization, targeting and documentation.

Personalization means that the content of the delivered information depends on the familiarity of the user. Targeting means that the information prescriptions are targeted according to patient's moment in care, i.e., information is delivered at right time. Documentation means that information prescriptions are documented as part of the provided treatment.

There is a variety of paper-based mediums for delivering information therapy such as handing out information pamphlets or sending them through the post. There are also many electronic infrastructures (such as electronic medical record systems, personal digital assistants, order entry systems and personal health records) that have been proposed for delivering information therapy.

IV. MANAGING PATIENTS' HEALTH DOCUMENTATION

A. *PHRs vs. EHRs*

A PHR is a record of a consumer that includes data gathered from different sources such as from health care providers, pharmacies, insures and the consumer [23]. It includes information about medications, allergies, vaccinations, illnesses, laboratory and other test results, and surgeries and other procedures [24]. PHRs are owned by the patients. They also allow individuals to access and coordinate their lifelong health information and make appropriate parts of it available to those that are authorized by the individual.

In medical consultation, as well as in any medical treatment, a complete and accurate summary of the health and medical history of the patient is of prime importance. A problem here is that as a patient may have lived in various places and a patient may have many healthcare providers, including primary care physician, specialist, therapists and other medical practitioners, patient's health documentation may be distributed over several healthcare providers.

PHRs provide a simple way for solving the problem of patients' scattered clinical documentation [25]. They differ from EHRs (Electronic Health Record) [26, 27] in that they are owned by the patients and they can only be accessed by the patient and those that are authorized by the patient while EHRs assume that the health records are designed only for use by health care providers and are owned by medical authorities.

Managing fragmented healthcare documentation by EHRs has been successful only in a very few industrialized countries, such as in Singapore and Denmark [28]. Instead successful results from the use of PHRs are reported from many countries and communities. Therefore we have also concluded that the use of PHRs instead of national EHR's in managing patients' health documentation is a more appropriate solution in the context of telemedicine.

B. *Web-Based PHRs*

PHRs can be classified according to the platform by which they are delivered. In web-based PHRs health information is stored at a remote server, and so the information can be shared with health care providers. They also have the capacity to import data from other information sources such as a hospital laboratory and physician office. However, importing data to PHRs from other sources requires the standardization of PHR-formats.

Various standardization efforts on PHRs have been done. In particular, the use of the Continuity of Care Record (CCR standard) of ASTM [24] and HL7's [29] Continuity of Care Document (CCD standard) [30] has been proposed. From technology point of view CCR and CCD-standards represent two different XML schemas designed to store patient clinical summaries [31]. However, both schemas are identical in their scope in the sense that they contain the same data elements.

Web-based PHRs are core components in our proposed ecosystem. However, it does not assume that all patients have a PHR, but it encourages patient to acquire a PHR. Using a PHR does not require patients to own any personal devices for internet connection nor any efforts for managing it. Rather it requires patient to authorize healthcare personnel to maintain and access their PHR.

Acquiring a PHR is a tempting opportunity as there are many freely available web-based PHRs available, and moving personal data between standard-based PHR is supported by the vendors. For example, as the support of the Google Health was retired on January 2012 Microsoft has developed a transfer process for the user of the Google Health for moving their health records into Health Vault.

Similar to Dossia and World Medical Card, it is a web-based system to store, maintain and share health and fitness information. They support a number of exchange formats including industry standards such as the CCR and CCD standards.

V. CCD STANDARD AND PHRS

A. CCD Standard

CCD standard (an XML schema) is nowadays increasingly used for specifying the structure of exchanged clinical documents as well as specifying the structure of the PHR. This feature simplifies the update of the PHR as well as the generation of the clinical documents that will be stored in the PHR.

The CCD standard is a constraint on the HL7 CDA standard. The CCD standard has been endorsed by HIMMS (Healthcare Information and Management Systems Society Though) [32] and HITSP (Healthcare Information Technology Standards Panel) [33] as the recommend standard for exchange of electronic exchange of components of health information.

Although the original purpose of the CCD documents was to deliver clinical summaries between healthcare organizations, nowadays the XML schema of the CDD documents is increasingly used for specifying the structure of the PHRs. The same schema can be used in message as all its parts are optional, and it is practical to mix and match the sections that are needed.

B. HL7 RIM and RMIM

We have given preference to CCD standard (an XML schema), which is based on HL7 RIM (Reference Information Model). The RIM is the cornerstone of the HL7 message development process and development methodology [34]. It expresses the data content needed in a specific clinical or administrative context and provides an explicit representation of the semantic and syntactical connections that exist between the information carried in the fields of HL7 messages [35].

The RIM is based on two key ideas [28]. The first idea is based on the consideration that most healthcare documentation is concerned with “happenings” and things (human or other) that participate in these happenings in various ways.

The second idea is the observation that the same people or things can perform different roles when participating in different types of happening, e.g., a person may be a care provider such a physician or the subject of care such as patient.

As a result of these ideas the RIM is based on a simple backbone structure, involving three main classes, Act, Role, and Entity, linked together using three association classes Act-Relationship, Participation, and Role-Relationship (Fig. 1). Note that HL7 uses its own representation of UML to reflect the use of these six backbone classes. Each class has its own color and shape to represent the stereotypes of these classes, and they only connect in certain ways.

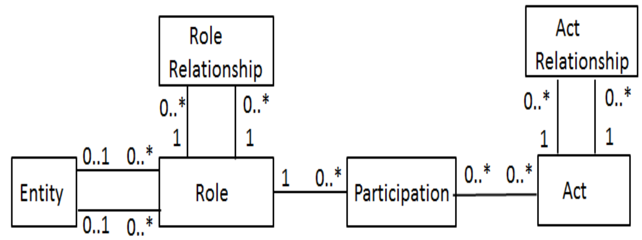


Figure 1. RIM backbone structure.

Each happening is an Act and it may have any number of Participations, which are Roles, played by Entities. An ACT may also be related to other Acts via Act Relationships. Act, Role and Entity classes have a number of specializations (subclasses), e.g., Entity has a specialization LivingSubject, which itself has a specialization Person.

The classes in the RIM have structured attributes which specify what each RIM class means when used in a message (exchanged document). For example, Act has structured attributes classCode and moodCode. The former states what sort of Act this is (e.g., observation, encounter, or administration of a drug). moodCode indicates whether an Act has happened, is request for something to happen, a goal or a criterion. The idea behind structured attributes is to reduce the original RIM from over 100 classes to a simple backbone of six main classes [28].

Note that the RIM is not a model of healthcare, nor is it a model of any message, although it is used in exchanged messages. The structures of exchanged documents are defined by constrained information models [28]. The most commonly used constrained information model is the RMIM (Refined Message Information Model). Each RMIM is a diagram that specifies the structure of an exchanged document.

A RMIM diagram is specified for a specific use case [35]. The diagram is derived from the RIM by limiting its optionality. Such specifications are called CDA Profiles [36].

In developing a RMIM diagram the RIM is constrained by omission and cloning. Omission means that the RIM classes or attributes can be left out. Note that all classes and attributed that are not structural attributes in the RIM are optional, and so the designer can take only the needed classes and attributes. Cloning means that the same RIM class can be used many times in different ways in various RMIMs. The classes selected for a RMIM are called clones.

The multiplicities of associations and attributes in a RMIM are constrained in terms of repeatability and optionality. Further, code binding is used for specifying the allowable values of the used attributes.

Although the semantics of all CDA documents is tractable through a RMIM back to the RIM, we neither can use the RMIM nor the RIM in formulating queries on patient’s health documentation as each RMIM only models one type of documents. Another reason is that there are no query languages specified for the information model used in the RMIM and RIM schemas.

C. HL7 CDA Levels

Each CDA document has one primary purpose (which is the reason for the generation of the document), such as patient admission, transfer, or inpatient discharge. Each Clinical Document Architecture (CDA) document is made up of the header and the body [28].

Depending whether the header and body of the CDA documents are based on the RIM they are classified into three levels:

- CDA Level 1: Only the header is based on the RIM while the body is human readable text or image.
- CDA Level 2: Only the header of the document is based on the RIM while the body is comprised of XML coded sections.
- CDA Level 3: Both the header and the body are based on the RIM.

The CDA header is common to all the three levels of CDA. The header contains basic metadata. These include information about what the documents is, who created it, when, where, and for what purposes. Its primary purpose is to provide unambiguous, structured metadata about the document itself, which can be used in document registers to classify, find and retrieve documents.

In HL7 CDA terminology the header is an instance of an Act called Clinical Document. This means that there is a Refined Message Information Model (RMIM) that models the headers of all HL7 CDA documents. To illustrate this, a simplified RMIM of the Header of CDA documents is presented in Fig. 2. The diagram presents classes of the RMIM but not all their attributes.

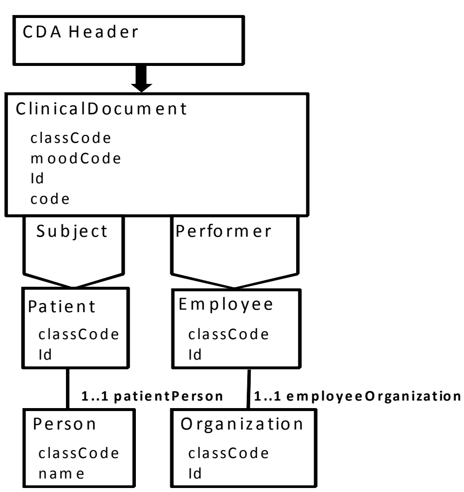


Figure 2. A simplified RMIM of CDA Header.

Note that HL7 uses its own representation of UML in RMIM diagrams: each class has its own color and shape to represent the stereotypes of these classes, and they only connect in certain ways.

The entry point of this diagram (CDA Header) is ClinicalDocument, which is specialization of the RIM class Act. Classes Patient and Employee are specializations (subclasses) of the RIM class Role. Person and Organization are specializations of the RIM class Entity. Subject and Performer are specializations of the association class Participation. Each specialization inherits all of the properties (attributes) of the generalization. For example, the class Patient is a specialization of Role with the addition of the optional attribute veryImportantPersonCode.

D. The Structure of a CCD document

Each CCD document have one primary purpose (which is the reason for the generation of the document), such as patient admission, transfer, or inpatient discharge. Further, each CCD document, as well all HL7 CDA documents, is comprised of the Header and the Body.

The sections that can appear in the Head and in the Body in a CCD document are presented in Fig. 3.

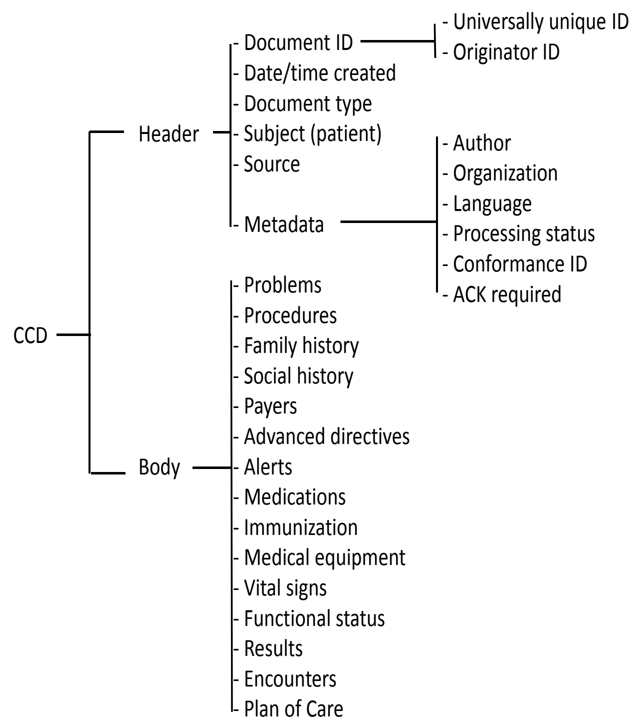


Figure 3. The parts of the CCD document.

A CCD document that includes a header and the Medications section from the Body is presented in Fig. 4. The content of the document is derived from the scenario presented in Section II, i.e., the document would be inserted in Nancy Taylor's PHR, which is based on the CCD-standard.

In a PHR the CCD documents are usually organized into hierarchical structures that simplify the search of documents, e.g., grouping together the documents by episode, clinical specialty or time period. Yet each clinical document is stored as a stand-alone artifact, meaning that each document is complete and whole in itself.

```
<CCDfile>
  <DocumentID>DOC_123</DocumentID>
  <Patient>
    <PatientID>AB-12345></PatientID>
    <PatientName>Nancy Taylor></PatientName>
  </Patient>
  <Medications>
    <Medication>
      <MedicationID>Medication.567</MedicationID>
      <DateTime>
        <ExactDateTime>
            2012-03-01TO12:00
        </ExactDateTime>
      </DateTime>
      <Source>
        <Actor>
          <ActorID>Pharmacy of  Health</ActorID>
          <ActorRole>Pharmacy</ActorRole>
        </Actor>
      </Source>
      <Description>
        <Text>Two tablets twice a day</Text>
      </Description>
      <Product>
        <ProductName>Aciclovir </ProductName>
        <BrandName>Zovirax</BrandName>
      </Product>
      <Strenght>
        <Value>400</Value>
        <Unit>milligram</Unit>
      </Strenght>
      <Quantity>
        <Value>40</Value>
        <Unit>Tabs</Unit>
      </Quantity>
    </Medication>
  </Medications>
</CCDfile>
```

Figure 4. A simplified example of a CCD document.

VI. Designing a PHIS

A. *Personal Health Information System PHIS*

Traditional PHRs fail in supporting information therapy, which is a key component in patient-centered care. Further, our studies have indicated that user-friendly health system should support the functionalities of many traditional e-health tools such as PHRs, remote patient monitors, health-oriented blogs, and health-oriented information servers. Further, by gathering these functionalities into one system we can achieve synergy, i.e., achieve functionalities that would not be obtainable by any of the e-health tools independently.

In gathering functionalities into a PHIS we have adapted the ideas of knowledge centric organizations to i.e., we have revolved the e-health tools around a health oriented knowledge base. So, all the e-heath tools share patient's health data. Further, by exploiting the characteristics provided by cloud computing [37] we can easily ensure the interoperation of patient's healthcare team: accessing the advanced PHR requires only internet connection.

A. *PHIS Ontology*

The architecture of the PHIS and its connections in the cloud are presented in Fig. 5. As the figure illustrates patient and the members of his or her healthcare team access the PHIS-server through the personalized health portal. It is a site on WWW that provides personalized capabilities for its users and links to other relevant servers.

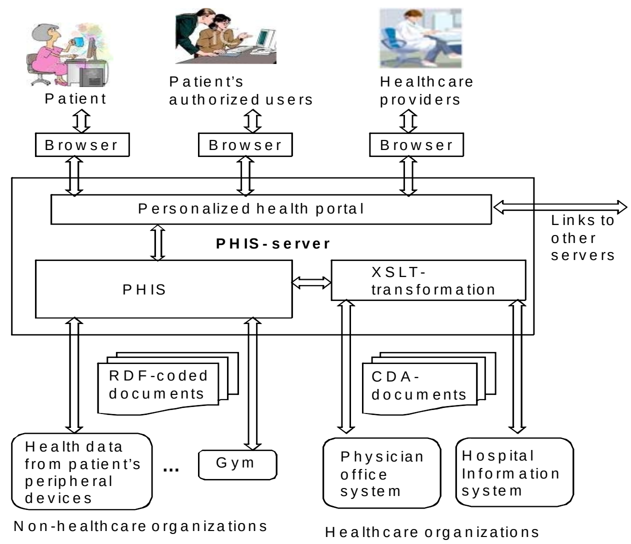


Figure 5. The component of the PHIS-server and its external connections.

In designing the PHIS we have followed the idea of knowledge oriented organizations [13], where the key idea is to revolve all applications around a shared ontology (stored in a knowledge base), which we call *PHIS-ontology*. It is developed by integrating the ontologies of the e-health tools supported by the PHIS. For now we have integrated the ontologies of the Blog manager, Information therapy (Ix) manager, Remote manager, and PHR manager. Such an internal architecture of the PHIS is presented in Fig. 6.

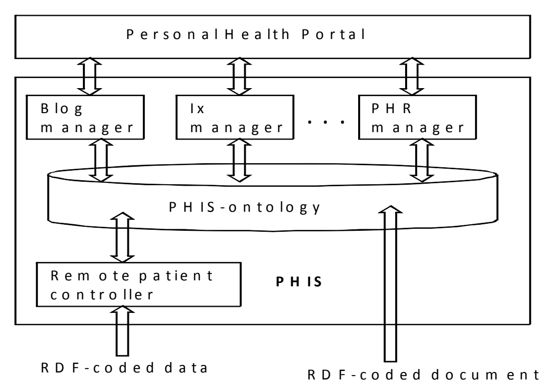


Figure 6. The e-heath tools accessing the PHIS-ontology.

Fig. 7 illustrates the idea of the knowledge base and the case where PHIS-ontology is developed by integrating the Blog-ontology, Ix-ontology, PHR-ontology and RM-ontology (Remote Monitoring ontology). In the figure ellipses represent OWL´s classes, rectangles represent OWL's data properties and the lines between ellipses represent OWL's object properties. Accordingly class A is shared by all the four ontologies.

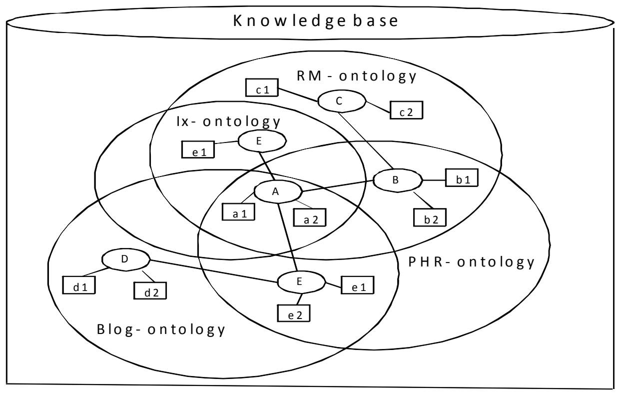


Figure 7. PHIS-ontology.

In order to illustrate shared classes, A could be class Disease, B class Patient, and C class Informal_entity. Further assume that object property A-B is suffer_from, object property A-E is deals, data property b1 is patient_name, and data property e1 is a url. In such as setting we could specify by RDF (Resource Description Framework) that John Smith suffers from diabetes and the educational material dealing diabetes is stored in a specific url.

A portion of the PHIS-ontology is graphically presented in Fig. 8. In this graphical representation ellipses represent classes and subclasses, and rectangles represent data and object properties.

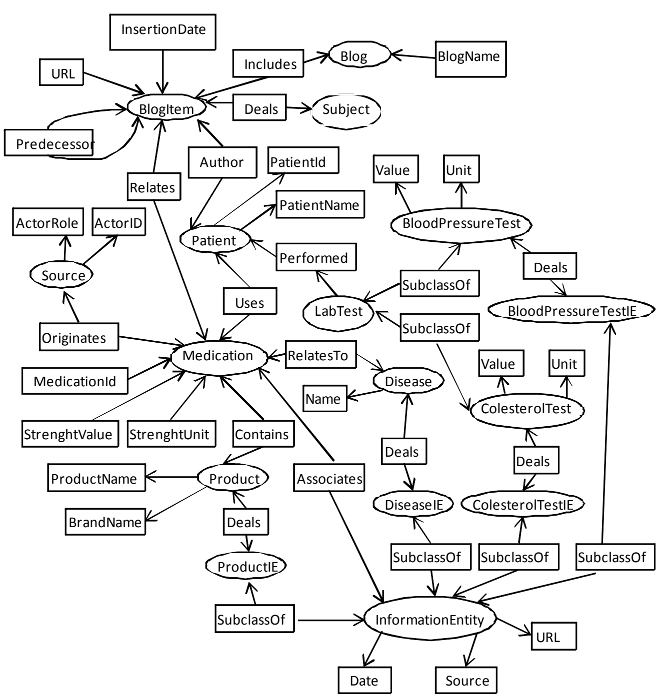


Figure 8. A portion of the PHIS-ontology.

A portion of the graphical ontology of Fig. 8 is presented in OWL in Fig. 9.

```
<rdf:RDF
        xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-nsl#
        xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#
        xmlns:owl=http://www.w3.org/2002/07/owl#>

        <owl:Ontology rdf:about="”PHA/>
         <owl:Class rdf:ID=“Blog/”>
         <owl:Class rdf:ID=“BlogItem/”>
        <owl:Class rdf:ID=“Patient/”>
        <owl:Class rdf:ID=“Medication/”>
        <owl:Class rdf:ID=“Source/”>
        <owl:Class rdf:ID=“Product/”>
        <owl:Class rdf:ID=“LabTest/”>
        <owl:Class rdf:ID=“BloodPressureTest”>
                <rdfs:subClassOf rdf:resource=“#LabTest”/>
        </owl:Class>
        <owl:Class rdf:ID=“ColesterolTest”>
                <rdfs:subClassOf rdf:resource=“#LabTest”/>
        </owl:Class>

        <owl:ObjectProperty rdf:ID=“Relates”>
                <rdfs:domain rdf:resource="#BlogItem”/>
                <rdfs:range rdf:resource="#Medication”/>
        </owl:ObjectProperty>

        <owl:ObjectProperty rdf:ID=“Uses”>
                <rdfs:domain rdf:resource="#Patient”/>
                <rdfs:range rdf:resource="#Medication”/>
        </owl:ObjectProperty>
                .
                .
                .
</rdf:RDF>
```

Figure 9. A portion of the PHIS-ontology in OWL.

In order to understand the relationship of XML, OWL and RDF note that XML (Extensible Mark-up Language) is just a meta language for defining markup languages. By a meta language we refer to a language used to make statements about statements in another language, which is called the object language. Accordingly RDF [38] and OWL [39] are object languages. Instead, XML says nothing about the semantics of the used tags. It just provides a means for structuring documents. Due to the lack of semantics we do not use XML for representing PHIS-ontology but instead we use ontology languages RDF and OWL.

B. Transforming HL7 CDA into OWL

Although the semantics of all CDA documents is traceable through a RMIM back to the RIM, we neither can use a RMIM nor the RIM in formulating queries as there are no query languages specified for the information model used in the RMIM and RIM schemas. For this reason we transform the RMIM of the CDA header into OWL.

Transforming a RMIM diagram into OWL is straightforward in the sense that both models are object-oriented although the notation used in RMIM diagrams slightly differs from the traditional UML notation. Yet their basic modeling primitives are the same, namely classes, subclasses, properties and values. The classes are also connected in a similar way through properties.

In order to illustrate the transformation of RMIM diagram into OWL we have presented the RMIM diagram of Figure 4 in OWL in Fig. 10.

Classes, subclasses, data properties and object properties are modeling primitives in OWL [40]. Object properties relate objects to other objects while datatype properties relate objects to datatype values [13]. For example, Performer is an object property. Its domain is clinicalDocument and range is employee. Note that, in Fig. 10 we have omitted most datatype properties. The only datatype property presented in the figure is "code". Its domain is clinicalDocumernt and its range is xsd:string, i.e., string in "XML-terminology".

```
<rdf:RDF
 xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-nsl#
 xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#
 xmlns:owl=http://www.w3.org/2002/07/owl#>
 xmlns:xsd =http://www.w3.org/2001/xml-schemal#>
 <owl:Ontology rdf:about="registryOntology"/>
 <owl:Class rdf:ID="act/">
 <owl:Class rdf:ID="role/">
 <owl:Class rdf:ID="entity/">
<owl:Class rdf:ID="participation/">
 <owl:Class rdf:ID="clinicalDocument">
      <rdfs: subClassOf rdf:: resource "#act"/>
 </owl : class>
 <owl:Class rdf:ID="patient">
      <rdfs: subClassOf rdf:: resource "#role"/>
 </owl : class>
 <owl:Class rdf:ID="employee">
      <rdfs: subClassOf rdf:: resource "#role"/>
 </owl : class>
 <owl:Class rdf:ID="person">
       <rdfs: subClassOf rdf:: resource "#entity"/>
 </owl : class>
 <owl:Class rdf:ID="organization">
       <rdfs: subClassOf rdf:: resource "#entity"/>
 </owl : class>
 <owl:ObjectProperty rdf:ID="subject">
         <rdfs:domain rdf:resource="#clinicalDocument"/>
         <rdfs:range rdf:resource="#patient"/>
 </owl:ObjectProperty>
 <owl:ObjectProperty rdf:ID="patientPerson">
         <rdfs:domain rdf:resource="#patient"/>
         <rdfs:range rdf:resource="#person"/>
 </owl:ObjectProperty>
 <owl:ObjectProperty rdf:ID="performer">
         <rdfs:domain rdf:resource="#clinicalDocument"/>
         <rdfs:range rdf:resource="#employee"/>
 </owl:ObjectProperty>
 <owl:ObjectProperty rdf:ID="employeeOrganization">
         <rdfs:domain rdf:resource="#employee"/>
         <rdfs:range rdf:resource="#organization"/>
 </owl:ObjectProperty>
 <owl:DatatypeProperty rdf:ID="code">
          <rdfs:domain rdf:resource="#lclinicalDocument"/>
          <rdfs:range rdf:resource="&xsd;string"/>
 </owl:DatatypeProperty>
        .
        .
        .
</rdf:RDF>
```

Figure 10. A part of CDA Header in OWL.

VII. CONSULTATION SERVER

A. The Architecture of the Consultation Server

In designing the internal architecture of the Consultation Server we have followed the idea of knowledge oriented organizations, where the key idea is to revolve all applications around a shared ontology. As illustrated in Fig. 11, in our solution all the applications of the Consultation Server are revolved around the Consultation Ontology.

The users access the Consultation Ontology through the Consultation Portal, which provides connections to the relevant cloud applications. The applications are based on the use cases of the various user groups. For example, Submit Consultation Request and Pick-up a Consultation Request are two typical applications developed for physicians. These applications interoperate through accessing the same data items included in the Consultation Ontology.

Note that although the Consultation ontology is specified in QWL and queried by SPARQL [41] (i.e., by a RDF-based query language), it is stored for the efficiency reasons in a relational database [42].

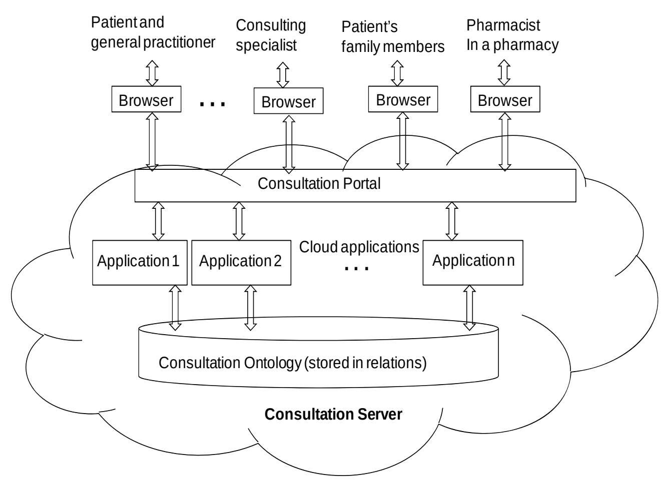


Figure 11. The components of the cloud-based Consultation Server.

B. The Structure of the Consultation Ontology

A portion of the Consultation ontology is graphically presented in Fig. 12. In this graphical representation ellipses represent classes and rectangles represent data type and object properties. *Object properties* relate objects to other objects while *data type properties* relate objects to datatype values. Classes, data type properties and object properties are modelling primitives in OWL.

Note that in Fig. 12 we have presented only a few of objects' datatype properties. For example, in the figure we have omitted most of the datatype properties from the classes Physician and Organization. Instead all the datatype properties of the class Consultation request are presented in the figure. The class Speciality and its datatype property Class represent a taxonomy. That is, except for the root node there is a link from each class instance to its parent.

The idea behind this taxonomy is that the symptoms and the specialities of the physicians are specified by the same vocabulary. This feature simplifies the matching of consultation requests and physicians' specialities.

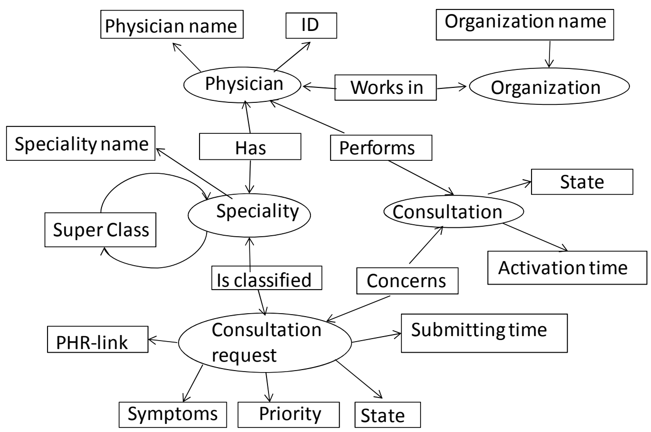


Figure 12. Graphical presentation of a portion of the Consultation Ontology.

The Consultation Ontology enables a variety of useful queries for physicians such as the followings:

- Is there any pending consultation request having classification "Skin diseases"?
- Is there (in the affinity domain) any physicians having speciality in diabetes?
- Is there any consultation request matching with my speciality?
- Is there any pending consultation request that is submitted by a physician of the University Hospital?
- Is there any consultation request that has been pending over ten minutes?
- Give me the names and specialties of the consulting physicians that work in a specific affinity domain.

Note that the Consultation Server can support more than one telemedicine affinity domain. Further, as most OWL ontologies, such as the Consultation Ontology, are usually stored in relational database systems, it is also possible to use the triggering mechanism of the SQL [43] in automating the management of the consultation requests. For example, a request can be automatically allocated to a physician having required speciality and having no ongoing consultation.

C. *Cloud Computing and SaaS*

Cloud computing is an appropriate choice for telemedicine consultation management as it allows organizations to use applications without installation. Moreover, in most cases cloud-based solutions reduces the cost of acquiring, delivering, and maintaining computing power [37]. However, in our vision the main goal behind cloud computing is to achieve synergy through controlled sharing of data.

In particular the Saas (Software as a Service) [44] model of cloud computing is appropriate for the Consultation Server. It is a software delivery model in which applications are hosted by service provider and made available to customers over the Internet. It provides access to software and its functions remotely as a Web-based service.

Further, there are various architectural ways for implementing the SaaS model. For example, in the case where the Consultation Server serves more than one telemedicine affinity domain we could use the following software architectures:

1. Each telemedicine affinity domain has a customized version of the Competence Server that runs as its own instance.
2. Many telemedicine affinity domains use separate instances of the same application code.
3. A single program instance serves all telemedicine affinity domains.

In our designed version the Consultation Server supports only one clinical affinity domain and so only one version and one instance is required.

VIII. TELEMEDICINE ECOSYSTEM

To succeed e-health systems should not be considered just as a technical infrastructure but rather as ecosystems having many interconnected parts [45].

So far we have considered the technical infrastructure and the services of our designed telemedicine oriented cloud-based ecosystem. The other key parts of the ecosystem are ethical aspects, governance regulations, financing and stakeholders. For now, we shortly consider what kind of new alternatives the introduction of cloud-based solutions gives for these parts of the ecosystem.

A. *Ethical Aspects in Telemedicine Resource Alloation*

Resource management is the efficient and effective deployment of an organization's resources when they are needed. Such resources may include financial resources, inventory, human skills, as well consultation resources in telemedicine.

Resource allocation is a part of resource management. It is used to assign the available resources in an economic way. In process management, resource allocation is the scheduling of activities and the resources required by those activities while taking into consideration both the resource availability and the process time.

Although the resource allocation issues in healthcare sector are economic by nature, they inherently raise issues relating to ethics. According to the principles of medical ethics physicians should merit the confidence of the patients entrusted to their care, rendering to each a full measure of service and devotion. However, often scarcity of resources makes it difficult or impossible to provide the full measure of service and devotion.

When the conditions of scarcity occur, we have to decide what considerations should guide decisions for tradeoffs in a fair and compassionate manner. Further, to carry out the decisions the resource needs or requests must be prioritized, and finally the competitive request must be treated according to their priorities.

B. Governance Regulations

E-health application that maintains patients' health documentation must adhere to national legislated policies and regulations, which concerns privacy and security issues. One problem is that in many countries that are just beginning to investigate on e-health application do not yet have enough mature legislation with respect to e-health. Thereby national governments have an important role in promoting the development of appropriate legislation concerning e-health.

In our developed telemedicine ecosystem patient's health documentation is not stored in the national archives but rather on PHR system, where the patient's health documentation is owned by the patient and only used by the physicians, patient's family members and healthcare personnel authored by the patient. As a result, patients' health documentation is not under the control of national healthcare authorities, and thus is not so tightly dependent on whether there is advanced legislation for e-health.

C. Financing Cloud-Based E-Health Ecosystems

In designing an e-health ecosystem it is important to ensure that appropriate funding is in place for its implementation and operation. Financing can come from a variety of sources, such as government or public-private partnerships.

Financing PHRs is not an actual problem as there are many freely available web-based PHRs available. Further there are many freely available PHRs for specific communities, e.g., for employees of specific organization, customers of a specific insurance company or for the customers of a specific healthcare provider.

D. Stakeholders of Global Ecosystem

In designing the implementation of an e-health ecosystem, it is of prime importance to involve in preparation all the key stakeholders, such as governments, public and private healthcare providers, patients, as well as patient advocacy groups [45].

As our proposed ecosystem is not nationwide but rather "internet-wide", the system itself as well as its stakeholders may span over many countries. For example, governments and healthcare providers from a variety of countries may be involved in the ecosystem, and each stakeholder has different objectives and motivations for participating in the ecosystem.

IX. CONCLUSION AND FUTURE WORK

The goal of our work has been to show that cloud-based global telemedicine ecosystems that support patient centered care can be implemented from technology point of view. Yet there are many problems that may jeopardize the success of such ecosystems. In particular, the introduction of new technology requires training: the incorrect usage of a new telemedicine technology, due to lack of proper training, may ruin the whole ecosystem. In addition, a consequence of introducing a new telemedicine practice is that it significantly changes the daily duties of healthcare personnel, and the role of patient and patient's family members. Therefore one challenging aspect is the changing the mind-set of the involved healthcare personnel.

The introduction of a new technology in telemedicine is also an investment. It includes a variety of costs including software, hardware and training costs. Introducing and training the staff on new technology is a notable investment, and hence many organizations like to cut on this cost as much as possible.

In order to minimize the changes that the introduction of the system will cause on healthcare personnel, in our future work we will investigate how telemedicine consultation can be linked into physicians' day-to-day work patterns. In particular, we will consider how the functionalities of the Consultation Portal can be integrated with other e-health tools used by the physicians.

REFERENCES

[1] J. Puustjärvi and L. Puustjärvi, "Towards Patient-Centered Telemedicine," in Proc. of the International Conference on eHealth, Telemedicine, and Social Medicinee (eTELEMED 2014), pp. 88-93, 2014.

[2] D. M. Angaran, "Telemedicine and Telepharmacy: current status and future implications," American Journal of Health System Pharmacy, vol. 56, pp. 1405-1426, 1999.

[3] W. J. Blyth, "Telecommunications, Concepts, Development, and Management," Second Edition, Glencoe/McCgraw-Hill Company, pp. 280-282, 1990.

[4] G. Kontaxakis et al., "Integrated telemedicine applications and services for oncological positron emission tomography," Oncology Reports, vol. 15, pp. 1091–1100, 2006.

[5] A. Bauman, H. Fardy, and H. Harris, "Getting it right; why bother with patient centred care?," Medical Journal of Australia, vol. 179, no. 5, pp. 253-256, 2003.

[6] K. Wiedenmayer, R. Summers, C. Mackie, A. Gous, M. Everard., and D. Tromp, "Developing pharmacy practice," World Health Organization and International Pharmaceutical Federation, 2006.

[7] M. S. Raisinghan and E. Young, "Personal health records: key adoption issues and implications for management," International Journal of Electronic Healthcare, vol. 4, no.1, pp. 67-77, 2008.

[8] J. Puustjärvi and L. Puustjärvi, "Providing Relevant Health Information to Patient-Centered Healthcare," in Proc. of the 12th IEEE Conference on e-Health Networking, Applications and Services, pp. 215-220, 2010.

[9] P. Little, H. Everitt, and I. Williamson, "Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations," British Medical Journal, pp. 908-911, 2001.

[10] D. Lewis, G. Eysenbach. R. Kukafka, P. Stavri, and H. Jimison, "Consumer health informatics: informing consumers and improving health care," New York: Springer, 2005.

[11] W. Tuil, A. Hoopen, D. Braat, P. Vries Robbe, and J. Kremer, "Patient-centred care: using online personal medical records in IVF practice," Hum. Reprod., November 1, vol. 21, no. 11, pp. 2955 – 2959, 2006.

[12] J. Puustjärvi and L. Puustjärvi, "Moving from Remote Patient Monitors to Cloud-Based Personal Health Information Systems: a Way to Practicing Patient-Centered Chronic Care Model," in Proc. of the International Conference on Health Informatics (HEALTHINF 2012), pp. 37-45, 2012.

[13] M. Daconta, L. Obrst, and K. Smith, "The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management," John Wiley & Sons. 2003

[14] A. Dogac, B. Gokce, T. Aden, and T. Laleci, “Eichelberg, M.. “Enhancing IHE XDS for Federated Clinical Affinity Domain Support,” IEEE Transactions on Information Technology in Biomedicine, vol. 11, no. 2, 2007.

[15] R. Gillespie, D. Florin, and S. Gillam, “How is patient-centred care understood by the clinical, managerial and lay stakeholders responsible for promoting this agenda?,” Health Expectations, vol. 7, no. 2, pp. 142-148, 2004.

[16] I. Mil. J. Schulz, and M. Tromp, “Pharmaceutical care, European developments in concepts, implementation, teaching, and research: a review,” Pharm World Sci, vol. 26, no. 6, pp. 303–11, 2004.

[17] C. D. Hepler, “Clinical pharmacy, pharmaceutical care, and the quality of drug therapy”, Pharmacotherapy, vol. 24, no.11, pp. 1491–98. 2004.

[18] K. Fiandt, “The Chronic Care Model: Description and Application for Practice”, Available at: http://www.medscape.com/viewarticle/549040

Last accessed 2014-11-10.

[19] C. Boult and L. Karm, “Improving Chronic Care: The Guided Care” Model”, Available at: http://www.guidedcare.org/pdf/Guided%20Care%20model_Permanente%20Journal_Winter%202008.pdf

[20] M. Mettle and D. Kemper, “Information Therapy: Health Education One Person at a Time,” Health Promotion Practice 2004, vol. 4, no. 3, pp. 214-217, 2004.

[21] L. Trevena, H. Davey, A. Barratt, P. Butow, and P. Caldwell, “A systematic review on communicating with patients about evidence,” Journal of Evaluation in Clinical Practice, vol. 12, no. 1, pp. 13-23, 2006.

[22] L. Butcher, “What the devil is information therapy?,” Managed Care June 2007. Available at: http://www.managedcaremag.com/archives/0706/0706.infotherapy.html

Last accessed 2014-11-10.

[23] J. Puustjärvi and L. Puustjärvi, “Exploiting Cloud-Based Personal Health Information Systems in Practicing Patient Centered Care Model,” Communications in Computer and Information Science, vol. 273, 2013.

[24] C. Angst, R., Agarwal, and J. Downing, “An empirical examination of the importance of defining the PHR for research and for practice”, Proceedings of the 41st Annual Hawaii International Conference on System Sciences, 2008.

[25] J. Habib, “EHRs, meaningful use, and a model EMR. Drug Benefit Trends”, May 2010;22(4):99-101. Available at: http://dbt.consultantlive.com/display/article/1145628/1581538

Last accessed 2014-11-10.

[26] D. Garets, and M. Davis, “Electronic Medical Records vs. Electronic Health Records:Yes, There Is a Difference,” Available at: http://www.himssanalytics.org/docs/wp_emr_ehr.pdf

Last accessed 2014-11-10.

[27] H., Fraser, P., Biondich, D., Moodley, S., Choi, B., Mamlin, and P. Szolovits, “Implementing electronic medical record systems in developing countries,” Informatics in Primary Care, pp. 83-95, 2005. Available at: http://groups.csail.mit.edu/medg/ftp/psz/EMR-design-paper.pdf

Last accessed 2014-11-10.

[28] T. Benson, “Principles of Health Interoperability HL7 and SNOMED,” Springer, 2010.

[29] “Continuity of Care Record (CCR) Standard,” Available at: http://www.ccrstandard.com/

Last accessed 2014-11-10.

[30] CCD, “What Is the HL7 Continuity of Care Document?,” Available at: http://www.neotool.com/blog/2007/02/15/what-is-hl7-continuity-of-care-document/

Last accessed 2014-11-10.

[31] J. Puustjärvi and L. Puustjärvi, ”Integrating and Querying Patient’s Medical Data”, ACEEE International Journal on Information Technology, vol 3, no. 3, pp. 39-49, 2013.

[32] HIMMS; “Healthcare Information and Management Systems Society,” Available at: http://himms.org/

Last accessed 2014-11-10.

[33] “Healthcare Information Technology Standards Pane,l” Available at: http://hitsp.org/

Last accessed 2014-11-10.

[34] R. Dolin, L. Alschuler, C. Beerb, P. Biron, S. Boyer, E. Essin, and T. Kimber, ” The HL7 Clinical Document Architecture”, J. Am Med Inform Assoc, vol. 8, no. 6, pp. 552-569, 2001.

[35] K. Boone, K., “The CDA Book,” Springer, 2011.

[36] R. Spronk, “CDA around the world: archive of CDA instance examples,” Availble at: http://www.ringholm.com/download/CDA_R2_examples.zip

Last accessed 2014-11-10.

[37] D. Chappel, “A Short Introduction to Cloud Computing: An Enterprise-OrientedView,” Available at: http://digitalebookden.com/a-short-introduction-to-cloud-platforms.html

Last accessed 2014-11-10.

[38] “RDF – Resource Description Language,” Available at: http://www.w3.org/RDF/

Last accessed 2014-11-10.

[39] “OWL – WEB OntologyLanguage,” Available at: http://www.w3.org/TR/owl-features/

Last accessed 2013-5-10.

[40] G. Antoniou and F. Harmelen, “A semantic web primer,” The MIT Press, 2004.

[41] “SPARQL Query Language for RDF,” Available at: http://www.w3.org/TR/rdf-sparql-query/

Last accessed 2014-11-10.

[42] “Introduction to Relational Databases,” Available at:

http://www.databasejournal.com/sqletc/article.php/1469521/Introduction-to-Relational-Databases.htm

Last accessed 2014-11-10.

[43] “SQL tutorial,” Available at: http://www.w3schools.com/sql/default.asp

Last accessed 2014-11-10.

[44] “Software as a service,” Available at: http://en.wikipedia.org/wiki/Software_as_a_service

Last accessed 2014-11-10.

[45] R. Shehadi, W., Tohme, J., Bitar, and S., Kutty, “Anatomy of an E-Health Ecosyste,” Available at: http://www.booz.com/media/__tracked/media/uploads/BoozCo-Anatomy-of-E-Heatlh-Ecosystem.pdf

Last accessed 2014-11-10.

Technology-Enhanced Language Learning: A Case Study of a Global Classroom in Second Life

Kevin Oh, Natalie Nussli
School of Education
University of San Francisco, United States
{koh2@usfca.edu, ncnussli@usfca.edu}

Abstract—**This study reports on a case study about conversation practice in Second Life, a three-dimensional (3D) virtual environment, between 12 English language learners and 18 special education teachers. The purpose of the study was to examine participants' perceptions of the usability of virtual worlds as a language learning platform and to identify the unique skills required for successful teaching in a 3D environment. Specifically, the anxiety-reducing features of the virtual, anonymous environment in which oral language skills can be practiced through avatars were investigated. Extensive challenges inherent to the lesson design and audio quality were identified in this study. Guidelines on how to address these challenges as well as ideas for alternative, more effective learning designs, such as the discussion group format, are outlined. The study resulted in nine suggestions for virtual conversation practice in 3D environments. These findings will be relevant to other language instructors who plan to use Second Life or a similar virtual world for oral fluency enhancement in collaborative language practice.**

Keywords-Second Life, virtual worlds, foreign language acquisition, oral fluency, collaboration, conversation partners, anxiety, oral production, teacher education

I. Introduction

This article reports on an investigation of the usability of Second Life, a highly compelling visual and immersive virtual world, to improve English language learners' (ELLs) oral fluency by engaging them in purposeful interaction with English native speakers [1]. Second Life (http://secondlife.com) is a popular example of a 3D interactive environment.

Oral fluency is a complex, difficult-to-measure construct. It has been defined as the ability to produce language without undue pauses or hesitations [2]. Four types have been identified [3]. These include the ability to (i) fill time with talk, (ii) produce logical and semantically dense language, (iii) have appropriate things to say in a wide range of contexts, and (iv) express oneself in a creative and imaginative way, using a wide variety of alternative linguistic devices and choosing the one that is most situation appropriate.

The current study is situated in a 3D immersive virtual world. Virtual reality has been defined as "a computer-generated display that allows or compels the user (or users) to have a sense of being present in an environment other than the one they are actually in, and to interact with that environment" [4]. Educational platforms can be found in the areas of astronomy, medicine, music, literature, biology, history, mathematics, forensic science, ecology, and tourism, to name a few. The Abyss Observatory [5], for example, is a museum and aquarium of earth and oceanic science, marine life, and underwater technology. Sploland [6] is a virtual science museum with more than 100 hands-on exhibits.

Second Life provides a powerful platform for situational language practice. Recent studies [7]-[11] indicate that a virtual world can assist language learners by extending the traditional language classroom. It provides an environment for stress-free, one-on-one oral practice through activities, such as role-playing, discussions, presentations, debates, games, and simulations. English language learners, in particular Chinese students, tend to be apprehensive of spoken communication for fear of being negatively evaluated [12]. The results of a study by Wehner et al. [13] suggest that virtual worlds may help to reduce student anxiety and increase their motivation to learn a foreign language. Virtual worlds could be an effective way to help Chinese ELLs overcome these fears and inhibitions.

In a group of 313 Chinese ELLs, Mak [14] identified five factors leading to speaking-in class anxiety. These include: "speech anxiety and fear of negative evaluation; discomfort when speaking with native speakers; negative attitudes towards the English classroom; negative self-evaluation; and fear of failing the class/consequences of personal failure" (p. 202). Appropriate wait-time has particularly strong implications for Chinese-speaking ELLs because it helps to reduce their anxiety. Chinese students appreciate a longer wait-time to speak up and respond than European language learners "because 'group unity' and 'face' are important elements of their culture" [14]. These two constructs may be at risk when the students are under pressure. Given the linguistic and cognitive demands on oral language production, learners should be given ample preparation time prior to giving a speech or presentation in front of an audience in order to reduce their anxiety. Large amounts of teacher talk and limited student talk obstruct students' speech in the classroom [15] [16]. While Chinese ELLs may be willing to participate in interpersonal conversations, for example with one conversation partner at a time, they may be reluctant to speak English in class, that is, in front of an audience, for fear of negative evaluation [17].

The current study contributes to the research about language learning in virtual worlds by identifying ELLs' and special education (SPED) teachers' perceptions of Second Life as a language learning platform, the unique skills required for teaching in a 3D space, and potential challenges. Recent studies in the field of second and foreign language acquisition are reviewed in the next section, followed by the purpose of the present study. The sample, the six data collection instruments, the procedures, and data analysis are described in the method section. Results are organized around the instruments and are presented from the perspective of both the ELLs and the SPED teachers. The discussion section is organized around the research questions and concludes with suggestions for alternative learning designs. The conclusion summarizes the impact of the study and concludes with the study's limitations and suggestions for future research.

II. Literature Review

This review starts with a discussion of recent findings suggesting the superiority of 3D virtual environments over traditional learning environments for specific tasks. The second part of the review elaborates on virtual worlds research in the field of second language acquisition.

A. *Objective Measurement of Learning Outcomes*

Overall, despite an abundance of studies examining teachers' and students' perceptions of their virtual world experiences, only few studies have objectively measured learning outcomes. Chau et al. [18], for example, evaluated whether 3D virtual environments could facilitate students' learning outcomes. The task required students in an introductory management information system course to identify information security issues in a virtual office set-up in Second Life. The experimental group using Second Life gained higher test scores in the tasks than students in the control group who watched a video showing the exploration in the 3D virtual environment. The difference reached a statistically significant level ($p < 0.0001$).

Using a true experimental design with two randomized groups, Farra [19] investigated the longitudinal effects of virtual reality simulation on the learning outcomes and learning retention of disaster training with nursing students. The experimental group completed Web-based modules and a virtually simulated disaster experience, whereas the control group only completed the Web-based modules. The main effect of the virtual simulation was found to be significant ($p < .0001$). Although both groups showed similar improvement of scores following the teaching interventions on the first post assessment, significant differences were seen in the scores at the two-months follow-up, indicating that the virtual reality simulation demonstrated stability over time, whereas the non-simulation control group showed declining scores.

Another recent example demonstrating the superiority of virtual worlds to traditional environments for specific tasks comes from Moskalium, Bertram, and Cress [20] who compared three training conditions (virtual condition, standard condition, and control condition). The training was designed for police officers whose communication skills were being trained for collaboration between ground forces and a helicopter crew during an operation. In terms of knowledge transfer, the findings showed that the virtual training was significantly more efficient than in the standard condition and the control condition ($p = 0.02$). Also, virtual training was equally efficient as standard training regarding knowledge acquisition, indicating that virtual training is an effective tool for training complex collaborative tasks that cannot be fully trained for in reality due to excessive risks.

B. *Language Acquisition in Virtual Worlds*

Specifically in the field of language learning situated in virtual worlds, a review of the literature by Peterson [21] has produced broadly encouraging findings, such as the reduction of constraints on learning (e.g., inhibition) through the use of avatars, an enhanced sense of presence and copresence, increased access to diverse groups of interlocutors, and enhanced participation for target language use, provided that teachers take full advantage of the unique affordances of a virtual environment. At the same time, Peterson also identified a number of challenges, for example, that the management of the virtual world interface can cause technostress, which, in turn, may hamper participation.

Through interactions with target language speakers in Second Life, ELLs can benefit from immersive language practice; tasks for learners at all levels can be designed [22]. Second Life has been shown to offer an "interactive, immersive and content-rich virtual environment for input, interaction, task-based learning and output production" [22]. Virtual worlds offer a suitable platform for synchronous language interaction and the opportunity for conversation and collaboration between language learners and native speakers of the target language.

Wang et al. [23] investigated effective and practical ways to integrate Second Life into an English as a Foreign Language (EFL) program for students in Chinese universities, in which Chinese students conversed with American native speakers. Their findings suggested that virtual worlds not only extend the language learning classroom but that they also offer practice opportunities that a traditional instructional setting does not allow. Second Life can complement face-to-face learning, for example, by conducting interviews, surveys, literary tours, or practicing giving directions in Second Life. Many study participants reported having a positive attitude towards using Second Life for language learning and perceived the collaboration with their American conversation partners as interesting, effective, and helpful in improving their English skills. Meaningful and authentic interactions with the American students were identified as key motivational factors. Once the Chinese students were immersed in Second Life, they demonstrated more active and sustained speaking events, which helped them to improve their communicative abilities. Their positive experiences were, however, tarnished by technical problems, including poor audio quality (echoing and interrupted audio) and frozen screens to an extent that they seriously interfered with smooth communication and task completion in Second Life. Wang et al.'s [23] recommendations for facilitating language learning events in Second Life include: preparing students for task completion, setting a time limit for any given task, closely monitoring student language performance, encouraging post-task reflection, and providing feedback. Despite technical issues, the collaboration was found to enrich the cultural and communicative experience.

Similarly, Knutzen and Kennedy [24] reported on a partnership between ELLs in Hong Kong and student teachers enrolled in a Teaching English to Speakers of Other Languages (TESOL) program at a university in the U.S.. The two groups met in Second Life at a virtual American diner and communicated through text-chat and voice. Among the conditions that resulted in the most productive interactions was the use of voice communication to practice speaking and listening, as well as the use of separate sound parcels in the form of Cadillac diner booths to allow private conversations.

Support also comes from Wehner et al. [13] who investigated the relationships between motivation, virtual worlds, and foreign language acquisition. One section of a Spanish course used Second Life as part of its instruction, while the other section participated in the traditional curriculum. Overall, the group using Second Life consistently reported more positive feelings in all areas of motivation and lower levels of anxiety than the traditional group.

The findings of Ishizuka and Akama [25] were also supportive of the potential of Second Life for second language acquisition. Good scenarios and controlling learning environments based on second language acquisition theories have the potential to change language teaching and learning. Several attempts to use Second Life for language learning have been made in the past. A number of EU-funded, large-scale projects include the Access to Virtual and Action learning Live Online (Avalon) project, the Networked Interaction in Foreign Language Acquisition and Research (NIFLAR) project, and the Talk with Me project. These projects aim to facilitate cross-cultural language learning by taking advantage of virtual worlds to simulate communicative acts and provide information on learning models and practices using Second Life as a language learning platform [25].

The current study reports on the findings of an exploratory case study with two groups. International students enrolled in an ELL program and SPED teachers studying at the same university in the United States met in Second Life for conversation practice. The overarching question that framed this research was to identify the usability of Second Life for oral fluency enhancement. The study was guided by the following research questions:

1. What are the English language learners' and the special education teachers' perceptions of Second Life as a language-learning platform?
2. What are the unique skills that a teacher should have to teach in Second Life and similar 3D virtual worlds?
3. What types of problems associated with language instruction in Second Life were identified?

III. METHODS

A. Needs Assessment

Prior to the conversation practice workshop, an extensive needs assessment was conducted to identify the ELLs' performance gap in terms of oral fluency. It revealed that many Asian, particularly Chinese students, at this university were experiencing a performance gap between their actual oral proficiency in American English and the proficiency they needed to fully contribute to class discussions and be understood when they speak. Multiple class sessions of courses intended for international students were observed, a self-report oral proficiency survey was administered to those students, and their instructors were interviewed.

Fifty-eight percent of Asian students (i.e., of 80% of international students), were from China, which explains the student homogeneity in some English as a Second Language (ESL) classes. Some oral skills classes consisted of Chinese students exclusively. Class observations indicated that the Chinese students tended to speak in English only when necessary (in front of the class or when speaking with the teacher), otherwise conversing in Chinese.

Oral comprehension was found to be much less of a challenge for these students than oral production. It was determined that the main reason for the performance gap was a lack of interaction with English native speakers and a resulting lack of oral fluency skills. The lack of more teacher time to work one-on-one with each student, in addition to the lack of time and frequent feedback in situations close to students' lives seemed to reinforce students' low self-confidence and their constant use of the mother tongue to communicate with each other.

A lack of motivation was identified as another cause for the performance discrepancy. Because the community of Asian students at this university is very large, these students feel no social pressure to improve their oral skills. They have the tendency to lead insular lives within their communities, where they speak their native language and where English is not a necessity. If their grades are sufficiently high to continue their studies, they may have little motivation to work on their oral English skills although they may be aware of their inadequate oral fluency.

Finally, a lack of knowledge and skill was identified as another cause for the lack of oral fluency. Many Asian learners tend to focus their efforts on reading and writing, whereas they neglect their comprehension and production skills. Only 19% of the Chinese students in oral skills classes have spent at least one year in an English speaking country, compared to 59% of the culturally more heterogeneous, higher-level classes. The short duration of their time spent in an English speaking country (6 months or less) explains the need to further develop their oral production skills.

B. Purpose

The purpose of the conversation practice workshop was to improve learners' oral fluency and comprehensibility by engaging them in purposeful, extended interaction with English native speakers in Second Life.

C. Sample

Twelve ELL undergraduate students at a university in California were teamed up with 18 SPED teachers enrolled in a graduate course at the same university for the purposes of English conversation practice.

D. Data Collection

Six different instruments were used for data collection. The two groups, ELLs and SPED teachers, each received a different set of the following instruments: a preliminary survey, a mid-reflection, and a post-survey (see Table I).

Table I. Overview of Instruments

Instrument	ELLs	SPED Teachers
Preliminary survey	37 items Multiple choice; rating and open-ended questions	17 items Multiple choice, open-ended and rating questions
Mid-reflection	6 items Open-ended and rating questions	7 open-ended questions
Post-survey	33 items Multiple choice; open-ended and rating questions	5 open-ended questions

The preliminary survey was completed after watching a 5-minute video showing a tour of the National Oceanographic and Atmospheric Administration (NOAA) Second Life region and prior to the two virtual meetings for conversation practice. In the preliminary survey, the ELLs and the SPED teachers were asked about demographic information, their technology background, and their perception of the usability of Second Life as a language learning platform. In addition, the ELLs were asked to share their perceptions of their oral fluency in English, their attitude toward English native speakers, and their perception of their motivation and self-efficacy in learning and speaking English. Participants replied to the mid-reflection prompts after the first of two meetings in Second Life. The prompts were designed around the research questions and provided an opportunity to reflect on the usability (i.e., practical applicability) of virtual worlds for language learning. The post-survey offered an opportunity to reflect on the experiences after the second virtual meeting. Respondents were expected to be able to make informed decisions about the usability of Second Life for language learning after the two virtual meetings. Both researchers kept a researcher journal to take field notes. Although most studies investigating the potential of Second Life as a language learning platform reported positive findings, the practical issues that make Second Life-based language learning "conceptually applicable but difficult to conduct" [26] remain unsolved. The purpose of the researcher journal was to identify such issues and develop appropriate solutions.

E. Procedures

The two groups, ELLs and SPED teachers, were each introduced to Second Life by their instructors, respectively. The SPED teachers spent 30 minutes in Sploland as a class to experiment with hands-on activities and then teleported to Spaceport Alpha for a short fieldtrip. The purpose of these preliminary fieldtrips was to provide students with a chance to familiarize themselves with navigation and voice communication. The instructor was physically present in the computer laboratory together with the SPED teachers, whereas the second author joined the group in-world as a technical facilitator. The following week, a small group of ELLs met in Second Life with a small group of SPED teachers on EduNation. ELLs and SPED teachers met twice for one to two hours each. Sound checks were conducted before the meetings. All participants were required to use USB headphones. The meetings were organized by an external English instructor, who was commissioned to design and lead the workshop on account of her specialization in teaching ELLs in Second Life. Examples of the activities include: playing domino to increase vocabulary (see Fig. 1), a scavenger hunt requiring the oral description of interactive household items in the instructor's virtual house, a TV quiz show in a TV studio, conversation practice in a conference room (see Fig. 2), playing taboo, a grammar rummy, brainstorming ideas at a lounge (see Fig. 3), and a murder mystery in a castle.

Figure 1. Playing domino.

Figure 2. Conversation practice in Second Life.

Figure 3. Collaborative brainstorming.

F. *Data Analysis*

In this exploratory case study, the qualitative data, which emerged from the preliminary survey, the reflection, and the post-survey, have been used to gain insight into the personal reflections and perceptions of the participants. Emerging themes were identified through open coding and combined into recurring patterns. Quantitative data from the surveys inform the story and substantiate the qualitative information.

IV. Results

The results have been arranged chronologically.

A. *Preliminary Student Survey (ELLs)*

(1) Demographic information: The sample consisted of 12 ELLs between 18 and 21 years old. Most students' mother tongue was Chinese. The sample consisted of 8 women (67%) and 4 men (33%), mostly (87%) ranging between 18 and 21 years old. All, but two, students had Chinese as their mother tongue.

(2) Technology background: Twenty percent of the ELLs reported using technology for less than an hour each day, a slight majority (53%) between 2 and 5 hours, and 27% for at least 6 hours a day. On a 10-point rating scale, they self-reported their technological expertise at 7.33 (1=lowest, 10=highest). In terms of 3D virtual worlds, 76% reported having no or little experience.

(3) Oral proficiency: While these learners reported relatively good aural comprehension in English, they had difficulty understanding conversations among their English-native peers. Most, however, reported struggling to express what they wanted to say, rarely speaking up voluntarily in English, and lacking the confidence to do so.

(4) Intelligibility: All, but one, reported that they believed they could be understood mostly well. When asked about the reasons why people may have difficulty understanding them, 87% reported that it was due to a lack of vocabulary.

(5) Motivation: On average, respondents' motivation to learn English was 5.4 on a 7-point rating scale (1=weak, 7=strong). On average, their attitude toward English native speakers was 5.7 (1=unfavorable, 7=favorable). When asked to rate how worried they were about speaking English outside of class, the average rating was 2.4 (1=very little, 7=very much). Their attitude about their English course(s) was, on average, rated at 5.7 (1=negative, 7=positive).

(6) Anxiety: Most respondents (67%) reported that they were not nervous when they had to speak English to someone they just met, whereas 33% reported being somewhat nervous. Seventy-three percent agreed that they did not have to worry about losing face in Second Life because their conversation partner could not see their real face. In the same vein, 80% replied that the use of an avatar in Second Life made them feel more at ease because it helped them disguise themselves.

(7) Usability of Second Life: A clear majority stated that they perceived Second Life as easy (88%) and interesting (100%) to use. In terms of the usability of Second Life for language learning, 95% perceived it as useful. Everyone reported being interested in communicating with others in Second Life. While most students (80%) were looking forward to the Second Life meetings with English native speakers, 20% were undecided.

B. *Students' Mid-Reflection (ELLs)*

The mid-reflection was completed after the first virtual meeting with their native English-speaking partners, that is, the SPED teachers. All ELLs reported finding the virtual meeting with their English native speaking partners useful. They all appreciated the opportunity to practice speaking in a relaxed environment. They suggested that the virtual environment may have helped them to overcome shyness, to save face even when mistakes were made, and that they felt more comfortable speaking in an online setting than face-to-face. Being able to make friends, engaging in interesting interactions in a relaxing, game-like, and visually appealing environment were mentioned as critical factors.

The activities in Second Life, however, could have been more interesting, entertaining, and interactive. There were too many people at the same place at the same time (lack of functioning sound parcels) and it was hard to understand each other and the teacher. For example, due to poor sound quality and interference, instructions were unclear so that tasks could not be completed accurately. Also, Second Life was perceived as being too complicated. The ELLs would have liked to practice pronunciation and grammar, have more activities, and have separate sound parcels for private conversations. Despite these challenges, the average rating of the usability of Second Life for language learning on a scale from 1 (useless) to 10 (excellent) was still quite high at 7.7. Most students were looking forward to the second meeting.

C. *Students' Post-Survey (ELLs)*

In contrast to their answers in the pre-survey, no one reported difficulties in understanding their native-English speaking partners. No one reported difficulties in expressing their own thoughts and opinions in English. Everyone was confident that they were able to express the full nuance of their thoughts and opinions to varying degrees (very confident: 33%, quite confident: 50%, moderately confident: 17%). On average, respondents' motivation on a 7-point rating scale (1=weak, 7=strong) was unchanged at 5.4. The statements about being nervous when they had to speak English to someone they just met were almost unchanged, compared with the pre-survey. Their attitude toward English native speakers was almost unchanged at 5.8 (1=unfavorable, 7=favorable). When asked to rate how worried they were about speaking English outside of class, the average rating was almost unchanged at 2.1 on a 7-point rating scale (1=very little, 7=very much). The perceptions in terms of losing face remained unchanged compared with the pre-survey.

Again, the majority (58%) reported technical difficulties. Most students reported having sound issues, although their English instructor had conducted a sound check with them in an Adobe Connect classroom. Even when they did have sound, some voices could not be heard clearly. In terms of the Second Life viewer software, some students were frequently logged off and lost valuable time having to log in again or having to restart their computer. One student reported motion sickness due to the navigation in a 3D immersive space. As in the preliminary survey and the mid-reflection, all students, but one, agreed that the use of an avatar in Second Life made them feel more at ease due to the anonymity. All students found Second Life easy and interesting to use and confirmed its usability for language learning, although 83% stated that Second Life was not necessary for the type of language practice they had experienced. The same learning effect could have been achieved through other means with less hassle. Despite various challenges, everyone reported having enjoyed interacting through their avatar.

Benefits of using Second Life for speaking practice were identified as: communicating without seeing each other, more opportunities to meet native speakers, reduced nervousness, increased confidence, and overall more opportunities to speak than in real life. When asked what the external ELL instructor, who had conducted the activities, could have done differently, several students stated that they would have liked more opportunities to speak with their partners and more interesting activities. All students reported being satisfied with the experience and encouraged their own instructor to use Second Life as a language-learning platform.

D. *Preliminary SPED Teacher Survey*

(1) Demographic information: The convenience sample consisted of 18 (16 female and 2 male) special education teachers. All teachers were between 20 and 30 years old, except for two older participants.

(2) Technology background: In terms of virtual worlds experience, 61% reported having no experience, 22% reported having little experience, and 17% reported using virtual worlds occasionally. Seventy-one percent reported using technology for up to five hours a day, while 29% were using technology for more than five hours a day. When asked to rate their computer expertise on a 10-point rating scale (0=lowest, 10=highest), the average rating was 7.1, which is a little lower than the ELLs' self-reported expertise. Respondents were also asked to describe themselves in terms of Roger's technology adopter categories [27] (see Table II).

Table II. Technology Adopter Categories

Adopter Category	Responses
Innovator ("techies", guaranteed to adopt technology as a pedagogical tool)	22%
Early adopter ("visionaries", will adopt technology earlier than majority)	28%
Early majority ("pragmatists, will adopt technology as soon as majority of teachers does)	44%
Late majority ("skeptical", reluctant to adopt technology)	6%
Laggard (unlikely to adopt technology as a pedagogical tool)	0%

(3) Usability of Second Life: Based on the NOAA video they watched, the average rating of Second Life for education on a 10-point rating scale (1=useless, 10=extremely useful) was 6.28. Overall, their first impression of Second Life was that it had potential for education and seemed user-friendly. When asked what they hoped to gain from the two virtual meetings with their ELL partners, most said they hoped to identify the best techniques to work with ELLs and to see some of the educational applications of Second Life from a more active role, rather than just being a student in it.

(4) Unique affordances: The experiential value emerged as one of the key affordances of virtual worlds for educational purposes. "Very usable for education. It appears to be a very professional approach to virtual reality, allowing people to interact in a social way while maintaining the capability to have an emphasis on education." The immersiveness of the 3D environment emerged as another unique affordance.

> *I think it is a great idea to use Second Life for education. It immerses the student into what they are learning about. In the video it says users can actually experience how a tsunami looks and how why it happens. This learning experience is more fulfilling than just reading from a textbook.*

E. SPED Teachers' Mid-Reflection

Nine out of 18 SPED teachers submitted the mid-reflection. Their usability ratings went down quite markedly from 6.28 (preliminary survey, N=18) to 4.33 (mid-reflection N=9) on a 10-point rating scale (1=useless, 10=extremely useful). The ratings ranged from 1 to 10, which reflects the wide variety of attitudes. Due to the frustrations of the first meeting, nine participants failed to submit their mid-reflection, but it can be assumed that the mean rating might have been even lower than 4.33 if everyone had submitted their comments. Only one participant found the first meeting useful. She described that when she and her conversation partner met at the ELL instructor's virtual house, they had to describe household items. Being able to interact with the objects that the ELL student was trying to describe was perceived as useful.

The remaining eight respondents provided several reasons for not finding the meetings useful in terms of getting teaching experience. The set up (i.e., planned activities) failed to encourage discussion between the two groups. There were too many people and it was hard to talk because everyone else could be heard too. Although each team worked in a separate room of a house, the lack of (functioning) sound parcels resulted in the participants hearing everyone speak, which made it challenging to identify one's team members' voices. Separate sound parcels would also have satisfied the participants' desire for private communication in a less threatening environment without public exposure. The following statements illustrate that the potential loss of face was a concern not only among ELLs, but also among SPED teachers, "I felt uncomfortable trying to talk when I knew the whole group would hear me" and "I could imagine how hard it was for the other students who were not native English speakers to have to speak English in front of everyone".

More interaction and better time management would have been appreciated. The following statement describes the amount of spoken interaction between SPED teachers and ELLs that was distinctive of all sessions that the authors observed (eight sessions totaling approximately 10 hours for all teams combined).

> *I asked the student I was working with questions about herself to get to know her, and she wouldn't even respond to my questions even though I tried to rephrase what I was saying to her to help her understand. She would just stay quiet. Maybe it's because it was online, but actually working with the ELL student in person may have been more helpful. Since it was online and you can't see her face that could be why she didn't respond at all. She then only read the prompt to me when asked to by her professor and only asked for help on three words. I asked the student if she needed help understanding anything in the prompt, and she said no. That was as much time as we talked with each other out of the whole hour and a half.*

As evidenced by the following statement, the virtual presence did not seem to promote interaction, rather, it encumbered interaction.

> *The ELLs seemed very apprehensive to engage, even laughingly reverting to chats in their native languages when they perceived no one to engage with in English. The facilitators encouraged all to engage freely. However, without the prompts, neither group seemed to readily approach the other.*

The second prompt asked how useful the virtual meeting had been in terms of expanding their experience of using a virtual world. Only a minority found the first virtual meeting useful. A unique affordance, however, was identified as, "Being able to virtually meet with my ELL partner was nice, and I couldn't imagine doing the activity over the phone." The remaining comments were mostly negative. Respondents were disappointed because mostly they "just stood there", without having a purposeful role assigned. Due to the lack of interaction, they did not have the impression that the meeting had helped to improve the ELL partners' oral fluency. It was suggested that a tour would have encouraged interaction more effectively.

The third prompt inquired about features they liked about the interaction and their perceptions of the unique benefits of Second Life for this type of learning activity. While some respondents liked nothing about it, several respondents appreciated that there were no limitations by location or by physics. The potential to establish personal rapport in a virtual setting was emphasized.

> *We actually had a few minutes at the end of the activity to explore the boathouses' kitchen and we both agreed that it (the kitchen) was very nice and we chuckled about that. It was nice being able to connect to my buddy about something we both appreciated.*

The virtual environment was described as offering language immersion to ELLs and as a way to introduce teachers to novel technologies.

> *I realized the usefulness and the possibilities for offering an environment that may be less intimidating than face-to-face meeting for ELLs. I liked interacting exclusively with a group that had a facilitator who helped to orient the Cohort to the space, and who provided us with a shared set of expectations and goals. The unique benefits were that it not only provided ELLs with a means for immersion, it also gave new teachers access to technologies that we may not have engaged before.*

The following comment illustrates the importance of having a back-up plan in case of technical issues.

> *It's hard to say what I liked about this interaction because I only spent about 10 minutes working with my ELL student while 60 minutes were spent on tech glitches from other students and 20 minutes were based on directions from the other professor. However, overall I think the concept of using Second Life is a great idea because you get to integrate technology in the lesson and stay at home while interacting with another student. You still have chances to communicate and interact with them (such as seeing the written cue cards the ELL student was reading) while online.*

The fourth prompt specifically inquired about challenges the SPED teachers had experienced. Although some spaces offered separate sound parcels, they did not offer privacy:

You and your ELL partner had to go to a separate corner to work but other people would come and sit with you, this loud TV background noise kept interrupting the conversation, and other noises were distracting so it was hard to work with this student.

All eight sessions were heavily impeded by technology glitches. "Some people's headphones, microphones, and speakers were not working, which is what took a majority of the time," thereby minimizing the time that could actually be spent productively. "All of the technology glitches caused us to only really work on the activity for half an hour." Getting everyone on the same page and understanding the directions of the activity were also perceived as difficult. The inability to see a partner's facial expression, the lack of sound parcels, the difficulty to access note cards, programming errors, and the apparent group divide were other challenges. One respondent who did not have sound wrote, "Also, as an observer, I noticed each group's inclination to remain comfortably divided."

The last prompt asked what could have been done to improve their virtual experience. Three respondents wrote that they would have liked to know exactly what was planned for them to do during their virtual experience. It was also criticized that the sound issues should have been resolved sooner. Some respondents would have preferred to be together at the school's computer laboratory, similar to the introductory fieldtrip to Sploland and the International Space Museum. "As new users of the program, it would be helpful for us to have support as we try to support ELLs." Finally, one respondent who had problems logging in suggested that there should have been a checklist of things to do beforehand. They had received a comprehensive manual but it is unclear how many of them had actually browsed the manual. Only two out of nine respondents were looking forward to the second meeting with their ELL partners.

Overall, the first virtual meeting was neither perceived as useful for the development of teaching skills nor as an expansion of their virtual worlds skills and experience. The unique affordances of virtual worlds were perceived as being useful for language immersion (given adequate settings and time and room for private discussions), distance learners, and for learning activities that would benefit from the absence of physical boundaries. The two major challenges were the lack of a setting and lesson design that would encourage discussion, the lack of effective private sound parcels, and the apparent technical glitches that consumed the major portion of the lessons. Due to these impediments, the SPED teachers' usability ratings of Second Life decreased dramatically after the first virtual meeting with their ELL partners, to the extent that the majority of SPED teachers were not looking forward to the second meeting.

F. SPED Teachers' Post-Survey

Nine themes emerged from the SPED teachers' post-survey.

(1) Maximized Interaction

All 18 SPED teachers completed the post-survey after the second meeting. The first prompt asked, "Do you think that the meetings between you and your ELL partner(s) have helped them to enhance their oral proficiency? How would you have designed the meetings to help them improve their oral proficiency?" Only two out of 18 participants stated that the meetings might have helped the ELLs to increase their oral fluency. Specific instructions and scenarios that would have guided the interactions would have been appreciated, such as guessing games, giving a tour, and speed rotation activities.

I would have designed the meetings to get more voice time with their English counter parts, or even have them do a "speed rotation" activity where they would talk to someone for a few minutes, practice common phrases "my name is, I am ___ years old, I live in, something I like doing is _____" etc. and then have them move onto the next English speaker.

(2) Purposeful Roles

The SPED teachers would have appreciated the allocation of specific roles to justify their presence and promote involvement. They sometimes perceived themselves as by-standers rather than active participants due to the lesson design.

An activity that might have been more successful would have been one that would have required not only speaking but actual interaction between the ELLs and the native speakers like a guessing game such as a version of 20 questions or Guess Who. These would give the ELLs a chance to practice speaking more than just two words. It would also give the native speakers a reason to be there and involved.

(3) Extensive Second Life Training

The following statement highlights the need for more extensive training in the use of Second Life and the need for an effective set-up, time management, and trouble-shooting.

I felt that the sessions were poorly run and inefficient. I think that more training was needed, especially on the ELLs' part, on how to use Second Life. They seemed very confused about how to navigate the software and it took most of each session just to get them set up. By the time everyone was set up and "ready to go" the session was over and nothing was accomplished besides both parties feeling frustrated.

(4) Clear Instructions and Prior Information

Similar to the mid-reflections, it was criticized that the instructions were confusing. Several respondents stated that a face-to-face meeting would have been more effective and doubted the effectiveness of Second Life for these purposes.

It was suggested that prior information about the activities and expectations would have been helpful to prevent the ELLs from disengaging.

The activities and set up were so confusing for English speakers I cannot imagine how difficult it must have been for ELLs to follow. Also, when we were all at the brink of frustration they would start speaking to each *other in Chinese. There should definitely have been some trouble shooting and background information given to all participants prior to the first session.*

(5) Reduced Anxiety

Most SPED teachers acknowledged that Second Life has great potential because it allows for more interaction than other media, such as a video call, and because communicating through avatars appears to be less intimidating and to reduce anxiety levels. It provides anonymity for the ELL student and allows them to relax.

(6) Technical Facilitator

"I think that using Second Life as a language learning platform makes things harder than they need to be," even to the extent that technology is a barrier to learning. One participant pointed out the risk of cognitive overload due to the dual focus on language learning and technical aspects. Extensive Second Life training and support by a technical facilitator can help to prevent cognitive overload.

There may be times where a student is unsure how to use the program (such as opening the notecards, using the programs to write on boards, etc.), which can make it hard for that student to focus on language-learning when they are too occupied trying to figure out the technical things.

(7) Compensating the Lack of Social Cues

The lack of social cues, such as facial expressions, was evident. It was unclear whether silence meant active listening, confusion, disinterest, or headphone issues. The design of the activities needs to be carefully thought out in order to compensate these deficits.

(8) Taking Advantage of the Unique Affordances

When asked if the two teaching sessions on EduNation had taken advantage of the potential of virtual worlds, most respondents reported that the tools, such as a game show set, were not used in a way to promote oral communication successfully and that the lesson design failed to allocate the SPED teachers a purposeful role. The following statement describes several of these challenges, for example, excessive teacher talk making the SPED teachers redundant.

The dominos game was a clever idea but there was almost no opportunity for interaction and the native speakers did not seem to have much purpose in being there. With the game show activity again the idea was clever but the reality was that the ELLs were the only ones participating and their participation consisted of saying only two numbers at a time. The native speakers just sat and watched and every once in a while were able to give definitions if we knew them but the majority of the time the host would give the definitions eliminating any need for the native speakers.

Among the benefits were the convenience of working from one's home and the integration of multimedia into Second Life. In addition, most respondents agreed that Second Life would allow for interaction in shared activities that they might not ever have the opportunity to do, but "in terms of actually taking advantage of the virtual worlds and letting two students explore something together was limited. The sessions were low contact, low learning environments and activities." One suggestion was, "We could have explored different parts of the island we were on, or gone swimming in the ocean to see what we could find, etc."

When asked to describe how this project had helped them to familiarize themselves with virtual worlds, most SPED teachers agreed that it had, indeed, helped them to see how Second Life could be used and, in particular, how it could have been used differently. Surprisingly, it also became evident that even experienced users of virtual worlds and videogames had difficulties with the voice chat configuration.

(9) Mastery of the Unique Teaching Skills Required for Teaching in 3D Environments

The final prompt inquired about the unique skills that a teacher should have to teach in Second Life. A long list of unique characteristics and abilities was generated: patience, kindness, understanding, non-judgment, technical expertise, strong communication skills, excellent Second Life expertise, creativity in order to make materials enjoyable and accessible to many different learners, the ability to anticipate student needs without being able to read body language and facial expressions, the ability to diagnose a technical problem and troubleshoot, the ability to plan and implement a virtual lesson, the ability to specify and explain expectations and give clear directions in a calm manner, the ability to be clear and concise, the ability and willingness to organize activities in a way that everyone gets an opportunity to engage in an extensive conversation in a private chat area, the ability to stay calm in the face of technical glitches and not let the participants feel one's frustrations, think through the language demands of any given activity, and the ability to accommodate the needs of ELLs in a culturally sensitive way in order to allay their anxieties. A teacher in Second Life needs a back-up plan if technological errors take over the lesson.

The following statement summarizes many of the unique skills required to teach in Second Life from the SPED teachers' perspective:

Know how to have everyone set up ahead of time, how to give clear, concise directions to everyone, both written and oral, give group wide instructions more frequently, keep everyone on the same page and interested, pick good material and topics that both groups of student can discuss and participate in, and create quick, friendly opportunities between each pairing/group with clear objectives of what the conversation should be based around with small sheets providing words to be used throughout the conversation.

Despite the numerous frustrations that these SPED teachers experienced, the post-survey reflected that, overall, the respondents found the experience interesting because it showed them practical examples of teaching in Second Life, although they thought that the unique affordances of Second Life were not fully taken advantage of. The potential of virtual worlds was recognized, provided that the activities are carefully planned and that technical glitches can be drastically minimized to enhance everyone's experience.

Fun experience. Frustrating at times, but awesome to see where technology can go in education. The thought of kids being able to do this on an iPad with more ease and perhaps a webcam feature with kids from across the planet is a very exciting concept!

V. DISCUSSION

The discussion of the results has been arranged around the three research questions.

1. What are the English language learners' and the special education teachers' perceptions of Second Life as a language-learning platform? Almost all ELLs perceived Second Life as a useful and interesting language-learning platform. A perception that was shared by many was that they had more opportunities to speak with native speakers than in real life. On the one hand, this statement is surprising, considering that all ELLs were actually studying at a university in the United States where the majority of students were English native speakers. On the other hand, it could be an indicator of the difficulties that these students may have in engaging in conversations with native English speakers. ELLs have frequently indicated that communicating through a virtual world helped them to alleviate their anxiety, which is in agreement with Wang et al. [26] and leads to the suggestion that learning activities be designed in a way that only the instructor knows the true identity of an avatar for assessment purposes.

2. What are the unique skills that a teacher should have to teach in Second Life? The SPED teachers provided a long list of qualities that they would like to see in a teacher teaching in Second Life, which were informed by their own experiences in this workshop. Many of these skills should not only apply to virtual worlds teachers, but should be present in all teachers. Examples include patience, kindness, and understanding. It is hypothesized, however, that a virtual teacher may need even stronger skills in these areas when teaching in a virtual environment where social cues and non-verbal gestures are mostly absent. To give directions, for example, an avatar cannot rely on supporting verbal directions by gestures. Especially the lack of smiling and eye contact makes it more challenging to establish rapport and to convey kindness and understanding. Among the most frequently mentioned skills of a virtual teacher were: the ability to give clear and concise directions in a 3D environment, the ability to stay calm in the face of technical glitches, and the willingness and flexibility to resort to plan B if the lesson is not working out the way it was planned.

3. What types of problems associated with the EFL program in Second Life were identified? Although the ELLs' perception of the usability of Second Life for language learning was much more positive than that of the SPED teachers, both groups identified the same two main challenges. First, the way that the lessons were set up failed to encourage interaction between the two groups. On the few occasions when they spoke with each other rather than listening to the instructor, the ELLs tended to give monosyllabic answers. It is possible that the Chinese students were reluctant to reply or elaborate on their answers because they were under pressure and had not been given appropriate wait-time [15] [16]. To reduce anxiety in ELLs, it is recommended that language-learning activities in Second life be designed in a way that opportunities for interaction in private sound parcels are maximized. The risk of negative evaluation by either the instructor or other listeners should be minimal. The visually stimulating and interactive environment in Second Life lends itself to extensive and engaging collaborative activities in relatively authentic and contextualized settings, such as a collaborative scavenger hunt at the Star Trek Museum of Science in Second Life [28].

Technical issues, mainly related to voice communication in Second Life, were the second major challenge. Instead of the use of Second Life voice chat, it is recommended that Skype be used instead to accommodate both text and voice chat [16]. It is also recommended that an in-world facilitator support the instructor. Prior to the actual Second Life assignments, students will wish to familiarize themselves with Second Life. The authors described an 11-Step Virtual Worlds Teacher Preparation Workshop with an introductory fieldtrip to five Second Life islands to ensure that students master the navigation skills required for the subsequent training steps [28].

Overall, nine guidelines emerged from the SPED teachers' post-survey (see Table III).

Table III. Nine Guidelines

#	Themes
1	Maximized interaction
2	Purposeful roles
3	Extensive Second Life Training
4	Clear instructions and prior information
5	Measures to reduce anxiety
6	Technical facilitator
7	Compensating the lack of social cues
8	Taking advantage of the unique affordances of virtual worlds
9	Mastery of the unique teaching skills required to teach in 3D environments

Virtual worlds have provided broad access to native speaking communities and virtual spaces for learning and collaboration [24] and provide the potential to address the five components of the National Standards for Communication, Culture, Connections, Comparisons, and Communities [24]. Collaborative project design, however, can be challenging in virtual worlds. The findings of this study concur with Warburton and Pérez-García [29] who identified a set of guidelines that address factors fostering collaboration in 3D environments, such as running a social event before the main activity, ground rules for communication, making collaboration intrinsic to the tasks, guidance and regular briefing in order to scaffold gradually increasing levels of task complexity, video tutorials, and live mentoring/assistance.

In conclusion, alternative activities and learning designs have to be found in order to make Second Life more appealing for ELLs and their native or nonnative English speaking conversation partners. In addition to the ideas brought forward by the SPED teachers, such as giving tours, speed rotation activities, and guessing games, the discussion group format may be an effective alternative. Morgan [30], for example, described the integration of the U.S. Holocaust Museum in Second Life into a history classroom using a student-led discussion format through voice chat. Similarly, Prude [31] described teaching Asian religions in Asian-inspired destinations in Second Life, also using a synchronous discussion format. Although the latter chose text chat for communication, the learning design could easily be adapted for voice communication in order to practice oral production in a foreign language. The discussion group format would allow ELLs sufficient preparation time, which would likely reduce their anxiety in oral production [15] [16]. Having students explore a virtual space prior to the discussion would allow them to take advantage of the unique affordances of virtual worlds, such as spatial representation, experiential learning, motivation, transfer, and collaboration [32].

VI. CONCLUSION

Despite serious challenges, the participants perceived Second Life as a useful, supplementary tool for instructors because it promotes contextualized language practice, provided that the lesson design follows the nine guidelines suggested earlier. Most importantly, the teacher should master unique teaching skills required to teach in 3D environments. Failing this, the same or possibly superior learning effects can be achieved in alternative environments (e.g., Skype) with less hassle. The time needed to prepare participants for the technical requirements of Second Life or other virtual worlds should not be underestimated. More than one fieldtrip as a class is recommended.

The findings, practical guidelines, and ideas for alternative lesson design will be relevant to other language instructors who plan to use Second Life for oral fluency enhancement. Virtual meetings between ELLs and English-native speakers in Second Life have the potential to offer an innovative, creative, and stimulating way to practice speaking English in contextualized settings, provided that the activities are framed by a pedagogical rationale that justifies the use of 3D technology. If ELLs are teamed up with native English speakers, specific roles should be assigned to the latter. But, even without the presence of native speakers, language learners can benefit from mutual interaction by practicing their language skills in content-rich virtual worlds [24]. Increased speaking opportunities with English-native speakers is likely to enhance ELLs' confidence and may encourage them to transfer the skills practiced in virtual worlds to the real world, specifically their university settings. Limitations of this research are that all instruments relied on self-reporting and the teams only met twice for one or two hours each, which made it impossible to measure potential gains in oral fluency. No statistical analysis of the usability ratings could be conducted because half of SPED teachers failed to complete the mid-reflection.

Recommendations for future research include a more in-depth investigation of the effectiveness of virtual worlds on the oral fluency of ELLs and how instructors can design in-world activities effectively to take full advantage of the unique affordances of virtual worlds.

ACKNOWLEDGMENTS

We thank Julie Johnson and Alejandro Rodriguez for their valuable contributions in the needs assessment of Chinese students in this university's English as a Second Language department as an integral part of an instructional design project.

REFERENCES

[1] K. Oh and N. Nussli, "Creating a global second language acquisition classroom using a virtual environment," Proceedings of The Sixth International Conference on Mobile, Hybrid, and On-line Learning (eLML), Spain, 2014, pp. 30-36.

[2] C. J. Fillmore, in C. J. Fillmore, D. Kempler and W.S.W. Wang (Eds.), Individual differences in language ability and language behavior, NY: Academic Press, 1979, pp. 85-102.

[3] P. Skehan. Second-language acquisition resarch and task-based instruction. In J. Willis, & D. Willis, Challenge and change in language teaching, Oxford: Heinemann, 1996, pp. 17-30.

[4] R. Schroeder, "Defining Virtual Worlds and Virtual Environments," Journal of Virtual Worlds Research, vol. 1, 2008, p. 2-3.

[5] H. Nishimura, K.Y.T. Lim, and K. Koyamada, "The Abyss Observatory designing for remote collaboration, self-directed discovery and intuition development in multi-user interactive 3D virtual environments," Journal of Virtual Worlds Research, vol. 5, 2012, pp. 1-11. Retrieved from: http://journals.tdl.org/jvwr/index.php/jvwr/article/view/6304

[6] R. Rothfarb and P. Doherty, "Creating museum content and community in Second Life". In J. Trant and D. Bearman (Eds.). Museums and the Web 2007: Proceedings, Toronto: Archives & Museum Informatics. Retrieved from: http://www.archimuse.com/mw2007/papers/rothfarb/rothfarb.html

[7] K. Yee and J. Hargis, "Jumping head first into Second Life for higher education," International Journal of the Scholarship of Teaching and Learning, vol. 6, 2010, pp. 1-15.

[8] S. Aydin, "Second Life foreign language learning environment: A review of research," Turkish Online Journal of Distance Education, vol. 14, January 2013, pp.53-63.

[9] C. Balcikanli, "Language learning in Second Life: American and Turkish students' experiences," Turkish Online Journal of Distance Education, vol. 13, April 2012, pp.131-141.

[10] S. Grant and R. Clerehan, "Finding the discipline: Assessing student activity in Second Life," Australasian Journal of Educational Technology, vol. 27, 2011, pp. 813-828.

[11] M. Henderson, H. Huang, S. Grant, and L. Henderson, "The impact of Chinese language lessons in a virtual world on university students' self-efficacy beliefs," Australasian Journal of Educational Technology, vol. 28, 2012, pp. 400-419.

[12] C. X. Wang, B. Calandra, S. T. Hibbard, and M. L. McDowell Lefaiver, "Learning effects of an experimental EFL program in Second Life," Educational Technology Research Development, vol. 60, 2012, pp. 943-961, doi 10.1007/s11423-012-9259-0.

[13] A. K. Wehner, A. W. Gump, and S. Downey, "The effects of Second Life on the motivation of undergraduate students learning a foreign language," Computer Assisted Language Learning, vol. 24, 2011, pp. 277-289, doi 10.1080/09588221.2010.551757

[14] B. Mak, “An exploration of speaking-in-class anxiety with Chinese ESL learners,” System, 39, 2011, pp. 202-214.

[15] X. Ma, “Teacher talk and EFL in university classrooms,” MA Thesis, School of Foreign Languages and Literature, Chonquing Normal, University and Yangze Normal University, China, 2006.

[16] X. Lei, “Communicative teacher talk in the English classroom,” English Language Teaching, vol. 2, 2009, http://www.ccsenet.org/journal/index.php/elt/article/viewFile/338/301 .. [retrieved: June 29, 2014]

[17] M. Liu and J. Jackson, “An exploration of Chinese EFL Learners' unwillingness to communicate and foreign language anxiety,” Modern Language Journal, vol. 92, Spring 2008, pp. 71-86.

[18] C. Chau, A., Wong, M. Wang, S. Lai, K.W.Y. Chan, T.M.H. Li, D. Chu, I.K.W. Chan, and W. Sung, “Using 3D virtual environments to facilitate students in constructivist learning,” Decision Support Systems, vol. 56, 2013, pp. 115-121.

[19] S. Farra (2012) “Effects of Disaster Training With and Without Virtual Simulation.” (Electronic Thesis or Dissertation, document number ucin1330024525). Retrieved from https://etd.ohiolink.edu/

[20] J. Moskaliuk, J. Bertram, and U. Cress, “Impact of virtual training environments on the acquisition and transfer of knowledge,” Cyberpsychology Behavior and Social Networking, vol. 16, 2013, pp. 210-214.

[21] M. Peterson, “Towards a research agenda for the use of three-dimensional virtual worlds in language learning,” CALICO Journal, vol. 29, 2011, pp. 67-80.

[22] M. T. Thomasy Blasing, “Second Language in Second Life: Exploring interaction, identity and pedagogical practice in a virtual world,” Slavic and East European Journal, vol. 54, 2010, pp. 96-117.

[23] C. Wang, H. Song, D. E. Stone, and Q. Yan, “Integrating Second Life into an EFL program in China: Research collaboration across the continents,” TechTrends, vol. 63, November/December 2012, pp. 14-19.

[24] B. Knutzen and D. Kennedy, “The global classroom project: Learning a second language in a virtual environment,” The Electronic Journal of e-learning, vol. 10, 2012, pp. 90-106.

[25] H. Ishizuka and K. Akama, “Language learning in 3D virtual world: Using Second Life as a platform,” eleed, vol. 8, 2011, http://nbn-resolving.de/urn:nbn:de:0009-5-31706 [retrieved: January, 2014]

[26] F. Wang and E. Shao, “Using Second Life to assist EFL teaching: We do not have to sign in to the program”, TechTrends, vol. 56, 2012, pp. 15-18.

[27] E.M. Rogers, Diffusions of innovations, Glencoe: Free Press.

[28] K. Oh and N. Nussli, “Teacher training in the use of a three-dimensional immersive virtual world: Building understanding through first-hand experiences,” Journal of Teaching and Learning with Technology, 3, 2014, pp. 33-58.

[29] S. Warburton and M. Pérez-Garcia, “3D design and collaboration in massively multi-user virtual environments”. In D. Russell (Ed.) Cases on collaboration in virtual learning environments: processes and interactions, Hershey, PA: IGI Global, 2009, pp. 27-41.

[30] E.J. Morgan, “Virtual worlds: Integrating Second Life into the history classroom,” The History Teacher, vol. 46, 2013, pp. 547-559.

[31] M.A. Prude, “A classroom of bunnies, blimps, and werewolves: Teaching Asian religions online in Second Life,” ASIA Network Exchange, vol. 20, 2013, pp. 1-12.

[32] B. Dalgarno and M.J.W. Lee, “What are the learning affordances of 3-D virtual environments?” British Journal of Educational Technology, vol. 41, 2010, pp. 10-32.

Early Stages of Business Modeling for Open Source Home Care Technology

Lessons learned from an initial inventory

Elles Gyaltsen-Lohuis, Marike Hettinga, Ruud Janssen, Jan M. Nauta, Sikke Visser
Research group IT Innovations in Health Care
Windesheim University of Applied Sciences
Zwolle, The Netherlands
{ege.lohuis, m.hettinga, tjwm.janssen,
j.nauta, s.visser} @windesheim.nl

***Abstract*— Many eHealth innovations never get beyond the project phase. Initiating a viable business model in an early stage of the development of eHealth innovations enhances the chance of structural embedding the innovations in routine health care. This paper presents the early stages of business model development for an innovative IT supported home care system based on open source software. After a literary review on open source business models and the home care market, the eHix method, a business model approach developed by research group ICT Innovation in Health Care to is used build up the business model for the home care system. The structure of the eHix, a method in which one is only allowed to move to the next phase if the previous phase is completed, ensures an efficient and effective route to business modeling. Going through the first stage of the business modeling, it is clear that the mapping of intended target groups and their accompanying needs in various scenarios becomes key to determining the right value proposition for the intended home care system. In-depth analysis of all the stakeholders and their interest in the network then provides the essential criteria for the feasibility of the various scenarios. It is only after all the scenarios of the care system are mapped that the revenue models can be identified efficiently and further selective steps can be taken towards feasible business models that show the potential of the innovation to its full advantage.**

Keywords- Home care technology; Open source software; Business modeling.

I. INTRODUCTION

This paper provides an extension of the work presented at the eTELEMED 2014: The Sixth International Conference on eHealth, Telemedicine, and Social Medicine [1]. The work centers on the Hightech@home project, a project aimed at developing technology and knowledge concerning open source IT supported home care.

Home care is this paper is defined as the whole of care, nursing, treatment and support of the person seeking help at home, performed with the aid of self-care, informal care, volunteering and/or (additional) professional care and specifically designed to enable the person seeking help to live independently in his or her own home [2].

With the use of open source software the project is geared towards increasing availability of IT supported home care and lowering costs of IT supported home care by avoiding vendor lock-ins. The project team aims to enhance the chance of structural embedding of the IT innovation in routine home care; the end of the project should not imply the end of the innovation. Therefore the project also includes a viable business model for the Hightech@home care innovation.

This paper focuses on an analysis of the business modeling in the early stages of the Hightech@home care technology development, reflects on the initial findings and posits further challenges to the business modeling of open source home care technology. The paper aspires to contribute to the body of knowledge regarding the early stages of business modeling for home care technology and to clarify the fundaments of viable business models for home care innovations in order to increase the likelihood of practical embedding of these innovations in routine home care.

A. The Hightech@home project

The Hightech@home project consists of five tracks. The first track, the technical design and development, focuses on a portal being generically available through any web browser. The portal will facilitate video contact and transmit and receive data from sensors located in the house, on a portable device or on the body of the user. Technical design and development is iteratively informed by the user research in the second track. The iterative cycles start with small-scale experiments with mock-ups leading to a field study when robust prototypes are available. During this field study data will also be retrieved to perform a cost-benefit analysis, which is the third track of the project. To ensure the structural embedding, viable business models are developed in the fourth track. Finally, the fifth track aims to facilitate Bachelor students in the Hightech@home project in particular and in eHealth research in general through the development of a digital learning environment.

B. The choice for open source software

The Hightech@home project's choice for open source software is based on the current lack of low-cost, high-quality, high-tech, plug-and-play home care technology. Open source software is defined in this project as software of which the source code is available at no cost, the user is free to customize this code and integrate it into other systems and products [3]. Open source software has the potential to enable system interoperability, which literally unlocks the consumer technology we already have in our homes and carry around.

If high-tech sensor and communication technology is available for home care, the costs are high. Potential users with the greatest need for this type of technology, e.g., elderly, chronically ill, often lack the required financial means. Vendor lock-in (mostly through the use of closed protocols) is one of the reasons for the high pricing. Customers are dependent on one vendor after an initial choice for home care technology. If there is, for instance, a wish for extending the current technology with additional sensors, customers appear to be dependent on the original vendor. Hence low-cost alternatives are not an option. Although completing a single project like the Hightech@home project will not break through an established vendor lock-in, the aim is to produce open source technology and knowledge, thus providing a starting point towards interoperability.

Existing home care technology often requires a technician for its installation. This not only results in extra costs, but also raises the threshold for people to start using it. By designing and developing plug-and-play technology by means of a co-creation process with potential end-users, the Hightech@home project ensures that customers (clients or patients), informal carers and care professionals are able to install the equipment. Furthermore, having interoperability at the center of development, allows for (re-)use of already existing technology.

By designing and developing an innovative home care application, the Hightech@home project aims to provide a starting point towards low-cost, high-quality, high-tech, plug-and-play home care technology. Being aware of the budding opportunities of this innovation, the project also strives to contribute to the structural embedding of this innovation in routine home care. The end of an innovative project like this should not imply the end of the innovation, which is often still the case with innovative eHealth applications [4]. In the reality of health and home care there is still a gap to be bridged between the opportunities offered by innovative applications in health and home care and the actual practice of routine health and home care. Although suggestions are constantly being made for the integration of innovations in health care, a comprehensive approach which supports the adoption of innovations into routine health care practice is still lacking [5].

C. Business modeling the project

IT-related business model innovations have been identified as a mayor factor in achieving structural innovation in health care [6]. Initiating a viable business model in an early stage of the development of the innovation enhances the chance of structural embedding the innovation in routine health care. Characterizations and definitions of business models are abundantly present in literature. Since the concept of service is a focus in the Hightech@home project and a number of partners – a foundation, several small and medium enterprises (SMEs), a professional care organization and a University of Professional Education - will work together to create and deliver the service, we apply the definition of Bouman et all. (2008) to characterize the business model envisioned for this project: ‘A business model is a blueprint for a service to be delivered, describing the service definition and the intended value for the target group, the sources of revenue, and providing an architecture for the service delivery, including a description of the resources required, and the organizational and financial arrangements between the involved business actors, including a description of their roles and the division of costs and revenues over the business actors [7].’

In earlier projects [5] the research group IT (Information Technology) Innovations in Health Care at Windesheim University of Applied Sciences developed a business model approach to be used as an instrument to bridge the gap of innovative eHealth ideas to successful IT-based care services. A key component of this approach is an online webtool: the eHealth Innovation Matrix [8]. Therefore theHix method is used during the Hightech@home project. The use of this method implies the application of relevant instruments and knowledge as included in all five phases distinguished in the eHix model

– this will be explained at section II Methodology.

Hightech@home’s aim to achieve a better chance of structural embedding the innovation poses the question of a feasible business model. Hence this paper sets out to address the following research questions:

What business model would be most feasible for the Hightech@home open source technology innovation ?

- What are the specifics of open source and business models for open source software?
- What characterizes the home care technology market, and specifically in the Netherlands?
- What are the particulars to be considered in the early stages of business model development for open source home care technology?

This introduction has provided a summary of the aim of the Hightech@home project and its focus on open source and business modeling open source technology. Section II gives an overview of literature on the area of open source, business modeling and home care technology in the Netherlands. This is followed by the overview of the methodology used to build a business model. The methodology is then applied to the early stages of the development of the home care technology. The preliminary outcomes regarding the business models are being discussed and the paper closes with a conclusion and the identification of further work.

II. Open Source Software and Business Modeling in Home Care

A. Open Source Software and Business Modeling

1) Open Source Software

The use of open source software is central to the Hightech@home project. In order to determine the implications of the use of open source software for the business model design of the home care innovation, it is important to have insight into the potential advantages and disadvantages of the use of open source software. The first advantages obviously are the low costs of purchase and flexible customizing; the source code of open source software is available at no cost and the user is free to customize and integrate the code into other systems and products.

A fast time-to-market is mentioned in literature as an advantage of open source software in that adjustments and improvements by the community are quickly picked up and shared [9]. Another advantage is the fact that there are many tools and components available in the communities. Open source software can also increase the flexibility in integration projects with other tools and solutions, since one does not have to deal with intellectual property [10]. This benefit reduces the possibility of vendor lock-in.

In addition to the low purchase costs, the flexibility, the ability to customize and a fast time-to-market, Krishnamurty (2005) mentions the advantage of the support of a community of developers and testers. According to Krishnamurty, this will increase the robustness and reliability [11]. An interesting observation from West (2007) is that organizations usually invest heavily in the protection of intellectual property, often leading to the redevelopment of existing solutions. The open source model with its unprotected adaptability implies an evolutionary development rather than redevelopment. This implies that initial innovation budgets can be used more efficiently in open source projects. In addition to the above benefits West (2007) also notes less marketing spending as open source technology generally "sells" through word of mouth.

There are a significant number of disadvantages of the use of open source software mentioned in literature. Martin (2013) gives an excellent up-to-date overview of these disadvantages – starting with the degree of uncertainty regarding the support of the open source software [10]. Dependence on the strength of the community of developers and testers is a risk. While standard license structure with traditional software gives the guarantee of the supplier, the availability and continuity of the community is not always adequately secured, causing a serious liability that has to be taken into account. Also new developments and improvements on the open source software do not have to be in line with the users' IT strategy. Besides, the more the open source products are being customized, the less one can make use of standard training. This means that the user should allow more resources for these aspects. Other disadvantages regarding the use of open source software include the control of version and usability. Versioning can be quite complex, since there are many different developers making adjustments and improvements. This can also make usability problematic, as it depends on the developers in the community and the available resources.

2) Business models for Open Source software

Considering the specifics of the open source software, its advantages and disadvantages, it is clear that these specifics pose a challenge in developing a viable business model. Revenue models for open source software in literature make the fundamental distinction between commercial open source and community open source [12]. Community open source is open source software that is owned by a community or a legal entity representing the community. The community members typically do not derive direct revenues from the software but subsidize it from complementary products and services. In contrast, commercial open source is open source software that is owned by a single legal entity with the purpose of deriving income from the software. The open source software used in the Hightech@home project is community open source software. The source code of the project is available under only one license, the GNU General Public License Version 2 or later [13], and anyone can enter the market and generate revenue from the project.

There is often still the notion that utilizing open source software development does not offer any chances for revenues from products and thus there cannot be any successful open source business models [14]. However, already two decades ago open source successes showed that fruitful business models can be built around open source projects and products [3]. There are various ways of generating revenue from community open source software to be found in literature – for a good overview see Korhonen (2013). Overall the open source business models seem to center on the sale of complementary products and service to complete a whole product solution. The three dominant ways to collect revenue from community open source software are usually consulting and support services around the software, derivative products built on the community open source software, and increased revenue in supplementary layers of the software stack [15].

Besides selling complementary products and/or services, companies also have ways of capturing value and indirect revenues from the open source software by creating and guiding the formal architecture of the open source software. Particularly for platform technologies, control of the interfaces regulates the supply of complementary assets [16]. Therefore, a company that waives ownership of the software code creates a shared platform that fits best its own internal architecture and suite of complementary products. Similarly, ongoing leadership of open source community software allows a company's developers to influence the code to be most compatible with a firm's own requirements and software architecture, even in independent projects [9]. Even without influence, companies that are actively involved in the open source development may gain technical knowledge, which is suitable for deployment and support or an expert reputation, which is useful in marketing.

While the models of revenue are developed with the open access of the software code as a focal point, over the years hybrid open source business models have been developed [17]. A hybrid model fuses open source licensing with commercial licensing of software. For example, customers want to extend a product they have acquired and this product comes with an open source license. The customers must then release the modifications they made under the same open source license. If customers do not want to do that, they can purchase a commercial license. The company using the hybrid model manages the licenses and competes against proprietary vendors. This business model requires the company to control very tightly the source code of the product. This might lead to a situation where open source developers and communities do not contribute much at all, defying the purpose and perceived benefits of open source community software.

Besides generating revenue, it is clear that the company has to address the challenges of working with open source in order to develop a successful open source business model. These challenges are mostly caused by the numerous actors in the open source software development [14]. The business model must be flexible enough when it comes to release dates and requirements implemented per release [18]. For a solid integration of open source business models into the company, the requirements of interacting with the open source development communities also need to be acknowledged and practiced by the company [19]. Developing the open source business model, the company's own needs also have to be taken into account. As Munga et al. (2009) remark, companies cannot use traditional business models as such with open source, but can adapt these models by making open source a fundamental part of them, while asking themselves what the future implications of this open source business model are.

The use of community open source software like the software in the Hightech@home project has considerable implications for the business model. The various models in literature do suggest creative solutions. Services and derivative products around the software, shared platform creation and hybrid models fusing open source benefits with commercial interests are all possible features to be taken into account while developing a sustainable business model. Because of the numerous actors in the open source software development, flexibility in the business model seems to be a key component for success.

B. Home Care Technology and the market in The Netherlands

In addition to the specifics of open source software, we also need to take the distinctive market into account while developing the business model for the Hightech@home project. So what is distinctive about the health and home care market, and specifically the Dutch market?

Dutch health care is accessible for everybody – this makes it unique [20]. Internationally there are only a few countries in the word where this is the case – only 5 % of the world population lives in a country where good care is accessible to everybody. Worldwide the health care system has been put under strain for the last decade. The increasing life expectation, improved survival of people with acute and long-term conditions and a greater array of available treatment options are placing an increasing burden on healthcare systems internationally. The most vulnerable groups in society such as the elderly, the chronically ill and people with disabilities are encouraged to live at home longer than before. In addition, the economic crisis that emerged in 2008 still stunts the economic growth, which used to compensate and finance the increasing costs of health care. Even with substantial developments in IT applied in healthcare over recent decades, the perceived increase in productivity does not seem to compensate the increased expenditure yet [20].

Regarding the Dutch home care market specifically, there is a large range of technological possibilities available [21]. The home care technology varies from disclosure of information in text and images, video communication, screen case, telemonitoring to other IT applications like ambient technology which enables monitoring in and around the home. Many projects haven been initiated. The drivers for these projects are mostly rising expectations of patients, increased efficiency, increased life expectancy and demand of care, treatment and prevention, improved availability of technology and safety and quality demand [22]. Nevertheless, there is not much evidence of structural scaling and implementation of these projects. Many obstacles are encountered here. For instance, there are many initiatives for which the health costs are not compensated or investments are too high or there is a lack of other means for structural financing. Care professionals often do not comprehend the added value or have clinical concerns against the implementation of home care technology. Failing change management seems to focus too much on technological implementation. Doubts have arisen about safety and privacy, and lack of clinical involvement and lack of scientific foundation are other barriers that hamper structural scaling [22]. Also the economic crisis contributes to short-term politics and leaves little room for investment in the necessary cultural and organizational change. In addition every Dutch health and home care facility seems to have its own in-house innovative technology to distinguish themselves in the health market, thereby increasing the costs and not realizing the potential economies of scale. Overall the Dutch home care technology market can be characterized as an unstructured and unorganized market with many parties providing part of the solution [21].

With this in mind, we take a closer look at the landscape of home care technology in the Netherlands. Fig. 1 illustrates the Dutch health care scene with its many parties. While the Dutch government regulates the financing of the health care system, the health insurance companies and the municipalities purchase the care for the clients through the professional care organizations within these constrains. The clients receive their care from these selected care organizations and in return pay their premium to their health insurance company and tax to their municipality. Besides professional care, many clients also rely on informal care; there are about 3.5 million informal carers in the Netherlands on a population of 16 million people [23].

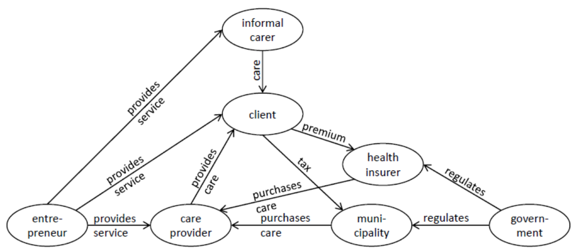


Figure 1. The Dutch health care landscape with its many parties

Since the beginning of 2014 a significant transition has been taken place, as a result of which a substantial part of the arrangements for home care will be implemented by the municipalities, who will take on this role for the very first time. The municipalities have considerable flexibility in determining what home care facilities and arrangements they deem necessary to organize or fund. It is expected that the municipalities will focus on cooperation with local parties (healthcare providers, day care facilities, community centers, social housing corporations, etc.). There is a lot of room for interpretation in the current framework contracts of the municipalities for their health and home care purchasing. It is not yet clear how the funding will be worked out by the municipalities, but heavy budget cuts are expected on all fronts [24].

Besides the technological developments and complex finance structure of the Dutch home care market, there is another important factor that defines the home care market: the dual client demand. Dohmen (2012) mentions the dual client-demand of home care technology - purchased by care organization but marketed to the end-user, the care consumer - as one of the most pressing aspects that makes the development and implementation of home care technology difficult [21]. Both parties can have different – often conflicting – requirements and demands. For instance, a care organization might be looking for a more efficient manner of care-giving, while its clients are happy enough with partial support. In addition to this, there is a gap between usage and expectation. The younger generation of care-professionals is often more used to working with technology than their older clients while the reverse can also be seen: a large part of the care-professionals is not trained in the use of information technology, while their younger clients are quick to master the technology.

With this context in mind, we see that the entrepreneur who is looking to market a home care innovation is dealing with multiple stakeholders in a turbulent landscape. The routes to the potential market are numerous and complex and entail different approaches in design, development and marketing of the innovation. To come to a viable business model in the home care market the method used to develop the model obviously has to incorporate these aspects.

It is clear from literature overview that the characteristics of the home care technology market in the Netherlands and the specifics of open source and business models for open source software pose challenges to the development of a business model for an open source home care system. The home care market is divers and complex and the use of open sources software asks for flexibility in the business model. On the way to develop feasible business models for the Hightech@home care system the particulars of the early stages of business model development for open source home care technology will come to the fore in the first methodological steps.

III. METHODOLOGY

In order to develop a feasible business model for the Highttech@home open source project in a structured manner and to bring about the specifics of the early stages of business modeling, the eHix framework is used. As mentioned in the introduction of this paper, in earlier projects the research group IT (Information Technology) Innovations in Health Care at Windesheim University of Applied Sciences developed a business model approach as an instrument to bridge the gap of innovative eHealth ideas to successful IT-based care services. This approach, the eHealth Innovation matrix (eHix), is well-suited to the characteristics of an IT service innovation like the Hightech@home home care services [25].

The starting point of the eHix method is the STOF model (Bouwman, De Vos & Haaker, 2008) [25]. This model provides a systematic approach for service and business modeling of innovative ICT services and offers a solid foundation for the design and examination of the different aspects of the business model, such as the service user, the technology, the organizations, finance and regulations. The STOF model describes the business model based on four perspectives or domains: the Service perspective - a description of the service, the value proposition (the value of the service for users) and the intended audience; the Technology perspective - a description of the required technical functionality and architecture to deliver the service; the Organization perspective - a description of the resources, activities and roles and structure in which the value network partners to provide the service - and the Finance perspective - a description of how the service generates revenue (earnings) and the distribution of costs and benefits among the parties involved in the value network.

Naturally these different aspects will change in the course of an innovation process. The service develops under the influence of input of target users, the choice of technology is changing as a result of learning from experiments and organizations play different roles at different stages. That is why the eHix combines the STOF model with a phasing that provides support for the lifecycle of the innovation process. The five phases of Hettinga (2009) form the basis for the phasing: inventory phase, design & development phase, experimental phase, pilot phase and implementation phase [26]. In the inventory phase the ideas of a new service are created and the needs and requirements of the users for example are analyzed. In the design and development phase, the technology is designed and developed, and thoughts about the business model of the service are initiated. In the experimental phase users try out the new application in a laboratory setting while in the pilot phase more users are involved testing the application of the service in their daily practice and giving input for a successful implementation in the deployment phase.

The eHix thus combines the five stages in the innovation process (concept phase, design phase, test phase, pilot phase, implementation phase) with the four aspects of the business model (service, technology, organization, finance), resulting in a matrix containing 20 (4 by 5) cells. Each cell contains the essential steps and choices to be made in the innovation process for a specific domain within the business model in a particular phase. For each cell keywords describe the essence of the cell. For each keyword checklists, tools, methods, approaches and examples were collected that support the decision making process to the next step in the innovation process. The content offered by the eHix cells varies. For example, there is a template for conducting a stakeholder analysis, a preparation of a business case, a template to aid in making a technical design and a format to evaluate a pilot.

For the development of the eHix content a thorough study of factors determining the success and failure for e-health innovations was done. The identified success and failure factors are assigned to the various components of the STOF model so the success criteria can be used for evaluation of the business model and also serve as building blocks of eHix itself. The list of success criteria was compiled based on a literature review and several expert meetings wherein insurance companies, entrepreneurs, researchers, health care institutions and government agencies participated. The results of the expert meetings and literature are assigned to the eHix products per fase. The eHix matrix structure is shown below in Fig. 2 with the main keyword in the cell displayed.

1) Designing the Hightech@home care business model with the eHix method

To guide the development of a viable business model, the eHix has been applied to the Hightech@home project, which has gone through the inventory phase and is currently in the middle of the design & development phase. The project's aim is to realize a home care service system as outlined in Fig. 3. This product is at the service of the clients (care receivers). As Fig. 3 shows the service system can be divided into two parts: the left part for the client and the right part for the professional carers / informal carers.

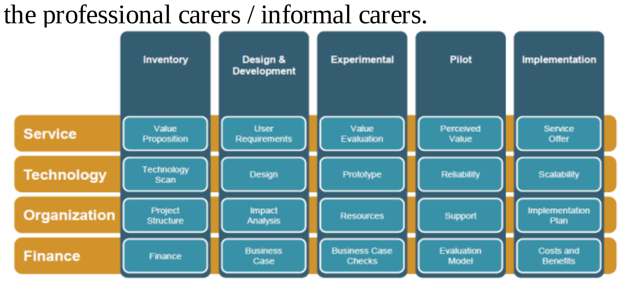


Figure 2. eHix matrix structure combining business model domains with innovation phases

The client's package consists of a smartphone or tablet with which the client can request the assistance of a carer either by 'normal' call or by video call. The two sensors, a GPS and a fall detection sensor in the client's device make it possible to detect the position of the client in case the client is in need of assistance outside of the home. The device of the carer is also a tablet or smartphone; communication with the client is possible and the necessary sensor information (fall detection sensor / GPS) is displayed. So there are three manners to obtain the client / carer information: (a) by video contact: personal communication, (b) GPS: location and (c) by the fall detection sensor: information. The fall detection sensor serves as an example for the possibilities to connect generic sensors. In addition, functions like data mining (eg. filling in data on client and carer) and viewing the logged information are provided for.

2) Service Domain

The first part of the STOF framework describes the service offered, the value proposition and the possible market segments of the Hightech@home care service. The value proposition of the home care service is still hypothetical in this inventory phase and needs to be demonstrated in the experimental phase and pilot phase. The home care service is offered as a technical support service so that the elderly and persons with a chronic illness and/or intellectual disability are able to continue living at home (longer). The support service enables the clients and their professional carer and informal carers to communicate with each other in the traditional way of using a telephone with voice communication, and adds an extra interface of video communication. This extra interface of video communication is expected to provide the client and carer with more information and reassurance.

The GPS and fall detection sensor offer real-time and accurate information on the physical state of the client to the carer. For the client the GPS and fall detection sensor anticipate in a greater sense of safetly and reassurance. With the combination of an extra interface and sensors the home care system is expected to result in a more efficient and effective way of communicating. Making use of open software is anticipated to lead to lower cost of development and higher ease of use for the client (plug-and-play). Appropriately translated by the developers of the home care system the low costs and ease of use could result in a higher quality of care for the clients.

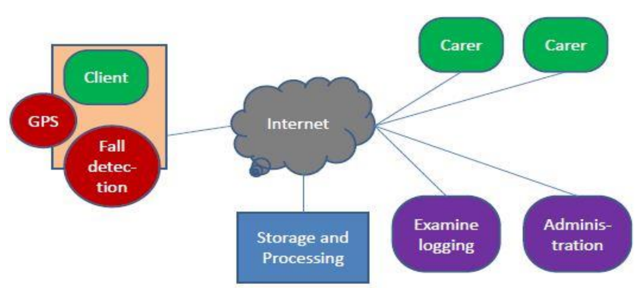


Figure 3. The envisioned Hightech@home care service system

The service also hopes to ensure that the clients will be able to do more themselves or with informal carers, thereby being more independent and requiring less care. This independence is expected to increase the quality of life for the client. The Hightech@home care service also provides extra information by logging the selected sensor data. Analysis of the logged data by the carers can ensure that care is better tailored to the clients' needs.

The Hightech@home service is expected to help professional carers and the informal carers to improve the effectiveness of their care. Expected efficiency, labor savings and increased service levels are all part of the value proposition for the professional care organization, as well as creating a competitive advantage compared to other care organizations with this home care service. The enhanced sense of security that the service offers through real-time monitoring is an important value proposition for both client and informal carer. The hypothetical value proposition has to be validated in the experimental phase and pilot phase.

3) Technology Domain

The next step in the eHix method is the description of the technical functionality required to realize the service. Relevant variables in this technology domain are: technical architecture, network accessibility, needed devices and application. The Hightech@home care service comprises a robust smart phone equipped with sensors (GPS and fall detection sensor), servers for storage and processing of information, and tablets or smart phones to display client messages and other information, see Fig 3. The technology used for the home care service is based on open source software. Even though the web application developed has been used in other projects, the combination with generic sensors has not been proven yet. Reliability is important because the service intents to deliver instant and stable communication between client and carer and real-time and accurate information about the movements of the client.

Beside an impeccable reliability, the home care service needs to meet a number of certifications. There are a number of European requirements for the introduction of home care service [27]. Since the home care service also logs patient data, the law on medical health informatics must also be met [28].

4) Organizational Domain

The third part of the business model contains a description of all the different actors in the network needed to deliver the service and the position of these actors in the value network. First we chart the actors, then we analyze their strategic interests to complete the stakeholder analysis.

The central actor in the Hightech@home care service is the client with a mental and/or physical disability living at home. The client receives care from a professional carer who is employed by a formal care organization and/or from an informal carer. The professional care organization provides care to people with physical and/or mental disabilities in their own living environment. The informal carer is often related to the client, either family or living in the neighborhood. The insurance company and/or the municipality finance or arrange the care and in doing so adhere to the laws and standards promulgated by the government.

The Hightech@home project team acts as the entrepreneur in this case, developing the home care service and the possibilities to market. The team consists of Stichting TriVici, VAC Thuistechnologie, Frion and Windesheim Research group IT Innovations in Health. Stichting TriVici is founded on new concepts and innovations bringing care to people. They provide innovative solution to the Dutch home care market. An example is the development of an open source portal. VAC Thuistechnologie is a consultancy firm in information technology which excels in the home care market based on many years of experience. Frion is a large professional care organization that provides support to people with intellectual disabilities. The Windesheim Research group IT Innovations in Health Care studies the development and use of ICT in healthcare.

When it comes to the interests of the various parties in the network, our starting point in any home care service is the client's concern. The client is interested in the provision of a certain degree of independent living, an effective means of communication, a good quality of life, better safety and more comfort. For the client the home care service obviously has to be easy to use and non-intrusive. With uncertain finance prospects the costs of the service are also a concern for the client. The professional care organization's interest and the formal carer's interests partly overlap with the client's interests: safety, independence, communication and quality of life are part of their main concerns. Putting the clients' welfare first at all times, Frion is looking for innovative ways to let their cients live their lives on their own terms, being in control of their own lives as much as possible. Efficiency, labor savings and increased service levels would be means to reach Frion's innovative objectives. Integration and embedding the service in their existing offerings would also add to reach the care organization's objectives. The insurance company's interest are guided by support among caregivers and clients, health benefits like higher quality of living and self-reliance, reduced health care costs, workload reduction in care and replacement of existing care (substitution). The insurance companies are mostly bound to the national agreements (the Dutch government decides which care is included in the basic package of insured care), but they do look for ways to enhance their image to attract and retain customers. An innovative home care service could be a good unique selling proposition (USP) to them. Besides the government's guidelines, the municipalities are expected to be led by efficiency, cost-effectiveness, necessity and practicality.

Stichting Trivici's interest in the Hightech@home care service is to develop and market a home care system which encourages low cost, interoperability and enables plug- and-play home care technology. Safety, independence, increased service levels and quality of life are the pointers where Stichting Trivici works with. VAC Thuistechnologie is looking for unique solutions to solidify its market position in the Dutch home care market. The Hightech@home project is a stable starting point for further development of open source home care concepts for this commercial partner. The Windesheim Research group IT Innovations in Health Care aims to provide and produce knowledge on how to design and implement innovative health care solutions and aims to spread this knowledge towards other projects and publications.

5) Financial Domain

On the basis of an analysis of the financial domain, the way the service intends to generate revenues and the way risk, revenues and investments are divided among the various actors in the network, it is clear that there are various revenue models possible, depending on the target group and buyer within this target group.

Central to the Hightech@home care service is the client as target group. This client can be the individual buyer of the home care system. In that case the service can take a revenue from for instance hosting, support service, extensions and other services as we have seen in the examples of open source business models. The fee can take different forms, ranging from subscription based on monthly / yearly fee to a usage fee based on actual services used. There are different levels of subscription possible (basic – premium) and price-wise the client's concern regarding price will probably be leading. The informal carer, another potential target group, can also be regarded as the potential individual buyer. In this case the revenue model could be quite similar to the model in which the client is the one who purchases the system.

Considering the other prospects identified in the service domain, the professional care organization comes to the fore. In case this party decides to purchase the home care service system, the revenue model will be quite different, since we are dealing with an organization instead of an individual. Possibilities of installation / implementation fees, hosting, consulting and support service and organizational customizing with open license or commercial license as shown in the various open source business models come into view. Revenues can be made through various forms from subscriptions to lump-sum delivery with service contracts. In case the care organization has an in-house IT service, a once-off implementation contract also belongs to the possibilities. Although the municipality and insurance companies are not part of the Hightech@home project, it is not unthinkable that they would be interested in purchasing the home care service as an extra facility for their citizens or customers. Since there are large-scale organizations too, revenue forms here could resemble the revenues to begotten from the professional care organization as buyer. The proven model of platform creation for shared services comes into view as a possibility here.

The amount and forms of revenues vary greatly among the different potential target groups and their investment potential. The scale of the implementation and intensity of the use of the services will partly determine the revenue models and thereby the business modeling.

6) Results

Working out the value propositions of the Hightech@home care service in the first step to business modeling, brings the awareness that this intended home care service can add value to more than just the client and carers. Although the client and the carers are the primary target groups, the home care service also has the potential to deliver value for the professional care companies and even to the insurers and the municipality. Straightaway there is the confrontation with the many parties in the home care landscape and with the fact that the target groups have not been determined yet. The intended target groups and their needs are central and largely determine the course of the development of the business model. Therefore there is the need to take action in this first step of the business modeling. Previous assessment showed how complex the Dutch home care landscape is and how different the routes the home care system can take to market. These observations call for the drafting of scenarios per target group to be written out at the outset of the project.

The mapping of intended target groups and their accompanying needs in the scenarios are key to determining the right value proposition per target group for the Hightech@home care system. It is no use to hurrying to phase two – the design phase - of service at this very moment; one is simply not able to determine the user requirements if the target groups and their needs are not known. Unambiguousness at the first phase is obviously fundamental.

The technology scan in the inventory phase of the technology domain focusses on the technical realization of the value proposition of the Hightech@home care system. The first thoughts were on the client and care organization as target groups, and it is clear there has not been enough though on the potential target groups of municipality and insurers yet. All the probable value propositions in the service domain have not been described so far, therefore the technical options for realization cannot be drafted properly yet. For how to know what to develop if it is not clear what the Hightech@home innovation has to deliver? The eHix framework sends us straight back to the first cell of value proposition as a no-go and puts us to the task of charting all the target groups, needs and values per scenario as detailed as possible.

Leaping to the organizational domain, the target groups are joined by the project parties in the value network of the Hightech@home care system. By charting the actors' interests in a complete stakeholder analysis one gets clarity about the joint interests of the parties and is able to value the expertise in the project. Since the technical requirements of the Hightech@home care system are not known in detail yet, one cannot establish whether or not there is enough expertise in the project team. This has to be re-evaluated and may be in need of adjustment after the scenarios in the inventory phase of the service domain have been worked out.

The stakeholder analysis of the project also makes it possible to specify the evidence needed to persuade the potential target groups for the Hightech@home care system. Since this evidence has to be gathered in the experiments and pilots, it is essential to be clear on the results before setting up the experiments and pilots for the Hightech@home project. The evidence needed will of course vary per scenario. If it turns out that particular evidence needed for a target group – for instance labor savings for the insurer - cannot be established, the evidence may become a determining factor in the selection of scenarios for the Hightech@home care system. The stakeholder analysis is also most useful in selecting the scenarios in the financial domain. The analysis shows the financial potency of the potential target groups within the Hightech@home network. The clients and informal carers obviously have a lower budget to spend than the care organizations, insurers and municipalities. If the Hightech@home project's aimed low-cost care service turns out to be out of reach for them, the other prospects come to the fore. In this manner the stakeholder analysis facilitates the selection of the likely scenarios.

The revenue of the Hightech@home care system and its various forms diverge among the different potential buyers. Purchase and use of the home care system by the individual client or individual informal carer ask for a very different implementation / service model and use than purchase by the professional care organization. When the purchase is made by insurers or municipalities, the scale of service and intensity of use even broadens. At this point in the project one can only make an inventory of the possible revenue models based on literature research of open source models. Only after all the scenarios of the Hightech@home care system are mapped, the revenue models can be identified efficiently and further steps can be taken towards viable business models.

IV. CONCLUSION

The aim of this paper is to elaborate on the factors that play a role in the initial development of a feasible business model for open source home care technology, using the Hightech@home care system as an example. First of all literature overview pointed out the specifics of open source software and business models for open source software. Based on open source software, there are various established business models, from adding services and products around the software and creating platforms to share to hybrid models of open and closed licensing. Creative solutions are already proven according to literature. What makes open source software especially challenging in tailoring it to a market is the numerous actors in the open source software development. The large community of developers and testers of the open source software has to be kept in mind by building flexibility into the business model for the home care technology market.

What characterizes the home care technology market? Literature overview describes the home care market as complex. Especially in the Netherlands the home care market is a fragmented landscape with demanding parties and various routes to this market. To make sure all options are explored, the open source business model for this scenery has to be developed in a thorough and methodical matter, making sure a solid foundation has been laid in the first stages of the business modeling.

What are the particulars to be considered in the early stages of business model development for open source home care technology? As the first inventory stage of business modeling, charting the value proposition for the Hightech@home care service shows, it is most efficient to first work out a complete stakeholder analysis, followed by all possible scenarios in a consistent manner. Charting the value proposition of the care service brings awareness of other potential target groups which cannot be put aside. The intended target groups are central to the course of the business model. Absolute clarity on these groups, their needs and value propositions is vital before jumping into further development of the business model.

Making choices in the inventory phase is not advisable. On the contrary, the inventory phase is best used to chart the full potential of the home care service. The writing out of scenarios per prospective target group ensures a comprehensive view of the potential of the home care service and a good foundation to develop a vital business model. The analysis of all the stakeholders and their interest in the network also provides the essential criteria for the feasibility of the scenarios. These criteria like income, scale of use or evidence needed for persuasion make it possible to make a preliminary selection of scenarios if necessary or preferred.

The question what business model would be most feasible for the Hightech@home open source technology innovation cannot be answered yet in these early stages of development. This is logical since since not enough research has been done into the full potential of the service. The use of a structured method like the eHix method to build up the business model, a method where one is only allowed to move to the next phase if the previous phase is completed, ensures an efficient and effective route to business modeling.

It is in the nature of business and entrepreneurs to fast track ahead on the road to success. However, laying foundation of business models by starting with an in-depth stakeholder analysis, working with probable scenarios and diligently following a business modeling method is the way to develop feasible business models that show the potential of the innovation to its full advantage.

ACKNOWLEDGMENT

The authors wish to thank TechForFuture, Centre of Expertise HTSM (http://www.techforfuture.nl) for its financial support. The center is an initiative of Saxion and Windesheim, Universities of Applied Cciences and was established with the support of the Province of Overijssel.

REFERENCES

[1] M. Hettinga, E. Gyaltsen-Lohuis, A. de Keizer and J.M. Nauta, (2014). "Open Source Home Care Technology. Technical design and development, user research, cost-benefit

analysis, and business modeling," The Sixth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2014) IARIA, March 2014, pp. 267 – 270, ISBN: 978-1-61208-327-8

[2] Raad voor de Volksgezondheid (NVR), Discussienota thuiszorg, Zoetermeer: NRV, 1989.

[3] F. Hecker, "Setting up de shop: The business of open-source software," Institute of Electrical and Electronic Engineers, pp. 45-51, Jan./Feb. 1999.

[4] T.H.F. Broens, R. Huis in't Veld, M. Vollenbroek-Hutten, H. Hermens, A. van Halteren and L. Nieuwenhuis, "Determinants of successful telemedicine implementations: a literature study," Journal of Telemedicine and Telecare, vol. 13, pp. 303-309, 2007.

[5] R. Menko, M. Hettinga, R. Janssen, S. Visser and T. Haaker, "Applying the STOF Business Model Framework in Ehealth Innovations," The Fifth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2013) IARIA, March 2013, pp. 108-113, ISBN: 978-1-61208-252-3

[6] T. Haaker, "Business modellen voor ICT-intensieve zorginnovatie," Nictiz Kennisartikel 10025, Den Hague: Nictiz, 2010.

[7] H. Bouwman, H. de Vos and T. Haaker, Mobile Services Innovation and Business Models, Berlin Heidelberg: Springer, 2008.

[8] www.ehix.nl [10-12-2014]

[9] J. West, "Value Capture and Value Networks in Open Source Vendor Strategies," 40th Hawaii International Conference on System Sciences (HICSS 2007) IEEE Press, Jan. 2007 , p. 176, doi:10.1109/HICSS.2007.600

[10] E.J. Martin, "Open Source: is it Right for you?" EContent, Jan./Feb. 2013. [Online]. Available from: http://www.econtentmag.com/Issues/3691-January-February-2013.htm

[11] S. Krishnamurthy, "An analysis of Open Source Business Models," [Online]. Available from: http:/papers.ssrn.com

[12] E. Capra and A. Wasserman, (2008). "A framework for evaluating managerial styles in open source projects," Proceedings of the 4th International Conference on Open Source Systems (OSS 2008) Springer, pp. 1-14, ISBN: 978-88-903120-1-4

[13] www.opensourcematters.org [10-12-2014]

[14] P. Korhonen, Business models and motivation of companies that take part in open sourse software development project, Oulu: University of Oulu, 2013.

[15] D. Riehle, "The single-vendor commercial open course business Model," Information Systems E-Business Management, vol. 10, pp.5–17, 2010.

[16] J. West and J. Dedrick, "Innovation and Control in Standards Architectures: The Rise and Fall of Japan's PC-98," Information Systems Research, vol. 11, pp. 197- 216, 2000.

[17] R.T. Watson, M. Boudreau, P.T. York, M.E. Greiner and D. Wynn Jr., "The Business of Open Source," Communications of the ACM, vol. 51, pp. 41-46, 2008.

[18] K. Gary, H. Koehnemann, J. Blakley, C. Goar, J. Mann, and A. Kagan, "A Case Study: Open Source Community and the Commercial Enterprise," Sixth International Conference on Information Technology: New Generations (ITNG 2009) IEEE Press , pp. 940-945, doi: 10.1109/ITNG.2009.313.

[19] N. Munga, T. Fogwill and Q. Williams, "The Adoption of Open Source Software in Business Models: a Red Hat and IBM Case Study," Proceedings of the 2009 Annual Research Conference on the South African Institute of Computer Scientists and Information Technologist (SAICSIT 2009) ACM Publications, pp. 112-121, doi: 10.1145/1632149.1632165.

[20] J. Kemperman, J. Geelhoed and J. Op 't Hoog, Briljante Business Modellen in de Zorg. Baanbrekende benaderigen voor betere en betaalbare zorg, Den Hague: Academic Services, p. 44, 2014.

[21] D.A.J. Dohmen, De 'e' van e-Health. Juiste energie voor optimale implementatie van moderne technologie in de zorg thuis Inspectie voor de gezondheidszorg, Enschede: Universiteit van Twente, 2013.

[22] KPMG, Accelerating Innovation: The Power of the Crowd. Global Lessons in eHealth Implementation, Manchester: University of Manchester, 2012.

[23] www.mezzo.nl/hoeveel_mantelzorgers_zijn_er_in_nederland [10-12-2014]

[24] www.innovatieroutesindezorg.nl/Veranderingen-2015 [10-12-2014] and www.zorgvisie.nl/PageFiles/80908/Jan%20Telgen.pdf [10-12-2014]

[25] S. Visser, T. Haaker, R. Menko and M. Hettinga. (2012). "Succesvol Innoveren in eHealth met eHix" [Online]. Available from: http://www.ehix.nl/Documents/Overig/whitepaper_ehix.pdf

[26] M. Hettinga, Telezorg: van Buzz naar Business? Lectorale rede, lectoraat ICT-innovaties in de Zorg, Zwolle: Christelijke Hogeschool Windesheim, 2009.

[27] www.dekra.nl [10-12-2014]

[28] www.nen.nl [10-12-2014] and www.qaeh.nl [10-12-2014]

Combining User-Centered Design with the Persuasive Systems Design Model;

The Development Process of a Web-Based Registration and Monitoring System for Healthcare-Associated Infections in Nursing Homes

Nienke Beerlage-de Jong[1], Andrea Eikelenboom-Boskamp[2,3], Andreas Voss[2,3], Robbert Sanderman[1,4], Lisette van Gemert-Pijnen[1]

[1]Center for eHealth Research & Disease Management
University of Twente
Enschede, the Netherlands
n.beerlage-dejong@utwente.nl,
j.vangemert-pijnen@utwente.nl

[2]Medical Microbiology and Infectious Diseases
Canisius Wilhelmina Hospital
Nijmegen, the Netherlands

[3]Medical Microbiology
Radboud University Medical Centre
Nijmegen, the Netherlands
eikelenboomandrea@gmail.com, vossandreas@gmail.com

[4]Department of Health Sciences
University of Groningen
Groningen, the Netherlands
r.sanderman@rug.nl

***Abstract* — To improve the usefulness and user-friendliness of eHealth interventions, a framework for the development of eHealth technology has been developed. It combines User-Centered Design with the Persuasive System Design model. The current paper is aimed at offering a (practical) method for the integration of these two design approaches. Via a case study, the paper demonstrates how User-Centered Design and the Persuasive Systems Design model can complement and mutually enrich each other.**

Keywords - User-centered design; Persuasive design; Web-based; eHealth technology

I. INTRODUCTION

In this paper, we present an extended version of work presented at the Sixth International Conference on eHealth, Telemedicine and Social Medicine (eTELEMED 2014) in Barcelona, Spain [1].

In healthcare, and specifically in infection prevention and –control, many apps have been developed. They are intended to improve quality of care and patient safety, but are definitely not always successful. They often don't fit with the way the intended end-users work or think, or they fulfil a non-existent need.

Therefore, the Center for eHealth Research and Disease Management, has developed a framework to develop apps that are successful in daily clinical practice: the CeHRes Roadmap [2]. This framework makes use of User-Centered-Design (UCD) and the Persuasive Systems Design (PSD) model.

Within UCD, during the entire development and design processes of a technology, close cooperation is sought with end-users and other stakeholders. This is done to achieve an optimal fit between the newly developed technology, the context in which it is used and the way the intended end-users work or think [3]. User problems can thus be recognized and prevented. For the prevention of such issues, and to motivate and support the end-users, the PSD model is used [4]. According to this model, technology can be developed with the aim of changing the users' behavior, without using coercion or deception [5][6]. Thus, the technology itself can make a substantial contribution to its own success. We, the Center for eHealth Research and Disease Management, are convinced that the combination of these two models can be of great added value for the development of successful eHealth technology.

This paper describes such a development and design process, with the aim of offering a tool for the development of persuasive and user-friendly eHealth technology. To do so, the different stages of the iterative development process (following the CeHRes Roadmap, see Fig. 1) of a single app is described.

Whereas most development processes start with the design of a technology, we consciously take a step back. The first stage of the CeHRes Roadmap is the *Contextual Inquiry*. In this stage, the developer seeks cooperation with known stakeholders of the technology to explore the context in which the technology must be used, and which preconditions it should meet [2].

Based on the outcomes of this inquiry, and again in cooperation with known stakeholders, it is then studied whether there are any potentially relevant stakeholders missing, and what values and needs they have. This is all done during the stage of *Value Specification* [2].

Then, in the *Design* stage, attention is given to the question of how solutions to the found problems and preconditions can be incorporated within the technology. A prototype of the technology is developed and evaluated by intended end-users [2].

As mentioned before, within this paper, the development process of a single app will be described, i.e., the Prevalence App. Aim of this app is to support Elderly Care Physicians in nursing homes during the registration of their clients during prevalence measurements of Healthcare Associated Infections (HAIs).

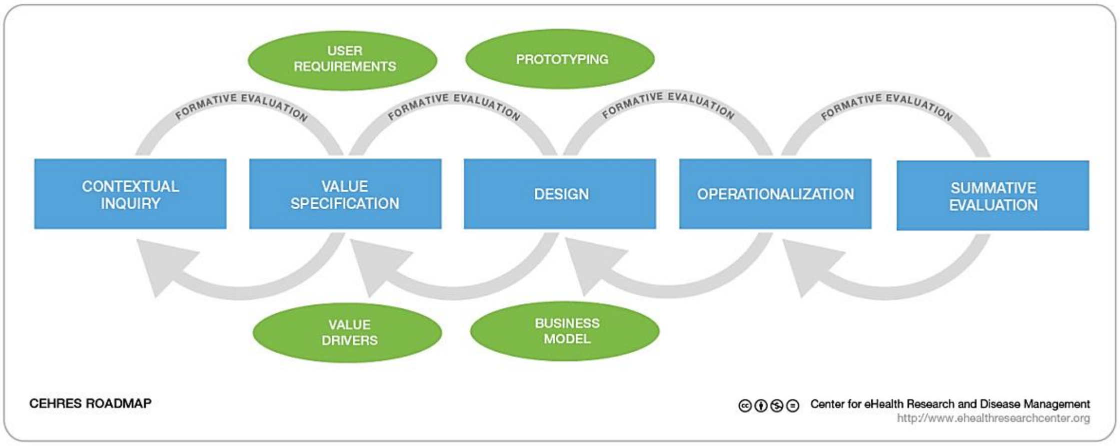


Figure 1. The CeHRes Roadmap [2].

HAIs form an increasingly pressing threat to patient safety [7]. To be able to adequately protect patients against this threat, one of the first steps is to gather knowledge on its occurrence [7]. For hospitals, already a vast amount of surveillance data is available [7][8][9][10][11][12][13]. For other healthcare institutions (such as nursing homes), prevalence studies have more recently begun to take place. The results of the first prevalence study of HAIs in Dutch nursing homes were published in 2011 [14].

To enable prevalence research in nursing homes, data must be collected on all clients that are present in the nursing homes at one point in time. The success of such data collection is entirely dependent on the willingness and capability of the elderly care physicians to register their clients in a correct and timely manner. Preferably, the registered data should be collected in a standardized way to fit other (e.g., nation-wide) surveillance programs [7][8]. The prior registration system consisted of many (often irrelevant) long and very complicated questions, that are prone to interpretation errors. Physicians sometimes doubted the exact meaning of the questions, which could harm the reliability of the data. Registration via this system also took its users quite some time.

The eventual aim of the new registration system is to use user-centered and persuasive design to optimally support nursing home physicians during the registration of their clients in HAI prevalence measurements. However, for the current paper, that is not the main purpose or focus. Here, we describe the process towards achieving such aim. To reach the end goal, our approach is applied. In this paper, we describe how UCD and PSD can together contribute to the development of a successful registration system for HAIs in nursing homes.

The rest of this paper is structured as follows. In Section II, the methods that are used in this study are described: an expert-discussion to gain insight into the current situation (Section II-A); a questionnaire study to analyze the users' needs (Section II-B); in-depth interviews and scenario-based user-tests to analyze the system's user-friendliness and persuasiveness (Section II-C). Then, in Section III, the results of the different parts of the study are described: Section III-A is about the current situation, Section III-B concerns the users' needs and values, and Section III-C goes into the development and evaluation of the prototype. Finally, the study is discussed (Section IV) and conclusions are drawn, about the added value of combining UCD and PSD (Section V).

II. METHODS

Here, the different steps, that are taken to follow the CeHRes Roadmap towards the development of a user friendly and persuasive eHealth technology are described.

A. *Contextual Inquiry – An Expert Discussion*

This project started with a request from iPrevent [15] to aid in the development of a new 'mobile' registration system. iPrevent is a regional infection control network within which healthcare institutions, elderly care physicians, medical microbiologists and Infection Control Professionals (ICPs) work together to structurally offer high quality care in the field of infection prevention [15].

Instead of immediately initiating the development and design of a technology, the CeHRes approach required us to take a step back, to perform a Contextual Inquiry [2]. Therefore, to gain insight into what iPrevent specifically wants or needs from this project, and into the prerequisites for the registration system, an expert discussion was held.

1) *Participants*

Participants in the expert discussion were invited to the meeting via e-mail. Two of them are project leaders (a medical microbiologist and an ICP) of the iPrevent network, they are also the initiators of the development of a Prevalence App. One of the project leaders is also the data analyst, who is responsible for collecting, analyzing and reporting the registration data (both for publication purposes and as feedback towards the participating nursing homes). Also, two behavioral researchers, specializing in UCD for persuasive technology, participated in the expert discussion.

2) Procedure

The expert discussion included relatively few people. Therefore, it was decided to keep the setting and discussion informal. The expert discussion thus took the shape of a conversation, during which the behavioral scientists did ask additional questions to acquire all the information that was necessary for the development of a new registration system, e.g., who have an interest in the performance or results of the prevalence measurements, and why.

3) Data-analysis

Notes were made by one of the behavioral researchers, during the entire discussion. These were used to form a 'working document' that would be used as a communication tool (between behavioral researcher and project leaders) during the entire development process. In the document, the outcomes of the discussion were complemented with literature about the registered HAIs and national surveillance programs. This was done to make sure the registration system would fit the context as described by the experts.

B. Value Specification – A Questionnaire

Then, following the CeHRes Roadmap, the Value Specification stage was initiated, to gain insight in what values end-users had (concerning a registration system) and whether the project aims match these values [2]. For this purpose an online questionnaire was developed, based on the results of the Contextual Inquiry.

1) Participants

Elderly care physicians were invited to participate, by one of the project leaders, via e-mail. They were informed about the Prevalence App that was being developed, and that their input was needed to make sure that it would fit their needs and work processes. A total of 24 elderly care physicians, who worked at different nursing homes within the iPrevent network, were willing to participate in the study. Their ages ranged from 30 to 61 (mean age 47 years). Most participants were female (19 female vs. 5 male).

2) Procedure

The invitation e-mail included a direct link to the questionnaire. The questionnaire consisted of some basic questions about the demographics of the respondent, his / her technology use and how the registration is currently being performed. Then, questions were asked about how relevant the physician feels the prevalence measures are, and how, when and via what kind of device the physician prefers to register clients. Additionally, questions were included about what information or feedback the physician expects to get from the registration system and what would motivate them to adopt a new system.

3) Data-analysis

Questionnaire results are analyzed to obtain descriptive statistics. Further statistical analyses were deemed redundant, given the developmental purpose of the study. Answers that are given to open-ended questions, were summarized if they overlapped and then their frequencies were analyzed as an indication of how many respondents share a specific opinion.

C. Design – In-Depth Interviews and Scenario-Based Tests

Next step in combining user-centered and persuasive design, is to start the Design stage of the CeHRes Roadmap [2]. Thus, to validate the questionnaire data and to optimize the user-friendliness and persuasiveness of the prototype, scenario-based user-tests and in-depth interviews with end-users are performed.

1) Participants

Participants were invited to the study by one of the project leaders. If they agreed, they were contacted by the researcher to schedule a meeting at a location that was convenient for the participant. Four female elderly care physicians who worked in different nursing homes within the iPrevent network participated. Their ages varied from 33 to 59 (mean age 45 years). One of the nursing homes they worked at, already used Electronic Client Files (ECFs), the others expected they would start to do so in the near future.

2) Procedure

First of all, a prototype of the Prevalence app was developed, in close cooperation with an ICP, using Balsamiq software [16]. It was based on the outcomes of the expert discussion, about the context of requirements of the registration. The prototype incorporated elements of the PSD model [4]. Not all elements of the PSD model were deemed relevant, only elements that are relevant in this specific context and for this specific app have been applied.

Then, two scenarios were developed to be used in the user-tests. They were developed in close cooperation with an ICP and made use of literature on the HAI definitions. The scenarios addressed critical issues for registration e.g., a client with multiple infections or a lab test that has been performed without the results being known yet.

The physician was instructed to talk out loud during the entire user-test, not only mentioning what she thought, but also what she saw or sought, did or wanted to do during the registration of the fictional client. The entire conversation was audio recorded, with permission of the participant.

3) Data-analysis

The conversations, including both interview and user tests, took about 45-60 minutes each. Audio recordings of the conversations were transcribed verbatim and analysed using a code book, which aided in the structuring of data. Some examples of the codes are given in Table 1. All codes were combined and the frequencies with which they were mentioned were analysed.

TABLE I. EXAMPLES OF CODES USED FOR ANALYSIS OF USER-TESTS

Examples of codes used for data-analysis		
Category	***Code***	***Description***
Contextual Inquiry	C1	Subject describes a problem that is experienced in the current work process
Value Specification	V1	Subject thinks working with the new system might be faster
Design	D1	Subject thinks the order of the items in the mock up is wrong
Operationalization	O1	Subjects talks about a possible barrier for using the new system

III. RESULTS

Aim of the current paper is to describe how UCD and the PSD model can simultaneously and complementarily be applied, to develop successful eHealth technology. For that purpose, the development process of the Prevalence app is used. In this section, we will describe the results that were rendered during the different stages of the CeHRes Roadmap.

A. *Contextual Inquiry – Description of the current situation*

First of all, the Contextual Inquiry generated more in-depth insight into the current situation (the situation prior to the development of a new registration system).

iPrevent has, over the last years, worked together with the approximately 30 nursing homes within their network, to perform annual prevalence measurements for HAIs in nursing homes. This implies that the elderly care physicians are required to once a year register all relevant data about the residents that live in their nursing homes. This not only entails information about the presence of HAIs, but also about the presence of risk factors for the occurrence or spread of infections. Risk factors are e.g., the use of antibiotics or catheters, or staying in a room with multiple other clients. Inherent to the fact that it is a prevalence measurement, all clients must be registered within a short timeframe around a reference date.

The content of the registered data is largely determined by the given definitions of HAIs. These definitions were developed by the regional network in cooperation with the national surveillance system (PREZIES) [17]. In the hospital setting, surveillance is performed by trained infection control nurses. The use of the definitions in the nursing home setting, where registration is performed by physicians with little or no experience with surveillance, registration is far more complicated. The prior registration system directly used the definitions of HAI as questions. It did not, however, offer additional explanation or clarification about their meaning, or any other kind of support with answering the questions. Physicians had indicated that they sometimes debated with their colleagues about how to interpret and answer a certain question (see Quote 1). This, of course, caused some issues with standardization and analysis of data.

> (Quote 1 – Originally in Dutch) "[…]The way the questions are asked. They are not always clear, and that's important. Now, I sometimes have to go to one of my colleagues to ask them, like: hey, how do you answer this? And it is important, that everyone registers in the same way."

The prior registration system consisted of an online questionnaire, developed by experts in the field of Infection Prevention and Control. The questionnaire consisted of a long list of complicated questions (see for example the screenshot of the prior registration system in Fig. 2). Many of them are irrelevant for most residents. For example, if a client does not use an antibiotic, the question about what an antibiotic is used for, is rather redundant. Also, most questions were presented on a single page. Physicians thus had to scroll down for quite some time. They had to read all questions, including irrelevant ones, to check whether they applied to the client.

Figure 2. Screenshot of the prior registration system

Furthermore, an increasing amount of nursing homes (>30) participate in the prevalence measurements that iPrevent performs. Thus, increasingly large datasets are collected. Data processing, data analysis and presentation of feedback of the results to the nursing homes were all performed by a single data-analyst. This will soon no longer be feasible. Therefore, project leaders (and the data-analyst) would like the system to perform these tasks automatically.

Finally, project leaders requested that the new system would be 'mobile' (to be used on a smartphone) to enable bedside registration of clients by elderly care physicians.

B. *Value Specification – Users' Needs and Values*

For the second stage of the CeHRes Roadmap, the questionnaire that was used resulted in insight in the users' values and needs.

First of all, most (83,3% out of 24) participants indicated that they did consider prevalence measurements of HAIs in the nursing home setting to be relevant. They found it important since it contributes to gaining insight into the current status of HAIs in nursing homes, and enables organizational policies to be adapted to the findings. One of the participants also considered it to be of great importance for the quality of care (see Quote 2).

> (Quote 2 – Originally in Dutch) “Thus far, too little research has been performed, among the nursing home population, to be able to act in a meaningful and evidence based way.”

When asked what kind of device they would like to use for the registration, most subjects indicated they preferred a PC (50,0%) or laptop (20,8%). The other subjects preferred to use a smartphone (12,5%), tablet (8,3%), or paper (8,3%).

Most important reasons for users to be willing to use the new registration system were: (1) if they can interrupt registration without losing data; (2) if the new system is more user friendly; (3) if it can be opened simultaneously with ECFs; (4) if clear insight is given in the results; and (5) if registration can be performed faster.

One of the prerequisites that were found during the expert discussion said that it would be desirable if registration could be performed at the residents’ bedside. This would enable the physicians to directly see how the resident is doing. However, the questionnaire showed that none of the physicians considered this to be desirable. Most of them (54,2%) did indicate that it would be of added value for them to be more flexible in the location in which they register their clients, but didn’t want to do so at bedside. Also, 41,7% said that it would not be of added value at all since they just liked to register their residents in their offices. One physician (4,2%) wanted to register her clients in the department’s office, with the client files at hand.

The in-depth interviews gave even more insight into the situation. The clients that are to be registered with the prevalence app are mainly elderly people, who have health issues. Unfortunately, these elderly clients are quite often also somewhat lonely. So, upon doing their rounds or checking up on their clients, the elderly care physicians had experienced that clients were in need of attention and wanted to interact with them. The physicians felt it would be poor bedside manners to be standing next to a client, while being entirely focused on the registration. At the same time, if they were to pay more attention to the client, they feared that the registration would take up to much time, or would be prone to errors due to a lack of concentration. They are therefore opposed to bedside registration of clients.

At the same time, physicians did indicate that being ‘mobile’ during registration would have advantages. They explained that their nursing homes were (going to be) using ECFs. This is software that contains highly personal and private information about the residents. Therefore, many safety measures have been taken to protect this information. Because of one of these safety measures it is impossible to simultaneously open the ECF and the World Wide Web. In practice, this meant that elderly care physicians had to open the ECF, and write down all the information about all of their clients that they needed to register in the prevalence measurements. Then they had to close the ECF, open the registration system and enter the information they had written down. Not to mention that if they had forgotten any information, the entire procedure had to be repeated. Therefore, some subjects did want registration to be possible on a mobile device, but for reasons that differed from what was expected, i.e., so they could simultaneously open the ECF on their pc and the registration system on the other device.

C. *Design – Developing and evaluating a prototype*

The results that were generated in the Design stage are twofold: (1) we can now describe how we applied certain elements of the PSD model to the prototype of the prevalence; and (2) we can make a brief analysis of the user-friendliness and perceived persuasiveness of the prototype of the app.

1) *The use of the PSD model in the Prototype*

Here, a description is given of how elements of the PSD model were incorporated in the prototype. This is intended to benefit other eHealth technology developers, who might use it as an example for how the theoretical constructs of the PSD model can be applied in practice.

a) *Primary Task Support*

One of the main concerns with working with the prior registration system was that working with it required the physician to read through many irrelevant questions for every client. One of the most important elements of PSD that were to be used in the new system was therefore *tunneling* [4]. Tunneling is defined by Oinas-Kukkonen and Harjumaa as ‘Using the system to guide users through a process or experience’ [4]. The prototype system was designed to guide the user through the process of registration: questions to be answered are dependent on the answers given to prior questions. Thus, the entire system is one big decision tree, to make sure every client is registered via the shortest (fastest) possible route.

Another concern with the prior registration system was the complexity of the used questions, a single question could actually consist of multiple individual questions (see for example Quote 3). To reduce the complexity, and thus the risk of interpretation errors, *reduction* was used. The aim of reduction is ‘to reduce complex behavior into simple tasks, to help users perform the target behavior’ [4]. This was done by translating complex and long questions into multiple shorter and easier questions with a routing structure between them. For example, the originally used question for Gastro-Enteritis, was rather lengthy and complicated (see Quote 3).

Does the resident have Gastro-Enteritis? The diagnosis Gastro-Enteritis is given if one of the following symptoms occurs in the client:

- Three times or more diarrhea (different from normal for this client, frequency is not applicable when using incontinence materials)
- Diarrhea and two of the following symptoms: fever, vomiting, nausea, stomach ache, stomach cramps, blood or mucus in feces.
- Vomiting three times within 24 hours, without any additional symptoms (if vomiting is not associated with medicine use)
- Vomiting and two of the following symptoms: fever, nausea, stomach ache, stomach cramps, blood or mucus in feces.

Quote 3. Question about Gastro-Enteritis in the prior registration system (Originally in Dutch)

The different elements of this one question were split up into multiple questions. These were one-by-one presented to the users. A build-in logic system directed the users to the appropriate follow-up questions on different screens. Some of these screens are shown in Fig. 3.

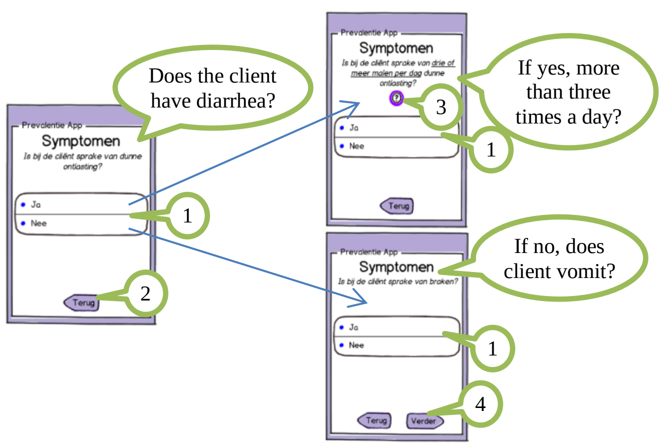


Figure 3. Part of the screens of the prototype used for Gastro-Enteritis; Example of Reduction: *(1) Answer options: Yes / No; (2) Button to return to previous screen; (3) Help-button for additional information; (4) Button to continue to next screen.*

As mentioned before, it was of great importance that registration would be faster and could be paused without losing data. To enable this, elements of *tailoring* are applied to the prototype. Tailoring, according to Oinas-Kukkonen and Harjumaa, stands for 'the adaptation of the offered information to the potential needs, interests, personality, usage context, or other factors relevant to a user group' [4]. In this case, the system tailored the information to the usage context factors. Every nursing home was given a unique log-in code and password. A physician had to log-in once, and was then able to continue registering their clients one after another. Moreover, when starting the system, physicians are given two options: to register a new client or to edit data of an existing client (see Fig. 4). For the latter, an overview was generated of all clients that had previously been registered by that specific nursing home. Clients of other nursing homes are not shown. By selecting a client, their data are shown and can be adjusted.

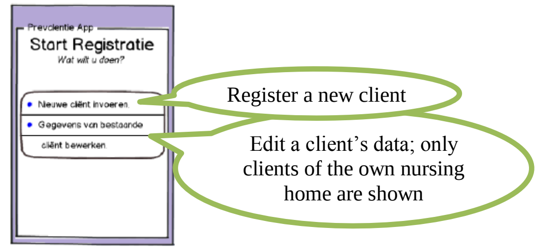


Figure 4. Example of a screen that leads to Tailored information

b) Dialogue Support

Dialogue support stands for the feedback that a system gives to its users, to help them keep moving towards their goal (completing the registration) [4]. In this case, the prototype registration system requires physicians to indicate which pathogens caused an infection. However, during the user tests it became clear that within nursing homes, little funding is available to perform the laboratory tests to acquire this knowledge. Furthermore, if a laboratory test was performed at the moment of registration, its results were not always known yet. Therefore, in the final registration system, questions are added to ask whether a laboratory test was performed. If yes, the system inquires whether its results are already known. If the latter question was answered with 'no' a pop-up screen appeared. This screen *reminder* was shown, which means that the system reminds its users of their target behaviour [4]. In this case, the reminder made the physician aware that lab results should be added later.

Also, *suggestion* [4] is added to the final registration system. According to Oinas-Kukkonen and Harjumaa 'systems offering fitting suggestions will have greater persuasive powers' [4]. Within the prevalence app, this is done by letting the system offer fitting suggestions for reaching the goal of registering clients. A suggestion might concern a suggested behaviour, but in the case of the prevalence app, it consisted of suggestions in the sense of finishing the words that the user started to type. For example, physicians have to register all antibiotics that are used by their clients. However, the variety of antibiotics that exist is enormous and their names are complex. Initially, the intention was to let physicians scroll through an alphabetical list of generic names of all available antibiotics (see Fig. 5).

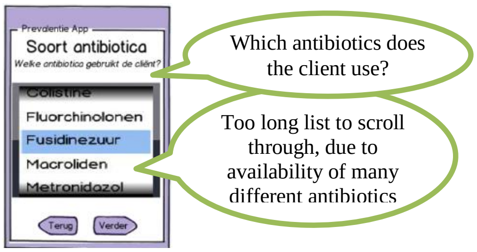


Figure 5. Example of the screen that subscribes the need for Suggestion

However, due to the large number of antibiotics that are available, this was still too time consuming. Therefore, in the final registration system, a search system in which the physician enters the first three letters of the antibiotic is added. The system then automatically generates only antibiotics that start with these letters. The same search system was then used for selecting the pathogens that have caused the infection.

c) System Credibility Support

Finally, in healthcare in general and for the registration of HAIs in nursing homes specifically, it is of great importance that the system is credible according to its users. Therefore, a website was created about the services provided by the registration system [18]. The Infection Manager website was developed within the (INTERREG IVa) EurSafety Health-net project [19]. This is a large and successful European cross-border project, involving many hospitals, microbiology laboratories and other healthcare institutions. It aims to increase patient safety by preventing (the spread of) HAIs. The part of the Infection Manager website about the registration system consists of e.g., background information about the project and the parties that were involved in the development process. This was intended to give more clarity about the *trustworthiness* and *expertise* of the project and its project members [4].

Lastly, in the final registration system, it was decided to add the EurSafety Health-net logo, to give the system more *surface credibility* [4]. This entails the initial assessment that users make about the credibility of the system, based on first hand inspection [4].

2) Analysis of the User-Friendliness of the Prototype

User-testing the developed prototype, is aimed at evaluating whether the elements of PSD have indeed contributed to the development of a user-friendly and persuasive eHealth technology.

Based on the performed user-centered scenario-tests, we found that issues with the prior registration system had indeed been resolved, or had at least been improved. For example, to implement the concept of tailoring, elderly care physicians now had to log-in to be able to register their clients. Besides allowing us to tailor the presented information, it also had other (unintended) benefits: some questions had become redundant, which could possibly speed-up the registration process (see Quote 4).

> (Quote 4 – Originally in Dutch) "Cause now, I have to indicate every time: Nursing home X, Department so-and-so. But we fill in the registration for one entire department. So it saves time if we would only have to fill it in once."

According to the physicians, an additional advantage of the log-in system, is that registration can be paused without losing the data (see Quote 5).

> (Quote 5 – Originally in Dutch) "Well, the advantage is that one can keep adding (new) clients. You could for example continue at a later moment. Such as when you only have time to complete the registration of 10 clients, then register 10 now, and just continue two hours later."

Still, major and minor adjustments had to be made in the mock ups. These concerned the clarity of wording, sequence of questions, completeness, user-friendliness, design and location of the buttons. For example, initially, there were two screens in our prototype for 'Aids', which asks whether the client uses any aids such as a catheter or tracheotomy; and 'Incontinence', which asks whether the client is incontinent. During the user-tests, the subjects had several comments about these screens. First of all, whereas we interpreted the term 'aid' as being a catheter of some kind, the subjects indicated that the term 'aid' to them meant 'walker' (see Quote 6). So, they suggested using a different term.

> (Quote 6 – Originally in Dutch) "Yes, we use the word 'Aid' for something completely different. We use this word for walkers. So I would try to come up with a different word here."

Furthermore, they found the screen about incontinence unclear. In one of the scenarios, a client was described who had a catheter. The participants indicated that, although incontinence is a possible reason for clients to get a catheter, they did not consider this client as being incontinent anymore (see Quote 7). Therefore, they said the option of having a catheter or stoma should be added to this screen.

> (Quote 7 – Originally in Dutch) "You see, this client is not incontinent, but has a urethra catheter… So this is strange. You should add catheter here I guess. Because with a catheter you are not really incontinent anymore."

As a result, the two screens where replaced by a new screen. This screen asks whether a client is incontinent. However, an additional answer option has been added, to indicate that a client has a catheter or stoma.

After the fourth user-test, no major issues where found anymore. Therefore, meetings with Information and Communication Technology (ICT) developers were held to further discuss how to incorporate the requirements into the registration system and to finally develop it.

IV. Discussion

As far as the authors know, this study was the first to demonstrate how User-Centered Design (UCD) and the Persuasive Systems Design (PSD) model can be combined, and can even complement each other, in the development of easy to use eHealth technology. It was intended to provide insight into the steps that might be taken for applying elements of the PSD model during the developmental stages of an eHealth technology. It did so, by describing the development process of a single app: the Prevalence App.

First of all, the use of UCD has proven its benefits: the constant and structural cooperation with end-users during the development process, gave us the opportunity to make it an iterative and reflexive process. This means that it was possible to evaluate the eHealth technology with end-users, in every stage of its development, and to (at any time) adjust the direction that it was going in. This aids in the dynamic development of an eHealth technology that fits its users' needs and context, and could potentially prevent high costs of re-design if major necessary adjustments are only found after final release of the technology.

Second, this paper has shown that the PSD model is indeed useful and applicable for the design of any kind of eHealth technology. However, and more importantly, we believe that it is the combination of the two theoretical frameworks above, within the CeHRes Roadmap, that has proven to be of the greatest added value. We feel that by 'simply' applying the PSD model to an eHealth technology, one would not be using it to its full potential. Merely incorporating elements of the PSD model in eHealth technology, does not necessarily mean that the end-users of the technology also have a need for these elements. If there is no such need, developers are at risk of using sledgehammers to crack nuts, to be greatly overreaching themselves. This paper has demonstrated that the success of a technology is not dependent on the quantity of persuasive elements that are used. Rather, the key to success is to use elements of the PSD model in a focused and user-centered manner. This is why the combination with UCD is so beneficial. Using UCD, for instance in the form of scenario-based user-tests, developers can expose which elements of the PSD model should be used, to cope with any kind of usability issues and to create an optimal fit with the end-users.

In their paper, Oinas Kukkonen and Harjumaa present the PSD model as being part of development process [4]. They indicate that before the actual design of the technology, the first step in that development should be to understand 'fundamental issues behind the persuasive systems' [4]. Then, 'the context for persuasive systems needs to be analyzed, recognizing the intent, event, and strategies for the use of a persuasive system' [4]. The integration of the PSD model within the Contextual Inquiry, Value Specification and Design stages of the CeHRes Roadmap offers a structure for the application of these steps in the development of a successful eHealth technology. It offers a practical framework by which elements of the PSD model can be used in practice.

We believe that the holistic approach using the CeHRes roadmap, has provided the opportunity to further ground the PSD model. It has proven to be a very suitable tool to integrate UCD and the PSD model in a way that developers not only profit of their individual strengths, but were they also complement and further strengthen each other. Via UCD, for which we used e.g., scenario-based tests, we found essential issues during the use of the prototype. Via the PSD model, we were able to address these issues, and to eliminate or at least decrease their presence.

Thanks to the CeHRes Roadmap, we developed a registration system to optimally support elderly care physicians in the correct and timely registration of their clients, taking into account the national prevalence studies [7][8] with which collected data should be compatible. Certainly, we still want to evaluate the speed, user-friendliness, fit with work processes, ease of use, clarity and persuasiveness of the final registration system. Thus, a summative evaluation is currently being planned. This evaluation will combine both qualitative and quantitative methods and will e.g., focus on user friendliness, speed of registration (both were found important by the end-users) and amount of errors that are made (important for the quality of the data). But for now, it can be said that the new system has already been used in two rounds of prevalence measurements, successfully registering over 3000 nursing home residents. It is web-based, so it can be used on any device capable of connecting to the world wide web. An example of a screen of the eventual registration system is given in Fig. 6.

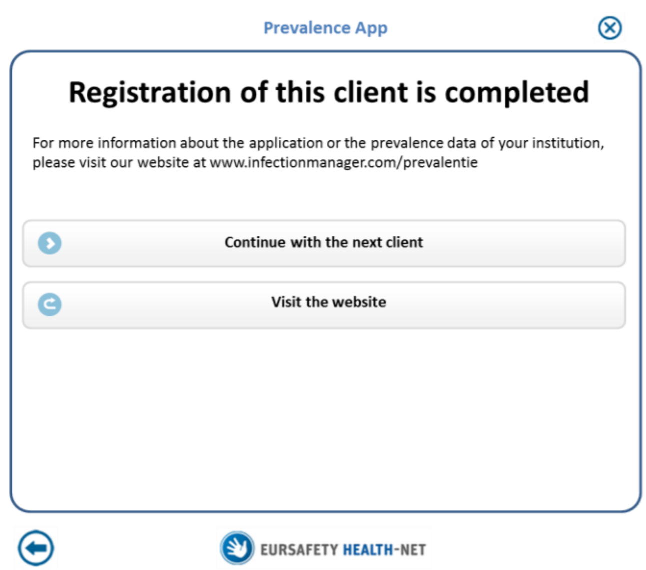


Figure 6. Example of the look and feel of the eventual registration system (Originally in Dutch)

A. Limitations

A limitation of this study is that it suffered from very strict and ambitious deadlines. This limited us in the amount of effort that we could put into the implementation stage of the development. Although this is thought to be a very relevant stage according to the CeHRes Roadmap, in this project its influence might be limited. The elderly care physicians who are the end-users of the registration system in this study, are obligated to use it if their nursing home participates in the prevalence measurements.

Another possible limitation might be that our study had a relatively low number of participants. However, aim of this study was solely the serve as an example of a development process of an eHealth technology, not to perform an evaluation of its effects.

Also, the prevalence measurements have to be suited to be used in national surveillance (by PREZIES). Therefore, the new registration system has to meet certain requirements (e.g., data that must be collected, or in what form). Thus, in this case, and in most cases where a new eHealth technology is being developed, legislation was part of our Contextual Inquiry and also may have had an influence on the design of the technology.

Finally, the given setting for this project (nursing homes) presented us with its very own challenges. The opportunities for using technology were limited, because of the technological infrastructure of Dutch nursing homes (wherein often outdated PCs are used), and the degree to which people are used to working with technology (e.g., only 47,1% of the physicians used a smartphone). However, this gave us an interesting opportunity to put ourselves and the possibilities of the CeHRes Roadmap to the test, to see how it and how we would cope with such limitations.

V. Conclusion

This article goes beyond the mere development of eHealth technologies. It has subscribed our strong believe that combining UCD and the PSD model is of paramount importance for the creation of successful and persuasive eHealth technologies, because (1) UCD gives insight in the needs and wishes of the end-users, that have to be met by the eHealth technology; (2) the PSD model offers opportunities to deal with the issues and needs that are found using UCD; and (3) although they might have used their own words to express themselves, end-users appeared to have very clear ideas about the their needs regarding Persuasive Systems Design.

Also, our approach, using the CeHRes Roadmap has allowed for the development to be an iterative process, which may prevent costly redesign to be necessary.

Acknowledgment

This project was funded by the Interreg IV "EurSafety Health-net" project. We also thank the elderly care physicians who participated in this study.

References

[1] N. de Jong, A. Eikelenboom-Boskamp, A. Voss, and J.E.W.C. van Gemert-Pijnen, "User-centered and persuasive design of a web-based registration and monitoring system for healthcare-associated infections in nursing homes" The Sixth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2014) IARIA, Mar. 2014, pp. 152-157, ISSN: 2308-4359, ISBN: 978-1-61208-327-8

[2] N. Nijland, "Grounding eHealth; towards a holistic framework for sustainable eHealth technologies," University of Twente: Enschede.

[3] J.E.W.C. van Gemert-Pijnen, J. Karreman, S. Vonderhorst, F. Verhoeven, and M.J. Wentzel, "Participatory development via user-involvement-a case study about the development of a web-based patient-communication system about Methicillin-resistant Staphylococcus aureus," Electronic Journal of Health Informatics, vol. **6**(4), pp. e28, 2011.

[4] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive Systems Design: key issues, process model, and system features," Communications of the Association for Information Systems, vol. 24, pp. 485-500, 2009.

[5] H. Oinas-Kukkonen, "Behavior Change Support Systems: a research model and agenda," in Persuasive Technology, T. Ploug, P. Hasle, and H. Oinas-Kukkonen, Eds. Springer Berlin Heidelberg. pp. 4-14, 2010.

[6] H. Oinas-Kukkonen and M. Harjumaa. "Towards deeper understanding of persuasion in software and information systems," First International Conference on Advances in Computer-Human Interaction (ACHI 2008), IEEE, Feb. 2008, pp. 200-205, ISBN: 978-0-7695-3086-4, doi: 10.1109/ACHI.2008.31

[7] Euopean Centre for Disease Prevention and Control, "Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals," Stockholm: ECDC, 2013.

[8] PREZIES. "Referentiecijfers maart 2007 t/m maart 2012: Prevalentieonderzoek," 2012 [cited 2014 6-11]; Available from: http://www.rivm.nl/Onderwerpen/P/PREZIES/Prevalentieonderzoek_Ziekenhuizen/Referentiecijfers_Prevalentieonderzoek_ziekenhuizen.

[9] R.T. Mayon-White, G. Ducel, T. Kereselidze, and E. Tikomirov, "An international survey of the prevalence of hospital-acquired infection," Journal of Hospital Infection, vol. 11 Suppl A, pp. 43-48, Feb. 1988, doi: 10.1016/0195-6701(88)90164-8.

[10] R.M. Klevens, et al., "Estimating health care-associated infections and deaths in US hospitals", Public health reports, vol. 122(2), pp. 160-166, Mar. 2007.

[11] E.T.M. Smyth, et al., "Four country healthcare associated infection prevalence survey 2006: overview of the results," Journal of Hospital Infection, vol. 69(3), pp. 230-248. Jul. 2008, doi: 10.1016/j.jhin.2008.04.020.

[12] D. Gravel, et al., "Point prevalence survey for healthcare-associated infections within Canadian adult acute-care hospitals," Journal of Hospital Infection, vol. 66(3), pp. 243-248, Jul. 2007, doi: 10.1016/j.jhin.2007.04.008.

[13] O. Lyytikäinen, M. Kanerva, N. Agthe, T. Möttönen, and P. Ruutu, "Healthcare-associated infections in Finnish acute care hospitals: a national prevalence survey, 2005," Journal of Hospital Infection, vol. 69(3), pp. 288-294, Jul. 2008, doi: 10.1016/j.jhin.2008.03.005.

[14] A. Eikelenboom-Boskamp, J.H. Cox-Claessens, P.G. Boom-Poels, M.I. Drabbe, R.T. Koopmans, and A.Voss, "Three-year prevalence of healthcare-associated infections in Dutch nursing homes," Journal of Hospital Infection, vol. 78(1), pp. 59-62. 2011, May 2011, doi: 10.1016/j.jhin.2011.01.024.
[15] iPrevent, www.i-prevent.net, [cited 2014 6-11].
[16] Balsamiq, https://balsamiq.com/
[17] iPrevent, "Definities van verpleeghuisinfecties," [cited 2014 6-11]; Available from: www.infectionmanager.com/definities
[18] Infection Manager. [cited 2014 6-11]; Available from: www.infectionmanager.com/prevalentie
[19] M.J. Wentzel, J. Karreman, and J.E.W.C. van Gemert-Pijnen, "Towards an internet-based infectious disease management platform to increase patient safety," The third International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2011) IARIA, Feb. 2011, pp. 47-50, ISSN: 2308-4359, ISBN: 978-1-61208-119-9

Active Surgeon Support during Orthopedic Surgery using the BOrEScOPE-Exoskeleton: System Design and First Results

Peter P. Pott, Markus Hessinger, Roland Werthschützky,
Helmut F. Schlaak
Institute of Electromechancial Design
Technische Universität Darmstadt
Darmstadt, Germany
p.pott@emk.tu-darmstadt.de
m.hessinger@emk.tu-darmstadt.de
werthschuetzky@emk.tu-darmstadt.de
schlaak@emk.tu-darmstadt.de

Eugen Nordheimer, Essameddin Badreddin,
Achim Wagner
Institute of Computer Engineering, Automation Lab
Heidelberg University
Heidelberg, Germany
eugen.nordheimer@ziti.uni-heidelberg.de
badreddin@ziti.uni-heidelberg.de
achim.wagner@ziti.uni-heidelberg.de

***Abstract*— A great number of medical robotics projects is driven by researchers all around the globe. Aim is to enhance surgery output, accelerate the procedure or to shorten post-operative convalescence. In most cases, the surgeon interacts with a machine directly by some kind of remote control in general soft tissue surgery or robotic systems recapitulate pre-programmed trajectories, e.g., during milling of cavities. One option to achieve a better acceptance in human robot interaction systems in operating theatre is to use exoskeletons for tight integration. This is widely accomplished in body rehabilitation to provide patients with continuous passive or active motion. However the way to commercial application is long for many systems. In this paper for an anthropomorphic upper extremity exoskeleton worn by the surgeon during orthopedic interventions (e.g., pedicle drilling) first results concerning control strategy and user guidance are presented. The system is intended to enhance overall task-precision as the surgeon is guided by optic, acoustic, and haptic perception. The parallel flux of forces and the inherently wearable robot base attached to his back allow the surgeon to directly maintain responsibility for surgery. The mechanical design as well as the control strategy are described briefly. The device provides seven concentric axes and uses conventional DC motors and wire gears to deliver torque. An optical tracking system is employed to provide low-latency absolute position data of the system and the patient. A User Guidance Opto-Acoustic Display is utilized to provide the surgeon with information on position and orientation of the tool in six degrees of freedom with respect to the desired trajectory. The control strategy is decomposed into several levels. First experiments have demonstrated the correlation between provided workspace and space requirements during pedicle screw placement and an intuitive handling of the user guidance system to follow a desired trajectory.**

Keywords- exoskeleton, orthopedic surgery, human-machine interaction, behavior-based system decomposition

I. INTRODUCTION

Parts of this work have been previously published at the 7th International Conference on Advances in Computer-Human Interactions (ACHI 2014), March 23-27 2014, Barcelona, Spain [1].

Medical robotic systems for the use in the OR have been under development for more than 20 years [2]. Early systems for neurosurgery [3] [4] and orthopedics [5] proved usefulness and even made it for commercialization. However, their impact was not as high as expected [6]. In the last ten years, many new robotic systems have been developed and even introduced to the market. The most successful is the daVinci Surgical System by Intuitive Surgical, Inc., Sunnyvale, CA, USA. Nevertheless, there are hundreds of different systems [2] and many specific reviews e.g., by Nguyen [7], Taylor and Stoianovici [8], Cleary and Stoianovici [9], Korb [10], Cepolina [11], Kuo [12], Vitiello [13], Dhumane [14] to learn more about the various fields of robotics.

The aim of our work is to develop and to design a robotic interaction system for orthopaedic surgery. Here, the surgeon has to fulfil delicate tasks like drilling the spine while maintaining high precision in the sub-millimetre range. Placing a cooperating robotic arm next to the OR table [15], the ceiling [16] or even on the patient [17] does not seem to be appropriate. The primary reason is that such systems either require rather large space next to the table, have to be rigidly installed in the OR reducing flexibility of use, or tend to hinder the approach to the situs due to their mounting in the third case. Earlier work of our group showed the high potential of placing the robot in the user's hand [18] [19] to compensate tremor and involuntary movements both from surgeon and patient [20]. This robotic system provides precise movement and ease-of-use. However, its size and weight are not appropriate for longer deployment. Instead of using a passive balancing system for the handheld device or even a separate cooperating robot placed next to the OR table, as both approached would be bulky and space consuming, we decided to develop a new system worn at the surgeon's arm near to his or her centre of gravity to improve ergonomic handling. This is intended to lead to a better acceptance of and control over the system by the surgeon.

In the following sections, we will present and describe the system's concept, basic components, the control strategy (all in Section II), first results (Section III), and a conclusion (Section IV).

II. SYSTEM DESIGN

In this paper, we provide an overview of the BOrEScOPE system. It comprises an external optical high-speed tracking system for six Degrees Of Freedom (DOF), position and orientation measurement fused with data from an Inertial Measurement Unit (IMU), the robotic system including actuation and sensor systems together with the mechanical part, the control hard- and software, and finally an opto-acoustic display unit for interactive user guidance. In the following sections, we will address these sub-systems and describe the control strategy.

A. Robotic System

The robotic system of the BOrEScOPE basically consists of an exoskeleton for the (right) arm of the surgeon including shoulder and wrist (Fig. 1). All together, seven active DOF are realized to provide good compliance with the human anatomy and the same dexterity. The range of motion of the shoulder (170° abd./add.; 150° flex./ex.; 180° inw./outw. rotation), elbow (100° flex.), and wrist (150° pron./sup.; 20° ulnar flex./ex.; 120° flex./ex.) joints has been derived experimentally. Shoulder elevation is not considered as the abduction angle is reduced to 80°. The arm is attached to a backpack that is carried by shoulder and hip harness.

To achieve a lightweight mechanism in the final version, the actuators are placed in the backpack and force is transmitted via Bowden cables. The actuators will be based on the twisted string-concept [21], using a bunch of at least two strings that are twisted axially by a DC motor. This causes the string-arrangement to shorten and produces a rather high force. Using a lightweight ∅ 17 mm DC motor (1741 024 CXR by Faulhaber, Schönaich, Germany) with 8 mNm nominal torque and three strings, a force of 130 N can be produced. Also, no traditional gear reduction is needed leading to very quiet operation.

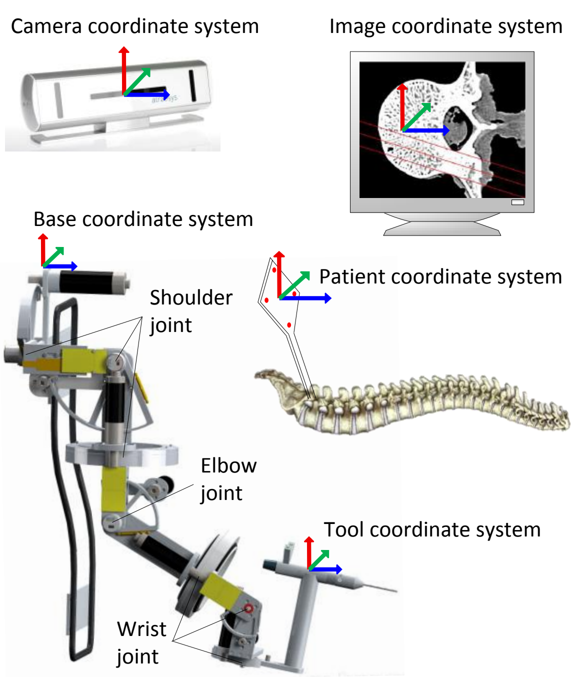


Figure 1. Overview of the robotic sub-system of the BOrEScOPE consisting of the actual exoskeleton, a 3D measurement camera, control computer, and patient (shown as spinal column only).

As only pulling forces can be produced, an antagonistic arrangement is used. Sensors are deployed at the string actuator to measure contraction and at the actuated joint to provide precise angle information. By doing so, the elasticity of the Bowden cable is used to derive a series-elastic actuator (SEA) [22]. Prior work of our group showed good results using SEA in human machine interaction [23]. The inherent compliance allows zero-torque control and robust reaction to dynamic external forces. This reduced stiffness "feels better" than a conventional robotic arm.

The final system will be designed to carry a 2.5 kg payload of a surgical device [24] and compensate the gravity force of the human arm up to a body weight of 80 kg. Shoulder and elbow joints can provide speeds up to 6 rad/s. The static force to guide the user can be up to 10 N at the handle.

B. Opto-Acoustic Display

One of the challenges in developing a user-friendly Graphical User Interface (GUI) for the Human Machine Interaction (HMI) is to facilitate an intuitive operation and control of the technical system. The basic requirements are to reduce the possible error occurring during user interaction with the machine and to navigate the user. Since the tool position is influenced by the human tremor (frequency range of several Hertz), and since latencies in the feedback-loop must be avoided, a dynamical tool tracking is proposed, consisting of a combination of optical and inertial motion measurements. Based on these data the User Guidance Opto-Acoustic Display (UGOAD) is realized, which navigates the user to the goal pose (position and orientation), displays the processing trajectory, and gives a feedback of pose errors. Display and measurement latency has to be kept low to reduce phase shift in the feedback loop and to provide stable overall system behavior. The goal 6D poses as well as the processing trajectories are provided from planning data, which are defined by the surgeon using 3D patient imaging (CT). According to the requirements, a first UGOAD functional prototype was realized (see Fig. 2).

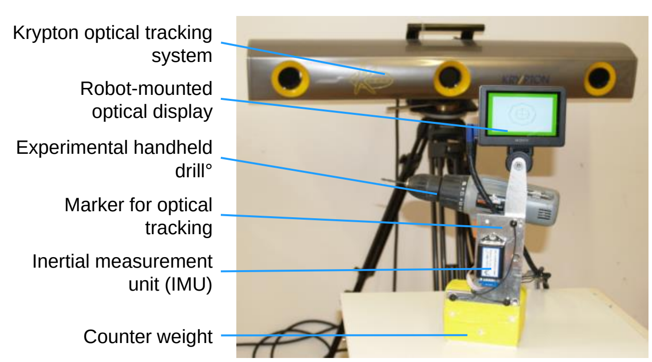


Figure 2. Experimental environment for the first handheld prototype of the opto-acoustic display (UGOAD) deployed in the BOrEScOPE system.

The experimental handheld drilling tool (Fig. 2), on the left) was equipped with three active optical LED markers and an IMU device (Crista IMU, Cloud Cap Technology, Inc.). The miniature monitor (Fig. 2, on the right) mounted at the distal end of the system can provide both optical and acoustic information. In the final implementation, a lightweight miniaturized screen will be attached directly to the tool. The optical tracking system (Krypton K600, Nikon Metrology, Inc.) (Fig. 2, in the background) is used in addition to the Crista IMU to collect the motion data of the handheld device. Data fusion is accomplished using Kalman-filter based methods [25]. The resulting filtered variables for position, orientation, velocities, angular rates, and linear acceleration are utilized for navigation purposes and provided to the lower levels. In later development stages the display can be mounted and aligned to the exoskeleton. The 6 DOF user navigation is realized by 2D representations of the tool pose on the UGOAD which is described below in detail.

C. Control Structure

The control system is developed according to Nested Recursive Behavior-based Control (RNBC) structure [26]. Accordingly, the hardware is realized as a number of components (Fig. 3) interacting on diverse behavioral levels. In contrast to a one-to-one mapping of the behavior levels, one single behavior level can be distributed on multiple hardware components. Several behavior levels may be aggregated in one single hardware device. In the latter case, behaviors are realized as software processes. For the BOrEScOPE realization, the upper levels, i.e., *mission*, *navigation* and *trajectory control*, are realized as software processes integrated into a QNX-based (QNX Software Systems Ltd.) real-time PC. The behavior levels for *position control*, *collision avoidance*, *velocity control* and *force control* are realized using an embedded PC based on xPC Target™ (The Mathworks, Inc.). The xPC™ Target PC is interconnected with the QNX PC via a serial link and to the motor controllers (type EL7342 by Beckhoff Automation, Verl, Germany) via EtherCAT. The motor controllers directly control the currents of the actuators. Position constraints for link actuation are calculated using the robot kinematics in order to avoid internal collisions. Additionally, external ultrasonic (US) sensors can help to avoid collisions of the robot with its environment. A milling tool can be aligned with the patient coordinate frame and a target bone can be processed with the preplanned trajectory.

To achieve compliance with the behavior of the operator, three interaction modalities are realized: The opto-acoustic display provides optical (1) and acoustical output (2) while the robot provides haptic feedback (3). The control algorithm's input is a virtual static force field generated around the main axis of the bore and depending on the actual distance, speed and direction of movement of the BOrEScOPE's end effector [27, 28]. When the patient is moving, this force field also moves in space. To achieve smooth and comfortable movement, the real force acting between BOrEScOPE and the wrist are measured. The user tries to minimize the forces following the BOrEScOPE system.

Using this algorithm, the 7 DOF redundant robotic system can be controlled easily and intuitively while maintaining

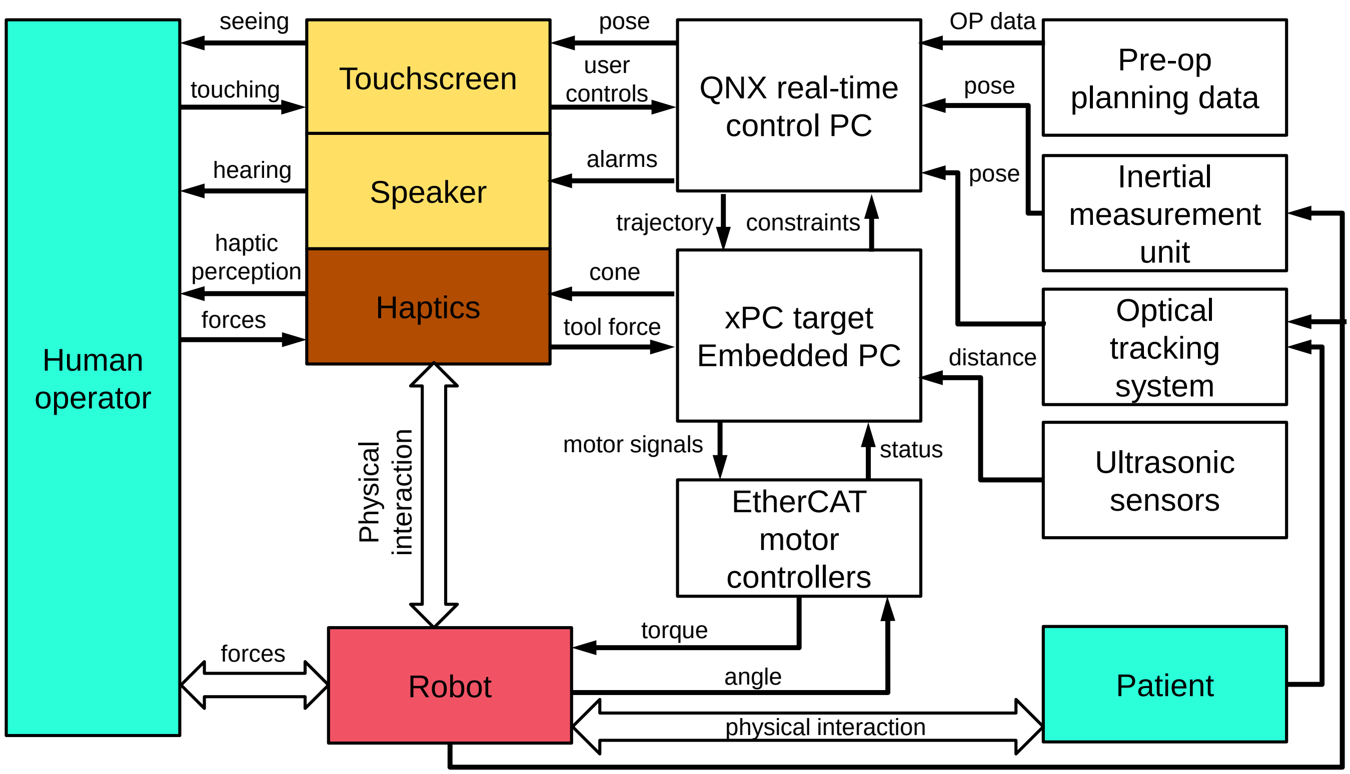


Figure 3. System architecture of the BOrEScOPE system. The human operator interacts with the robotic system, which interacts with the patient. This latter interaction is measured by a number of sensors while the first is based on audio, visual, and haptic effects.

Figure 4. First Prototype of the BOrEScOPE system as a CAD drawing.

the human's dexterity. As both, the linear displacement at the actuators and the angular displacement in the actual joints are measured and controlled, serial-elastic actuation is achieved.

III. RESULTS

The BOrEScOPE system is still under development. The two main sub-systems *opto-acoustic display* and *robotic system* show first and promising results that are described in the following.

A. Robotic System

The mechanical sub-system of the BOrEScOPE is shown in Fig. 4. The device features the same seven axes as a human arm and can be adjusted to persons between roughly 165 and 200 cm body height and a BMI under 30. It is worn around the arm and thus provides congruent axes. To allow this for the axial rotation of upper and lower arm, special wire ball bearings (LEL 1.5/5 by Franke GmbH, Aalen, Germany) have been chosen for a lightweight, strong, and backlash-free solution for these two DOF. For first experiments, conventional DC gear-motors (shoulder joint: 3890 048 CR+38/2 S, elbow joint: 3272 048 CR+32/3 S, wrist joint: 2657 048 CR+26/1 S) and wire-gearing have been chosen to reduce development effort while still accounting for backlash-free smooth motion with constant friction. The range of motion of the shoulder (100° abd./add.; 90° flex./ex.; 180° inw./outw. rotation), the elbow (105° flex.) and wrist (150° pron./sup.; 35° ulnar flex./ex.; 90° flex./ex.) fits in the requirements of the operational task. Its force is capable to maneuver payload up to 1 kg safely within the complete range of motion. The device can be worn by the surgeon using a backplate and a rucksack-like arrangement of straps and belts.

A first realization of the 3-DOF wrist of the robot is shown in Fig. 5. Here, especially the wire-geared actuation principle and the structural integration of the torque sensors is visible. The structural integration allows high-stiffness measurement with no additional masses or elasticity. It is achieved by integrating full-bridge strain gauges to the wire-gearing mechanism in a way that the pulling force of the wire is measured.

B. Opto-Acoustic Display

The measured peak response time of hand movement as a result of optical stimuli amounts to around 250 ms. The requirement of visually provided information should be adapted on this process time. The reaction time of the UGOAD as well as the robot must be kept within a limit of 10-20 ms (10 to 20-fold faster). Thus, the calculation of graphical contents and of the control algorithm should terminate within this time. Based on this knowledge, the sensor data acquisition, the global-control loop, and UGOAD were implemented as real-time processes in the QNX Neutrino operating system.

The first display prototype was realized by a 2D representation of the 6 DOF pose data. Accordingly, the actual and the reference pose of the tool are shown in the x-y-plane of the display. The z-axis is perpendicular to the display plane. In order to intuitively capture the 6 DOF contents in the 2D image a two body projection metaphor is

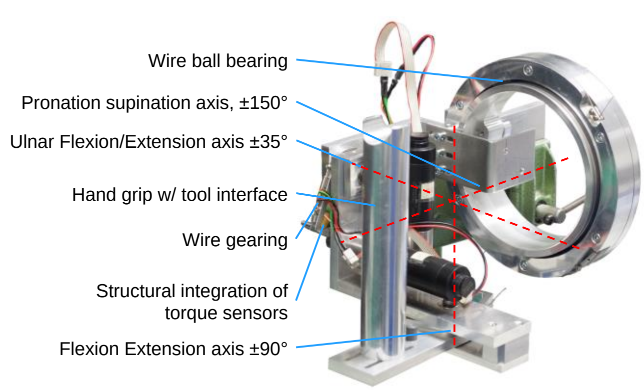


Figure 5. First realisation of the wrist. The three axes intersect in one point in the centre of the human wrist (not shown).

realized. In this imagination one small colored octagon is mounted virtually at the tool tip and one large colored octagon at the rear of the tooling machine. Looking from above in direction of the drilling tool (z-axis) corresponds to looking through the large octagon and through the small octagon on the tool tip, which is in the center of both. The small black octagon with crosshairs and large black octagon are virtually mounted at the target (reference) pose. If the tool is aligned (Fig. 6a), the small octagons are aligned and the large colored octagon has its original size in the central position. If the tool is misaligned in the x-axis (Fig. 6b) the large colored octagon is shifted correspondingly in the x-direction. The same holds for the y-axis. A misalignment in the z-axis is represented by the size of the large colored octagon. A deviation in the positive z-direction means that the tool is too far away from the user, which is shown by the reduced size relatively to the large black octagon. Negative deviation displays an increased size of the octagon to report that the tool is too narrow. A deviation in the orientation is displayed as shift of the small colored octagon. For example, if the tool is turned around the y-axis (Fig. 6), the tool tip is moved in x-direction, displayed as x-axis-shift of the small colored octagon. The corresponding principle holds for the orientation error around the x-axis. Here, a y-shift of the small octagon can be observed. The orientation error around the z-axis is directly displayed as a rotation of the colored octagons around their centers.

As additional element, a rectangular border is shown in green color, which indicates that the pose is in the desired workspace. If the tool approximates the limit positions for at least one axis, the color changes firstly from green to orange, showing that a user intervention is required. In critical vicinity to the constraints the color changes to red (Fig. 6d) asking for urgent motion actions. The color change is supported by changing the waveform of the acoustic channel.

C. *User Experiments*

Several experiments with subjetcs (users) have been performed resulting in a first performance test of the UGOAD. The goal was to keep the handheld drill (Fig. 2) still in position and orientation in six DOF while looking at the UGOAD only and at the tool-tip exclusively in guided (assisted) and unguided (unassisted) case, respectively. The user himself chose the holding comfortable pose. Substantial results of these trials are presented for two users exemplary.

User1:

The first human operator (user 1) was requested twice to hold the tool calmly during 60 seconds. Firstly, without UGAOD assistance and secondly after short instruction and training with the UGOAD. The results point out that it was possible to keep the tool in the defined workspace (±5 mm in position-axes and ±5° in orientation-axes) in assisted attempt (blue trajectory, Fig. 7). Unassisted the workspace was left after short time and the trajectory was drifted in all axes (red trajectory, Fig. 7). To evaluate the position errors the histo-

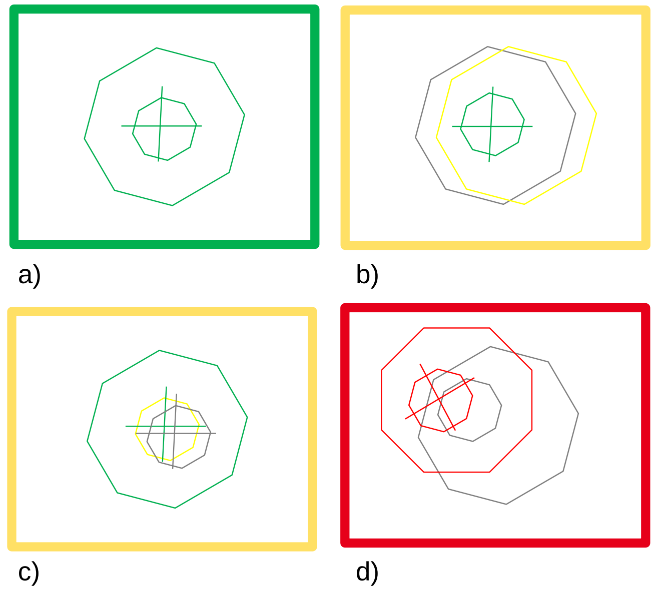


Figure 6. Display content of the UGOAD (refer to Fig. 2): a) Correct position and orientation, b) Translational displacement in x-axis, c) Rotational displacement around x- and y-axis, d) Displacement in all view axes.

grams for both cases with and without UGOAD are represented in one plot (Fig. 8).

Without assistance, the position is distributed over a wide range according to the non-stationary process. With UGOAD feedback the human-in-the-loop position control reaches as stationary condition, while the remaining position error has a distribution with an approximately Gaussian

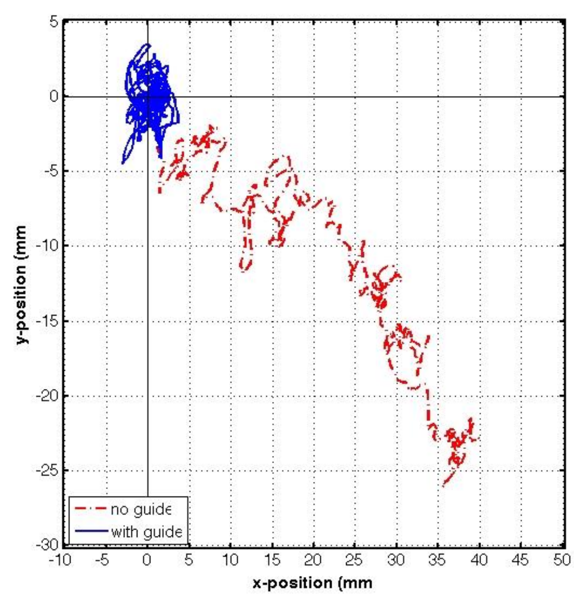


Figure 7. Position deviation in xy-direction with and without guide with the human operator (user1) in the loop (without the robotic sub-system). z-axis deviation is not shown for reasons of brevity.

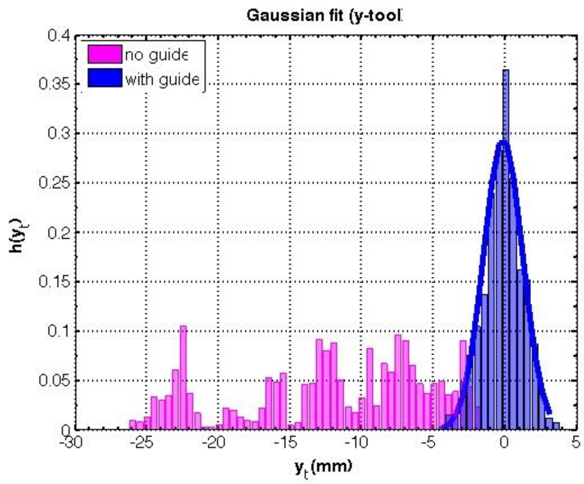


Figure 8. Position y-deviation histogram for guided and unguided trial of user1 (instructed).

shape. The histograms show the posed distribution together with a fit using the normal distribution in one dimension while the standard deviation and the mean of the distribution are $\sigma = 1.3529$ and $\mu = -0.2039$ for the case with UGOAD and $\sigma = 6.606$ and $\mu = -12.0661$ for the case without UGOAD, respectively.

User2:

Another human operator (user 2) was also requested twice to hold the tool also calmly during 60 seconds. In both cases, the UGOAD was used. In contrast to the former experiment, user2 was not instructed about the operating principle of the UGOAD so that the user had to find it out by himself during trial 1. Nevertheless, it was possible to keep the tool inside the workspace (blue trajectory, Fig. 9). The second trial shows the learning effect in operation (red trajectory, Fig. 9).

The histograms show also Gaussian shapes for both trials with UGOAD while the standard deviation and the mean of the distribution are $\sigma = 1.8460$ and $\mu = -3.9976$ for trial 1 and $\sigma = 0.8879$ and $\mu = -0.4686$ for trial 2, respectively (Fig. 10).

It is obvious that using the opto-acoustic display a strong improvement of the pose deviation was achieved (here presented for two different axes and operators).

IV. Discussion and Conclusion

To set up a robotic system with close human-machine interaction in a medical environment is a delicate task. However, the project is still in progress and work starting from the presented concept to the final realization is still ongoing. We managed to define interfaces between the robotic system and the human operator not only mechanically, but also visually and using the audio channel. Smooth and comfortable working with the system is strongly dependent on low latency, high update rates, and actually predictable behavior. Here, our system will have to deal with some drawbacks as the force field generation is depending on data quality of the optical tracking system which tends to jitter and noisy signals. This will be addressed in future by using redundant LED markers and by combining data of an inertial tracking system. Furthermore, the quality of real-time data transfer will be improved.

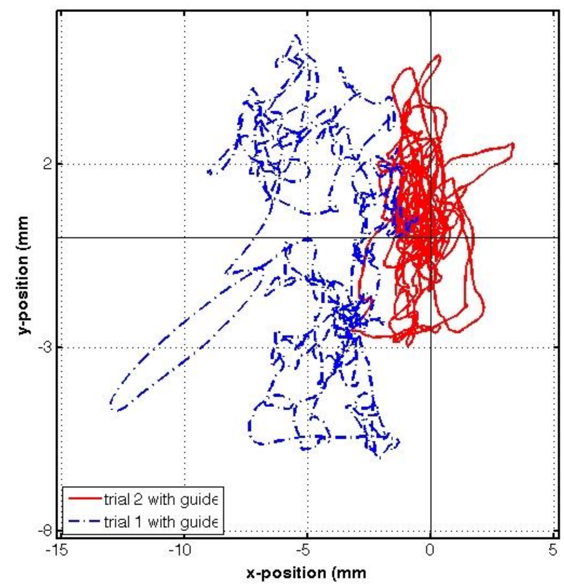


Figure 9. Position deviation in xy-direction for two trials with guide with the human operator (user2) in the loop (without the robotic sub-system).

Mechanically, the robot will have to cope with force-depended friction wire gearing and residual backlash in the gearing. This issue will be addressed by a model-based controller with individual parameters for each axis. The first prototype shown in Fig. 4 differs slightly from the initial concept due to time restriction during development. However, the schedule of the first tests of the complete system is set and first promising objectives have been reached.

First experiments with users demonstrate that the 6 DOF-guidance can be captured by the majority of subjects

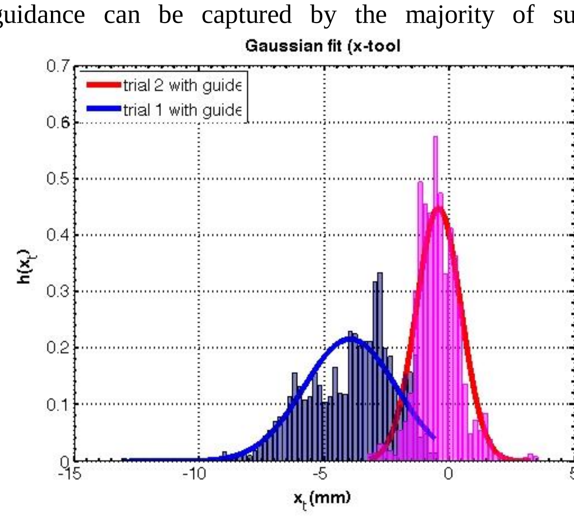


Figure 10. Position x-deviation histogram for guided trials of user2 (uninstructed).

without further explanation (Fig. 6). Thus, the usage of the UGOAD as a feedback in the human interaction with the machine implicates a massive improvement of human performance to achieve the common tasks and there is every indication that the developed UGOAD insure an intuitive operation and an intuitive control.

ACKNOWLEDGMENT

The work is funded by the German Federal Ministry of Education and Research under grants 16SV5773 and 16SV5774.

REFERENCES

[1] P. P. Pott, M. Hessinger, R. Werthschützky, H. F. Schlaak, E. Nordheimer, E. Badreddin, and A. Wagner, "BOrEScOPE – Exoskeleton for Active Surgeon Support during Orthopedic Surgery," Proc. 7. International Conference on Advances in Computer-Human Interactions (ACHI 2014), Barcelona, pp. 377-380, 2014.

[2] P. P. Pott, H.-P. Scharf, and M. L. R. Schwarz, "Today's State of the Art of Surgical Robotics," Journal of Computer Aided Surgery, vol. 10, pp. 101-132, 2005.

[3] Y. Kwoh, I. Reed, J. Chen, H. Shao, T. Truong, and E. Jonckheere, "A new computerized tomography-aided robotic stereotaxis system," Robotics Age, pp. 17-22, 1985.

[4] J. Doll, W. Schlegel, O. Pastyr, V. Sturm, and W. Maier-Borst, "The use of an industrial robot as a stereotactic guidance system," Proc. International Symposium and Exhibition on Computer Assisted Radiology (CAR 87), Berlin, D, 1987.

[5] R. H. Taylor, *et al.*, "An Image-Based Roboitc System for Hip Replacement Surgery," Journal of the Robotics Society of Japan, pp. 111-116, 1990.

[6] C. Caetano da Rosa, Operationsroboter in Aktion. Bielefeld, Germany: transcript Verlag, 2013.

[7] K. Cleary and C. Nguyen, "State of the art in surgical robotics: Clinical Applications and Technology Challenges," Computer aided Surgery, vol. 6, pp. 312-328, 2001.

[8] R. Taylor and D. Stoianovici, "Medical Robotics in Computer-Integrated Surgery," IEEE Transactions on Robotics and Automation, vol. 19, pp. 765-781, 2003.

[9] K. Cleary, D. Stoianovici, V. Watson, R. Cody, B. Hum, and D. Lindisch, "Robotics for Percutaneous Spinal Procedures: Initial Report," Proc. Computer Aided Radiology and Surgery, San Francisco, USA, 2002.

[10] W. Korb, R. Marmulla, J. Raczkowsky, J. Mühling, and S. Hassfeld, "Robots in the operating theatre—chances and challenges," Int. J. Oral Maxillofac. Surg., vol. 33, pp. 721-732, 2004.

[11] F. Cepolina, B. Challacombe, and R. C. Michelini, "Trends in robotic surgery," J Endourol, vol. 19, pp. 940-51, Oct 2005.

[12] C.-H. Kuo and J. S. Dai, "Robotics for Minimally Invasive Surgery: A Historical Review from the Perspective of Kinematics," in *International Symposium on History of Machines and Mechanisms*, H.-S. Yan and M. Ceccarelli, Eds., ed: Springer Science+Business Media B.V., 2009, pp. 337-354.

[13] V. Vitiello, S.-L. Lee, T. Cundy, and G.-Z. Yang, "Emerging Robotic Platforms for Minimally Invasive Surgery," IEEE reviews in Biomedical Engineering, vol. 6, pp. 111-126, 2013.

[14] P. W. Dhumane, M. Diana, J. Leroy, and J. Marescaux, "Minimally invasive single-site surgery for the digestive system: A technological review," J Minim Access Surg, vol. 7, pp. 40-51, 2011.

[15] M. Jakopec, S. J. Harris, F. R. Baena, P. Gomes, J. Cobb, and B. L. Davies, "The First Clinical Application of a "Hands-on" Robotic Knee Surgery System," Computer Aided Surgery, vol. 6, pp. 329-339, 2001.

[16] H. Mayer, I. Nagy, A. Knoll, E. U. Schirmbeck, and R. Bauernschmitt, "A robotic system providing force feedback and automation for minimally invasive heart surgery," Proc. 20th International Congress and Exhibition on Computer Assisted Radiology and Surgery CARS 2006, Osaka, Japan, pp. 265-267, 2006.

[17] M. Shoham, M. Burman, E. Zehavi, and Y. Kunicher, "MARS: miniature bone-mounted robot," Proc. ISRACAS'2003, Tel Aviv, Israel, 2003.

[18] P. P. Pott, *et al.*, "ITD - A handheld manipulator for medical applications - Concept and design," Proc. 3rd annual meeting of CAOS, Marbella / Spain, pp. 290-291, 2003.

[19] P. P. Pott, A. Wagner, E. Badreddin, H.-P. Weiser, and M. L. R. Schwarz, "Inverse Dynamic Model and a Control Application of a Novel 6-DOF Hybrid Kinematics Manipulator," Journal of Intelligent and Robotic Systems, vol. 63, pp. 3-23, 2010.

[20] A. El-Shenawy, A. Wagner, P. P. Pott, and E. Badreddin, "Disturbance Attenuation of a Handheld Parallel Robot," Proc. IEEE ICRA, Karlsruhe, 2013.

[21] T. Würtz, C. May, B. Holz, C. Natale, G. Palli, and M. C, "The Twisted String Actuation System: Modeling and Control," Proc. IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Montréal, Canada, pp. 1215-1220, 2010.

[22] K. Kong, J. Bae, and M. Tomizuka, "Control of Rotary Series Elastic Actuator for Ideal Force-Mode Actuation in Human-Robot Interaction Applications," IEEE/ASME Transactions on Mechatronics, vol. 14, pp. 105-118, 2009.

[23] M. Grün, R. Müller, and U. Konigorski, "Model Based Control of Series Elastic Actuators," Proc. IEEE Biomedical Robotics and Biomechatronics (BioRob), Rome, Italy, pp. 538-543, 2012.

[24] M. Hessinger, J. Hielscher, P. P. Pott, and R. Werthschützky, "Handheld surgical drill with integrated thrust force recognition," Proc. E-Health and Bioengineering Conference (EHB), Iasi, RO, 2013.

[25] N. Sadaghzadeh, L. Zouaghi, A. Wagner, J. Poshtan, and E. Badreddin, "Cascaded Error Estimation and Compensation of an Inertial Measurement Unit using the Fusion of Optical and Inertial Sensors," Proc. COMADEM, Helsinki, 2013.

[26] E. Badreddin, "Recursive Control Structure for Mobile Robots," Proc. International Conf. on Intelligent Autonomous Systems 2 (IAS.2), Amsterdam, pp. 11-14, 1989.

[27] A. Wagner, E. Nordheimer, and E. Badreddin, "Hierarchical constraint-based singularity avoidance," Proc. System Theory, Control and Computing (ICSTCC), Sinaia, 2012.

[28] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," The International Journal of Robotics Research, vol. 5, pp. 90-98, 1986.

Business Process Evaluation in Agile Business Process Management Using Quality Models

Michael Gebhart
iteratec GmbH
Stuttgart, Germany
michael.gebhart@iteratec.de

Marco Mevius, Peter Wiedmann
HTWG Konstanz
Constance, Germany
mmevius/pewiedma@htwg-konstanz.de

Abstract—**Agile business process management is an adaptation of the agile methodology, which is well-known in the area of software development. It supports the tight and efficient collaboration between customer and process analysts when designing processes. However, this methodology requires a governance approach that ensures the high quality of the processes. This article shows the usage of quality models to evaluate business processes in agile business process management environments and its specific challenges. To illustrate the approach, an existing quality model is reused and formalized. Afterwards, a business process in the context of offer management is captured and evaluated by means of this quality model.**

Keywords-business process; design; quality; agile; business process management; ISO 25000.

I. INTRODUCTION

This article is an extended version of [1]. Fast and efficient adaptability to business processes becomes an increasingly important competitive factor for organization [2]. For that reason, high business process flexibility is an important requirement for organizations, such as administration sections within companies, to counter current challenges. Among other things, explicit knowledge about the structure and functionality of business processes is essential for the understanding of organizational sequences [3]. For example, business processes, in which citizens and experts of public administration participate together, should be accessible for both sides easily. For that purpose, a targeted enhancement of Business Process Management (BPM) with the help of agile advantages generates new significant potential for the automation, modeling, interaction, and optimization of business processes. The idea of agility is described as the ability to balance flexibility and structure [4] and to minimize risks for instance by conforming project changes rapidly [5].

In the past, different (agile) approaches have been developed. One approach that is also chosen in this article is BPM(N)$^{Easy1.2}$ [6][7]. It describes a combination of Business Process Management and Business Process Model and Notation (BPMN) with the ambition of making BPM easier. In reference to the agile software engineering, which is already well-known and used in a lot of organizations and projects, the agile BPM method adapts aspects from it. The approach extents and supports the interaction between every participant with focus on more coherency without confronting them with unneeded complexity. Furthermore, it follows an empirical, incremental and iterative concept to increase predictability of the process quality and to reduce project risks [6]. For instance, the efficiency and effectiveness of capturing a business process can be enhanced significantly.

However, within the prediction and control phase of the business process quality, the participants have to know what constitutes a good process and how to evaluate processes [8]. Today, there are no general rules, which define what a good business process is. The literature suggests aspects, such as the customer value, process standardization, and the employee well-being to measure quality [9]. But this information is not sufficient to perform a systematic or even automatic quality analysis of business processes. Aggravating this situation, contradictory constraints and needs – for instance speed and quality – generate the need to focus on the delivering value [10].

To enable a systematic quality assurance in agile BPM from a functional point of view, this article introduces the application of quality gates and quality models. Quality gates define a specific point within a project to evaluate determined maturity and sustainability [11]. These quality gates appear frequently during the application of an agile BPM approach, such as BPM(N)$^{Easy1.2}$ and ensure the synchronization and acceptance of all participants. For instance, an automated business process has to correspond with all predefined requirements and expectations. The quality gates considered in this article are supposed to close the gap especially at the beginning and during the business process modeling step. The requirements and expectations are represented by quality models that enable a systematic evaluation of business processes. In this article, existing work in the context of business process evaluation is reused. I.e., existing quality models are evaluated and the preferred one is adapted [12] for applying and measuring it in BPM environments. Especially agile environments with short iterations and high interaction are suited for the continuous monitoring and improving of the business process quality.

The article is organized as follows: Section II analyses relevant literature regarding business process modeling. Furthermore, quality models for BPM and their application for agile BPM are discussed. Section III introduces the BPM(N)$^{Easy1.2}$ method and demonstrates where and how quality models can be applied within an agile approach. In Section IV, the application of a certain business process quality model in agile BPM is evaluated by means of a scenario from offer management. Section V introduces the tools, which have been applied during the evaluation. Especially, the internals of the quality analysis tool are revealed. Section VI presents the conclusion and discusses future research work.

II. BACKGROUND

This section describes the fundamental terms and existing work in the context of modeling and evaluating business processes. Regarding the quality of business processes, work that targets the quality from both a functional and a technical point of view is considered. Furthermore, this work is examined in detail regarding its applicability in agile BPM environments. As the approach chosen in this article was already applied in related disciplines, the according results are presented.

A. Modeling of Business Processes

In the following, work in the context of modeling business processes with focus on agile BPM is presented. The introduced approaches address especially the involvement of all participants within modeling business processes.

The process of continuous improvement and involvement of stakeholders is well-known in the area of quality management. Total cycle time or Six Sigma are examples cf. [13]. A variant of Six Sigma, the Lean Sigma approach, represents the combination of Lean and Six Sigma. This approach helps to improve the product and process quality on the one hand, and on the other hand it increases process performance. Audits, which are conducted internally or externally, are used to interview the user directly and to draw from the interview results conclusions about possible improvements in quality. These improvements lead not only to, e.g., an increased quality of business process activities execution, moreover, through the interviews the understanding of business processes can be enhanced.

In the context of BPM, Cheng et al. describe [14] the problem, which exists between the connection of business rules and business models. For instance, business rules can have compulsory possibilities of a sequence. These possibilities depend on business rules, which in the first step are often formulated in common language, e.g., “The customer is not allowed to order articles without logging in”. Cheng et al. suggest a framework, which supports the mapping of business rules and business processes and assists the stakeholders in identifying inconsistencies. Hereby, the usage of this framework simplifies the “connections” of business rules and models, but takes not account of their quality itself explicitly.

Antunes et al. [15] focus on the integration of end users. They developed an approach and tool to support business process modeling from the perspective of end-users. For this purpose, Antunes et al. use concepts of representation and visual composition. In contrast to other approaches, there is not only a common language support, e.g., by adding annotations to a business process activity. The approach extends these annotations by adding pictures to a business process element. This additional illustration leads to a higher understanding of the modeled business processes.

The approach of Bittmann [16] adds additional information as natural language artefacts to business processes. Every business process activity can be described more in detail by adding a written text. Within this text, the identifier label of a business process activity, such as “entering data”, has to be mentioned and linked. The approach fosters the higher integration of employees, especially of the operative business (non IT specialists).

B. Quality of Business Processes

In this section, work regarding the quality of business processes is introduced. This work is considered from both a functional and technical point of view. Furthermore, the applicability in agile BPM as introduced in the section before is examined.

The International Organization for Standardization (ISO) and the International Electro technical Commission (IEC) have created standards regarding the quality of software products. Both ISO/IEC 9126 [17] and the successor ISO/IEC 25000 ff. [18] define relevant terms for software product quality. Furthermore, they describe quality characteristics, their subcharacteristics, and their final quality measure elements. They hereby provide a wide overview of measuring the quality of software products. In order to apply these standards on business processes, the term “business process” has to be distinguished from “business process model”. As the standards refer to software products, they can only be directly applied on business process models as software artifacts. Also, in this case, only a subset of described characteristics is applicable. Heinrich et al. [19], Sánchez-González et al. [20], and YeongSeok et al. [21] show the adaptation of these standards on business process models. However, according to the introduction, we focus on the quality of business processes and their content instead of the models as software artifacts and their syntactical correctness etc. For that reason, the standards cannot be applied. Nevertheless, they give hints about characteristics that might be important for business processes as well.

Further standards regarding quality management focus on quality management systems. Examples are ISO 9000 ff. [22], or branch-specific manifestations, such as the European Norm (EN) 9100 for aerospace. There also exist standards for the quality management in projects, such as ISO 10006. Even though they consider the quality in business domains and in some cases also describe business processes, the quality of the business processes themselves is not explained in detail. This is also the case when choosing Capability Maturity Model Integration (CMMI) or IT Infrastructure Library (ITIL).

In [9], Krogstie describes criteria for so-called good processes. He introduces dimensions of value that is valid for most customer groups. Furthermore, he summarizes heuristics for good business processes. Even though no metrics are provided, these heuristics can be good starting points to derive more concrete quality aspects that again enable a systematic and automatic evaluation of business processes. In addition, this work helps to understand the purpose of business processes and why it is important to have good processes. Thus, it forms the framework for a quality model as it focuses on the motivation and strategic goals of business processes.

In order to enable a more systematic quality analysis of business processes, Kneuper created the quality model Gokyo Ri based on existing standards, such as ISO 9000, CMMI, and ITIL [23]. It refines the quality of business processes so that their quality can be determined. Even though this quality model focuses on business processes and their content, the quality model is still too abstract to be used in agile business process management environments. In agile projects the quality has to be determined in short intervals best automated based on modeled business processes. Thus, Gokyo Ri has to be further refined until at least a subset of the quality attributes can be determined automatically or with short user interaction intervals.

Similarly, Lohrmann et al. introduce quality attributes for business processes [8]. Also, in this case the quality attributes are derived from business-related quality concerns and focus on the content of the business process and not the artifact. Lohrmann et al. distinguish between the efficacy and efficiency of business processes that can be either determined on basis of business process models and running instances. Former is called business process design and implement efficacy and efficiency. Latter is called business process enactment efficacy and efficiency. Even though Lohrmann et al. do not describe an entire quality model, they introduce quality attributes that are relevant for the business process quality as considered in this article. Nevertheless, similar to the quality model introduced by Kneuper the quality attributes are still too abstract to be applied in an agile environment. They first have to be refined so that they can be determined either based on business process models or by answering simple questions.

Regarding a more technical point of view, Suarez et al. [24] describe best practices for modeling business processes using certain languages, such as the Business Process Model and Notation (BPMN). Even though this article also focuses on BPMN as modeling language, these best practices mostly consider syntactical correctness of created models or related issues. The content of the processes and their quality from a functional point of view is not considered. The described best practices are also not aligned with a holistic quality model. So, the impact of these best practices on abstract quality characteristics is not obvious. The best practices can increase the quality of modeled business processes. They are also applicable in agile BPM environments as they can be easily determined or can be even measured automatically by tools. Nevertheless, they do not target the kind of business process quality considered in this article.

C. *Related Application in Service-Oriented Architectures*

The approach chosen in this article has already been successfully applied in other areas. One example is the evaluation of web services in service-oriented architectures: Today, a lot of best practices exist that describe how to design web services so that they fulfill wide-spread quality attributes, such as loose coupling and autonomy. For example, Erl [25][26], Cohen [27], and Josuttis [28] describe such best practices. Whenever the IT architect or a developer has to decide, how to design a web service, these best practices should be considered. However, due to the increasing complexity of architectures, it is nearly impossible to consider them all without mistakes. Sometimes, the best practices require an understanding of the entire architecture of the considered system. And in addition, in some cases, they refer to technical details that have to be completely understood to apply the best practices successfully. Furthermore, similar to the quality models for business processes, these best practices are described informally and require interpretation effort. This hampers their efficient application in real-world projects. IT architects and developers have to map the textually described best practices onto the technology used in the concrete project. This again may result in wrong applications of the best practices.

For that reason, in [28], Gebhart et al. introduced metrics that reuse the mentioned work of Erl et al. to evaluate web services regarding the introduced best practices in a comprehensible and repeatable manner. These metrics are formalized so that they can be systematically measured on web services described using the Service oriented architecture Modeling Language (SoaML) [29] as profile for the Unified Modeling Language (UML) [30]. By means of these metrics, IT architects and developers know exactly, what elements of the modeled architecture have to be considered to evaluate a service. Based on mapping rules introduced by Gebhart et al. in [31], the metrics could be mapped to concrete implementation artifacts. This enables the systematic evaluation of implemented web services using Web Services Description Language (WSDL) [32] to describe the service interface, XML Schema Definition (XSD) [33] for the data types, and Service Component Architecture (SCA) [34] to implement the internal logic. The application of the metrics is demonstrated for a geographical information system by Gebhart et al. in [35]. Also, in this case, existing abstract quality attributes were refined to enable a fully or partially automated quality assurance. As result, a solution was created to ensure the systematical creation of a flexible and maintainable architecture [36].

D. *Summary and Need for Action*

The overview shows that there exists a lot of work in the context of modeling and evaluating business processes as required in this article. However, some work focuses on best practices or quality attributes from a functional point of view that are too abstract to be measured directly and especially too heavyweight to be determined in agile environments. Other work considers fine-grained quality aspects, such as syntactical correctness that can be easily determined, however does not provide value for the quality of business process from a functional point of view. This article shows how to fill this gap by reusing existing quality models that focus on quality attributes from a functional point of view and breaking these quality attributes down into aspects that can be either directly measured on business process models or easily answered by process analysts. The approach has already been successfully applied in the context of service-oriented architectures. For that reason, the methodology to derive measurable quality aspects from quality models was reused. Furthermore, when evaluating the approach, the same quality analysis has been applied.

III. Business Process Evaluation in Agile Business Process Management Using Quality Models

In order to evaluate business processes in an agile BPM project, on the one hand, the point of time when an evaluation of the business process is possible and useful has to be determined. For that purpose, the approach of agile BPM has to be examined in detail. On the other hand, an appropriate quality model has to be identified and adapted so that the quality of the business process can be determined in a systematic and repeatable manner.

A. Agile BPM

As mentioned before, there exist different approaches of agile BPM. To identify the point of time, when to apply a quality model in an agile BPM project, these approaches have to be analyzed in detail. For instance, Meziani [37] introduces a method called AGILIPO. This method describes an approach, which focuses especially on the integration of software systems and organizational knowledge. To automate business processes, the method uses concepts as used in agile software engineering. In addition, AGILIPO suggests the usage of social media tools to interact with all stakeholders. For example, the stakeholders can write comments (common language) on existing business process models. This feedback can be used to optimize current models in a next iteration.

Another approach is provided by Schnabel et al. [38]. They outline the Lightweight Process Modelling process and the Language for Lightweight Process Modelling (LLPM). The formal semantic of LLPM is focused on a simple graphical form. LLPM “introduces goals as unbound activities that are bound to a particular service either at design time or at runtime by composition tools” [38]. As a result, business process models can be maintained more flexible and agile.

In the following, the agile approach BPM(N)$^{Easy1.2}$ [6] is used to show when (time of application) during the methodology the quality model is expected to be applied. This approach has been chosen, because of this holistic concept behind. BPM(N)$^{Easy1.2}$ enables highly sophisticated agile BPM. It covers all aspects of BPM – from process design and process execution to process controlling with focus on the integration of all process participants. The following Fig. 1 provides an overview of the approach and the including quality gates. Latter are displayed as stars:

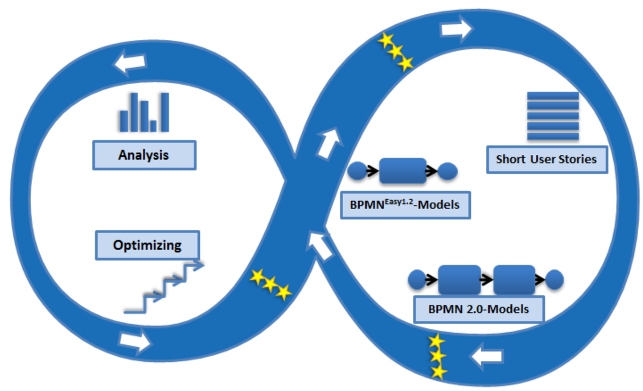


Figure 1. Illustration of the BPM(N)$^{Easy1.2}$ approach.

The approach consists of two connected cycles. One cycle is used to capture new BPMN$^{Easy1.2}$ models and short user stories. BPMN$^{Easy1.2}$ represents the modeling language of BPM(N)$^{Easy1.2}$ and a subset of BPMN. Both BPMN$^{Easy1.2}$ models and short user stories formulate the requirements of the activities within a business process. The models are used to design the flow in general and set up a first model very easily. The short user stories describe further information, e.g., additional business rules. The formulated requirements are the basis for the modeling and implementation of an enriched BPMN 2.0 business process. For the enrichment, a BPMN$^{Easy1.2}$ model and a number of user stories are selected to work on. Furthermore, the business process is modeled on the business user’s point of view. In addition, in consultation with a business user, an IT expert is able to use the business process model to automate the process. Once the modeling and implementation stages are completed, the resulting BPMN 2.0 models are transferred to a final control. Within this control all participants assure that the result, e.g., an automated business process corresponds with the BPMN$^{Easy1.2}$ models and formulated short user stories (synchronization and acceptance).

Immediately after the acceptance, new requirements can be taken and transformed into a business process model or implementation. If defined key performance indicators show optimization potential (analysis and optimizing cycle), new BPMN$^{Easy1.2}$ models or short user stories will be generated. The several iteration and high collaboration between every participant allows the continuous monitoring of the business process quality [6].

However, in general there are still different weak points in agile methods. Mohammad [39] says that short response times and high interaction during the agile development do not require the writing of documents which can lead to a reduced quality of documentation. Furthermore, Mohammad [39] mentions the increased collaboration time of the participants. But in fact, in some circumstances there is not enough time for the required coordination or the participants are not at the same (physical) location [39]. In [40], agile methods are described as a risk of large or complex projects. The magnitude of uncertainty is increased. Therefore, agile methods are mistrusted in most organizations. To counteract these disadvantages and related lack of quality, it is required to introduce quality checks during the application of an agile approach. In [41], quality checks are suggested to be applied to different steps of agile approaches.

According to [41] and with the assumption that software engineering has the same goals as BPM, e.g., cost reduction and collaboration enhancement, the quality gates listed in Table I are suggested for agile BPM approaches:

TABLE I. QUALITY GATES

Quality Gate	Time of Application	Comment
1	Formulation of user stories	Continuous feedback and collaboration between every participant
2	Modeling of business process	
3	Automation of business process	Test of process application
4	Acceptance testing	

Today, some of the quality gates have already been implemented to assure the determined quality. For instance, the quality gate 1 can be applied by a continuous feedback process between every participant or by means of standard assurance tests of the process application [42]. For quality gates 1, 3 and 4, methods already exist, which can be used to assess the quality, e.g., real tests of a process applications.

Therefore, in this article, the quality gate number 2 that is applied during the modeling of business process is considered to improve and guarantee the expected quality.

B. *Quality Model Choice and Adaptation*

In the previous section, quality gates during an agile methodology have been identified. One quality gate considers the quality of modeled business process. This section shows how this quality gate can be supported by IT.

In order to support this quality assurance, an appropriate quality model has to be prepared. For that purpose, first the most appropriate existing quality model has to be identified. Afterwards, its direct applicability has to be verified. As described in Section II, appropriate quality models are those introduced by Lohrmann et al. [8] and Kneuper [23]. However, in both cases, the introduced quality attributes have to be adapted for requirements in agile environments: As mentioned before, the quality of business processes has to be determined in short intervals, which again requires a quality analysis to be easy and lightweight. This requirement cannot be fulfilled by these existing quality models and the contained quality attributes. They are not formalized using metrics which hampers their automatic determination based on business process models. Furthermore, the informal description requires interpretation effort that can result in misunderstandings and wrong measures. This is a typical issue when performing quality analyses and has already been identified for other domains, such as the quality analysis of service-oriented architectures by Gebhart et al. [43].

Thus, after choosing a certain quality model, the quality attributes have to be refined if necessary until more fine-grained and comprehensible quality attributes are identified so that no interpretation is necessary any longer. They are called quality indicators, formalized as metric, and return a measure. It is not necessary that a quality indicator can be fully automatically measured on process models. If this is not possible as they require further knowledge, such as domain knowledge, the only condition is that it is possible to formulate unambiguous questions that can be answered by experts and do not require interpretation. Summarized, for every function and variable used within a metric, the criteria listed in Table II have to be fulfilled.

TABLE II. CRITERIA FOR FUNCTIONS AND VARIABLES IN METRICS

Criterion	Description
Technology Representation *for variables and functions*	A variable or function represents a certain aspect within the considered technologies, i.e., business process models in this case. This enables an automatic measurement.
Comprehensible Question *for variables and functions*	If Technology Reflection is not fulfilled, for example if expert knowledge is necessary, a comprehensible and unamabigious question can be formulated that can be answered by experts and does not require interpretation.
Composition *for functions*	If the previous criteria are not fulfilled, the considered function is composed of other functions using automatically measurable operators.

IV. EVALUATION

In this section, the methodology introduced in the previous section is applied and evaluated by means of a real-world example. For that purpose, first, the scenario is introduced. Afterwards, the quality criteria for business processes are exemplarily derived from a quality model. They constitute the basis for the business process analysis. In a next step, the business process is modeled. Finally, this business process is analyzed and revised using the criteria established before. The approach is shown in Fig. 2.

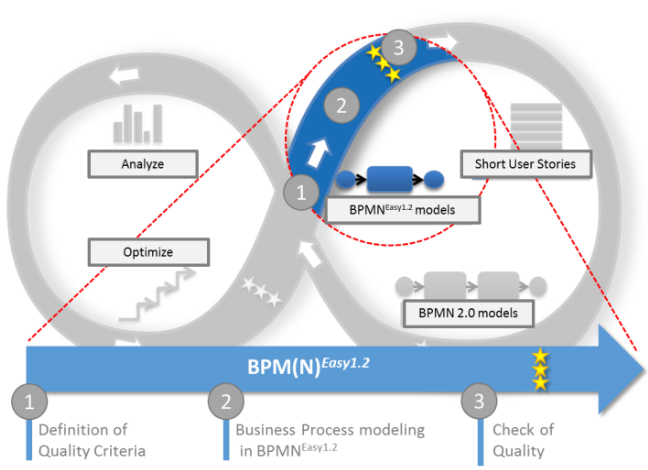


Figure 2: Quality gate and evaluation approach.

A. Scenario

The considered scenario "offer creation" originates from one of our industry partners. The activities in this scenario describe how to create a new offer. Participating roles are Sales, Accounting, and Calculation.

The corresponding business process is expected to be modeled with $BPMN^{Easy1.2}$. Besides the agile BPM methodology $BPM(N)^{Easy1.}$, $BPMN^{Easy1.2}$ is also an appropriate business process modeling language which uses BPMN 2.0 [44] but reduces its complexity so that it fits an agile BPM approach. After modeling the process using $BPMN^{Easy1.2}$, the resulting model can be enhanced, e.g., to create a model for business process automation.

To verify the correctness of the business process model, participants interacted with each other closely. All definitions were moderated and monitored by a BPM expert. If for example the IT experts had some understanding issues, these could be dealt with by direct dialog with the other participants.

B. Quality Criteria

To ensure the high quality of the business process model that is expected to be created for the scenario, first, the quality criteria have to be defined. These criteria constitute the basis for the evaluation. As described before, an existing quality model should be reused when possible. Based on this model, quality indicators should be derived as exactly formalized measures. Furthermore, for each quality indicator the expected target value has to be determined. The combination of quality indicators and target values represent the quality criteria.

In this article, the quality model and its attributes introduced by Lohrmann et al. [8] are chosen. In the following, two quality attributes and their correlating quality predicates are refined to derive quality indicators and target values.

1) Controlled resource consumption in activities: According to Lohrmann et al. [8], a business process fulfills this predicate when activities within the process are designed to avoid materials waste and capacity waste. This information is too abstract to be comprehensible on a certain business process model as it is not explained how this waste is reflected in process design. For that reason, the predicate and its quality attribute have to be refined into quality indicators.

For this purpose, best practices that could be identified in earlier projects are tested for their suitability to represent the considered predicate and its quality attribute. One best practice suitable in this case is that for each role at least two persons have to be available. This ensures that in case of a person being absent still another person can continue the work and other persons do not have to wait and to be idle, which represents a capacity waste. As the predicate refers to the business process as a whole, also the refinement has to be measured on the entire process. Thus, the indicator measures the degree to which the participating roles have more than one person assigned. This indicator can be formalized as metric (1) similar to the ones introduced by Gebhart et al. in [45]. Table III describes the used elements.

$$PAR(bp) = \frac{|\,F(R(bp), r, HSP(r))\,|}{|\,R(bp)\,|} \qquad (1)$$

TABLE III. VARIABLES AND FUNCTIONS USED FOR PAR (1)

Element	Description
PAR(bp)	*Person Availability of Roles:* Degree to which roles in business process bp have more than one person assigned
R(bp)	*Role of Business Process:* roles used in business process bp
F(e, v, c)	*Filter:* filter the elements e by condition c that uses the variable v as iterator
HSP(r)	*Role Has Several Persons:* true if role r has more than one person

TABLE IV. FULFILLED CRITERIA FOR PAR (1)

Element	Fulfilled Criteria
bp	Technology Representation: The considered business process is represented by the BPMN process file
PAR(bp)	Composition: This function is composed of other functions and all operations can be automated.
R(bp)	Technology Representation: The roles are represented by the pools and lanes within the BPMN business process model
F(e, v, c)	Composition: This function is requires other functions as input and the filter operation can be automatically performed.
HSP(r)	Comprehensible Question: This aspect is not measurable on standard BPMN 2.0 artifacts. Thus, it has to be answered by an expert, but the question is easily to understand, unambigious and comprehensible: "Are more than one person assigned to role r?" As input, a boolean value is expected.

Based on the quality indicator, Table V can be derived, which shows the possible values of the quality indicators and their interpretation. The value 1 is the desired one as it represents the case that all roles within the business process are filled with at least two persons. Thus, 1 is the target value for the quality indicator PAR.

TABLE V. INTERPRETATION OF VALUES FOR PAR (1)

Value	Interpretation
0	No role within the business process is filled with at least two persons
Between 0 and 1	Some roles are filled with less than two persons
1	All roles within the business process are filled with at least two persons

In order to prove the suitability of this quality indicator as quality indicator in an agile environment, in Table IV for each element used in the formalization the criteria introduced in Table II are checked. The table shows that for each element the criteria are fulfilled. Thus, the quality indicator is a valid indicator for an agile environment. As mentioned before, we assume business process models using BPMN 2.0 [44], respectively the reduced language $BPMN^{Easy1.2}$.

Another best practice mentioned by Lohrmann et al. [8] that influences the controlled resource consumption is the usage of work item lists for all user groups. We assume that the user groups are represented by the roles participating in the business process. Therefore, the variables and functions applied for the previous metric PAR (1) can be partially reused. For each participating role, the usage of work item lists has to be requested. In a modeled business process, this cannot be recognized. This is the reason, why this information has to be answered by an expert.

$$WILU(bp) = \frac{|\, F(R(bp), r, UWIL(r)) \,|}{|\, R(bp) \,|} \quad (2)$$

TABLE VI. VARIABLES AND FUNCTIONS USED FOR WILU (2)

Element	Description
WILU(bp)	*Work Item List Usage:* Degree to which roles within the business process bp use work item lists
UWIL(r)	*Role Uses Work Item List:* true if role r uses work item list

The used functions fulfill the criteria described in Table II as WILU is a composition of other functions and UWIL can be answered by a business analyst. Based on this quality indicator, Table VII shows the interpretation of values for WILU. According to this table, a value of 1 represents the case that all roles within the business process use work item lists. Thus, this value is the desired one.

TABLE VII. INTERPRETATION OF VALUES FOR WILU (2)

Value	Interpretation
0	No role within the business process uses work item lists
Between 0 and 1	Some roles use work item lists
1	All roles within the business process use work item lists

2) Controlled skill employment: A business process can only be efficiently performed when skill employment is controlled. According to Lohrmann et al. [8], this quality attribute or predicate is fulfilled when all activities are documented and trained. This refinement can be used as measurement. In BPMN, these activities are represented by manual tasks or tasks that are not further specified yet.

$$CSE(bp) = \frac{DT(bp) + TT(bp)}{2} \quad (3)$$

$$DT(bp) = \frac{|\, F(MT(bp), t, D(t)) \,|}{|\, MT(bp) \,|} \quad (4)$$

$$TT(bp) = \frac{|\, F(MT(bp), t, T(t)) \,|}{|\, MT(bp) \,|} \quad (5)$$

Also, in this case, all used functions and variables are described in Table VIII. They fulfill the required criteria defined in Table II as the manual tasks represent certain aspects within the technology and the other functions are either composed of others or comprehensible questions can be formulated as for D(t) and T(t).

TABLE VIII. VARIABLES AND FUNCTIONS USED FOR CSE (3, 4, 5)

Element	Description
CSE(bp)	*Controlled Skill Employment:* Degree to which skill employment is controlled in business process bp
DT(bp)	*Documentation of Tasks:* Degree to which manual tasks in business process bp are documented
D(t)	*Documentation:* true if task t is documented
TT(bp)	*Training of Tasks:* Degree to which manual tasks in business process bp are trained
T(t)	*Training:* true if task t is trained

According to Table IX, all manual tasks within the business process are documented and trained, when CSE (3) returns the value 1. Thus, this value is the desired and expected one.

TABLE IX. INTERPRETATION OF VALUES FOR CSE (3)

Value	Interpretation
0	No manual task within the business process is documented or trained
Between 0 and 1	Some manual tasks within the business process are documented and trained
1	All manual tasks within the business process are documented and trained.

In this article, we focus on this best practice as solely quality indicator for the considered predicate. If further best practices, standards, or guidelines can be identified as influencing quality indicators, they can be added later and have to be weighted.

Thus, in this article, four quality indicators with appropriate quality criteria are exemplarily identified. These concrete and partially automatically measurable quality indicators again influence more abstract quality attributes. Fig. 3 shows the derived quality model.

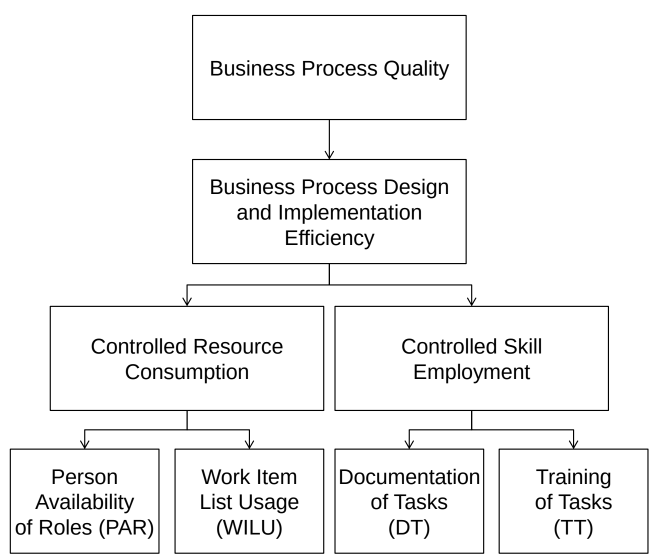


Figure 3. Derived quality model.

C. Business Process Modeling

After defining the quality criteria for the business process, in a next step, the business process can be modeled. The business process has been captured using the mobile application BPM Touch [46] (see more details in Section V). The mobile application supports agile BPM approaches by help of new information input concepts, e.g., the capturing of business process activities such as “Enter Offering” (c.f. fig. 4) can be done by audio recording instead of losing time writing this information down. The core of the user interface paradigm focuses on a simple touch technique. The modeling of business processes is carried out exclusively by touching a mobile device such as smartphone or tablet. For recording audio files participants can use built-in features such as the camera function or microphone of their mobile devices. Applying this mobile tool, a first draft of the business process has been modeled from the start event to end event. Fig. 4 shows the business process model.

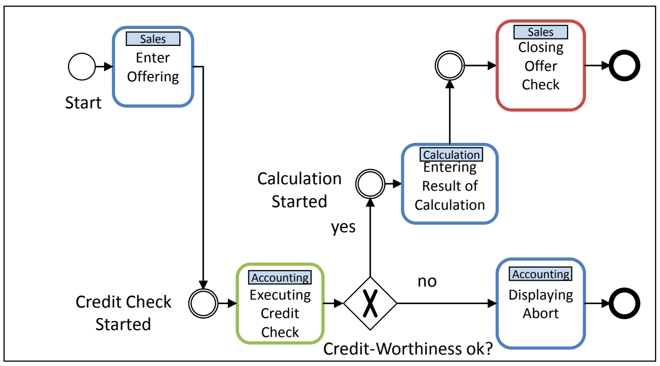


Figure 4. Offer creation business process.

The business process requires three roles: Sales, Calculation and Accounting and follows two different paths. In case of a successful credit check, the Sales can finalize the offer otherwise the business process will be aborted. During the scenario, the first draft of the business process model has been designed. BPMN$^{Easy1.2}$ provides three different activity types: manual (green form), semi-automated (blue form) and automated (red form). For instance, the “Enter Offering” activity is computer-aided and can be defined as a semi-automated activity. In addition, the required user stories have been described according to an agile methodology.

D. Analysis

When the business process is modeled, it can be analyzed regarding the prior defined quality criteria. For that purpose, the identified quality indicators are applied on the business process model and the results are compared to the expected and desired target values. To analyze the business processes, we adapted and applied the QA82 Analyzer to calculate the metrics. Furthermore, this tool enables to create questions for metrics that cannot be calculation but have to be answered by business analysts. Details about this analysis tool are provided in Section V.

1) Controlled resource consumption in activities: Applied on the modeled business process, the metric PAR (1) returns a value less than 1 as we realized by means of interviews that not every role is filled by at least two persons yet, i.e., HSP(r) is not true for all roles. Table V shows how to interpret this value. In order to fulfill the predicate of controlled resource consumption in activities, the metric is expected to return 1 as desired value. Thus, the business analyst is made aware to ensure that some further persons have to be assigned to roles with only one person. Even though if this is not possible, the business analyst gets the information that this fact represents a critical point for the efficiency of the business process.

Furthermore, in our scenario, only the Accounting uses with work item lists. This is the reason, why the second quality indicator WILU also returns a value less than 1. As described before, a value of 1 is the desired one. Thus, the business analyst is also made aware, how the controlled resource consumption and thus the business process design and implementation efficiency as part of the business process quality can be increased.

2) Controlled skill employment: In our scenario, all tasks represent manual tasks as automation has not been specified yet. When the metric is calculated, the business analyst has to answer, whether all of these tasks are documented and trained. In our scenario, the business analyst realized that this is not the case. Only some tasks are documented and trained. Thus, the metric returns a value less than 1. The interpretation of this value is shown in Table IX.

By applying the refined metrics, the business analyst is made aware that the documentation and training is important for the efficiency of the business process. If the metric returns a value less than 1, the analyst gets the information that further documentation and training effort is necessary.

E. Revision

Based on the analysis results, the business analyst can revise the business process. In our scenario, the business analyst was made aware that in some cases only one person is responsible for a certain task, that more work item lists should be used, and that not all tasks are documented and trained. By means of this information, the business analyst can revise the business process and the organization of the company.

In our scenario, the application of formalized quality indicators enabled the business analyst to systematically increase the quality of the business process. Furthermore, by means of the exact formalization, misunderstandings and interpretation effort could be avoided. The time for evaluating the business process could be reduced and the analysis is repeatable which enables a comparison of results over time. By applying appropriate tools, the efficiency of the evaluation could be further increased. The business process could be modeled in a lightweight manner and the analysis tool guided the business process through the entire evaluation by asking necessary questions. For that reason, in the next section, the applied tools are further described.

V. Tool Support

This section describes the functionality of the used tools in detail. During the evaluation two tools have been applied: BPM Touch [46] and QA82 Analyzer [47].

A. *Modeling Business Processes*

Existing solutions mainly represent a transformation of existing desktop BPM tools on mobile devices. In contrast, the described mobile application implements innovative concepts. The application implements user-friendly features and potentials of mobile devices were consistently emphasized.
Fig. 5 shows the main screen for modeling a BPMN$^{Easy1.2}$ business process.

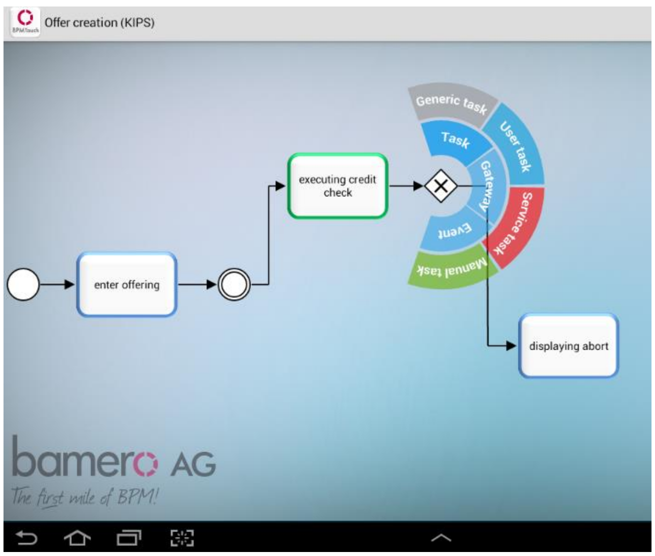


Figure 5. BPM Touch to model business process mobile

After a business process has been selected by a simple click on the sidebar, the business process model appears on the screen. The sidebar (on the left) contains all stored processes and allows a quick navigation between the business processes. The menu bar on the top provides basic functions to, for instance, create or save a new business process. In addition, a “share”-button allows the direct distribution of BPMN$^{Easy1.2}$ models to other participants. For example, business process models can be shared via e-mail, cloud service or automatically generated PowerPoint slides. The menu bar changes automatically in an “edit mode” if a business process item is touched. Furthermore, the handling of the canvas is very intuitive. For example, the start of each business process is already defined by a default start event. In addition, a "long touch" calls a highly innovative pie menu. This menu offers all possible elements of BPMN$^{Easy1.2}$.

All these functions of the mobile application address and support every activity during the agile BPM approach. For instance, BPMN$^{Easy1.2}$ models can be modeled directly from every participant without losing a great deal of time on training or coaching.

B. *Quality of Business Processes*

In order to increase the efficiency of quality analyses especially in agile environments, an appropriate tool support is necessary. For that purpose the already existing QA82 Analyzer [47] (Fig. 6) was applied. It is suited for agile environments and hybrid quality indicators that combine full-automated analyses and questions that have to be answered by the user for the following reasons: First, it supports the integration of custom quality models and combines the measure of model elements with questions that can be answered by experts, i.e., process analysts in this case. Second, the QA82 Analyzer can be integrated in business process modeling tools, such as BPMN$^{Easy1.2}$, using web services. This enables the display of quality analysis results directly in existing environments. Finally, the QA82 Analyzer allows the provisions of advices about how to improve the quality. As result, process analysts can model business processes using their modeling tool and directly get hints about how to design the process to improve their quality based on the custom quality model.

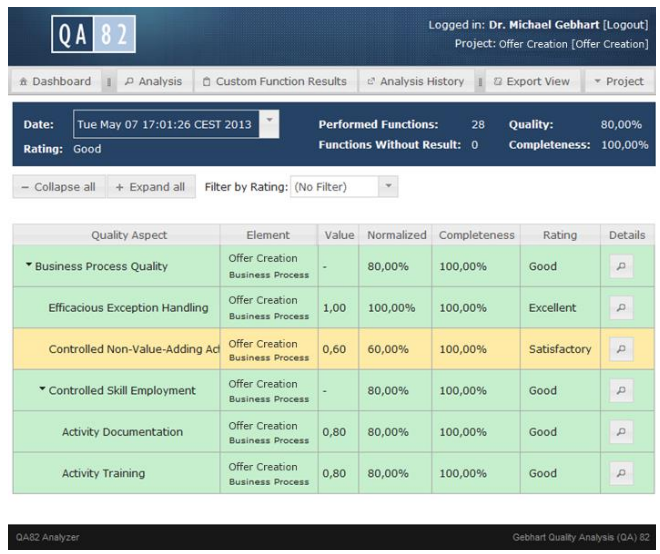


Figure 6. QA82 Analyzer to analyze business process.

To adapt the QA82 Analyzer for the evaluation of business processes in agile BPM, the quality model based on the quality attributes of Lohrmann et al. and the derived quality indicators were formalized and integrated into the QA82 Analyzer. This includes the mapping of functions to technology, i.e., to BPMN 2.0 artifacts, and the formulation of appropriate questions if necessary.

The QA82 Analyzer uses the concept of a query-based static analysis introduced by Gebhart in [48]. For that purpose, so-called information providers are utilized. When the user wants to query the quality of a certain business process, this information need is sent to the Analyzer component of the QA82 Analyzer. This component then tries to satisfy this information need. For that purpose, it looks up and queries available information providers. An information provider is a software component that is able to receive and to possibly answer a query. The concept of an information provider is further refined into technology providers and refinement providers. A technology provider on the one hand is able to answer the query directly by means of technical information. For example, a BPMN technology provider is able to answer queries like "Return the number of tasks within the business process" or "Return all manual tasks within the business process". There might further technology providers exist. In the context of the evaluation of services in service-oriented architectures, a WSDL technology provider and SCA technology provider have been developed. These information providers are able to answer queries like "Return all provided operations of a certain service". A refinement provider on the other hand is able to refine a query into several sub-queries and to compose the answers of the sub-queries to the answer for the original query. For example, when the roles performing manual tasks are required, first, the manual tasks of the business process are queried. Afterwards, for each manual task returned by the BPMN technology provider, the performing roles are queried. This is again answered by the BPMN technology provider. The approach is illustrated in Fig. 7

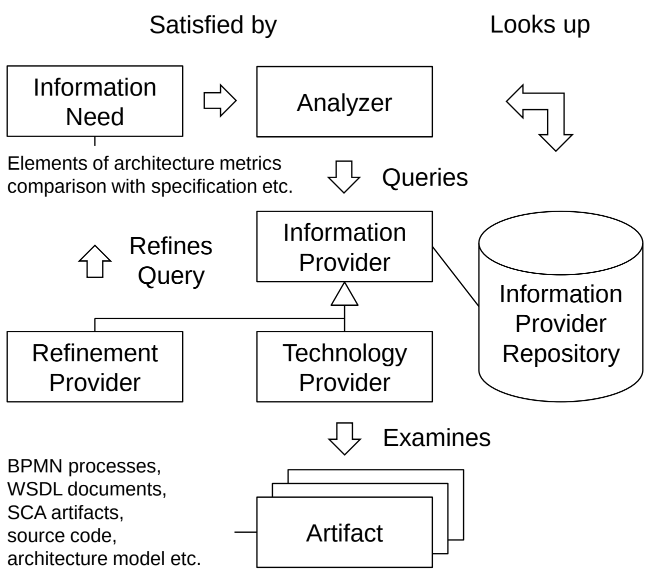


Figure 7. Query-based analysis approach.

The advantage of this approach is that with minimal effort new technologies and queries can be supported. To support the quality model introduced in this article, a BPMN technology provider was developed. Furthermore, to implement the quality model, a BPM refinement provider was developed that represents the quality attributes of the quality model. For example, the quality indicator Controlled Skill Employment (CSE) described in Table VI is contained by this refinement provider. If the Controlled Skill Employment is queried for a certain business process, the Analyzer sends the query to all information providers. The information providers try to answer the query, but there is only the refinement provider representing the quality model that is able to answer it. The refinement provider knows that to answer the query it first has to get the results for Documentation of Tasks and Trainings of Tasks for the considered business process. Thus, it sends these queries sequentially to the Analyzer component. The Analyzer component again tries to answer the queries by sending it to all available information providers. In this case, again the refinement provider is able to understand and answer the query. For Documentation of Tasks it knows that it has to find out how many of the manual tasks within the business process have been documented. For that purpose, the refinement provider first sends a query to the Analyzer component to receive all manual tasks in the business process. The Analyzer component tries to answer this query by asking all available information providers. In this case, the BPMN technology provider is able to answer the query. It returns all manual tasks in the process. Next, the Analyzer component returns this result to the refinement provider. Now, the refinement provider tries to answer the Documentation for each returned manual task. Thus, for each of the manual tasks, a query is sent to the Analyzer component. Once again, this component tries to answer the queries. However, in this case, there is no information provider able to answer the query. If this is the case, the Analyzer component does something special: It creates a question for the analyst to answer this query manually. Thus, in this case, for each manual task, the analyst is asked whether this one is documented or not. The same happens to the query about the training of the manual tasks. For each task, a question is generated and the analyst has to answer whether the task is trained or not. When the analyst has answered these questions, the analysis can be started again. This time, the answers are considered within the analysis. This means, when the Analyzer component tries to answer the queries Documentation and Training for each manual task, the answers of the analyst are returned as result. These results are used by the refinement provider to create the answer for the original query Controlled Skill Employment. Fig. 8 illustrates the interaction between the Analyzer component and the BPM information provider.

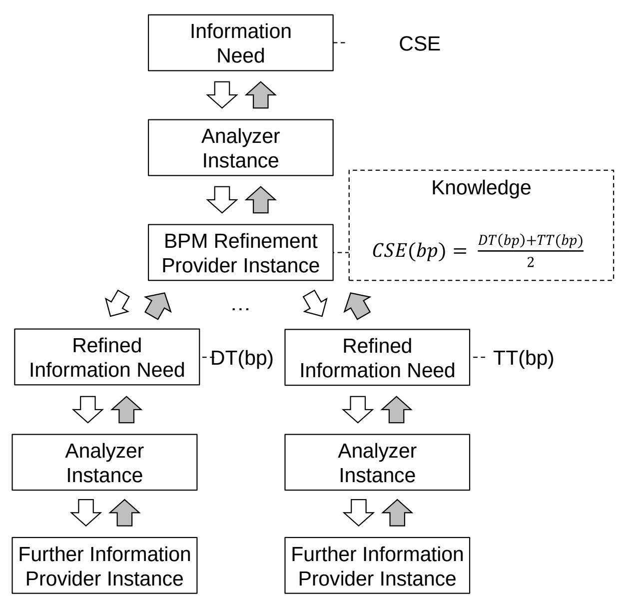


Figure 8. Interaction between Analyzer and BPM refinement provider.

As result, the QA82 Analyzer can be used to apply the identified quality indicators on any BPMN 2.0 or BPMN$^{Easy1.2}$ compliant business process. The concept of business process quality is step-by-step refined until information can be automatically measured on concrete business process elements, such as BPMN tasks, or a concrete question can be asked. The answers for these questions are stored and considered during the next quality analysis iteration. By this means, the business analyst does not have to answer the questions again and again. Only the first time certain information has to be added manually, the question has to be answered. This increases the efficiency of the evaluation and reduces the effort. Furthermore, the QA82 Analyzer stores the analysis results. This enables a comparison of results over time as previous results are accessible.

VI. Conclusion and Outlook

In this article, we demonstrated the application of business process quality models to support agile business process management (BPM) and to assure a high quality of created business process (applications). For that purpose, we provided a literature overview and described the general connections between quality assurance and agile BPM. Furthermore, we chose the agile BPM method BPM(N)$^{Easy1.2}$ and the quality model introduced by Lohrmann et al. [8] exemplarily. After the identification of the challenges, e.g., the degree of employee skills, we have shown how to address them.

First, the application of business quality models was aligned with an agile methodology. As essential deficit, the abstraction of available quality attributes was identified. To solve this issue, we demonstrated how these quality attributes can be refined to be applicable in agile environments. Hereby, we focused on the end users mainly and explicated how common language can be used to generate higher quality easily. Finally, we illustrated necessary tool support to increase the efficiency of quality analyses.

To illustrate our work, an evaluation in the context of a real business process was chosen. We described the scenario, the specification of quality criteria, the modeling of the business process, and finally the analysis and revision of the created business process model. The refined quality attributes enabled the systematic analysis of this process and the results helped the process analysts to revise the process and its environment in a quality-oriented manner. Even though the quality of a business process includes a lot of further aspects not covered in this article, the application of a fine-grained quality model increases the awareness of relevant aspects and supports the creation of high-quality business processes.

Thus, our approach enables companies and their analysts to increase the quality of created business processes whilst reducing at the same time effort and costs for quality assurance. All participants, such as business analysts, can create business process models using their preferred modeling tool and manually analyze created models in a systematic manner. When integrating the quality model and appropriate quality analysis methods into the tools, the analysts can even receive feedback about the quality of created models directly. Finally, derived advices could be shown and help the analysts to improve the created business models with regard to quality attributes that influence business-related goals.

Next, we will consider further quality attributes and derive appropriate quality indicators to enhance the created quality model. As described in this article, we will focus on reuse of existing quality attributes. In addition, we will investigate how all participants can communicate and interact more efficiently to increase the quality continuously during the entire business process lifecycle.

Also, the used tools will be investigated more in detail focusing on improving and extending the tools itself. For instance, the integration with existing modeling tools has to be enhanced. Finally, the approach is expected to be applied in further business process management projects to identify advantages and also weaknesses that have to be examined.

References

[1] M. Gebhart, M. Mevius, and P. Wiedmann, “Application of business process quality models in agile business process management,” The Sixth International Conference on Information, Process, and Knowledge Management (eKNOW), 2014, pp. 152-158, ISBN: 978-1-61208-329-2.

[2] T.M. Bekele and Z. Weihua, “Towards collaborative business process management development current and future approaches,” in Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference, 2011, pp. 458-462.

[3] J. Eder and H.Pichler, “Business Process Modeling and Workflow Design,” in D.W. Embley and B. Thalheim (eds), Handbook of Conceptual Modeling, Springer, 2011, pp. 259-286.

[4] J. Highsmith, “Agile software development ecosystems,” Pearson Education, 2002, pp. 26-34.

[5] Agile Manifesto, http://agilemanifesto.org. [accessed: May 21, 2013]

[6] M. Mevius, R. Stephan, and P. Wiedmann, “Innovative approach for agile bpm,” The Fifth International Conference on Information, Process, and Knowledge Management (eKNOW), 2013, ISBN 978-1-61208-254-7.

[7] M. Mevius and P. Wiedmann, „BPM(N)$^{Easy1.2}$ – Gebrauchssprachliche gestaltung IT-basierter prozesse,“ 8. Workshop „Bewertungsaspekte service- und cloudbasierter Achitekturen (BSOA)“ der GI Fachgruppe „Software-Messung und -Bewertung“, 2013, pp. 31-46.

[8] M. Lohrmann and M. Reichert, “Understanding business process quality,” in Business Process Management - Theory and Applications, Springer, 2013, pp. 41-73.

[9] J. Krogstie, “Model-Based development and evolution of information systems – a quality approach,” Springer-Verlag, London, 2012.

[10] J. Highsmith, “Agile project management – creating innovative products,” Addison-Wesley Professional, 2009, pp. 15-16.

[11] F. Salger, M. Bennicke, G. Engels, and C. Lewerentz, “Comprehensive architecture evaluation and management in large software-systems,“ Fourth International Conference on the Quality of Software Architectures, 2008, pp. 205-219.

[12] E. Mnkandla and B. Dwolatzky, “Defining agile software quality assurance,” International Conference on Software Engineering Advances, IEEE 0-7695-2703-5/06, 2006.

[13] J. vom Brocke and M. Rossmann, “Handbook on business process management 1: introduction, methods, and information systems“, Springer Verlag, Heidelberg, 2010.

[14] R. Cheng, S. Sadiq, and M. Indulska, “Framework for business process and rule integration: a case of BPMN and SBVR,” Business

Information Systems Lecture Notes in Business Information Processing Volume 87, 2011, pp. 13-24.

[15] P. Antunes, D. Simões, L. Carriço, and J. A. Pino, "An end-user approach to business process modeling," Journal of Network and Computer Applications 36 (2013), pp. 1466 -1479, 2013.

[16] S. Bittmann, D. Metzger, M. Fellmann, and O. Thomas, "Additional information in business processes: a pattern-based integration of natural language artefacts," Lecture Notes in Informatics, Modellierung 2014, ISBN 978-388579-619-0, pp. 137-152.

[17] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), ISO/IEC Standard No. 9126: Software Engineering – Product Quality; Parts 1-4.

[18] International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), ISO/IEC Standard No. 25000: Software Engineering – Software Product Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE.

[19] R. Heinrich and B. Paech, "Defining the quality of business processes," in Lecture Notes in Informatics Vol. P-161, 2010, pp. 133-148.

[20] L. Sánchez-González, F. Ruiz, F. García, and M. Piattini, "Improving quality of business process models," in Evaluation of Novel Approaches to Software Engineering, 2013, pp. 130-144.

[21] L. YeongSeok, B. JungHyun, and S. Seokkoo, "Development of quality evaluation metrics for BPM (business process management) system," in Proceedings of the Fourth Annual ACIS International Conference on Computer and Information Science, 2005, pp. 424-429.

[22] International Organization for Standardization (ISO), ISO Standard No. 9000: Quality Management Systems – Fundamentals and Vocabulary.

[23] R. Kneuper, „Gokyo ri: messung und bewertung der qualität von entwicklungsprozessen am beispiel des v-modell xt," in O. Linssen, M. Kuhrmann (Eds.): Qualitätsmanagement und Vorgehensmodelle, Shaker, 2012, pp. 25-34.

[24] G. N. Suarez, J. Freund, and M. Schrepfer, "Best practice guidelines for BPMN 2.0," in BPMN 2.0 Handbook, 2011, pp. 153-165.

[25] T. Erl, SOA – Principles of Service Design, Prentice Hall, 2008. ISBN 978-0-13-234482-1.

[26] T. Erl, "SOA – Design Patterns", Prentice Hall, 2008. ISBN 978-0-13-613516-6.

[27] S. Cohen, "Ontology and Taxonomy of Services in a Service-Oriented Architecture," Microsoft Architecture Journal, 2007.

[28] N. Josuttis, SOA in Practice – The Art of Distributed System Design, O'Reilly & Associates, 2007. ISBN 978-0-596-52955-0.

[29] OMG, "Service oriented architecture modeling language (SoaML) – specification for the uml profile and metamodel for services (UPMS)," Version 1.0.1, 2012.

[30] OMG, "Unified modeling language (UML), superstructure," Version 2.4.1, 2011.

[31] M. Gebhart and J. Bouras, "Derivation of web service implementation artifacts from service designs based on soaml," International Journal on Advances in Software, 6(1&2), 2013, pp. 170-180.

[32] W3C, "Web Services Description Language (WSDL)," Version 1.1, 2001.

[33] W3C, "XML Schema Part 0: Primer Second Edition," 2004.

[34] Open SOA (OSOA), "Service component architecture (SCA), SCA assembly model V1.00," http://osoa.org/download/attachments/35/SCA_AssemblyModel_V100.pdf, 2009. [accessed: January 04, 2011]

[35] M. Gebhart and S. Sejdovic, "Quality-oriented design of software services in geographical information systems," International Journal on Advances in Software, 5(3&4), 2012, pp. 293-307.

[36] M. Gebhart, "Service identification and specification with SoaML," in Migrating Legacy Applications: Challenges in Service Oriented Architecture and Cloud Computing Environments, Vol. I, A. D. Ionita, M. Litoiu, and G. Lewis, Eds. 2012. IGI Global. ISBN 978-1-46662488-7.

[37] R. Meziani and I. Saleh, "Towards a collaborative business process management methodology," International Conference on Multimedia Computing and Systems (ICMCS), 2011.

[38] F. Schnabel, Y. Gorronogoitia, M. Radzimski, F. Lecue, N. Mehandjiev, and G. Ripa, "Empowering business users to model and execute business processes," BPM 2010 Workshops, Springer, 2011, pp. 433-448.

[39] A. H. Mohammad, T. Alwada'n, and J. Adabneh, "Agile software methodologies: strength and weakness," International Journal of Engineering Science and Technology, 2013, ISSN : 0975-5462, pp. 455-459.

[40] J. Barlow, J. Giboney, M. Keith, D. Wilson, R. Schuetzler, P. Lowry, and A. Vance, "Overview and guidance on agile development in large organization," Communications of the Association for Information Systems, Vol. 29, No. 2, 2011, pp. 25-44.

[41] M. Huo, J. Verner, L. Zhu, and M. Babar, "Software quality and agile methods," 28th Annual International Computer Software and Applications Conference, IEEE 0730-3157/04, 2004.

[42] A. Janus, A. Schmietendorf, R. Dumke, and J. Jäger, "The 3C approach for agile quality assurance," Emerging Trends in Software Metrics (WETSoM), 2012 3rd International Workshop, 2012.

[43] M. Gebhart and S. Abeck, "Quality-oriented design of services," International Journal on Advances in Software, 4(1&2), 2011, pp. 144-157.

[44] OMG, "Business process model and notation (BPMN)", Version 2.0, 2011.

[45] M. Gebhart and S. Abeck, "Metrics for evaluating service designs based on soaml," International Journal on Advances in Software, 4(1&2), 2011, pp. 61-75.

[46] bamero AG, "BPM Touch® - The Revolution," http://www.bamero.de., accessed on 08/05/2014.

[47] Gebhart Quality Analysis (QA) 82 GmbH, QA82 Analyzer, http://www.qa82.com. [accessed: May 21, 2013]

[48] M. Gebhart, "Query-based static analysis of web services in service-oriented architectures," International Journal on Advances in Software, 7(1&2), 2014, pp. 136-147.

Teleconsultation/Telediagnosis using Teledentistry Technology: a Pilot Feasibility Study

Rodrigo Mariño
Oral Health CRC, Melbourne Dental School
University of Melbourne
Melbourne, Australia
rmarino@unimelb.edu.au

Matthew Hopcraft
Oral Health CRC, Melbourne Dental School
University of Melbourne
Melbourne, Australia
m.hopcraft@unimelb.edu.au

Utsana Tonmukayakul
Dental Health Services Victoria
Melbourne, Australia
utsana.tonmukayakul@deakin.edu.au

David Manton
Oral Health CRC, Melbourne Dental School
University of Melbourne
Melbourne, Australia
djmanton@unimelb.edu.au

Parul Marwaha
Oral Health CRC, Melbourne Dental School
University of Melbourne
Melbourne, Australia
pmarwaha@unimelb.edu.au

Andrew Stranieri
Centre for Informatics and Applied Optimisation,
Federation University
Ballarat, Australia
a.stranieri@federation.edu.au

Richard Collmann
Victorian eResearch Strategic Initiative
University of Melbourne
Melbourne, Australia
rcollmann@unimelb.edu.au

Ken Clarke
Institute for a Broadband-Enabled Society
University of Melbourne
Melbourne, Australia
clak@unimelb.edu.au

***Abstract*—This study assessed the feasibility of a teledentistry model for teleconsultation and telediagnosis in Residential Aged Care Facilities. Study feasibility was defined by the ability to develop remote treatment plans. Reliability of the remote assessments was assessed by comparing with those performed by traditional face-to-face oral examinations. An intraoral camera was operated by trained teledentistry assistants with the aim of screening residents for oral diseases and pathological conditions. The model was supported by traning and an instructional kit for the introral camera operators. The structure, content and delivery of the program, was evaluated. Residents' views about the structure, content and delivery of the program were also evaluated. A total of 50 residents participated in this assessment. Results indicated that the proposed teledentistry approach for oral health screening is feasible and reliable as an alternative to traditional oral health examination. Residents expressed high levels of satisfaction with the teledentistry service. This study provides an innovative solution towards closing the service delivery gap in the provision of sustainable oral health care services to underserviced populations (e.g., nursing homes, rural areas).**

Keywords - oral health, teledentistry, intraoral camera, nursing homes

I. Introduction

Improvements in oral health in Australia over the past 50 years have translated into a greater proportion of older individuals retaining more of their natural teeth, increasing the prevalence of caries and periodontal disease. Older people living in Residential Aged Care Facilities (RACFs) have been identified as a significant risk group for oral diseases in Australia. In addition, the changing demography and oral health needs of older Australians will present many challenges for the dental profession over coming decades. As an initial stage to address these challenges, a project was organized to test whether improvements in accessibility and appropriateness of oral health services can be achieved by utilizing advanced ICT techniques to screen for oral disease in older people living in RACFs [1].

In 2005 more than 41,000 residents in the state of Victoria, Australia lived in high or low-care residential facilities on a permanent basis; with just over half being dentate and having high dental treatment needs [2][3]. Significant barriers to accessing dental services exist. Residents are often physically and cognitively impaired, medically compromised and dependent on others to maintain their oral hygiene.

Face-to-face patient examinations are regarded as the most accurate method for correct oral health diagnosis. However in Victoria only 11% of aged care facility residents have seen a dentist in the past 12 months, as there are few dentists available to provide dental care for residents [4][5]. In fact, only half of Victorian dentists reported providing care to residents of RACFs, and those dentists spent on average only one hour per month providing care in this setting [5].

As the capability of information and communication technology (ICT) has risen, the use of ICT for data collection has increased. Expanded use of ICT has provided clinicians with alternatives to the traditional face-to-face oral examinations. This shift in focus has resulted in a vast increase in the number of published articles that include some form of either synchronous (real-time) or asynchronous (delayed image examination), teleconsultation/ telediagnosis [6].

Teledentistry can be successfully implemented in different oral health settings to deliver improved health outcomes and positive health professional-patient experiences [1][6]. Teledentistry has the potential to reduce the number of inappropriate referrals by screening patients to ensure that only those who need to see a specialist go on waiting lists. This ensures efficient use of scarce health resources, increasing access to specialist care, improving specialist productivity and supporting enhanced oral health across society.

Potentially, teledentistry could benefit an expanding segment of the population in relative and absolute terms; namely, older people living in RACFs and rural areas. According to the 2006 Australian Census, 55-64 year-olds made up 11.8% and those 65 years and over 13.7% of the total Victorian population [7].

Older people are proportionately over-represented in rural and regional Australian communities and these communities are ageing more rapidly than their metropolitan counterparts. Therefore, a teledentistry approach could potentially target a rapidly expanding segment of the population with special oral health needs. Other parts of the world have similar demographic and geographical problems making this study equally relevant to them.

A three-stage study was designed to address priorities established by Australia's National Oral Health Plan 2004-2013 for ‘Older People’ targeting older people living in RACFs, (an underserviced, high-risk population and one with major oral health needs) [8]. This project was conceived in an effort to promote affordable, timely oral health care and to test an oral health care model in which ICT is used with the aim of extending clinical care to residents who are physically separated from the examining oral health professional. This study was also a response to serious dental workforce shortages in caring for this group and provides opportunities to supplement traditional methods of oral diagnosis, care delivery and health promotion.

The study builds on a University of Melbourne Institute for a Broadband Enhanced Society (IBES), Project Seed Grant, which tested the technology under laboratory conditions (proof-of-concept) and developed the instructional material for non-oral health professional operators [9]. That study demonstrated that the proposed teledentistry approach for oral health screening using an intraoral camera was feasible and reliable as an alternative to traditional oral health examination. The long-term goal of the project is to test whether improvements in accessibility and appropriateness of oral health services can be achieved by utilizing advanced ICT techniques to screen for oral disease in older people living in RACFs. Stage 2 of this three-stage study involved the field testing of this teledentistry technology and is the subject of this article.

In the last few years, economic evaluations (EE) have acquired greater importance in decision-making in health. Health service managers, programmers and planners are required to select the interventions with the highest impact, based on evidence and prioritizing of high-risk groups. Analyses of this type assist decision-makers in determining which intervention (or combination of interventions) maximizes outcomes, given the available resources. Despite this, apart from one example of economic evaluations in teleorthodontics [10], the use and application of economic evaluations in teledentistry remains limited [5].

This paper is organised in six sections. The first Section provides the foundation for understanding the need for teledentistry in this particular population. Section II is concerned with the aims and objectives of the paper. Section III describes the methodology used in this field trial of teledentistry. Section IV presents the results of the trial, including interviews with personnel and patients involved in the program. Sections V and VI discuss the results of the trial and conclude on its findings, respectively. Further steps are also discussed in last section.

II. Aims and Objectives

This paper outlines the results of the second stage of this project, which aimed to assess, on a small scale, but under real conditions, the safety of the procedures, their feasibility, as well as patients’ and health practitioners’ experiences with the technology. A comparison was conducted between face-to-face-examinations and remote examinations using an intraoral camera.

The specific objectives of this study were to:

1) assess the feasibility of using teledentistry to screen for oral diseases and conditions and to develop treatment plans for older people living in RACFs;
2) identify barriers to the adoption of a teledental approach. These included: a) general staff workload; b) professional culture and acceptance (e.g., morale, motivation, resistance to change, etc.); and c) availability of appropriate equipment,
3) test the utility of an instructional training kit,
4) assess the residents’ views of their experiences during delivery of the program, as well as feedback and information provided during the teledentistry consultation.

However, in order to assist in the development of a consensus statement about the costs of oral health interventions, and the need for economic data which can be used as a reference for national programs, this study aims to determine the costs of implementing a teledentistry program to develop treatment plans for rural RACF residents in Victoria, Australia.

III. Methods

Three RACFs within Victoria, Australia, were successfully approached to participate in this stage; two in metropolitan Melbourne (suburbs of Brunswick and South Morang) and one in rural Victoria (Stawell). These clinics worked in partnership with the University of Melbourne's Melbourne Dental School, which acted as the central co-ordination and examination site.

Five non-oral health professional teledental assistants (e.g., registered nurses - RN) in these facilities were trained to manipulate an intraoral camera and use existing and introduced ICT infrastructure to transmit video images for remote examination and diagnosis. Video was streamed over the open Internet to enable the service in the first instance, as it would otherwise be necessary to overcome firewall issues at the RACF's where it can be difficult to enlist the aid of local IT support to change port settings for such short-term trials. If ongoing services were established in the future then VPN's (Virtual private Networks) could be used to ensure the security of patients' data. An oral health professional at the Melbourne Dental School performed a 'virtual dental examination', recorded findings and developed a treatment plan for a group of selected residents

Purposive sampling was used to recruit participants. To participate in this teledentistry study, the resident was required to have the ability to understand and to provide independent informed consent, the ability to communicate with the health professional and to undergo a 15-20 minute oral examination. They could be dentate or edentate. However, because of the nature of this trial, special care was taken to select residents who had some of their natural dentition.

Although sample size calculations are not strictly necessary for a pilot study [11], a sample size of 50 residents was considered to be adequate to meet the general aims of this study. A 20% attrition rate was expected over the six month duration of the field component of this study therefore 62 residents were recruited initially.

Patients were introduced to the study by the local RN. When the patients, or their primary carers, expressed interest in participating, each received a Plain Language Statement describing the study and a Consent Form. Once informed consent was obtained, patients underwent an oral health assessment.

A SOPROLIFE® intra-oral camera was used to capture video via a custom video streaming software platform designed for the project [12]. Simulations were conducted in the IBES test-bed facility [13]. The intra-oral camera was connected via a USB cable to a laptop or mobile tablet used for bed side evaluations, containing the software that compressed and encoded the 25 frame-per-second video into an mpeg4 video stream of at least 3Mbit/sec bandwidth, and preferably a 5Mbit/s stream if network conditions allowed. This bandwidth was found to give the clinician sufficient quality to interpret the images received and removed blurring due to the motion of the camera [8].

The clinician viewed the incoming video via a PC connected to a large monitor. A large screen facilitated simultaneous viewing of both the intra-oral camera video as well as that from a second web-cam, a high definition Logitech model C920 model, capturing the overall interaction between patient and the intra-oral camera operator. This was also streamed as an mpeg4 video of minimum 3Mbit/sec bandwidth. Mpeg4 audio was also transmitted at 128kbit/s along with the images via the use of Clear One Chat 50 model microphone/speaker units also connected via USB cables. This allowed excellent quality audio communications between patient and clinician ends. For test sites that could not accommodate a 3Mbits/s stream reliably a Store and Forward version was developed that enabled the Mpeg4 file to be stored on a central server for asynchronous download by the dentists. Each examination lasted approximately 15 minutes and each minute of video created a file of approximately 1GB. Thus, the video files were large (i.e., 15 minutes produces a 15 GB file).

Using a teledentistry installation each participant received a 'virtual' oral examination, including dental and oral mucosal assessments conducted with the assistance of a trained RN at the RACF using an intraoral camera operated in communication with a remotely located oral health professional.

Intraoral camera operators should be trained in the use of the hardware. In the present study, this training took a variety of formats including seminars, simulation activities, demonstration modules, and self-training. A sixty-six page training manual including diagrams, with content organized in five modules was also prepared for these purposes.

Intraoral camera operators training consisted of six hours delivered by a dentist from the Melbourne Dental School, University of Melbourne. The first three-hour session was an introduction to teledentistry, and on oral and dental anatomy. The second three hours was a demonstration of how to operate and manipulate the intraoral camera and send files over the Internet. After these sessions, the trainees were asked to have 10 hours of self-practice with an intraoral camera in their own time

The remote oral health professional was able to communicate in real-time with both the resident and the intraoral camera operator (i.e., the RN) via a video link to assist in taking a history, and to direct the RN where necessary in the use of the intraoral camera. To have communication in real time we used Skype® and Vidyo®. However, there were several problems with Vidyo due to firewall settings.

The information obtained from this examination was recorded and transmitted to a server for review of the 'virtual dental examination' to be performed remotely at a later time. Information was registered on a conventional Dental Health Services Victoria chart for the generation of treatment plans by qualified clinicians at the Melbourne Dental School, University of Melbourne.

On completing the virtual oral examination the residents were asked to complete a seven-item teledentistry assessment questionnaire to assess his/her views on the approach. As a further verification of the approach, the ability to understand communications between the oral health professional and the resident was determined both for the conventional face-to-face and remote communications. The evaluation form consisted of statements that participants rated on a five 5-points Likert scale, depicting their level of agreement with the statement (1 'Strongly agree'; 3 'Neutral'; 5 'Strongly disagree'). The summary evaluation also contained four open-questions, so participants could include their thoughts about their experience and critiques.

Ten residents received a second oral examination by a different oral health clinician. This was a traditional face-to-face examination (the clinician present in front of the patient) with findings recorded on a conventional dental chart.

Furthermore, we wanted to explore how the introduction of teledentistry was experienced by those in charge of teledentistry on site. RNs who collected the information for this project completed a questionnaire to assess their initial attitude to, and acceptance of the practice, and their overall experience with the teledentistry approach. The utility of the instructional training kit and any other issues associated with the project were also assessed.

Due to the small sample size, only descriptive analysis was used to illustrate the participants' views about the format, content and delivery of the teledentistry program. In some cases, categorical and ordinal data were analyzed utilizing Chi square analysis (χ^2) to compare results between different oral conditions and the distribution of socio-demographic and outcome variables were performed. The level of inter-examination reliability for the degree of consistency of the two sets of examinations was assessed using the kappa statistic.

To complement this analysis, and to obtain a better understanding of the usefulness of teledentistry approach, an economic evaluation was conducted. The form of economic evaluation used in this study was cost description. A cost description is the most basic type of economic evaluation, which is a partial form of economic appraisal because it looks only at the costs of a program and provides no information on the health outcome of interest [14]. It answers the most commonly asked question when considering introducing a new program: "how much will it cost?", and provides decision makers with important information on the resources needed to introduce or expand a service. It also allows program managers to determine a unit cost or cost per unit of service (i.e., for a treatment plan).

Teledentistry, as in any telehealth model, can work according to two consultation methods: real-time and store and forward. Both methods require the same basic infrastructure connected to telecommunications networks with sufficient bandwidth. However, as low-speed, high-latency connections are more typically found in rural and remote regions, this may impact the ability to perform real-time consultations. Therefore, practitioners without adequate bandwidth would have to rely upon the store and forward modality to send data to the specialist for later review. Store and forward provision can be just as effective at presenting cases as the real-time modality.

Thus, this analysis calculated the cost per unit of a treatment plan under Australian conditions. The program or intervention under evaluation in this study involved:

1) Oral examination data recorded at the RACF using an intraoral camera operated by a trained RN who transmitted the files to a remotely located dentist for asynchronous review and treatment planning (asynchronous model).
2) Remote real-time oral examination and treatment planning with the assistance of a trained RN at the RACF using an intraoral camera operated in communication with a remotely located oral health professional (real-time model).

The question being examined was: "what would be the cost of implementing an asynchronous and a real-time teledentistry model per resident if we provide a treatment plan for a cohort of 100 residents?"This figure was based on capacity of the largest RACF in Stawell.

The costs of running the program were based on this study, expert opinion, together with other relevant sources. A micro-costing approach, or direct cost measurement from a healthcare perspective, was applied to quantify and value all resources in this program, except where costs are small and unlikely to make any difference to the study results. Costs were grouped under three main categories:

1) Training. Calculations under this heading were done under the following assumptions. It was assumed that the asynchronous and the real-time models require a trained RN to manipulate the intraoral camera. Training of the intraoral camera operator involved six hours of direct contact and ten hours of practice examinations.
   Transportation. The Australian Tax Office travel reimbursement schedule for a 2.2 litre engine was used to quantify travel cost [15]. Average distance between the aged care facilities located in Stawell and the Melbourne Dental School, Univeristy of Melbourne was applied to determine the travel costs of the dental clinician to train the RNs.
2) Salaries. The salary of oral health personnel involved in the oral examination was based on the Dental Health Services Victoria scales [16]. The RNs' salary was based on Australian Department of Health and Aging recommendations [17].
   Each examination in the asynchronous model (store and forward*)* was calculated as lasting 15 minutes, whereas examination time in the real-time model was 20 minutes. It was assumed that in the real-time model there would be more commmunication with the remote examiner to prepare and finalise the treatment plan.
3) Teledental device. The investments necessary to conduct the teleconsultations were the intraoral camera with an integrated USB dock to connect to a computer. Camera costs were computed at market value [18]. Digital images obtained by the camera can be used to

support real-time or store-and-forward teledentistry consultations.

The life span of the intaroal camera and its direct USB Camera dock was estimated at 10,000 hours (about 5 years) of normal use as per its handbook [19].

All costs were computed in Australian dollars (AUD) (1.00 AUD =0.70 EUR) at 2014 price levels. No cost discounting applied to this short-term study.

Data collection extended between October 2012 and June 2013. Ethical approvals to conduct this study were sought and obtained from the University of Melbourne.

IV. RESULTS

At the end of data collection, 50 residents from three RACFs participated in the trial from the 62 initially recruited; with 58% being female. Three RNs conducted the examinations. Twenty-two examinations were conducted in Stawell, twenty-one in Brunswick and another seven in South Morang. The majority of participants (70%) had, at least part of their natural dentition, while fifteen residents (30%) had no natural teeth. By location, nine edentulous residents were from Stawell, and six were from Brunswick (See Table I).

A teledentistry installation enabled five trained intra-oral camera operators (registered nurses) to record, use and transmit video images for the generation of treatment plans by qualified clinicians at the Melbourne Dental School, University of Melbourne. Information from the remote examination was compared with a real-life dental examination. The intra-examiner agreements for dental examination parameters determined by the Kappa index reflected an 'Excellent' agreement (Kappa=0.83) [20].

When residents were asked to rate their satisfaction with the examination, the majority of the residents were either very satisfied: (46%) or slightly satisfied (38%) with the format of the remote dental examination.

TABLE I. RESIDENTIAL AGED CARE FACILITIES RESIDENTS' SOCIODEMOGRAFIC CHARACTERISTICS

	Residential aged care facilities		
	Brunswick n (%)	*South Morang n (%)*	*Stawell n (%)*
Sex			
Male	10 (47.6)	3 (42.9)	8 (35.4)
Female	11 (52.4)	4 (57.1)	14 (63.6)
Dentate Status			
Dentate	15 (72.4)	7 (100.0)	13 (60.0)
Edentulous[a]	6 (28.6)	- (0.0)	9 (40.0)
Total	21	7	22

a. total absence of natural dentition

The majority would also recommend (strongly: 46%; or slightly: 46%) a remote dental examination to other people of their age and background. When asked about the reason for not recommending this assessment, the comments were related to the lack of provision of oral health services, in particular preventive care (See Table II).

TABLE II. RESIDENTS' RESPONSES TO TELEDENTISTRY ASSESSMENT QUESTIONNAIRE (%)

1. How satisfied were you with the remote dental examination?[a]				
Strongly satisfied	Slightly satisfied	Neutral	Slightly Dissatisfied	Strongly dissatisfied
46.0	38.0	14.0	2.0	-
2. If remote examinations were available for patients, would you recommend them to other people?				
Strongly recommend	Slightly recommend	Neutral	Slightly not Recommen d	Strongly not recommend
46.0	46.0	4.0	2.0	2.0
3. How appropriate was the format of the remote dental examinations?				
Very appropriate	Slightly appropriate	Neutral	Slightly Inappropriate	Strongly inappropriate
46.0	46.0	6.0	2.0	-
4. How satisfied were you with the review of your dental needs by the remote dentist?				
Strongly satisfied	Slightly satisfied	Neutral	Slightly Dissatisfied	Strongly dissatisfied
46.0	32.0	6.0	16.0	-
5. Were instructions from the examiner in the face-to-face exam clear and easy to understand?				
Very easy	Slightly easy	Neutral	Slightly Difficult	Very difficult
86.0	12.0	2.0	-	-
6. Were instructions from the examiner in the remote examination clear and easy to understand?				
Very easy	Slightly easy	Neutral	Slightly Difficult	Very difficult
46.0	46.0	4.0	2.0	2.0

a. n= 50

When asked about how satisfied they were with the review of oral health needs, although the majority was either satisfied (46%) or slightly satisfied (32%), three residents (6%) were neutral and, more importantly, 16% were slightly dissatisfied. Asked about the reason for this dissatisfaction, most residents' comments were related to the lack of immediate feedback on the examination.

On comparing residents' opinions on the clarity of the communications received with the face-to-face examiner (i.e., the RN), 86% of the respondents found it "Very easy", and another 12% "Easy" to understand. Residents also found it generally easy to understand remote communications (46% "Very easy" and 46% "Easy"), and another 4% were neutral about it. Nonetheless, the remainder 4% found it "Difficult" or "Very difficult" to understand remote communications, but comments were related to the foreign accent of the oral health professional that provided feedback on the examination, rather than the technology used.

Over one quarter of the residents (28%) commented that the most valuable element of the remote dental examination was its convenience. For example, by taking video images in the RACF, residents could avoid the disruption, difficulty and cost of arranging travel to visit a dentist.

Three of the five RNs that had been recruited and trained conducted intraoral examinations with the RACF's residents. Despite having written instructions and being able to successfully transmit files during the training sessions, most of the intraoral examinations (n = 28) required an oral health professional to manipulate the intraoral camera. In another eight examinations the RNs were assisted, either remotely or at the RACF, by an oral health professional on how to properly manipulate the intraoral camera and transmit the video images. The RN's performed the examination and transmitted the videos unsupported in only 14 examinations (See Table III).

TABLE III. NUMBER OF TELEDENTAL EXAMINATIONS COMPLETED BY INTRAORAL CAMERA OPERATOR

	Residential aged care facilities		
	Brunswick	*South Morang*	*Stawell*
Nurse no supervision	6	4	4
Nurse under supervision	4	0	4
Oral health professional	11	3	14
Total	21	7	22

a. Residential aged care facility

The three participating nurses provided feedback on the training material presented (i.e., a hard-copy, on-line manual and demonstrations). There was general agreement that the material presented was clear and relevant to the purposes of this project. RNs also agreed that the length of the material was right. Nonetheless, they considered that the information about oral health in older adults was too long and less relevant to their work.

Estimated unit costs of the proposed oral examination and treatment planning for a RACF's resident population, including current training costs, instrument cost and staff time are presented in Table IV. The cost of training is common between the two options while costs related to intervention delivery were different. The average cost of a 16 hour training session was AUD 8.88 per resident (ranged from AUD 8.59 to AUD 9.18 per resident). The intervention delivery cost of the Storage and Forward model was AUD 54.93 per resident with the estimated range of AUD 44.88 - AUD 66.11. The total cost per resident of the real-time oral model was AUD 54.42 (ranged from AUD 44.01 to AUD 63.86).

Staff time costs accounted for 80% of the total intervention delivery cost in both options. The staff time cost of Storage and Forward option is slightly higher than the real-time mode because of a small increase in time a nurse spent on recording the video. The average oral examination time in the Storage and Forward model was 20 minutes, whereas the nurse took 15 minutes for the real-time oral examination.

TABLE IV. COST DESCRIPTION OF TELEDENTISTRY

	Average AUD[b]	*Minimum AUD*	*Maximum AUD*
TRAINING COST			
3-hour introduction at RACF[a]			
a) trainer			
training preparation & delivery	4.99	4.99	4.99
travel cost	0.22	0.22	0.23
b) trainee time cost	0.37	0.31	0.42
3-hour demonstration			
a) trainer			
training preparation & delivery	1.50	1.50	1.50
travel cost	0.22	0.22	0.23
b) trainee time cost	0.37	0.31	0.42
10-hour self-practice			
Trainee time cost	1.22	1.04	1.40
Total	8.88	8.59	9.18
Option 1 Storage and forward			
Intraoral camera and USB dock	0.34	0.31	0.38
STAFF TIME COST			
Nurse time	27.45	21.03	34.70
Dentist time	18.25	14.96	21.85
Total	*45.70*	*35.99*	*56.55*
Grand total	54.93	44.88	66.11
Option 2 Real-time			
Intraoral camera and USB dock	0.26	0.23	0.28
STAFF TIME COST			
Nurse time	26.03	20.23	34.70
Dentist time	18.25	14.96	21.85
Total	44.28	35.19	54.39
Grand total	54.42	44.01	63.86

b. Residential aged care facility

c. Australian dollars (1.00 AUD =0.70 EUR)

V. DISCUSSION

The present study both tested and proved the technical feasibilty and acceptance, by both users and residents, of an alternative model to the traditional face-to-face oral health examination using a teledentistry installation. The observed concordance of remote and face-to-face exams was high and residents expressed acceptable levels of satisfaction with the teledentistry model.

The 'virtual dental examination' can provide general and specialist oral health care support to local aged care facilities. It can assist in providing regular and timely oral health checks using trained non-oral health professional assistants in the first instance.

Additionally, there is anecdotal evidence from RACF staff that the stress imposed by travel to a dental surgery can lead to complete non-compliance with the dental examiner, to the point where attempts at oral examination are abandoned. This leads to further travel and dentist rebooking costs and often reluctance on the part of resident and practitioner to repeat the process. By using the teledentistry approach, the RACF avoids the disruption and difficulty of arranging travel for the patients for dental treatment.

A successful translation of this technology into clinical practice would extend the provision of health care/oral health care to remote and difficult-to-serve locations, and improve access for care to additional patient populations at reasonable cost. The ability to view examination results at their desk will enable oral health professionals to see and screen more residents per time unit in their catchment area. Further development of the procedures is warranted to allow for high-care resident assessment. Specialist dental services can subsequently be provided when the required treatment is identified.

Oral health professionals will also be able to triage and prioritize appointments, rather than travelling to each home without knowing beforehand what treatment each resident will require. Visiting domiciliary oral health professionals will be aware of the exact nature of the oral problem before they arrive. The oral health professional will also be able to plan a visit to treat other residents in the area, improving efficiency and meaning that more residents can be treated over the course of the year. The oral health professional can better identify older adults who require a diagnostic examination by a dental specialist.

Furthermore, by performing an in-RACF examination, the confidence that both the residents and their families have in the RACF will increase. From the health care system and societal perspective, a key impact will be in the satisfaction of knowing that residents have been well looked after, and that scarce resources are being well utilized. Additionally, the case for extension of funding would be bolstered. It will improve oral health for underserviced communities through education, diagnosis, treatment, health promotion and disease prevention.

Data collected from this project could also be useful as a starting point for a large oral health record repository, which would combine a digital record with 2-D and 3-D stills and video images, as well as radiographs.

Additional research should explore and address some technical and training aspects of this study, as a means to further verify this approach. Firstly, when the interaction during the conventional face-to-face exam and remote communications was examined, residents indicated levels of dissatisfaction. The face-to-face seems to provide a more effective mean to achieve clarity and easy to understand communication between the oral health professional and the resident. King and his collaborators [21] reported diminished quality of communication with videoconferencing. However, in the present study it appears that this was due to language and not technical aspects, but this was not explored. Secondly, it was expected that after a while the RNs would be able to operate the intraoral camera and send the files. However, despite having adequate training, written instructions, and receiving material compensation, some RN's still failed to fully engage with the study. This is despite the successful use of RNs in other areas of dentistry [22][23].

The main aim of implementing teledentistry would be the improvement of the clinical outcome as well as achieving the satisfaction of residents and of those in charge of the program. The latter is important, as problems and even failures of the telehealth initiative may result from discrepancies in the expectations of health care stakeholders, as well as any limitations of the health care technology [24]. Possible changes due to introduction of new technology also need to be assessed [24].

In the present study, although RNs understood the capabilities of this technology, they questioned its effectiveness in the context of a RACF. They felt that the approach did not recognize the realities of the RACF. They did not acknowledge its use as an alternative to face-to-face examinations when used on high care residents. The perception that this technology is of limited use in RACFs could explain, at least in part, the unwillingness to take up this new technology [25]. In any case, it seems that barriers to implementation are largely due to human factors. Further training and analyses of how different types of constraints operate to support or undermine the adoption of a teledentistry model need to be explored and addressed. Nonetheless, while these are important concerns, this study was not designed to explore in depth the experience of RNs and patients with teledentistry, but to give some preliminary insights on its technical feasibilty,acceptance and cost. A longitudinal study would allow assessment of users experience over time.

Estimating the costs of teledentistry programs is crucial for efficiently allocating resources to provide oral health care services for underserved communities, such as RACF residents. While the economic analysis has inherent limitations as a result of its reliance on a range of assumptions, the results provide important information for further economic evaluation, and help local program managers to determine the average cost per examination under the conditions prevailing in Victoria, Australia. Findings suggest that, the cost of the teleconsultation / telediagnosis has an overall cost per resident ranging from AUD 44.01 to AUD 63.86, with no significant differences between real-time and the Store and Forward model.

VI. CONCLUSIONS

Findings for this field trial indicate that using a teledentistry installation is an appropriate alternative to traditional oral health consultation, and could provide benefits to an expanding segment of the population in relative and absolute terms. This population comprises older people living in RACFs and older people living in regional, rural, and even outer-metropolitan areas.

An increasing proportion of older people are living in rural and regional Australia and these communities are demographically ageing more rapidly than their metropolitan counterparts. Teleconsultation / telediagnosis projects such as the present one have the potential to target this rapidly expanding aged segment of the population with special oral health needs. There is also potential for wider scale application for the provision of sustainable oral health care in rural areas. Nonetheless, involvement of a wider range of stakeholders will be necessary, as they all influence adoption. A recent review of factors influencing the implementation of telehealth highlighted some challenges at different levels [24]. These challenges need to be specifically targeted.

This study will lead to third Stage of this project, which would involve a multi-State, community-based trial of the technology which could be extended as an integrated part of the general adoption of telemedicine / telediagnosis. The long-term goals of such a project will be to test whether improvements in accessibility and appropriateness of oral health services can be achieved by utilizing ICT techniques to screen for oral disease and delivering oral healthcare services for older people. The project would deliver specifications and toolkits for new products and services, and develop markets for teledentistry in the public and private health sector. However, this will require further research into the acceptability of teledentistry by participants, including the format, content and delivery of the program, as well as the relevance and appropriateness of the information provided. It will also need to explore the sustainability of oral health care services in underserved areas and provide evidence for a first cut business plan for national and state programs in oral health.

People living in rural or underserved areas are amongst the most in need of oral healthcare. Teledentistry may be especially useful in remote rural areas or in other areas where there are few dental practitioners. Teledentistry may enhance the quality of services provided and reduce costs. It can benefit oral health care by enhancing early diagnosis, facilitating timely treatment of oral diseases. It can provide patient health education and health promotion and early detection of disease [25]. Eventually, face-to-face examinations could be replaced by the collection of patient information and data remotely and integrating dentistry into other health care services provided locally.

Despite this, there is currently no active teledentistry service in operation delivering private or public clinical care. The aim would be that all the fragmentary trials occurring across the world will provide the basis for fully funded service provision. The impact of such a model needs to be assessed further. Further research will also be required to undertake economic analysis and modeling to determine the intervention's productivity compared to the traditional model of oral health examination.

Acknowledgments

This project was funded by a grant from Dental Health Services Victoria's Research and Innovation Scheme and the Victorian Department of Business and Innovation BEIP program. We would also like to acknowledge the RACFs, residents and nurses involved in this study.

References

[1] R. Mariño et al., "Field testing of remote teledentistry technology," The Sixth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2014) IARIA, March. 2014, pp. 23-28, ISSN: 2308-4359, ISBN: 978-1-61208-327-8. Available from: http://www.thinkmind.org/index.php?view=article&articleid=etelemed_2014_2_10_40017/ 2014.11.25.

[2] Australian Institute of Health and Welfare. National survey of adult oral health 2004-2006. AIHW cat.no.DEN 165, Canberra, Australian Institute of Health and Welfare, 2007.

[3] M. Hopcraft, M. Morgan, J. Satur and F.A.C. Wright, "Edentulism and dental caries in Victorian residential aged care facilities," Gerodont, vol. 29, pp. E512-519, 2012.

[4] M. Hopcraft, M. Morgan, J. Satur, F.A.C. Wright and I. Darby, "Oral hygiene and periodontal disease in Victorian residential aged care facilities," Gerodont, vol. 29, pp. 220-228, 2012.

[5] M. Hopcraft, M. Morgan, J. Satur and F.A.C. Wright, "Dental service provision in Victorian residential aged care facilities," Aust Dent J vol. 53, pp. 239-245, 2008.

[6] R. Mariño and A. Ghanim, "Teledentistry: a systematic review of the literature," J Telemed Telecare, vol. 19, pp. 189-194, 2013.

[7] Australian Bureau of Statistics. *2006 Census QuickStats: Victoria*. [Online]. Available from: http://www.multiculturalaustralia.edu.au/doc/2006census-quickstats-victoria.pdf / 2014.07.31.

[8] Australian Health Ministers' Advisory Council, Healthy mouths healthy lives. Australia's national oral health plan 2004-2013. South Australian Department of Health, 2004.

[9] P.A. Scuffham and M. Steed, "An economic evaluation of the Highlands and Islands teledentistry project," J Telemed Telecare, vol. 8, pp. 165-177, 2002.

[10] R. Mariño et al., "A teledentistry approach for oral health screening in nursing homes," [Online]. Available from: http://www.slideshare.net/ifa2012/3-mariooral-health-if-apresentation29may2012/ 2014.07.31.

[11] L. Thabane et al., "A tutorial on pilot studies: the what, why and how," BMC Med Res Methodol, vol. 10, pp. 1, Jan. 2010, doi: 10.1186/1471-2288-10-1.

[12] SOPROLIFE: *Light induced fluorescence evaluator*. [Online]. Available from: http://www.soprolife.com/ 2014.07.31

[13] Institute for Broadband Enabled Society (IBES). Available from: http://broadband.unimelb.edu.au/ 2014.07.31

[14] M.F. Drummond, M.J. Sculpher, G.W. Torrance, B.J. O'Brien and G.L. Stoddart, Methods for the economic evaluation of health care programmes, 3rd edn. Oxford: Oxford University Press; 2005.

[15] Australian Taxation Office. *Deduction for business*. [Online]. Available from: https://www.ato.gov.au/Business/Deductions-for-business/Motor-vehicle-expenses/ 2014.07.31.

[16] Dental Health Services Victoria, "Human resources at Dental Health Services Victoria," unpublished.

[17] Department of Health and Aging, *Caring for older Australians*. [Online]. Available from: http://www.pc.gov.au/__data/assets/pdf_file/0006/109293/residential-care-trends.pdf/ 2014.07.31.

[18] Pegasus Dental Services, *Celebrating 10 years servicing the dental industry*. [Online]. Available from: http://www.pegasusdentalservices.com.au/images/april14specials.pdf/ 2014.07.31.

[19] ACTEON North America, *SOPRO Intraoral cameras*. [Online]. Available from: http://us.a-dec.com/en/~/media/Adec/Document%20Library/Product%20Information/Sales%20Information/Brochures/85551300.pdf. /2014.08.30.

[20] World Health Organization, Calibration of examiners for oral health epidemiological survey, Publication ORH/EIS/EPID.93.1, Geneve, WHO, 1993.

[21] G. King, H. Richards and D. Godden, "Adoption of telemedicine in Scottish remote and rural general practices: a

qualitative study," J Telemed Telecare, vol. 13, pp. 382-386, 2007.

[22] J.M. Chalmers and A. Pearson, "A systematic review of oral health assessment by nurses and carers for residents with dementia in residential care facilities," Special Care in Dentistry, vol. 25, pp. 227-233, 2005.

[23] L.J. Walsh, G. Hill, G. Seymour, and A. Roberts, "A scoring system for the quantitative evaluation of oral mucositis during bone marrow transplantation," vol. 10:190-195, 1990.

[24] U. Sharma and M. Clarke, "Nurses' and community support workers' experience of telehealth: a longitudinal case study," BMC Health Services Research, vol. 14, pp. 164, 2014, doi:10.1186/1472-6963-14-164.

[25] R.N. Patel and R.N. Antonarakis, "Factors influencing the adoption and implementation of teledentistry in the UK, with a focus on orthodontics," Community Dent Oral Epidemiol, vol. 41, pp. 424-431, 2013, doi: 10.1111/cdoe.12029.

[26] R. Mariño, P.Marwaha, S. Barrow and H. Baghaie, "Web-based oral health promotion program for older adults," 10th World Congress on Preventative Dentistry, October 2013. Available from: https://iadr.confex.com/iadr/wcpd13/webprogram/Paper182509.html/ 2014.07.18.

Personalized Virtual Coaching for Lifestyle Support: Principles for Design and Evaluation

Olga Kulyk*, Rieks op den Akker†, Randy Klaassen†, Lisette van Gemert-Pijnen*
*Center for eHealth Research and Disease Management, University of Twente, Enschede, The Netherlands
†Human Media Interaction, University of Twente, Enschede, The Netherlands
{o.a.kulyk, h.j.a.opdenakker, r.klaassen, l.vangemert-pijnen}@utwente.nl

Abstract—**There is a fast growing number of behavior change support systems (BCSS) aiming at supporting a healthy lifestyle. Existing lifestyle coaching services offer individual users access to web portals where they can communicate about a growing number of ingredients of everyday life concern: physical activity, nutrition, medication, mood, sleep and sexual health. Mobile technology in combination with body worn sensors support user's awareness of their physical condition and lifestyle. Despite the large number of available lifestyle interventions and pilot trials, only very few are successfully transferred into the real health care practice. Low usability and lack of transparency on the reliability and trustworthiness of the information are just a few examples of the major barriers for successful implementation.Traditional metric for measuring effects of behavior change support system are not suited for early stage health technologies. Professionals from the field of health and social psychology, and potential end users should be involved not only in the design and effects evaluation of BCSS, but also in the iterative process evaluation of these systems. Qualitative evaluation studies focused on the user experiences with technology can help researchers to understand what persuasive features can enhance adherence, motivate people and how this technology should be further developed to optimally match the needs of real users in daily healthcare practice. Based on two user studies, we present general guidelines for design and evaluation of lifestyle support systems with personalized virtual coaching. The first field study focuses on design and evaluation of a mobile physical activity coaching system. The second user study focuses on design factors that influence the attitudes of high-risk adolescents towards virtual coaching in mobile eHealth applications and social media. We present a new approach that integrates an animated digital coach in an activity monitoring lifestyle change support system. The main contribution of this paper are practical recommendations for persuasive design and evaluation methodology combining established methods from Human Computer Interaction and eHealth.**

Keywords-Mobile Activity Monitoring, Personalized eHealth; Persuasive Feedback, Usability; Virtual Coaching; Behavior Change; Lifestyle Support; Human Computer Interaction

I. Introduction

Recent massive media attention to the obesity epidemic worldwide and a growing number of patients with chronic diseases raise the demand for encouraging physical activity and for raising health awareness [1][2]. Next to classical web-based interventions, eHealth behavior change support systems for healthy lifestyle promotion aim to motivate patients to healthy behavior change [3][4][5]. Some systems become proactive and provide real time information and feedback to their users based on data gathered through various sensors and personal devices [6][7].

Despite the large number of existing lifestyle interventions and pilot trials, only very few are successfully transferred into the real healthcare practice [8][9]. Users often have problems to navigate through the system, they get lost or they do not find the information they are looking for [10]. Low usability and lack of transparency on the reliabilty and trustworthiness of the information are some of the major barriers for successful implementation [5][8][9]. There is also a lack of standardization for interoperability between various parts of the systems and a lack of connection between the feedback, the actual usage patterns and the task a user is involved in [11]. These problems are often caused by a design that does not meet the actual needs of the target users while using the system and a lack of connection with offline, daily, activities. A holistic design approach for eHealth intervention development, which we use in our research has proven to contribute in overcoming these barriers [9].

This paper presents new insights and general guidelines for design and evaluation of lifestyle change support systems that uses personalized virtual coaching [1]. The two user studies discussed in this paper represent different perspectives on lifestyle support systems, combining methods from Human Computer Interaction and eHealth. The first field study focuses on design and evaluation of mobile physical activity coaching for diabetes patients and office workers. This study takes a Human Computer Interaction (HCI) perspective and focuses on the effects of using a virtual animated character at the user interface. It was performed at the Human Media Interaction group of the Computer Science Department at the University of Twente in the context of the EU Artemis project Smarcos.

The second user study focuses on design factors that influence the attitudes of adolescents towards virtual coaching in mobile eHealth applications and social media. This study was performed at the Center for eHealth Research and Disease Management, Psychology Health and Technology group of the Behavioral Sciences Department at the same university. The data analysis for both studies is integrated using the persuasive system design model [12]. We present persuasive factors and general design guidelines for personalized virtual coaching in mobile health applications.

In the next section, we first highlight findings from related work on physical activity monitoring, virtual coaching, mobile eHealth applications and serious gaming for lifestyle support. After that, we present a new approach for a multi-device coaching system and general design guidelines based on the outcomes of two user studies on virtual coaching for lifestyle support. The first study describes results of the user evaluation of the mobile physical activity coaching system for office workers and diabetes patients. The second study focuses on persuasive factors and attitudes of high-risk adolescents towards virtual coaching, social media and mobile apps for sexual health promotion. Summarizing the main outcomes, we then present recommendations for the design and evaluation of lifestyle support systems with personalized virtual coaching. Finally, we present the main conclusions and discuss future work.

II. State of the Art

There have been various attempts in categorizing eHealth technology [6][9][13][14][15]. In this paper, we focus on monitoring physical activity and health related parameters (blood pressure, weight) in lifestyle interventions for preventive professional care support. A categorization is based on the type of platform that the eHealth technology is realized on: stand alone devices; integrated web-based interventions and personal mobile devices; or a combination of various devices to monitor online and offline activities of a target user.

A. Physical Activity Monitoring and Coaching

Wearable health technology, such as activity sensors, is often used as a surveillance tool to objectively assess physical activity patterns [7][16][17]. They provide an inexpensive measure of physical activity by counting the number of steps taken per day, enabling the accumulative measurement of occupational, leisure time, and household activity, along with activity required for everyday transportation. Real time monitoring offers the user immediate feedback on the accumulated step count which is important for motivating him to sustain or improve his activity level[17]. For a thorough review on this topic see [18].

Engaging patients requires user friendly interfaces and user friendly interaction with the systems. Patients often have to cope with various physiological measurements instruments (either active or passive): blood pressure, blood sugar and weight. Willingness to measure these parameters strongly depends on the complexity of the user interface of the measuring device or sensor, as well as the data transfer process [19].

Based on an extensive literature study, H. op den Akker et al. [20] identified six key areas for research to improve digital coaching for physical activity by tailoring to the individual user. Two of them are of interest here: advanced Human-Computer Interaction (HCI) and pervasive coaching. To increase perceived intelligence of a coaching system, a virtual coach offers an interesting opportunity as an interface metaphor. Bickmore et al. [3] studied the effects of interventions for multiple health behaviors using conversational agents as a coaching system. This study showed that virtual conversational agent as a coach can have a positive effect on perceived relationship of a patient with an eHealth system.

Computer tailoring and personalized eHealth offer great potential for motivating people by providing personal information and feedback [21]. Characteristics of an intervention, such as enabling personal goal setting and providing tailored feedback are thought to be among the important factors related to the use of lifestyle change support systems. Next to tailoring, personal feedback needs to be dynamic to provide new information and real time feedback on the daily activities. The user study of Consolvo et al. [16] reports that negative feedback or paternalism has a negative impact on the users.

B. Mobile eHealth and Coaching

Mobile devices providing personalized feedback to influence physical activity behavior are gaining more and more popularity [22]. There are few examples of mobile health applications (apps) specific for behavior change and physical activity support [6][22][23]. Despite a huge range of health-related apps on the market, there is little in depth research on user experiences and views on a wide range of features that apps can provide.

Fanning et al. [22] present extensive review on efficacy of mobile devices in the physical activity and recommendations for implementation. This study concludes that mobile technology applied in behavior change interventions is an effective tool for increasing physical activity. User studies in mobile health research are rarely performed with young adults, though adolescents are forerunners of mobile technology. Dennison et al. [23] present the findings of a focus group study with students on the use of mobile apps to support a healthy lifestyle, the attitude of adolescents on the usefulness of various features of such apps. The results suggest that the most important factors influencing the use and uptake of mobile apps are accuracy, legitimacy, security, effort required, and immediate effects on mood. Another features that young adults valued were ability to record and track own behavior and goals, as well as the ability to receive advice and real time information. Interesting finding from this study is that context-sensing capabilities of mobile apps and social media features were perceived as unnecessary.

Consolvo et al. [16] reports a long-term user evaluation with the UbiFit system, which aims at raising individual awareness on physical activity level. The results show that glanceable representations of information on personal, mobile displays can stimulate the person to do more exercises. These findings are consistent with another study [24].

C. Serious Gaming for Lifestyle Support

New forms of entertainment media such as serious gaming are used for promoting healthy lifestyle [25]. The general purpose of serious games can be defined as: *'games to train, educate, and persuade'* [26](pp. 14). The latest research shows that the use of game mechanics for supporting non-leisure activities has grown beyond serious games. This trend has being defined by the term gamification: *'the use of game design elements in non-game contexts'* [27](pp. 2). Serious gaming and interactive gaming elements embedded in eHealth technology offer great potential in innovative opportunities for engaging adolescents and other patients in interventions promoting healthy nutrition habits and physical activity changes that can contribute to obesity prevention and healthier lifestyle [25].

Existing applications that incorporate some gaming elements and personal visual feedback demonstrate the lack of a multidisciplinary approach in designing technology applications for care [28]. The main drawback of the current gamification applications aiming at behaviour change is the lack of actual benefits for the end user and the lack of focus on positive user experience. In general, there is a lack of research and evidence on the long-term user evaluations and implementation of such health-oriented gaming technologies [29].

D. Evaluation of Lifestyle Support Systems

The goal of every (health) behavior change support system is to change a certain behavior or a habit. User evaluations of a behavior change support system are usual focused to show that the intended change in behavior did actually took place. Measuring change in behavior requires multi-year studies with repeated follow-ups [30]. Changes in behavior on short-term are hard to measure and if they occur these changes are often short-lived. According to Klasnja [30] evaluating behavior change in the traditional clinical sense is not the right metric for early stage technology that are developed in the context of human computer interaction. Evaluating behavior change support systems in the context of the field of human computer interaction can focus on narrower notion of efficacy by looking at the outcomes of particular interventions strategies and whether the system is doing what it intended to do, even in a early stage of the development of the system. Qualitative studies focused on the experiences with technology can help researchers understand who and why a system is working and how the system should be further developed.

Technologies that promote a healthier lifestyle are gaining popularity in the human computer interaction discipline. Paradoxically, the evaluation of such technologies remains rather unclear [31]. Different frameworks, like the Framework for Ubiquitous Computing Evaluation Areas (UEAs) by Scholz and Consolvo [32], the persuasive system design model (PSD) by Oinas-Kukkonen and Harjumaa [12] and the Fogg Behavior Model (FBM) [33] could be used to evaluate behavior change systems and the features and functionalities of these systems. Evaluating the the perceived persuasiveness of a behavior change support system is an ongoing challenge [31]. Klasnja et al. [30] propose multi-methods approaches for evaluating behavior change support systems where combining qualitative and quantitative methodologies should be combined to provide deeper insights into users experiences with technologies.

III. User evaluations of a mobile animated activity coaching system

As humans interact with many different devices during the day, cross media systems offer the opportunity for the activity coach to travel with the user across those devices. Depending on the needs and context of the user, coaching can thus be provided on the most suitable device (e.g., smartphone, PC, smart television) [34].

A. Digitial Coaching Architectures

A multi-device digital coaching can have a more *centralized* or a more *decentralized* architecture. The main difference is in the measure of autonomy of the mobile coaching application. The Continuous Care & Coaching Platform (C3PO), developed at Roessingh Research and Development (RRD) in the Netherlands, enables continuous remote monitoring of elderly patients and patients with chronic disorders [35]. In the C3PO platform, there is only one device with which the patients users interact, the smartphone. The care givers can view patient data that is uploaded to a server. An activity monitoring and feedback system was designed to guide patients in reaching a healthy daily activity pattern. Objective daily activity is assessed using an inertial sensor node that captures and communicates wireless. The sensor can store large amount of data and send the data over Bluetooth to a PDA (an Android based HTC Desire) where further processing and communication to the patient is handled. The users receive feedback on their smartphone at scheduled times or if their activity level calls for this. In this *decentralized* architecture the coaching rules reside on the client-side mobile. In contrast to this, in the *centralized* architecture of the Smarcos platform the coaching rules reside on the server (see Fig. 1 for an overview). Based on server side stored sensor data or on fixed times, the server sends a message to the client, who can receive the message on the device of his own choice. To upload activity data the user has to connect his sensor to the internet.

In the Smarcos system, feedback is a reminder to connect the activity monitor to upload data, a motivating message when activity is less than the target or an overview of daily, weekly and monthly scores. A drawback of the Smarcos system compared to the RRD system is that feedback is not real-time. The RRD system allows immediate feedback based on recent data collected from the sensor. A call for urgent medication intake (e.g., in a diabetes I medication coach) requires real-time feedback. In the Smarcos system, scores are presented as percentages of a user set target score and in terms of kCal. The user can get these activity overviews on his mobile device app, as well as via a web portal.

B. User Evaluations: Animated Virtual Coach

User evaluations were performed throughout the development of the Smarcos coaching system [34] for physical activity support and diabetes II patients. We performed short user evaluation with diabetes patients and office workers. We focused on the graphical user interface, on the interaction and on personal feedback in particular. We looked at timing, content, modality and presentation format. We are particularly interested in the application of animated virtual humans and multi-modal natural dialog as a means for interaction between the user and the digital coach. At the Human Media Interaction group we developed mobile technology for responsive animated Embodied Conversational Agents (ECAs) [37] for the presentation of feedback on a mobile app. The system can produce ECA behaviors (eye blink, eye gaze, head movements, lip sync with natural speech and facial expressions) specified in the Behavior Markup Language (BML) [38]. In the centralized Smarcos system the server sends a feedback message to a client containing a BML specification. The BML contains the

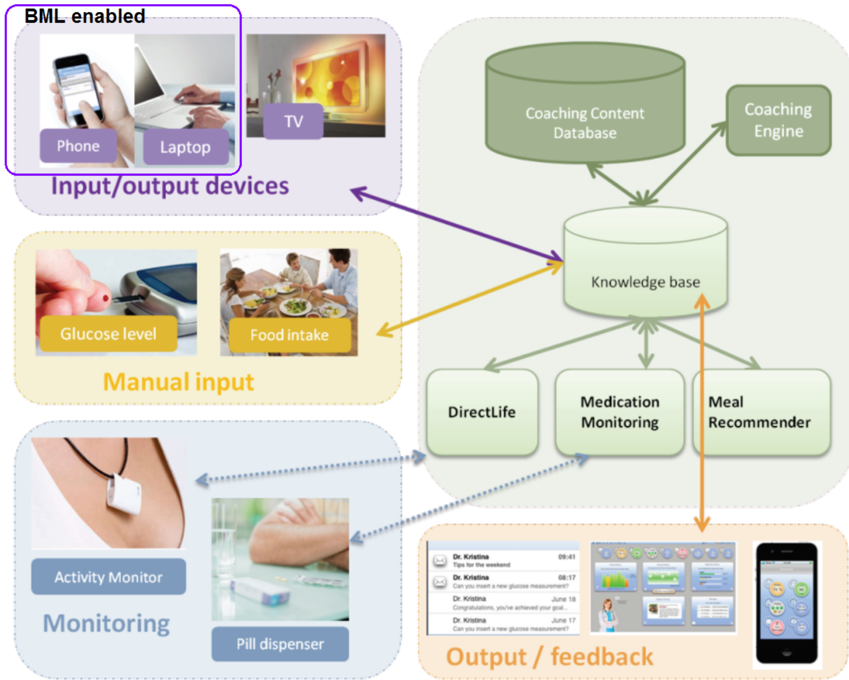


Figure 1: Overview of the multi-sensor multi-device digital coaching system developed in the Smarcos project [36].

text to be pronounced by the ECA as well as the non-verbal embodiments.

The mobile animated virtual coach was used in several short user experiments with the Smarcos architecture, as well as with the decentralized coaching platform developed at Roessing Research and Development [34].

In the second, a 'long-term' user evaluation Smarcos system was continuously evaluation by office workers during the six weeks period.

1) Subjects: Long-term Evaluation of Smarcos Coaching System: Participants were asked to join the user evaluation by email, social media and face-to-face communication. Participants had to be office workers (sedentary profession) and had to own an Android smartphone with an operating system Android version 2.3 or higher. Sixty office workers indicated they are willing to join the experiment. Participants were randomly assigned to one of the three groups: one group (N=19) received feedback as text (text group: TXT), a second group (N=15) received feedback by a virtual human coach represented as an embodied conversational agent (embodied conversational agent group: ECA). A third control group (N=9) did not receive feedback message on the smart phone. Participants of the control group could only get feedback via the web portal. The distribution over the three groups was random, but participants needed to meet some requirements to be selected for the ECA group. The inclusion criteria for the embodied conversational agent group was the ability to buy a Dutch text-to-speech engine from the Google Play store.

2) Procedure: Long-term Evaluation of Smarcos Coaching System: The duration of the complete evaluation was seven weeks. This included one assessment week at the start of the user evaluation. Before the start of the evaluation each participant was visited at home or at the office. The participants were required to own a smartphone with Android 2.3 or higher. Each participant received the activity sensor and an information sheet with details about the user evaluation. The goal of the study and the procedure was discussed. After reading all the information material and making sure no questions were left, participant had to sign two copies of the consent form (one for the participant and one for the researcher). Software, applications for the smartphone, user manuals about the installation of software and links to questionnaire were provided via email.

It was explained that the first week was an assessment week in order to establish their normal activity level to tailor the system and generate a personal goal. During the assessment week no feedback messages were given by the system. After the assessment week participants received their personal goals and used the system for six weeks. In these six weeks feedback (updates, requests, reminders and overviews) about their progress was provided. Participants were asked to upload their activity data at least one time per day by connecting their activity monitor to their computer. Halfway the evaluation the system offers users how over or under performed in the first three weeks a new (higher or lower) goal. Participants were free to accept this new goal.

Participants were asked to fill in three questionnaires during the evaluation. One in the first week (assessment week), one after the third week and one after the sixth week. Questionnaires were available online and participants received an email when it was time to fill in a questionnaire including a link to access the questionnaire. Shortly after the evaluation the participant was visited again to collect the materials. During this visit the participant was invited for a post-interview to discuss their experiences. This interview was voluntarily.

3) Results: Long-term Evaluation of Smarcos Coaching System: Forty three participants completed the user evaluation by finishing the assessment week, using the system for at least six weeks and completed all the questionnaires. Participants were between 21 and 57 years old, and worked 36,1 hours per week on average. All participants except two owned an Android smartphone. Those two participants were included in control group. All the participants were familiar with mobile internet, 26 subjects indicated to have very much experience with mobile internet, eleven participants indicated to have much experience and four participants indicated to have less experience with mobile internet. Most of the participants do sports, 35 indicated to do sports every week for more than one hour, five indicated to do sports every week for less than one hour and four participants indicated not to do sport. All participants, except one, were not familiar with behavior change support systems at all. One participant (from the control group) indicated to have a lot of experience with behavior change support systems.

All the participants of the user evaluation were invited for a semi-structured post-interview to discuss the system and their experiences during the user evaluation. Twenty-one participants accepted the invitation (text group: ten participants, virtual human group: ten participants, control group: one participant). During the interviews we asked the participants about their general impression (good and bad experiences), problems they have experienced during the evaluation, if they think that such a coaching system could help people to be more physical active, if they should recommend the system to others and the two different ways of presenting feedback were discussed. At the end there was some time left to discuss other remarks from the participants. During the post-interview, participants were asked about their general impression of the coaching system and their own experiences while using the Smarcos system. Other questions included: awareness on the amount of physical activity; content and visual representation of the feedback (text based / virtual human based); usefulness of the system regarding motivation to live a healthier life; recommendation regarding further use of the system and general remarks.

Interviews were transcribed and additional data analysis was carried out in order to identify the persuasive features according to the persuasive system design model [39]. It was important to identify which persuasive features might influence the response of participants regarding their experience with using the personal coaching system. Several crucial persuasive features were identified and categorized into (1) *stimulating*: features that are perceived as having positive influence on motivating the user, perceived usefulness and adherence, (2) *blocking*: features that are perceived as having a negative influence on motivating the user, perceived usefulness and adherence, and (3) *neutral* features that are perceived as having no effect on motivating the user, perceived usefulness and adherence. Table I presents the stimulating features and quotes from participants. Table II gives an overview of the stimulating persuasive features that should be improved.

Next to the features presented in both tables, several other important design factors have been identified based on the extra interview data analysis. First, not only personal feedback is important to motivate users, but also the *timing* of the *feedback* messages on personal goals (achieved yet or not, encouragement to continue): *"No, the timing of the feedback is wrong. You should be able to have more control over the system and your data."*; *"Feedback (timing and content) is predictable. Feedback should be more targeted and timing is important."*Second, the *'fun'* factor was mentioned several times as an important feature influencing the motivation of the users and interactivity with the personal coach: *"For myself it was fun to use and I would buy the system. It is good to motivate people to be more physical active."*

TABLE I: RESULTS LONG-TERM EVALUATION PERSONAL COACHING: *STIMULATING* PERSUASIVE FEATURES

Persuasive feature	Quote
Self-monitoring	*"It is nice to see your daily score of physical activity and week overviews." (text group)*
Social role	*"The system (virtual human coach) is like an external motivation to be more physically active."*
Praise	*"It was nice to receive feedback about your own behavior, also on the sensor itself." (text group)*
Liking	*"I think a virtual human will be more fun. Text is static and open for multiple interpretations. A virtual human can help to present the right interpretation by showing some empathy." (virtual human group)*
Similarity	*"The virtual human will add some social pressure to the feedback." (virtual human group)*

TABLE II: RESULTS LONG-TERM EVALUATION PERSONAL COACHING: *STIMULATING* PERSUASIVE FEATURES TO IMPROVE ON

Persuasive feature	Quote
Tailoring	*"Feedback (timing and content) is predictable. Feedback should be more targeted and timing is important. (virtual human group)"*
Personalization	*"I want to have more control over my goals. It is my goal and not the goal of the system. Goals can change day by day.. (text group)"*

Results of a six weeks user evaluation with the physical activity coaching system show that Physical Activity Level (PAL) values do not differ between the ECA and TXT condition. Thus feedback by means of an ECA has no added value over feedback by means of a text message if we look at the target objective. The control group that did not receive feedback and no reminders to upload their sensor data performed worse compared to both ECA and TXT group. In particular, in the control group the mean of the PAL values dropped from week 4 onwards. We used a Mann-Whitney U test to compare the results between the different groups. At the end of week 6 the PAL level of the ECA group wa significantly less than the mean for the TXT group $((Mdn = 1,61), U = 36,00, p = 0,014, r = -0,460)$. The number of uploads of PAL data in the control group was significantly less than in the TXT group during all six weeks $((Mdn = 3,00), U = 51,00, p = 0,045, r = 0,378)$.

Interviews and questionnaires reveal that on smartphones users prefer glanceable presentation of feedback messages.

Reading the short text message is faster than listening to the spoken message. Smart phones are for quick access and glanceable presentation of feedback fits the message and the use context. In line with the conclusions of Lisetti et al. [40], we believe that the opportunities offered by the technology of animated conversational characters are exploited fully in multi-modal spoken personalised emphatic dialogs with the user. The user evaluations reported by Lisetti et al. had a similar objective as but the research methods differ in a number of ways. The most important is that Lisetti applied a lab user test. whereas our study included a real-life evaluation. A second important difference is that our ECA platform is based on the BML framework and runs on mobile platforms. This offers new opportunities for coaching systems in clinical applications.

A cross media or multi-device coaching system can support the execution of an intervention program in which a team of human and virtual coaches work towards a negotiated goal, or to sustain a certain lifestyle. Such a blended format combining virtual coaching with real-world coaching is a novelty in HCI design.

IV. User Evaluations: Persuasive Design Factors for eHealth and Social Media Apps

Sexual health is a specific sensitive subject in many cultures and there is little research on the effects of prevention-focused interventions in this domain. The exploratory user study aimed at identifying the design features interventions have to possess to facilitate qualitatively well-designed and tailored eHealth interventions in the future and to evaluate currently available ones. We investigated which design factors are important using focus group discussions with high-risk adolescents. The user study focused on social media, serious games, mobile applications and the use of personal virtual coaching for lifestyle support (see Fig. 2). Primary research question was which persuasive design factors influence the use and adoption of various eHealth interventions in public sexual health services.

Participants explored and gave feedback on a number of existing and new social media applications and modern media applications in a focus group setting. All sessions were audio recorded with participants permission.

Participants were also asked to express their opinion about the three new concepts for new media applications integrated in modern social media, namely, (i) a serious game embedded in the existing social network, e.g., Facebook, (ii) a mobile application functionality embedded in Facebook, and (iii) a personal virtual coach embedded in either a social network, a website or a mobile application. Each concept represented certain persuasive features, which were discussed using clear visual examples without explicitly naming which persuasive feature was represented.

In the last part, adolescents shared their own ideas and tips about promoting public sexual health services and healthy lifestyle via eHealth applications and social media. The script, the power point slides and the duration of various parts of the focus group session were first tested during a pilot session and adjusted based on the outcome. An assistant took notes and answered questions about the group assignments.

During the data analysis, audio files were fully transcribed, analyzed, coded and categorized. The persuasive features have been coded according to the persuasive system design model [39]. An analysis of the influence of persuasive features on the response of adolescents towards various types of presented media was done.

A. *Result of the Focus Group Study*

In total, thirty seven young adults with low socio-economical and various ethnical backgrounds (51, 4% male and 48,6% female) participated in four focus group discussions. Participants are considered as having ethnic origin in case if at least one parent is born in a different country. Participants were between 12 and 24 years old (M=17,4 SD=3,1), and owned a mobile phone. Thirty four out of thirty seven participants had internet access at home, and thirty two participants had mobile internet access. From the thirty seven adolescents who participated in a focus group study, 89,2% participants were low-educated. All participants indicated they use social media daily. Most popular social media among adolescents were Facebook (27), Twitter (15), Instagram (20). WhatsApp and Youtube are also popular.

The results of the focus group study showed that adolescents have positive attitude towards the use of a personal virtual coach for health promotion, as long as they perceive there is a real human behind the virtual character. Several important persuasive features were identified during the data analysis. The identified features were categorized into (1) *stimulating*: features that are perceived as having positive influence on motivating the user, perceived usefulness and adherence, (2) *blocking*: features that are perceived as having a negative influence on motivating the user, perceived usefulness and adherence, and (3) *neutral* features that are perceived as having no effect on motivating the user, perceived usefulness and adherence. Table III presents the stimulating features and quotes from participants. Table IV presents the blocking feature and quotes from the focus group study.

TABLE III: RESULTS FOCUS GROUP STUDY: *STIMULATING* PERSUASIVE FEATURES

Persuasive feature	Quote
Trustworthiness	*“I would use the coach if the answers are reliable.”*
Expertise	*“I would like it (virtual coach) to be made by someone who has the experience.”*
Liking	*“Mostly if it looks nice, you are more likely to play with it (serious game).”*
Tunneling	*“If you just look on the Internet, for instance stuff about sex, lots of things come up. With the app it is much easier.”*

TABLE IV: RESULTS FOCUS GROUP STUDY: *BLOCKING* PERSUASIVE FEATURES

Persuasive feature	Quote
Recognition	*“I really don’t want anyone to know that I have been using this app (personal coach on facebook).”*

Based on the identified persuasive features, the following important design factors were formulated. *Anonymity* was found as the most important factor, which has important

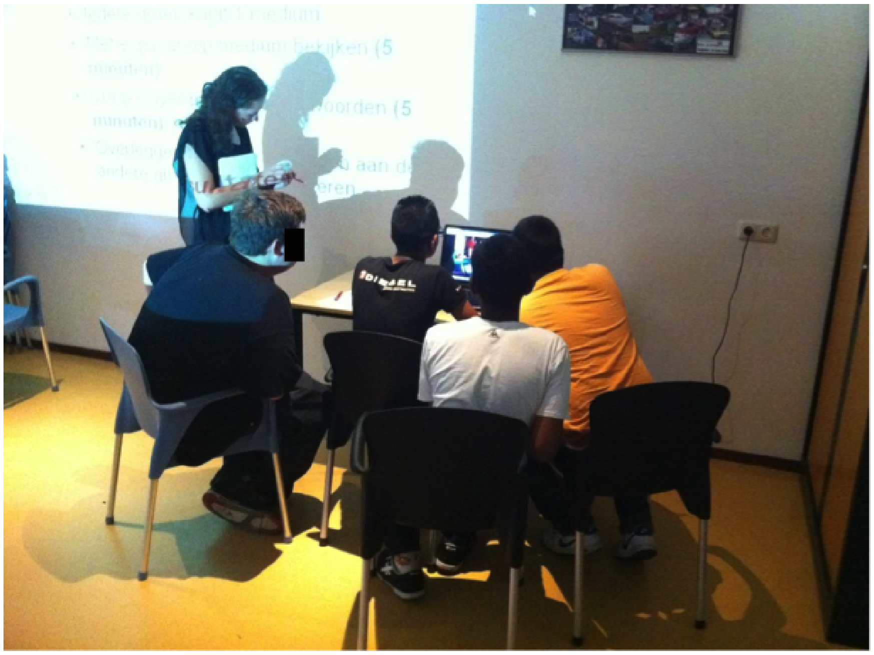


Figure 2: Focus group session: subgroup task where participants explore a serious game [1]. Picture was taken with participants' permission.

implications for the use of social networks for sexual well-being enhancement. Social media networks lack privacy and therefore eHealth applications, for example on Facebook, are not a recommended media for enhancing sexual well-being of high-risk adolescents. Instead, social networks can be used to increase the familiarity of the target group with the existing interventions. The next factor, level of *interactivity* was identified as indispensable. Serious games and mobile applications are expected to have a high level of interactivity to better engage users and thus increase uptake of lifestyle interventions. The type of platform the eHealth technology is realized on was another essential factor. Personal mobile devices, and smart phones in particular, were most preferred due to the high level of privacy and familiar user experience. In addition to the factors mentioned above, the *reliability* of the information *source* was clearly an important issue across all media types. Participants stated the importance of the clear visibility of the information source, as well as the logo or the name of the health organization behind the intervention. Another factor, namely, *support* for *visual aids*, was also identified across all types of applications. Specific to adolescents with low socieconomical background, lifestyle change support systems have to be more visually aided. The language use in the content has to be simple, low threshold and preferably in several languages to reach various ethnic groups. Applying these factors in the design of eHealth technologies for lifestyle support should increase their uptake and usefulness for enhancing well-being of high-risk adolescents and contribute to healthier lifestyle.

V. Lessons Learned: Design Guidelines for Personal Coaching and Lifestyle Support

What have we learned from studies about the effects of a virtual coach in lifestyle coaching systems? In general, user studies showed positive attitude towards the use of a virtual coach for lifestyle change support. Several stimulating persuasive features were identified in both studies. One persuasive feature *liking* overlaps. Visual appeal is a crucial design factor as it determines whether a user is will get motivated to start with the lifestyle behavior change program. Various visual aids and user-centered design methods can ensure that visual representations of the interface are appealing to the target user group. Personal tailoring plays an important role here, as one of the participants stated even variation in how the virtual coach looks like is desirable: *"I think it is more human like to receive a feedback message from a virtual human, but these messages need to be less predictable. Messages should be a kind of surprise (change the appearance of the virtual human every now and then)"*. This feature can be easily enabled by allowing user to choose from various avatars with different appearances and even different voices for the virtual coach.

Motivation is another important factor when it comes to the willingness to use a particular lifestyle coaching system. It makes a difference if an eHealth intervention is supported by a real human healthcare professional. Effective coaching and *tailored feedback* in terms of its *timing*, content and interaction design are crucial elements in affecting behavioral change. The next generation of the lifestyle coaching systems will be able to predict the optimal timing for providing feedback by analyzing previously given feedback messages. A *personal target goal* has to be challenging and reachable, step by step, within a set period of time.

Users prefer to be *in control* of how, when (right timing) and on what device they want to receive personal feedback from a virtual coach. Feedback messages have to offer enough *variation* so that the users stay motivated. In addition, we learned that users want to monitor their history and progress:

what they have done and what they should do next. They expect a connection between the personal goal-setting features for behavioral change and the personal coach to provide visual feedback.

Next lesson is the need to provide personalized (tailored) feedback: show *progress towards target*, adjust target, motivate, *suggest actions*, provide real time information. In addition, fun element has to be integrated in user interface to make interaction with the coach more playful. Users get bored by same messages and standard feedback. Variety in output modalities can facilitate this by providing user the choice to set the feedback presentation mode him/herself: test-based / virtual human and speech / light, etc. Facilitating navigation through information that user needs is also found important, for example by offering a user to search information by *alternative interaction modalities* such as speech input. However, the coach should not talk to the user when the head phone is unplugged. Therefore, a context-sensitive smart sensors technology is needed to enable this feature.

Online coaching also needs to be better integrated with offline feedback to stimulate the participation and commitment of the user to a lifestyle change support program [41]. Combining real-time usage behavior data with personalized virtual coaching and timely *persuasive feedback* can contribute to higher engagement and better uptake of lifestyle change support system by patients as well as healthcare professionals.

A. Challenges in Evaluating Lifestyle Change Support Systems

Evaluating lifestyle change support systems remains an ongoing challenge [31]. One of the main future challenges in evaluation of eHealth technology in general [11] is developing a mixed methods approach and standards for evaluating the effects of eHealth from a user perspective [6][9]. Validated instruments for user evaluations are needed to measure the effects of personalized eHealth interventions, such as changes in lifestyle or other behavior change [30]. The field of eHealth technology and telemedicine can benefit from adopting design and evaluation methods from the field of HCI. Professionals from the field of health and social psychology, and potential end users should be involved not only in the design and effects evaluation of eHealth applications for lifestyle support, but also in the iterative process evaluation of these applications. Health professionals and social scientists could augment about the effectiveness of the content of a persuasive message, and end users could provide feedback for constant improvements of the applications. Researchers and experts from the field of human computer interaction could play a vital role in better designing and evaluating persuasive features of the technology [31].

HCI evaluation methods are well suited for the short-term user evaluation to measure the effects of intervention before the long-term implementation. Naturally, HCI evaluation methods need to be adjusted to the specific goal of the eHealth technology and multi-device lifestyle support systems in particular [32]. In return, HCI field can benefit from the active logging methods and eHealth techniques for analyzing usage behavior patterns for better tailoring of personalized feedback [10][11][42]. In addition, psychological and motivational factors need to be incorporated into HCI evaluation techniques and measurements in order to identify how these factors influence the uptake and use of technology for lifestyle support.

VI. Conclusion and Future Work

Applying user-centered design techniques can significantly improve the appeal of the user interface and thus the engagement with personal lifestyle coaching system [32]. Next to the ease-of-use, visual appeal and clear presentation, the user experience has to be enjoyable and rewarding. Active engagement of the user in interaction with lifestyle intervention is crucial to ensure prolonged use of an intervention.

There is a need for guidelines and standardization for evaluation of eHealth technology in general [11], as well as for the lifestyle support systems, in particular. In addition, the multi-device architecture is necessary to enable easy exchange of monitoring data between web-based and private mobile parts of lifestyle support systems. Qualitative evaluation studies focusing on the process evaluation and user experiences with technology can help researchers to understand what persuasive features can enhance adherence, motivate people and how this technology should be further developed to optimally match the final design with the needs of real users in daily healthcare practice. This paper combines methods from HCI and eHealth and makes several contributions: (1) new insights into the existing lifestyle change support systems with personal virtual coaching, their limitations and recommendations for improvements (lack of unanimity in interface and interaction, etc.); (2) integrated approach by considering a multi-device eHealth system with motivational design features (such as timely feedback), which increases the involvement of patients as well as healthcare professionals and (3) practical recommendations for persuasive design and evaluation methodology combining validated and widely used methods from eHealth and HCI fields.

To conclude, lifestyle behavior support systems need to be be evaluated throughout all stages of the eHealth technology development cycle. Furthermore, it is also essential to evaluate the effects of eHealth interventions that have already been disseminated, using multidisciplinary approach and by independent evaluators.

VII. Acknowledgments

We thank all participants of the focus groups and other user studies, Municipal Health Services and social workers. Special thanks are due to Irina Lehmann and Cristel Boom for assistance in recruitment and data analysis. Part of the work presented is funded by the EU, within the framework of the ARTEMIS JU SP8 SMARCOS project 100249. Focus group study is a part of Sense eHealth project funded by the Ministry of Health, Welfare and Sport (VWS) and is carried out in collaboration with the Dutch National Institute for Public Health and Environment (RIVM).

References

[1] O. Kulyk, R. op den Akker, R. Klaassen, and J. van Gemert-Pijnen, “Let us get real! integrated approach for virtual coaching and real-time activity monitoring in lifestyle change support systems,” in Proceedings of the International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED’14). IARIA, 2014, pp. 211–216.

[2] F. Yekeh and J. Kay, "Hypothesis evaluation based on ubicomp sensing: Moving from researchers to users," in Persuasive'13, ser. CEUR Workshop Proceedings, S. Berkovsky and J. Freyne, Eds., vol. 973, 2013.

[3] T. W. Bickmore, D. Schulman, and C. Sidner, "Automated interventions for multiple health behaviors using conversational agents," Patient education and counseling, vol. 92(2), August 2013, pp. 142–148.

[4] T. W. Bickmore, L. Caruso, and K. Clough-Gorr, "Acceptance and usability of a relational agent interface by urban older adults," in ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'05). New York, NY, USA: ACM, 2005, pp. 1212–1215.

[5] H. op den Akker, M. Tabak, M. Marin-Perianu, R. Huis in 't Veld, D. V. M. Jones, D. Hofs, T. M. Tönis, B. W. van Schooten, P. M. M. Vollenbroek-Hutten, and P. H. J. Hermens, "Development and evaluation of a sensor-based system for remote monitoring and treatment of chronic diseases - the continuous care & coaching platform," in 6th International Symposium on eHealth Services and Technologies, EHST 2012. Portugal: SciTePress - Science and Technology Publications, July 2012, pp. 19–27. [Online]. Available: http://doc.utwente.nl/83432/ [retrieved: Nov 30, 2014]

[6] L. Van Velsen, N. Nijhof, and O. Kulyk, "Health 2.0 emerging technologies," in Improving eHealth, J. van Gemert-Pijnen, O. Petersen, and H. Ossebaard, Eds. Boom Publishers, den Haag, The Netherlands, 2013, pp. 111–126.

[7] S. Ananthanarayan and K. Siek, "Persuasive wearable technology design for health and wellness," in 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'12), 2012, pp. 236–240.

[8] N. Nijland, "Grounding ehealth: towards a holistic framework for sustainable ehealth technologies," Ph.D. dissertation, University of Twente, Enschede, the Netherlands, January 2011. [Online]. Available: http://doc.utwente.nl/75576/ [retrieved: Nov 30, 2014]

[9] E. J. van Gemert-Pijnen, N. Nijland, M. van Limburg, C. H. Ossebaard, M. S. Kelders, G. Eysenbach, and R. E. Seydel, "A holistic framework to improve the uptake and impact of ehealth technologies," J Med Internet Res, vol. 13, no. 4, Dec 2011, p. e111. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22155738 [retrieved: Nov 30, 2014]

[10] M. S. Kelders, E. J. Van Gemert-Pijnen, A. Werkman, N. Nijland, and R. E. Seydel, "Effectiveness of a web-based intervention aimed at healthy dietary and physical activity behavior: A randomized controlled trial about users and usage," J Med Internet Res, vol. 13, no. 2, Apr 2011, p. e32.

[11] K. D. Ahern, M. J. Kreslake, and M. J. Phalen, "What is ehealth (6): Perspectives on the evolution of ehealth research," J Med Internet Res, vol. 8, no. 1, Mar 2006, p. e4.

[12] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive systems design: Key issues, process model, and system features," Communications of the Association for Information Systems, vol. 24, no. 1, 2009, p. 28.

[13] S. Pingree, R. Hawkins, T. Baker, L. duBenske, L. Roberts, and D. Gustafson, "The value of theory for enhancing and understanding e-health interventions," American journal of preventive medicine, vol. 38(1), 2010, pp. 103–109.

[14] H. Oh, C. Rizo, M. Enkin, and A. Jadad, "What is ehealth (3): A systematic review of published definitions," J Med Internet Res, vol. 7, no. 1, Feb 2005, p. e1.

[15] A. Black, J. Car, C. Pagliari, C. Anandan, and K. e. a. Cresswell, "The impact of ehealth on the quality and safety of health care: A systematic overview," PLoS Med, vol. 8(1), 2011, e1000387. doi:10.1371/journal.pmed.1000387 [retrieved: Nov 30, 2014].

[16] S. Consolvo, P. Klasnja, D. W. McDonald, D. Avrahami, J. Froehlich, L. LeGrand, R. Libby, K. Mosher, and J. A. Landay, "Flowers or a robot army?: Encouraging awareness & activity with personal, mobile displays," in Proceedings of the 10th International Conference on Ubiquitous Computing, ser. UbiComp '08. New York, NY, USA: ACM, 2008, pp. 54–63.

[17] S. A. Clemes and R. A. Parker, "Increasing our understanding of reactivity to pedometers in adults," Medicine & Science in Sports & Exercise, 2009.

[18] H. op den Akker, V. M. Jones, and H. J. Hermens, "Tailoring real-time physical activity coaching systems: a literature survey and model," User Modeling and User-Adapted Interaction (available online), 2014.

[19] M. Kasza, V. Szucs, A. Vegh, and T. Torok, "Passive vs. active measurement: The role of smart sensors," in The Fifth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2011.

[20] H. op den Akker, R. Klaassen, R. op den Akker, V. Jones, and H. Hermens, "Opportunities for smart and tailored activity coaching," in Proceedings of the 26th International Symposium on Computer-Based Medical Systems (CBMS2013). USA: IEEE Computer Society, 2013, pp. 546–547, iSBN=978-1-4799-1053-3.

[21] W. Brouwer, A. Oenema, R. Crutzen, J. de Nooijer, N. de Vries, and J. Brug, "An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: A delphi study approach," J Med Internet Res, vol. 10(2):e10, 2008.

[22] J. Fanning, P. S. Mullen, and E. McAuley, "Increasing physical activity with mobile devices: A meta-analysis," J Med Internet Res, vol. 14 (6), Nov 2012, p. e161.

[23] L. Dennison, L. Morrison, G. Conway, and L. Yardley, "Opportunities and challenges for smartphone applications in supporting health behavior change: Qualitative study," J Med Internet Res, vol. 15, no. 4, Apr 2013, p. e86.

[24] S. Bosch, M. Marin-Perianu, R. Marin-Perianu, P. Havinga, and H. Hermens, "Keep on moving! activity monitoring and stimulation using wireless sensor networks," in Smart Sensing and Context, ser. Lecture Notes in Computer Science, P. Barnaghi, K. Moessner, M. Presser, and S. Meissner, Eds. Springer Berlin / Heidelberg, 2009, vol. 5741, pp. 11–23.

[25] T. Baranowski and L. Frankel, "Let's get technical! gaming and technology for weight control and health promotion in children," Childhood Obesity, vol. 8(1), 2012, pp. 34–37.

[26] S. Deterding, "Gamification: Designing for motivation," Interactions, vol. 19, no. 4, 2012, pp. 14–17. [Online]. Available: http://doi.acm.org/10.1145/2212877.2212883 [retrieved: Nov 30, 2014]

[27] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining 'gamification'," in Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, ser. MindTrek '11. New York, NY, USA: ACM, 2011, pp. 9–15. [Online]. Available: http://doi.acm.org/10.1145/2181037.2181040 [retrieved: Nov 30, 2014]

[28] W. Kamin and A. Kreplin, "Teaching the inhalation manoeuvre to asthmatic children by means of visual feedback," Pneumologie, vol. 61(3), 2007, pp. 150–156.

[29] H. Kharrazi, A. Faiola, and J. Defazio, "Healthcare game design: Behavioral modeling of serious gaming design for children with chronic diseases," in Proceedings of the 13th International Conference on Human-Computer Interaction. Part IV: Interacting in Various Application Domains. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 335–344.

[30] P. Klasnja, S. Consolvo, and W. Pratt, "How to evaluate technologies for health behavior change in hci research," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 3063–3072.

[31] S. Lanfrial, "From digital interventions to behavior change support systems: Understanding persuasive systems development and evaluation of process behavior change technologies," in Proceedings of IRIS 2012, 2012, pp. 1–16.

[32] J. Scholtz and S. Consolvo, "Toward a framework for evaluating ubiquitous computing applications," IEEE Pervasive Computing, vol. 3, 2004, pp. 82–88.

[33] B. Fogg, "A behavior model for persuasive design," in Proceedings of the 4th International Conference on Persuasive Technology, ser. Persuasive '09. New York, NY, USA: ACM, 2009, pp. 40:1–40:7. [Online]. Available: http://doi.acm.org/10.1145/1541948.1541999 [retrieved: Nov 30, 2014]

[34] R. op den Akker, R. Klaassen, T. Lavrysen, G. Geleijnse, A. van Halteren, H. Schwietert, and M. van der Hout, "A personal context-aware multi-device coaching service that supports a healthy lifestyle," in Proceedings of the 25th BCS Conference on Human-Computer Interaction, ser. BCS-HCI '11. Swinton, UK, UK: British Computer Society, 2011, pp. 443–448. [Online]. Available:

http://dl.acm.org/citation.cfm?id=2305316.2305397 [retrieved: Nov 30, 2014]

[35] H. op den Akker, M. Tabak, M. Marin-Perianu, M. H. A. Huis in 't Veld, V. M. Jones, D. H. W. Hofs, T. M. Tönis, B. W. van Schooten, M. M. R. Vollenbroek-Hutten, and H. J. Hermens, "Development and evaluation of a sensor-based system for remote monitoring and treatment of chronic diseases - the continuous care & coaching platform," in Proceedings of the 6th International Symposium on eHealth Services and Technologies, EHST 2012, Geneva, Switzerland. Portugal: SciTePress - Science and Technology Publications, July 2012, pp. 19–27.

[36] R. Klaassen, R. op den Akker, and H. op den Akker, "Feedback presentation for mobile personalised digital physical activity coaching platforms," in Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments, ser. PETRA '13. New York, NY, USA: ACM, 2013, pp. 64:1–64:8. [Online]. Available: http://doi.acm.org/10.1145/2504335.2504404 [retrieved: Nov 30, 2014]

[37] R. Klaassen, J. Hendrix, D. Reidsma, R. op den Akker, B. Dijk, and H. op den Akker, "Elckerlyc goes mobile-enabling natural interaction in mobile user interfaces," International journal on advances in telecommunications, vol. 6 (1-2), 2013, pp. 45–56.

[38] H. Vilhjalmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thorisson, H. van Welbergen, and R. J. van der Werf, "The behavior markup language: Recent developments and challenges," in Intelligent Virtual Agents, ser. LNCS, C. Pelachaud, J.-C. Martin, E. Andre, G. Collet, K. Karpouzis, and D. Pele, Eds., vol. 4722. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 99–111.

[39] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive systems design: Key issues, process model, and system features," Communications of the Association for Information Systems, vol. 24(28), 2009.

[40] C. Lisetti, R. Amini, U. Yasavur, and N. Rishe, "I can help you change! an empathic virtual agent delivers behavior change health interventions," ACM Trans. Manage. Inf. Syst., vol. 4, no. 4, Dec. 2013, pp. 19:1–19:28.

[41] J. Freyne, I. Saunders, E. Brindal, S. Berkovsky, and G. S. 0003, "Factors associated with persistent participation in an online diet intervention." in Extended Abstracts of the ACM Annual Conference on Human Factors in Computing Systems, 2012, J. A. Konstan, E. H. Chi, and K. Höök, Eds. ACM, 2012, pp. 2375–2380.

[42] A. Barak, L. Hen, M. Boniel-Nissim, and N. Shapira, "A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions," Journal of Technology in Human Services, vol. 26(2-4), 2008, pp. 109–160.

Impact of Population Size, Selection and Multi-Parent Recombination within a Customized NSGA-II and a Landscape Analysis for Biochemical Optimization

Susanne Rosenthal, Markus Borschbach
University of Applied Sciences, FHDW
Faculty of Computer Science, Chair of Optimized Systems,
Hauptstr. 2, D-51465 Bergisch Gladbach, Germany
Email: {susanne.rosenthal, markus.borschbach}@fhdw.de

***Abstract*—The main task in the drug design process is the prediction of the peptide structure and the bioactivity with the focus on simultaneously optimization of molecular peptide features. The synthesis and laboratory screening are the conventional but cost-intensive steps for optimization. Multi-objective genetic algorithms provide a range of methods for an efficient design of drug peptides. A customized NSGA-II has been especially evolved for biochemical optimization with the focus on producing a great number of very different high quality peptides within a very low number of generations e.g., under 20, termed early convergence. The focus of this work are an insight into the impact of the interdependence between the selection procedure and the population size, the empirical verification of the early convergence behavior within a limited range of population size and the influence of multi-parent recombination on the algorithm performance. These purposes are exemplary investigated on two different dimensional biochemical optimization problems, which are concrete, but as generic as possible. A landscape analysis is performed to gain an insight into the characteristic features and difficulties of the multi-objective optimization problems. The performance is assessed on the basis of a convergence indicator especially evolved for our preference of comparing the convergence behavior of populations with different sizes.**

***Index Terms*—multi-objective biochemical optimization, population size, landscape analysis, multi-parent recombination.**

I. Introduction

A customized Non-dominated Sorting Genetic Algorithm (NSGA-II) has been evolved with a considerable low number of generations and population size, termed early convergence for the molecular optimization of peptide sequences [1] [2]. Small peptides are of special interest in the area of drug design as they have some favorable features like conformational restriction, membrane permeability, metabolic stability and oral bioavailability [3]. Nevertheless, for this purpose these peptides have to optimize several molecular features at the same time. As both the synthesis and the laboratory characterization of peptides is very cost-intensive [4], moGAs provide an economical and robust method for peptide identification.

The NSGA-II is customized with regard to the encoding and the components mutation, recombination and selection. Different mutation and recombination methods have been evolved for this purpose and are introduced in [5][6]. These components and their parameter are not only inter related, but are also responsible for the performance of a GA. So far, less work has been done to gain an insight in the influence of the population size on the performance and in the interdependence with the selection operator and its parameters in the case of moGAs. The population size is an important value in influencing the performance of evolutionary algorithms [7]. Small population sizes tend to result in poor convergence and large populations extend the computational complexity of a GA in finding high quality solutions [8]. Therefore, an adequate population size that results in good performance is challenging. Diverse results have been presented regarding the choice and the handling of the populations size for single-objective GA: Yu et. al [9] study the connection between selection pressure and population size and ratify the concept of interdependence of parameters and operators in GA. The concept of self-adaption is used to overcome the problem of determining the optimal population size. Two forms of self-adaption are used: First, Bäck et al. [10] uses self-adaption as a previous setup and configuration step for evolutionary strategies. The population size then remains the same over all iterations. Second, Arabas et al. [11] introduces a GA with varying population size. The self-adaption of the population size is used throughout the whole GA run and depends among others on different parameters like the reproduction ratio. Eiben et al. [12] provide empirical studies that self-adaption of selection pressure and population size is possible and further rewarding regarding algorithm performance. In this case study, the global parameters tournament size and population size are regulated.

Several works have been proposed studying the effect of different numbers of parents for recombination in a range of 2 up to 10 parents in Evolutionary Algorithms (EA) e.g., [13][14][15], among others. These studies show that the optimal number of parents for recombination depends on the optimization problem as well as on the recombination method. Eiben represents in [13] a very extensive series of tests - in total 23000 test runs. In most of these cases the largest algorithm performance improvement can be obtained when the number of parents is increased from 2 to 3 parents. The experiments support two kinds of conclusions: Firstly, increasing the number of parents improves the performance continuously, but the degree of improvement is decreasing. Secondly, the performance improves by increasing the parent number until a certain parent number and decreases or oscillates afterwards. Another effect of a larger parent number is that less information are inheritance of the same solution and the generated offspring is more different from its parents.

The questions that we consider in this paper are:

1) Do large populations speed up the convergence behavior of the customized NSGA-II for a three-dimensional biochemical minimization problem?
2) Is there a predictable impact between population size and selection?
3) Is there a range of population size, which is able to perform well?
4) Do a variation of the parent number within the recombination procedure influence the algorithm performance?

The questions 1.-3. have been investigated and answered in [1] on a three-dimensional biochemical minimization problem and is further part of this work for comparison and completeness. The following question 5 is logical consequence and in the focus of this work:

5) Are the results of question 2.-4. transferable from the three-dimensional to the four- dimensional biochemical minimization problem?

These questions are answered in an empirical way: The performance of the customized NSGA-II is assessed regarding its early convergence and a high diversity within the solutions. Some metrics have been proposed to evaluate the convergence behavior of a moGA [16]. These metrics, generally, measure the distance of non-dominated solution sets to the true Pareto front [16]. This makes a comparison of generations with different sizes impossible. Therefore, a convergence indicator is introduced especially for the comparison of the generations with different sizes based on the hypervolume. The favorable features of this indicator are also discussed. A landscape analysis is performed to determine the characteristic properties of the biochemical landscape and to gain an insight into the difficulties of the optimization problems. Furthermore, we will discuss available open source Java tools that allow an easy implementation of the customized NSGA-II to solve multi-objective biochemical optimization problems.

The remainder of this paper is organized as follows: Section II describes the components of the customized NSGA-II. Section III provides a comparison of open source Java frameworks focused on a most simple implementation of the customized NSGA-II. Section IV presents a review on landscape analysis methods and the results of the landscape analysis performed on the biochemical objective functions. Section V introduces the new convergence metric and discusses the motivation for its evolution and the indicator features. Section VI provides the performance results of the configurations with different population sizes and multi-parent recombination assessed on the three- and four-dimensional optimization problem. Furthermore, this section responses the questions raised in this section. Section VII provides the conclusion of this work and gives an outlook on the future work.

II. The customized NSGA-II for peptide optimization

In this section, the customized NSGA-II is described as used in the presented experiments. In the previous work [2][5], we have assessed the performance and interaction of different recombination and mutation operators. In these experiments, we have determined the optimal onset of recombination and mutation method that is used within the following experiments. Additionally, we have customized the encoding and selection for the purpose of peptide optimization. The procedure of the customized NSGA-II corresponds to the procedure of the traditional NSGA-II [5] and is depicted in Fig. 1:

At first, the procedure initializes the start population with

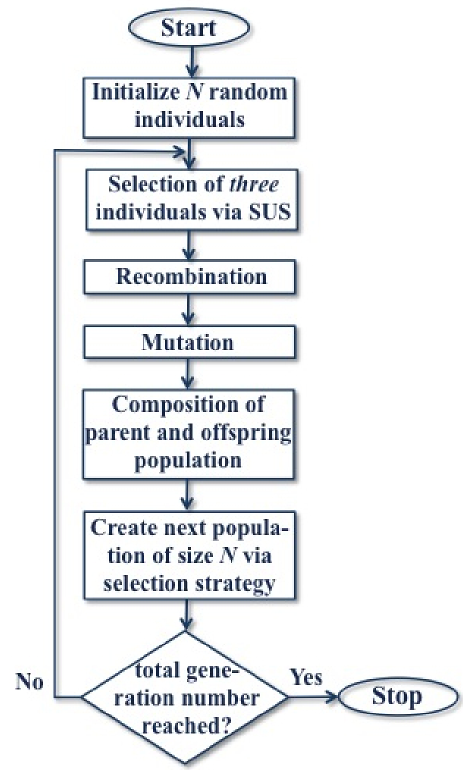


Figure 1. The procedure of the customized NSGA-II

a size of N. The main loop of the customized NSGA-II starts with the loop comprising the selection of the parent individuals for recombination and mutation via Stochastic Universal Sampling (SUS) based on roulette wheel selection: In each selection step, k individuals are selected via SUS to create k offsprings by recombination and mutation. This loop is repeated until the offspring population consists of N individuals. Then, the start and offspring population are shuffled together and the next population is created via a selection strategy. In the case, that the total number of generations is achieved, the main loop stops, otherwise it is repeated. The components of this procedure are motivated and described in detail below.

A. The encoding

The individuals are encoded as 20-character strings symbolizing the 20 canonical amino acids. The 20 characters are adopted of the single-letter code for amino acids. This encoding is firstly motivated by the idea of a most intuitive way of peptide encoding. Secondly, several tools predicting physiochemical or structural peptide properties make use of this encoding regarding the input data (e.g., see [17][18]). This avoids the data transformation before every fitness function evaluation.

B. The molecular fitness functions

Four molecular fitness functions are proposed to constitute the benchmark problems. These functions are selected under the aspect of predicting physiochemical properties, which are of importance for drug design [19]. Therefore, this combination of fitness functions allows conclusions on a range of important peptide properties [20]. These functions are associated to the three structural levels of peptides: The first two physiochemical functions refer to the primary structure. The third provides information about the secondary structure and the last one makes use of the primary structure of a peptide to provide information about a possible early tertiary structure disruption or an inadequate folding.

1) Molecular Weight (MW): The first fitness function is the calculation of the MW that is an important peptide feature for the purpose of drug design [3] and refers to the primary structure of a peptide. This fitness function is selected from the open source library BioJava, this library and a detailed description of this function are provided on the homepage [17].

2) Hydrophilcity (hydro): The second fitness function is the determination of the hydrophilicity (hydro) of a peptide. A hydrophilicity value is assigned to each peptide via the hydrophilicity scale of Hopp and Woods with a window size of the peptide length [21]. This fitness function refers also to the primary structure.

3) Needleman-Wunsch Algorithm (NMW): The third fitness function determines the optimal global similarity score provided by NMW [22] that is also part of the BioJava library [17]. The motivation for NMW is the identification of similarities between peptides regarding biochemical functionality and structure via a global sequence alignment to a pre-defined reference peptide. NMW makes use of a scoring model and in this case the BLOcks SUbstitution Matrix (BLOSUM) [18] in form of the percentage identity 100 (BLOSUM100) is used.

4) Instability Index (InstInd): The last fitness function is the InstInd and is used to analyze the primary structure of peptide sequences to predict a potential intracellular instability of peptides. This function is also provided by the BioJava library and a description is also given on the homepage [17]. These four fitness functions act comparatively: The fitness values of an individual are determined by the difference between the fitness function values of this peptide to the fitness function values of a predefined reference-peptide. Therefore, these four objective functions have to be minimized.

C. The recombination operator

In the previous work [5], different recombination operators are benchmarked on a three-dimensional molecular minimization problem. The linear $n-$point recombination operator achieved the best performance, where the number of recombination points n are determined by a linearly decreasing function:

$$x_R(t) = \frac{l}{2} - \frac{l/2}{T} \cdot t, \quad (1)$$

which depends on the peptide length l, the total number of the GA generations T and the index of the actual generation t. The motivation of this recombination operator is a preferred high explorative search behavior in the early generations and a high motifs-maintaining encouraging the local search in later generations. For this purpose, the number of recombination points in the first generation is $l/2$ and decreases linearly until one recombination point in the last generation. The recombination points themselves are determined randomly. One recombination point in the last generation guarantees a motif preservation of at least 50% of the peptide sequences. The recombination operator is usable as multi-parent recombination, where the default number of parent is three according to the results of Eiben [13]. The impact of the number of parents on the multi-dimensional biochemical minimization problem is challenging and in the focus of the following experiments.

D. The mutation operator

In combination with LiDeRP, an adaption of the deterministic dynamic operator of Bäck and Schütz revealed the best performance as reported in [5]. This mutation operator is motivated by the idea that a high number of mutations in the early generations provides a good exploration, whereas a low number of mutations in later generations leads to good exploitation. The mutation rates are determined via the decreasing function

$$p_{a,BS} = (a + \frac{l-2}{T-1}t)^{-1}, \quad (2)$$

with $a = 2$, l describes the peptide length, T the total generation number of the GA and t the index of the actual generation number. This decreasing function has been adapted to a lower start mutation rate as a high start mutation rate as well as a high start recombination rate results in a too high exploration in the early generations.

E. The Aggregate Selection

The flow diagram in Fig. 2 depicts the selection methods. The Aggregate Selection is tournament-based. From the tournament set individuals are chosen from the first front with a probability p_0 and with a probability $1-p_0$ the individuals are chosen via Stochastic Universal Sampling (SUS). The number N of pointers is the number of fronts and the segments are equal in size to the number of individuals in each front.

Therefore, the selection method has two parameters, the tournament size and the probability of choosing individuals

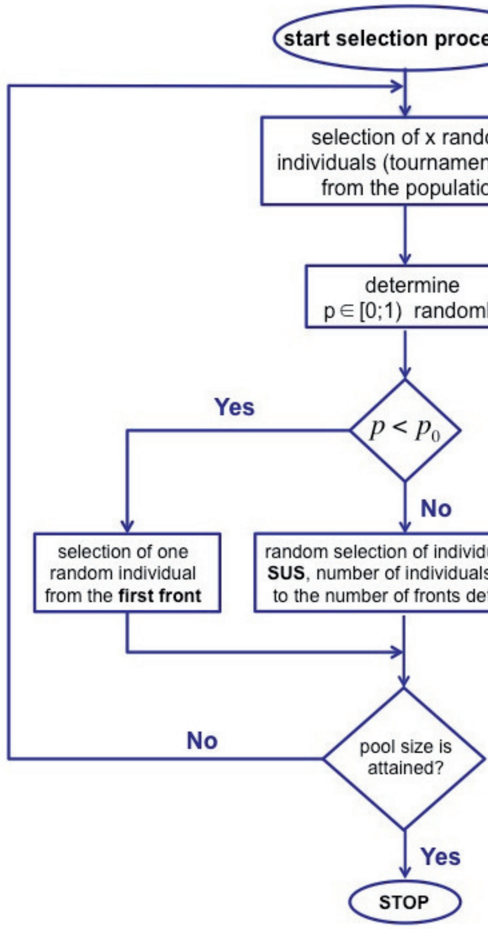


Figure 2. Aggregate selection strategy

from the first front. The tournaments size of 10 has proven to be an optimal choice. The parameter p_0 is challenging regarding the population size.

III. Open source Java frameworks

In this section, we summarize and describe different open source Java tools that provide Genetic Algorithm implementations. The summarization is focused on Java frameworks for a most simple implementation of BioJava, which provides several physiochemical properties via APIs. The main goal of this framework analysis is the selection of a tool that allows an easy implementation of the proposed customized NSGA-II.

The framework Java API for Genetic Algorithm (JAGA) in its current version 1.0 beta is a research tool developed and supported by the Computer Science Department of the University College London [23]. This tool does not include any moGAs, but it provides a protein string sequence encoding using 20 different characters symbolizing the 20 canonical amino acids. Among others, eight different physiochemical properties like hydrophobic, aliphatic, aromatic and polar are pre-defined for each canonical amino acid. In addition, it contains for each genotype a parameter-depending crossover and mutation method and allows a peptide or protein representation by the pre-defined amino acid patterns. The user interested in a moGA application has to extend this tool for this purpose, but the amino acid character encoding is a clear benefit. Other useful functions are the opportunity of creating a random initial population of protein sequences and the implementation of the Needleman-Wunsch or Smith-Waterman Algorithm.

The framework Metaheuristic Algorithms in Java (jMetal) in its current version 4.3 is an extensive and complex tool especially for moGA applications [24]. It contains beneath NSGA-II the moGA variants: Pareto Envelope-based Selection Algorithm (PESA), improved Strength Pareto Evolutionary Algorithm (SPEA2), improved PESA (PESA2), S-Metric Selection Evolutionary Multiobjective Evolutionary Algorithm (SMS-EMOA), Indicator-Based Evolutionary Algorithm (IBEA) and Multiobjective Evolutionary Algorithm based on Decomposition (MOEA/D). Further, different variation operators are implemented like single-, two- point, Simulated Binary Crossover (SBX) and polynomial, uniform and swap mutation. 'Ranking&crowding selection' is included as the traditional NSGA-II selection method as well as tournament and PESA2 selection. Additionally, jMetal provides several established metrics to evaluate the performance like the hypervolume, Inverse General Distance (IGD), General Distance (GD) and a measure for diversity. A definite advantage of jMetal is the intuitive and clear program construction, which allows an easy algorithmically extension. The disadvantage is a missing character or string encoding.

The framework Java-based Evolutionary Computation Research System (ECJ) in its current version 21 is comparable with jMetal in the issues functional complexity and potential extension. ECJ is developed at George Mason University's Evolutionary Computation Laboratory [25]. It includes the moGAs NSGA-II and SPEA2. Furthermore, different vector representations with corresponding variation operators are included as well as SUS and tournament selection, among others. Moreover, it proposes the potential to read populations from files. It does not provide an intuitive and clear program structure like jMetal.

The Multi-Objective Evolutionary Algorithm framework (MOEA framework) in the current version 2.1 is a Java framework for multi-objective optimization [26]. It provides a very wide range of MOEA variants as it includes the jMetal library in the version 4.3. Therefore, MOEA framework has the same features like jMetal regarding the benchmark problems, performance metrics and available variation operators. The MOEA provided by jMetal also only support binary, real-values and permutation encoding. On the other hand, the MOEA framework allows a new genotype implementation.

Evolutionary Algorithms workbench (EvA2) is a Java framework developed by the department of computer science at the Eberhard Karls University in Tübingen [27]. It is not only intended for research, but is also deployed for industrial applications and is available under LGPL license. Its specificity is its easy-to-use graphical user interface and provides a MATLAB interface to optimize functions in MATLAB with standard algorithm implementations in EvA2. It also has a client-server structure and provides NSGA-II, PESA and SPEA2 as moGA implementations. A string or character encoding is not implemented and an implementation afterwards is challenging, because encoding affects all parts of the toolbox.

The framework Java Class library for Evolutionary Computation (JCLEC) in the current version 4 includes the evolutionary features NSGA-II and SPEA2. It proposes different encodings with various variation operators except string or character encoding, but provides an expendable program structure. Further, JCLEC includes fitness proportionate selection strategies like tournament and SUS.

The modular framework for metaheuristic optimization (OPT4J) contains a set of multi-objective optimization algorithms including SPEA2 and NSGA-II [28]. OPT4J has been evolved under two aspects: A simple evolutionary optimization of user-defined problems and the potential of an arbitrary optimization algorithm implementation. Further, the common benchmark problems ZDT, DTLZ, WFG and Knapsack problem are available as well as the genotype encoding binary, integer, real-values and permutation. Unfortunately, the module-based structure and the use of the GUI makes an extension for the purpose of implementation of the proposed customized NSGA-II more complicated. A special feature is the graphical visualization of the optimization process regarding the convergence and potential of a Pareto plot.

Fig. 2 gives an overview of the reviewed Java frameworks. These frameworks are compared under the aspects of: (i) configuration of a character or string encoding as an option, (ii) an implementation of NSGA-II, (iii) potential of a simple extension, and (iv) an intuitive program structure according to the moGA components.

TABLE I
OVERVIEW OF THE SPECIAL FRAMEWORK ASPECTS

	JAGA	jMetal	MOEA	ECJ	EvA2	JCLEC	OPT4J
(i)	x						
(ii)		x	x	x	x	x	x
(iii)		x	x	x		x	
(iv)		x	x				

Table I reveals that none of the open source Java frameworks attains all required aspects in an adequate level. As a consequence, the proposed customized NSGA-II is implemented within jMetal. MOEA framework is a possible alternative as it provides the jMetal library. Nevertheless, some programming effort is necessary regarding this implementation. Furthermore, the protein string sequence encoding of JAGA serves as a model to the targeted peptide encoding. The experiments of the four-dimensional optimization problem are performed with an implementation of the proposed customized NSGA-II in jMetal, the experiments on the three-dimensional optimization problem have been performed with an own implementation as presented in [1].

IV. Molecular landscape analysis

Fitness landscape analysis is a common and important methodology to gain an insight in the complexity and difficulty of an optimization problem with the aim of designing a search algorithm with optimized performance [29]. The components of a landscape are the configuration set X of all feasible solutions of the optimization problem. According to the organization of X, a notation on neighborhood, nearness distance or accessibility on X is required. The third and essential component are the fitness functions $f : X \to \mathbb{R}$, [30].

The aim of the landscape analysis is the investigation of the landscape structure and the determination of the landscape characteristics that have a strong influence on the search algorithm e.g., see [31]:

- **Modality:** The modality provides the tendency of the fitness landscape to produce local optima. Therefore, the number and distribution of local optima are indicators for the modality.
- **Correlation:** This is an indicator for the dependence between two solutions of X. In the area of multi-objective landscapes, the correlation is of particular interest as it provides information about the actually optimization problem dimension and therefore about the problem difficulty. In the single-objective case, the autocorrelation is commonly used as an indicator for fitness diversity.
- **Ruggedness:** This is a feature describing the fitness variation between the fitness values of a solution and its neighbors. The modality and the correlation provide hints for the level of ruggedness.
- **Plateaus:** These are areas referring to neutrality, constituted by a set of solutions with equal fitness values. The size of these areas and the probability of existence are used for description.

Different techniques have been proposed to analyze the characteristic features of a fitness landscape. These techniques are classified into two categories both based on solution sequences obtained by random walks: The statistical analysis and the information analysis. [31]

Statistic analysis techniques investigate the fitness landscapes in a qualitative manner. Therefore, several correlation metrics have been proposed to measure the ruggedness of a landscape. Weinberg proposed the random walk correlation function $r(s)$ that measures the autocorrelation between two sets of fitness points separated by s solutions [32]:

$$r(s) = \frac{\sum_{i=1}^{n-s}(f_i - \bar{f})(f_{i+s} - \bar{f})}{\sum_{i=1}^{n}(f_i - \bar{f})^2}, \qquad (3)$$

where f_i is the fitness function value of the random walk solutions $\{f_i\}_{i=1...n}$ and $\bar{f}$ is the average fitness value of all solutions. High autocorrelation values indicate that the fitness values are similar and the landscape is less rugged. Otherwise, a small autocorrelation value indicates that the fitness values are uncorrelated and the landscape is rugged. Another correlation metric based on the autocorrelation is the correlation length that measures the distance beyond which two sets of fitness points becomes uncorrelated [31]:

$$l = -\frac{1}{ln|r(1)|}, \quad (4)$$

with $r(1) \neq 0$. The higher the correlation length the smoother is the landscape.

Jones proposed the Fitness Distance Correlation (FDC) [33]. This coefficient measures the relation of the fitness values and the distance of the solutions s_i to the nearest optimum x^* in the search space:

$$FDC = \frac{cov(f(s_i), d(s_i))}{\sqrt{var(f(s_i)) \cdot var(d(s_i))}}, \quad (5)$$

where d is the distance function to x^* and $cov(x, y)$ as well as $var(x)$ are the common statistical measures covariance and variance. The coefficient values are in the interval $[-1; 1]$. Jones further introduced a prediction of these values regarding the GA effectiveness in solving the optimization problem:

- $FDC \geq 0.15$: The fitness increases with the distance. The GA is potentially not effective or the problem is misleading.
- $-0.15 < FDC < 0.15$: There is virtually no correlation between fitness and distance. The problem is categorized as difficult.
- $FDC \leq -0.15$: The fitness increases as the optimum approaches. The GA is potentially effective or the problem is straightforward.

A great disadvantage of FDC is that the nearest optimum or at least the best-known solution has to be known in advance. Compared to the statistical analysis, the information analysis is a quantitative investigation and provides more detailed information about the landscape structure like a measurement of the optima density and plateaus as well as the degree of the random walk regularity [34]. Vassilev et al. [34] provides three information measures to determine the modality and the level of ruggedness. For each of these three measures, the random walk path $\{f_t\}_{t=0...n}$ is transformed into a string $S(\epsilon) = s_1 s_2 ... s_n$ with $s_i \in \{-1, 1, 0\}$, where

$$s_i = \begin{cases} -1 & \text{if } f_i - f_{i-1} < -\epsilon \\ 1 & \text{if } f_i - f_{i-1} > \epsilon \\ 0 & \text{if } |f_i - f_{i-1}| \leq \epsilon \end{cases} \quad (6)$$

and $\epsilon \in [0; l]$, where l is the maximal difference between two fitness values. The indicator is more sensitive to the steps of the random walk the smaller the value for ϵ. Then, the Information Content is defined via:

$$H(\epsilon) = -\sum_{p \neq p} P_{[pq]} log_6(P_{[pq]}), \quad (7)$$

where $p, q \in \{-1, 1, 0\}$, $P_{[pq]} = \frac{n_{[pq]}}{n}$ are the probabilities presenting frequencies of possible blocks pq and $n_{[pq]}$ is the number of occurrences of the blocks pq in $S(\epsilon)$. The base of the logarithm is chosen as 6. This is the number of all possible blocks pq. The information content depends on the parameter ϵ that is responsible for a more global or local view on the random walk according to the magnitude of ϵ. The Partial Information Content is a measure for the degree of ruggedness. The string $S(\epsilon)$ is transformed into a string $S(\epsilon')$ by deleting the elements 0 and blocks of equal elements are reduced to only one of these elements. The partial information content is defined by:

$$M(\epsilon) = \frac{v(\epsilon)}{n}, \quad (8)$$

where $v(\epsilon)$ is the length of $S'(\epsilon)$ and n the length of $S(\epsilon)$. Furthermore, $v(\epsilon)$ indicates the number of extrema along the landscape path. In the case $M(\epsilon) = 0$, the landscape path is nearly flat or monotonously increasing or decreasing. Otherwise, $M(\epsilon) = 1$ indicates that the landscape path is maximal rugged.

The Information Stability as the third indicator for information analysis proposed by Vassilev is an indicator for the highest difference between neighboring points in the landscape path. The information stability is defined as the smallest value of ϵ for which the landscape path becomes flat. In this case, the string $S(\epsilon)$ comprises only zeros.

Another information indicator was proposed by Leier et al. [35]. This indicator gives information about the density as well as length of flat areas. Therefore, it is an indicator for the ratio between flat and smooth parts of a landscape path and therefore an optimal measure for neutrality. It is defined as:

$$h(\epsilon) = -\sum_{p \in \{-1,1,0\}} P_{[pp]} log_3(P_{[pp]}), \quad (9)$$

where $P_{[pp]}$ is the frequency of blocks pp in $S(\epsilon)$.

A. Landscape analysis of the molecular fitness functions

The configuration set X of the molecular fitness landscapes are all feasible peptides of the length 20 consisting of the 20 canonical amino acids. Therefore, the landscape is of a high complexity 20^{20}. Furthermore, the search space is discrete as there are real-valued solutions that have no corresponding feasible peptides in the search space. The neighborhood of a solution is defined by all peptides differing to this solution in one amino acid in exactly one position [36]. Therefore, the move operator of the random walk is the mutation operator that changes one amino acid in a position in the solution. The mutation of an amino acid in the way that the same solution is achieved is excluded to prevent the random walk from stagnation.

The four fitness landscapes NWM, MW, hydro and InstInd are analyzed according to the important landscape properties: Modality, correlation and ruggedness. The basis of the landscape analysis is random walks of 100 steps that are repeated 30 times for statistical reasons. The starting solution of the random walk is initialized randomly.

In Fig. 3-6, six random walks are exemplarily depicted of the molecular fitness functions for a first global view on the landscapes. All four molecular fitness functions provide large variations of the fitness values over the 100 random walk steps and therefore indicate rugged landscapes. From

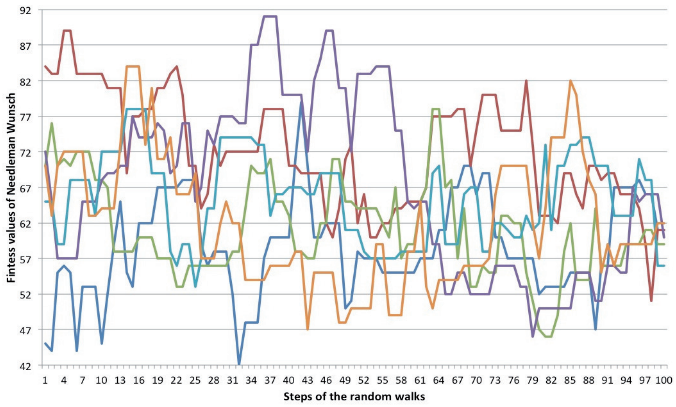


Figure 3. Exemplary presentation of six time series of the molecular fitness function NMW.

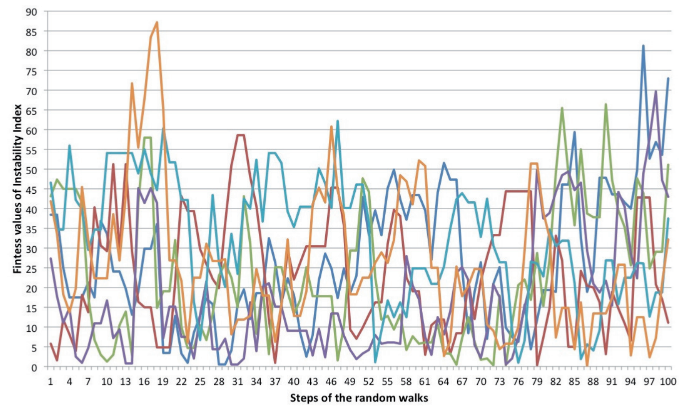


Figure 6. Exemplary presentation of six time series of the molecular fitness function InstInd.

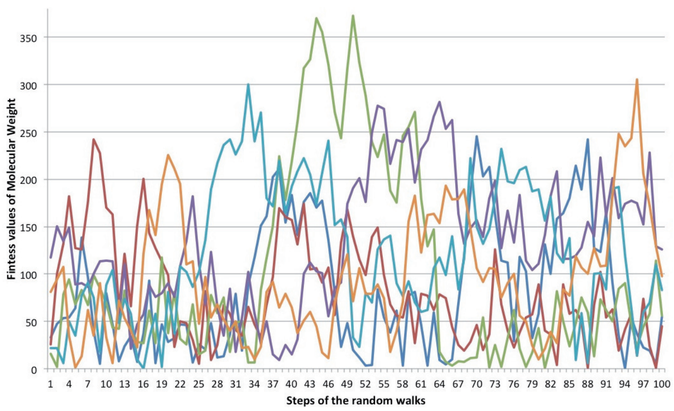


Figure 4. Exemplary presentation of six time series of the molecular fitness function MW.

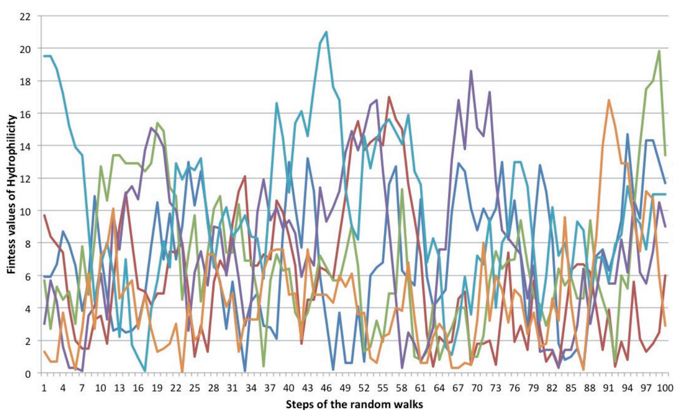


Figure 5. Exemplary presentation of six time series of the molecular fitness function hydro.

this global point of view, NMW is the only function revealing some plateaus or flat areas over two to five random walk steps (Fig. 3). The InstInd function also reveals some flat areas and plateaus, but in a lesser extent and averaging over a lower number of random walk steps (Fig. 6). The fitness values of the MW function are scaled by a factor of 10 and achieved some large jumps of the fitness values as well as some areas with oscillating parts (Fig. 4). The hydro fitness function appears similar to MW regarding the jumps and the oscillating parts (Fig. 5). Otherwise, it also reveals some isolated flat areas or plateaus. To quantify the rugged landscape properties of these four fitness functions, the autocorrelation of all solution (100 random walks repeated 30 times) is calculated after the model of Lee [36], which is an adaption of the autocorrelation function of Weinberg (eq. 3) by determining the average value and the standard deviation of all solutions and applying the average value on the starting point of the random walks:

$$p_s = \frac{\frac{1}{n+1}\sum_{i=0}^{n}(x_{i0}-\mu)(x_{is}-\mu)}{\sigma^2}, \tag{10}$$

where μ is the average value calculated by

$$\mu = \frac{1}{n}\sum_{i=1}^{n} f(x_i), \tag{11}$$

σ is the standard deviation determined by

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(x_i)-\mu)^2} \tag{12}$$

and x_{i0} is the starting point and x_{is} is the $s-$step of the i-th random walk. The self-correlation value of the starting point p_0 has to be 1 or at least approximately 1. This adaption is motivated by the fact that the random walk length of 100 is relatively small compared to the search space complexity, which empirically leads to a time-varying volatility - meaning, the average value and the standard deviation are very different between the random walks [36]. The autocorrelation of the time series of the four molecular fitness functions depicted in Fig. 7 confirm the time-varying volatility as the values for p_0 are differing from the value 1. The high ruggedness of the four molecular landscapes is visible by the fast decrease of the autocorrelation values to 0 within the first 20 random walk steps. Furthermore, most of the autocorrelation values of all four molecular functions are in the range of -0.3 to $+0.3$, which indicate a weak correlation. The autocorrelation values

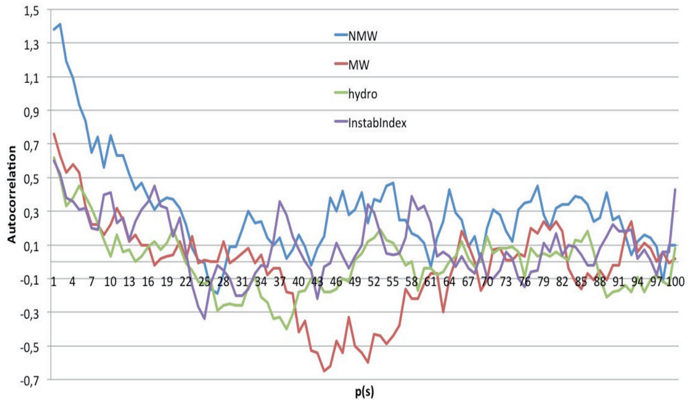


Figure 7. Autocorrelation functions of the random walks on the four molecular landscapes: NMW, MW, hydro, InstInd

of MW indicate a moderate autocorrelation between the random walk steps 40 to 57. Similarly, the autocorrelation values of NMW reveal some very moderate correlations between the random walk steps 45 to 88. Some outlier of very moderate correlations exists also for hydro and InstInd.

In addition to the autocorrelation, the correlation between the molecular fitness functions is also of great interest regarding the combination of these four molecular functions to a Multi-Objective Problem (MOP) as the high correlation between two time series of different fitness functions theoretically reduce the optimization problem dimension and makes the MOP less challenging. The correlation matrix indicates s potential linear relationship between different functions:

$$M_{corr} = \begin{pmatrix} 1 & corr(f_1,f_2) & \dots & corr(f_1,f_k) \\ corr(f_2,f_1) & 1 & \dots & corr(f_2,f_k) \\ \vdots & \vdots & \ddots & \vdots \\ corr(f_k,f_1) & corr(f_k,f_2) & \dots & 1 \end{pmatrix}, \quad (13)$$

where M_{corr} is symmetrical and consists of the Pearson correlation coefficients of the fitness function f_i and f_j:

$$corr(f_i, f_j) = \frac{\sum_{i=0}^{n}(f_i - \bar{f}) \cdot (f_j - \bar{f})}{\sigma_{f_i} \cdot \sigma_{f_j}} \quad (14)$$

Correlation values of $|corr(x,y)| < 0.3$ indicate a weak correlation between x and y, $0.3 \leq |corr(x,y)| \leq 0.8$ indicates a moderate correlation and $|corr(x,y)| > 0.8$ indicates a high linear correlation. The correlation matrix for the four molecular functions NMW (f_1), MW (f_2), hydro (f_3) and InstInd (f_4) according to eq. (13) is given by:

$$M_{corr} = \begin{pmatrix} 1 & 0.047 & 0.252 & 0.09 \\ \cdots & 1 & -0.014 & -0.032 \\ \cdots & \cdots & 1 & -0.266 \\ \cdots & \cdots & \cdots & 1 \end{pmatrix}. \quad (15)$$

This matrix is calculated of the 30 random walks consisting of 100 steps for each molecular function. The matrix entries reveal that there is no linear relationship between the time series of each two fitness functions: There is a weak relationship between NMW and MW (eq. (15): $corr(f_1, f_3) = 0.252$) as well as InstInd and hydro (eq. (15): $corr(f_3, f_4) = -0.266$) and no correlation between the other combinations.

Another important landscape property investigated in the following is the modality. The examination of the single molecular fitness functions according to the local optima is not advisable for the purpose of a MOP as the most of the local optima of the single functions are no optima in the sense of the MOP [13]. The optima in the multi-objective sense are the non-dominated solutions.

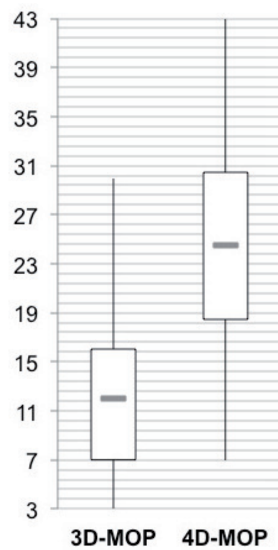


Figure 8. Number of the non-dominated solutions achieved by the random walks on the 3D and 4D molecular landscapes

Fig. (8) depicts the optima or non-dominated solution density of the 3D-MOP composed of NMW, MW and hydro as well as the 4D-MOP composed of NMW, MW, hydro and InstInd. In general, the 4D-MOP reveals nearly 50% more non-dominated solutions than the 3D-MOP within the 30 random walks of 100 steps. Further, the range of the non-dominated solutions achieved within the random walks is higher in the case of the 4D-MOP. As a consequence, the non-dominated solution density is higher in the case of the 4D-MOP and the corresponding landscape more rugged. Further, the 4D-MOP is more difficult for the customized NSGA-II then the 3D-MOP because of the MOML structures: The 3D-MOP has a higher front diversity and provides less solutions in the optimal front than the 4D-MOP.

Plateaus in this area of multi-objective real-valued landscapes are identified by consecutive equal fitness values for each molecular function. In the 30 random walks of the 3D-MOP, 20 plateaus have been identified: Two plateaus of each two consecutive equal fitness values have been identified in five random walks, a plateau of three consecutive equal fitness values have been found in one random walk and the remaining 9 plateaus have been identified in different random walks each consisting of two consecutive equal fitness values. In the case of the 4D-MOP, 8 plateaus have been identified: These plateaus consist only of two consecutive fitness values and only one random walk achieved two of these plateaus. Consequently, the 3D-MOP reveals statistically more plateaus than the 4D-MOP and a local search is of a greater interest in this case.

Transferring these results of the molecular landscape analysis on design considerations of a metaheuristic allows the conclusion that the search process has to be guided in direction of at most optimal solutions, which are spread over the landscape. Then, the neighborhood has to be searched for further high quality solutions. Transferring these results on the explorative- exploitive balance of the metaheuristic associates that a high exploration of the search process in the early generations and a more exploitive search behavior in the later generations is beneficial for such rugged landscapes with statistically only a low number of flat regions.

V. Evaluation measures for convergence and diversity

Firstly, the convergence measure is introduced, which has been especially evolved to evaluate generations with different sizes. Subsequently, the features of this indicator are discussed followed by the presentation of measurement for diversity.

A. Introduction of the average cuboid volume

In the past, several metrics have been proposed to evaluate the convergence behavior of populations produced by a moGA. Usually, they act on the distance of the non-dominated solution set of a generation to the true Pareto front. The hypervolume or the S-metric measures the overlapped space of the non-dominated solution set to a predefined anti-optimal reference point [37]. The hypervolume is a very established convergence metric with its favorable mathematical properties as one reason. Another convergence metric is the D-metric [16]. The D-metric makes use of the hypervolume and calculates the coverage difference of two solution sets. A reference set is needed to assess the convergence to the true Pareto front. The C-metric is an appropriate measure to compare the dominance of two Pareto optimal sets [37]. The Error Ratio (ER) is a percentage measure for the number of solutions in a set that are to be found on the true Pareto front [16]. GD is a measure of the average distance between a Pareto optimal solution set to the true Pareto front [38]. It includes the minimal component-wise distance of a solution set to the nearest one on the true Pareto front. The convergence metric of Deb also measures the distance between a solution set and a reference set of the Pareto front [39]. It calculates the average normalized distance for all solutions in the solution set. A recently published convergence metric is the Averaged Hausdorff Distance Δ_p [40]. It is based on GD and the IGD [41].

The reasons for the evolution of a new convergence metric in this paper and in the scientific community in general are multiple: The disadvantage of the metrics D-metric, ER, GD, Δ_p and the convergence metric of Deb is their dependency on the true Pareto front or at least a reference set of Pareto optimal solutions that are usually unknown in the case of real-world MOPs. Furthermore, these metrics are not useful indicators for an entire ranking between generations of different sizes. However, the populations in moGAs are generally limited in size. From a more global point of view, the evaluation and comparison of the global convergence behavior of whole populations - not only the non-dominated solution set of a generation - is required with respect to the influence of the population size or the selection pressure.

For this purpose, a new metric is presented that reflects the convergence behavior of a whole population and is a 'fair' indicator for comparison of generations of different sizes. This Average Cuboid Volume (ACV) is evolved according to the model of the hypervolume. The motivation for the exploitation of the hypervolume model is to profit from its preferable properties as mentioned above. The benefit of this new metric compared to the hypervolume is the low computational complexity as no point ordering is required.

In the following, we assume that the underlying optimization problem is to minimize. The metric calculates the average cuboid volume of the cuboids spanned by the solution points to a pre-defined reference point r:

$$ACV(X) = \frac{1}{n}\sum_{i=1}^{n}\left(\prod_{j=1}^{k}(x_{ij} - r_j)\right), \qquad (16)$$

where n is the population size, k is the number of objectives, x_i are the solutions on the population X and x_{ij} is the $j-th$ component of a solution x_i. It holds $(x_{ij} - r_j) > 0$ as the pre-defined reference point is chosen as the theoretical minimal limit of the true Pareto front. The lower the indicator values the more positive is the global convergence behavior as the reference point is chosen as a theoretical optimal point.

Obviously, ACV is not a suitable indicator to compare the experimental results of different dimensional optimization problems.

B. Discussion of the average cuboid volume

The question regarding the suitability of a metric for evaluation depends on the intention of the investigation object and the preferences. ACV is intended to evaluate the global convergence behavior of a whole population with the ultimate aim of comparing solution sets of different sizes according to the proximity to the true Pareto front.

The first expectation that is important for the use of ACV is that the convergence quality shall not change if the number of equally solutions increases. ACV does not fulfill this averaging strategy: Let $x \in \mathbb{R}^k$ be a solution of the optimization problem and $X = \{x\}$. Further, $Y = \{x, ...x\}$ is a set of n equally copies of the solution x, then

$$ACV(Y) = \frac{1}{n}\sum_{i=0}^{n}\left(\prod_{j=1}^{k}(x_j - r_j)\right) = \frac{1}{n}\cdot n\prod_{j=1}^{k}(x_j - r_j)$$
$$= \prod_{j=1}^{k}(x_j - r_j) = ACV(X). \qquad (17)$$

The second expectation is described by the following observation: An intuitive indicator reflecting the quality of approximation sets of different Pareto front refinements requires 'better' indicator values for the finest approximation set. This effect is demonstrated for ACV by an example also used in [13]:

Example 1: The Pareto front is given by the bounded convex function $f(x) = 1/x^2$ between the points $y_1 = (0.1, 100)$ and $y_2 = (1.1, 0.826)$ meaning

$$PF_{true} = \{(x, y)|y = 1/x^2 \text{ with } x \in [0.1, 1.1]\}. \quad (18)$$

We consider the following three approximation sets of increasing refinement of the Pareto front

$$Y_1 = \{(0.1 + 0.2 \cdot i, 1/(0.1 + 0.2 \cdot i)^2)|i \in \{0, 1, ..., 5\}\},$$
$$Y_2 = \{(0.1 + 0.1 \cdot i, 1/(0.1 + 0.1 \cdot i)^2)|i \in \{0, 1, ..., 10\}\},$$
$$Y_3 = \{(0.1 + 0.01 \cdot i, 1/(0.1 + 0.01 \cdot i)^2)|i \in \{0, 1, ..., 100\}\}.$$

Table I depicts the indicator values of ACV for the three approximation sets with the reference point $(0, 0)$.

TABLE II
ACV VALUES FOR THE APPROXIMATION SETS $Y_1 - Y_3$ WITH THE REFERENCE POINT $(0, 0)$.

X	Y_1	Y_2	Y_3
ACV(X)	3.13	2,75	2.43

The third preferable expectation of this indicator is the averaging effect. It is trivial that a dominating solution x yields better indicator values than the dominated one y, because $ACV(\{x\}) = \prod_{i=1}^{k}(x_j - r_j) < \prod_{i=1}^{k}(y_j - r_j) = ACV(\{y\})$. From this observation it can be interpreted that if one dominated solution x_1 in the solution set $X = \{x_1, x_2, ..., x_n\}$ is replaced by a dominating one $\bar{x}_1$, then $ACV(\{x_1, x_2, ..., x_n\}) > ACV(\{\bar{x}_1, x_2, ..., x_n\})$. The averaging effect of ACV is illustrated by the example, which has also been used for Δ_p [40]:

Example 2: The true discrete Pareto front is described by $P = \{p_i|p_i = (0.1\cdot(i-1); 1-(i-1)\cdot 0.1) \text{ with } i = 1, ..., 11\}$. Two solution sets are given by $X_1 = \{x_{1,1}, p_2, ..., p_{11}\}$ and $X_2 = \{x_{2,1}, x_{2,2}, ..., x_{2,11}\}$ with the elements $x_{1,1} = (\epsilon, 10)$ and $x_{2,i} = p_i + (\frac{\epsilon}{2}, 5)$ with $i = 1, ..., 11$. For the outlier $x_{1,1}$ the values $\epsilon = 0.001$ is used for numerical evaluations. X_1 is a better approximation of the true Pareto front, but it contains the outlier $x_{1,1}$. On the other side, X_2 is close to the true Pareto front and the difference of each element to the Pareto front is less than the one of the outlier. As we are interested in an averaging effect, the indicator values of X_1 has to be better than the one of X_2. This is true for ACV as $ACV(X_1) = 0.15$ and $ACV(X_2) = 2.65$.

The use of $ACV(X)$ as a convergence and as a diversity metric is not within our preferences. $ACV(X)$ is not a reliable indicator for diversity. A solution set with clustered solutions does not always achieve better indicator values demonstrated in the following example:

Example 3: Once more PF_{true} is described by eq. (18) and the solution set

$$Y_4 = \{(0.29, 11.89); (0.3, 11.11); (0.31, 10.4); (0.32, 9.77); (0.33, 9.18), (0.34, 8.65)\}$$

contains clustered solutions on the true Pareto front, then $ACV(Y_4) = 3.18 \approx ACV(Y_1)$. Though the solutions of Y_4 are much more clustered than those of Y_1, Y_4 receive nearly the same indicator values than Y_1.

C. The diversity measure

The measure for diversity calculates the average distance of all pairs of solutions (see [5]):

$$\Delta = \sum_{i,j=1, i<j, i\neq j} \frac{|d_{ij} - \bar{d}|}{N} \quad \text{with } N = \binom{n}{2}, \quad (19)$$

where $d_{i,j}$ symbolizes the Euclidean distance of two solutions x_i and x_j, $\bar{d}$ is the mean of all measured distances and n is the population size.

VI. RESULTS AND DISCUSSION

In this section the simulation onset for the test runs are described. The results are further depicted and discussed.

A. Simulation onsets

The test runs are performed for different configurations. The configurations are composed of a different population size (30, 50, 70, 100, 130, 150) and the selection parameters $p_0 = 0\%, 30\%, 50\%$. These parameters have been emphasized by previous experiments. The selection parameter $p_0 = 0\%$ stands for SUS exclusively. Each multi-objective configuration is repeated 20 times until the 18th generation - for statistical reasons. The test runs are evaluated via the convergence indicator ACV and the diversity measure as introduced in the last section. ACV uses the theoretical minimal limit $(0/0/0)$ of the Pareto front as an optimal reference point. Therefore, a good performance is achieved if the ACV value is as low as possible and the diversity value is as high as possible. Boxplots are created for each configuration and for each objective of evaluation as boxplots provide a good overview of the location parameter as well as the spread. The values of ACV and diversity are scaled under the same criterion for a better graphical presentation. The figures are ordered according to the population size. The standard population size within the customized NSGA-II is 100 [2][5] (Fig. 12, Fig. 23, Fig. 24). Therefore, the results are discussed regarding an increase and a decrease of this size.

B. Experiments on the 3D-MOP

The 3D-MOP comprises the molecular objective functions NMW, MW and hydro. In general, a decrease of the population size down to 70 and 50 results in an increase of the ACV values and a decrease of the diversity values (Fig. 10, Fig. 11). This means that the convergence and the spread within the solutions is reduced caused by decreasing the population size. The ACV values decrease for a population size of 30

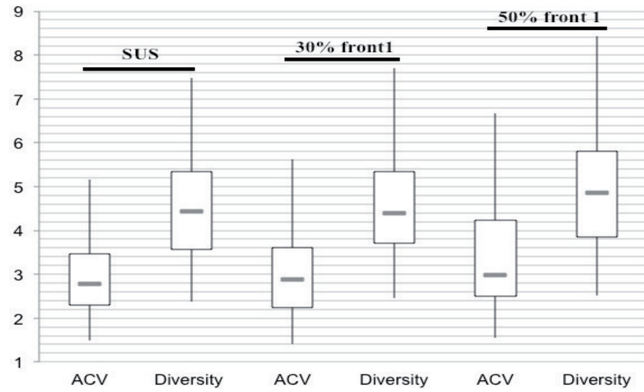


Figure 9. Population size 30

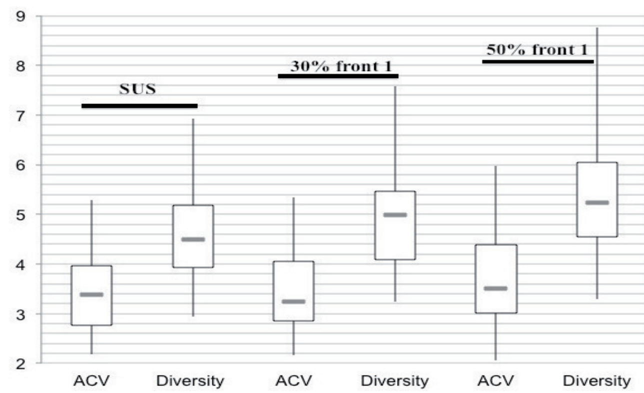


Figure 10. Population size 50

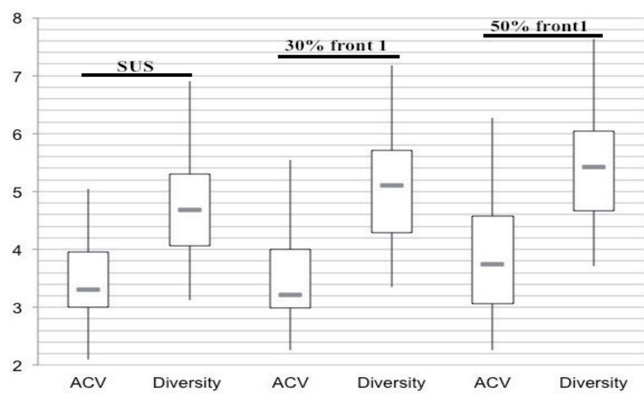


Figure 11. Population size 70

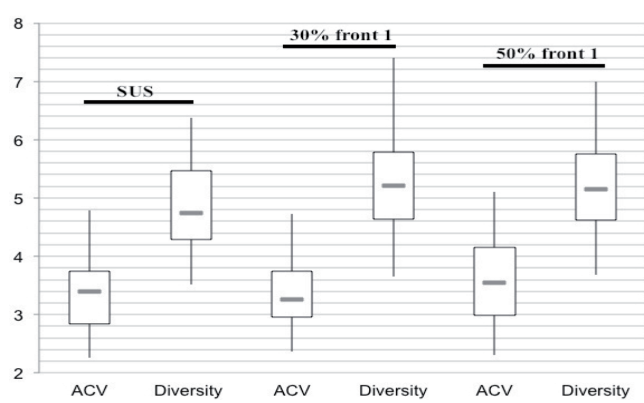


Figure 12. Population size 100

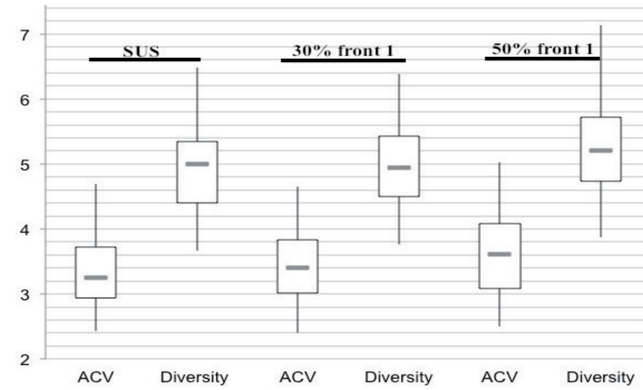


Figure 13. Population size 130

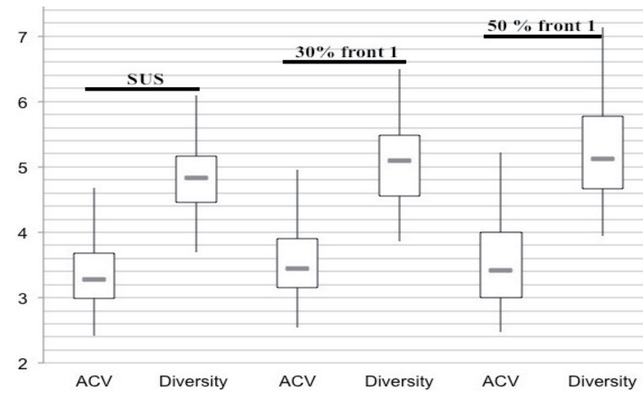


Figure 14. Population size 150

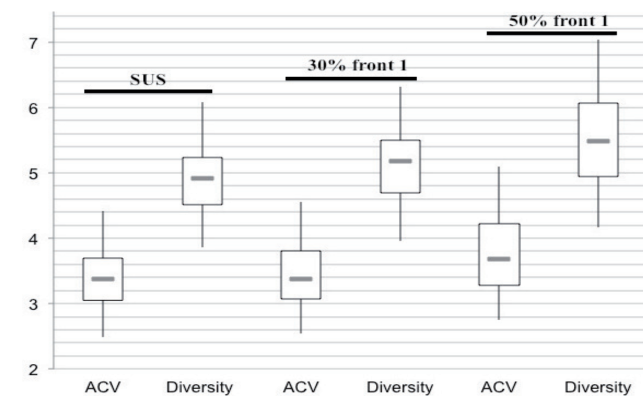


Figure 15. Population size 200

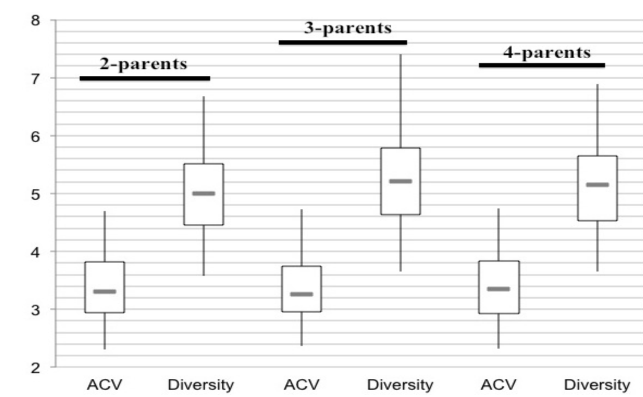


Figure 16. Multi-parent recombinations, population size 100, $p_0 = 30\%$

(Fig. 9) independent of the choice of the selection parameter. Moreover, the diversity also decreases and results in the lowest diversity among all configurations. An increase of the population size to 130 results in a decrease of ACV and in an increase of the diversity, once more independent of the selection parameter (Fig. 13). A further increase of the population size up to 150 and 200 results in a stagnation of the ACV and diversity values (Fig. 14, Fig. 15).

Further, the effect of the selection parameter is evaluated: Varying the population size from 50 to 100 (Fig. 10- Fig. 12), the ACV values are comparable for $p_0 = 0\%$ (denoted as 'SUS' in the figures) and $p_0 = 30\%$ (denoted as '30% front 1' in the figures), though the diversity improves evidently for $p_0 = 30\%$ compared to SUS. Independent of the population size, $p_0 = 50\%$ (denoted as '50% front 1' in the figures) results in a remarkable increase of the ACV values and only a slight improvement of diversity compared to SUS and $p_0 = 30\%$. For the population sizes from 130 to 200, the influence of the selection parameter is reduced (Fig. 13- Fig. 15): There is only a slight improvement to report in diversity for $p_0 = 30\%$ compared to SUS. The convergence is remarkable reduced for $p_0 = 50\%$, though the diversity is improved.

The best performance of the configurations is received with a population size from 70 to 100 and a selection parameter of 30% as the values for ACV are at most low, whereas the diversity values are at most high. At least, the performance of the configurations with a population size from 50 to 100 with $p_0 = 30\%$ are comparable in convergence and diversity with the performance of the configuration population size of 130 and SUS. Concluding, the best configuration is expectable with a population size in the range from 70 to 100 and a selection parameter of $p_0 = 30\%$.

The variation of the parent number within the recombination procedure reveals no effect on the ACV-values and therefore on the convergence (Fig. 16). A very slight increase of the diversity values is achieved by an increase of two parents to three parents. A further increase of the parent number results in a slight decrease of the diversity values. The variation effect is tested for the previously detected optimal algorithm settings of population size and selection parameters.

Regarding the questions presented in the introduction we conclude that an increase of the population size does not result in better performance. The customized NSGA-II provides good performance regarding convergence and diversity within a limited range of population size for the presented three-dimensional minimization problem. Empirically, there is no interdependence between population size and selection: The choice of $p_0 = 30\%$ usually results in the best performance independent of the population size. Therefore, it is not possible to speed up the convergence by increasing or decreasing of the population size and a suitable adaption of the selection parameter.

C. Experiments on the 4D-MOP

The 4D-MOP comprises the molecular objective functions NMW, MW, hydro and InstInd. The results are once more

Population size 30

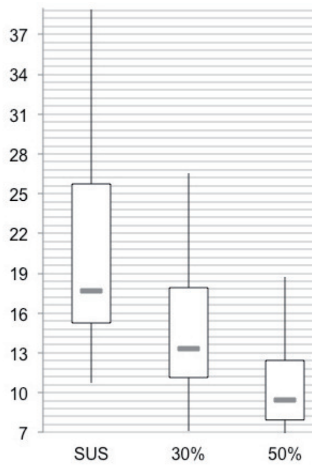


Figure 17. ACV

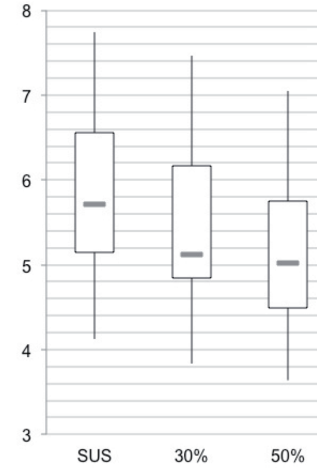


Figure 18. Diversity

Population size 50

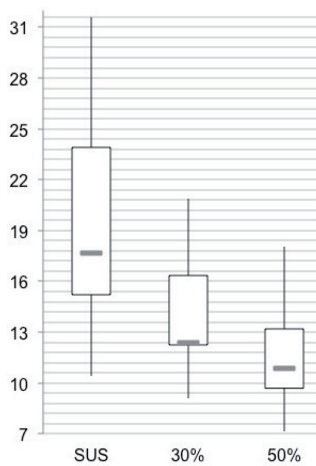


Figure 19. ACV

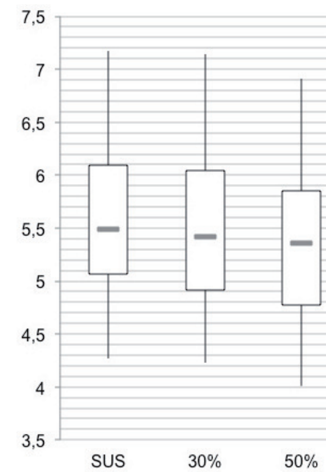


Figure 20. Diversity

Population size 70

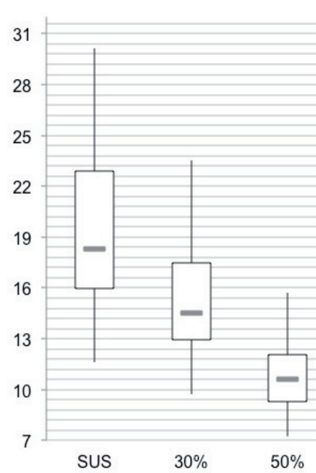


Figure 21. ACV

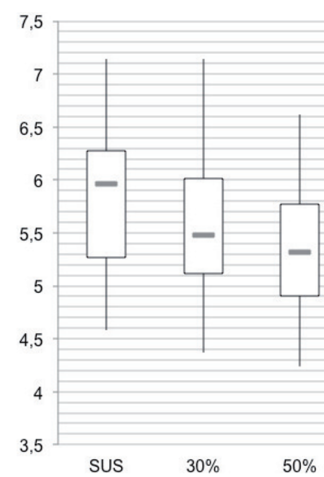


Figure 22. Diversity

Population size 100

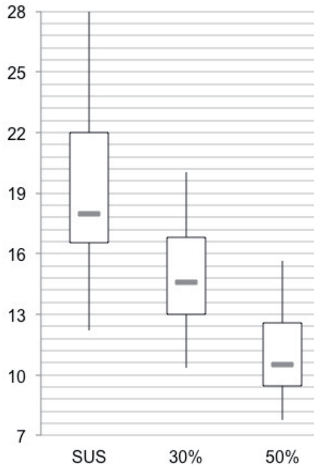


Figure 23. ACV

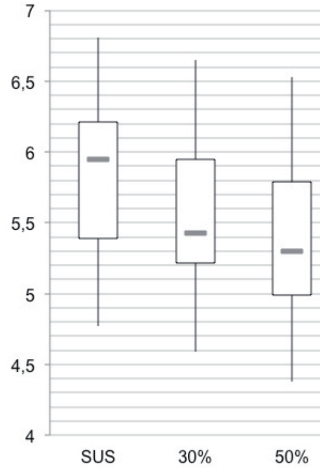


Figure 24. Diversity

Population size 130

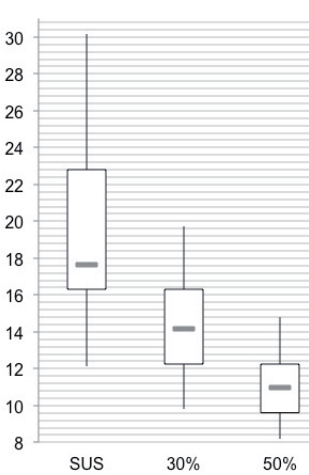


Figure 25. ACV

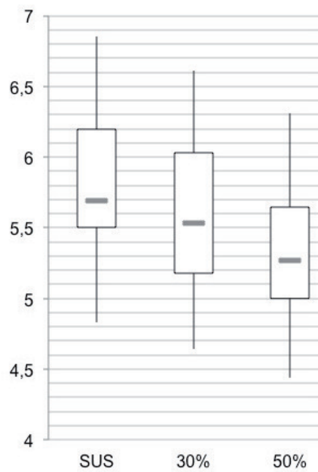


Figure 26. Diversity

Population size 150 and 200

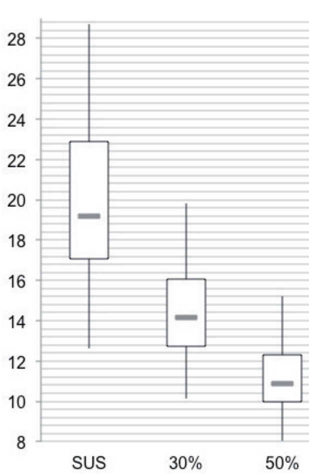


Figure 27. ACV

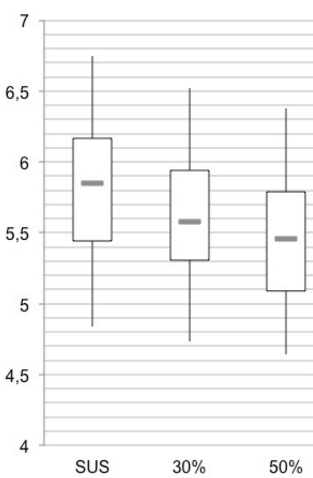


Figure 28. Diversity

Multi-parent recombination

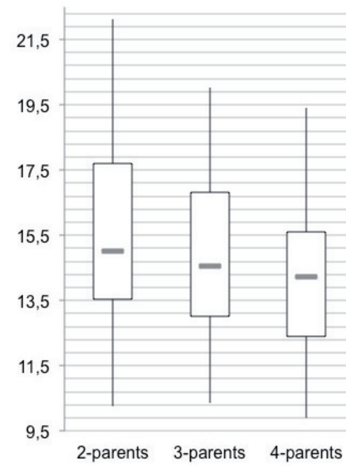


Figure 29. ACV

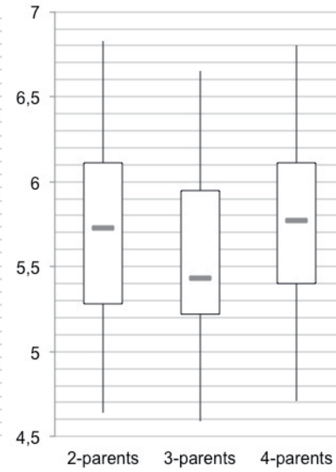


Figure 30. Diversity

discussed regarding an increase or decrease of the standard population size of 100 (Fig. 23, Fig. 24)). In general, a decrease of the population size from 100 to 70 (Fig. 21, Fig. 22), 50 (Fig. 19, Fig. 20) and 30 (Fig. 17, Fig. 18) results in an increase of the range for the ACV as well as the diversity values, which indicates an increasing spread of the indicator values. The tendency of the indicator spread lies in direction of higher values in the case of the convergence (Fig. 17, Fig. 19, Fig. 21). Otherwise, the tendency of the indicator spread for the diversity is more in direction of lower values (Fig. 18, Fig. 20, Fig. 22). This indicates a global convergence reduction with a decrease of the diversity at the same time. This effect is stronger for the selection parameter $p_0 = 0\%$ and is reduced with the increase of this parameter to 30% and 50%.

The increase of the selection parameter results clear decrease of the indicator values in general for all population sizes. This observation is quite different to the results of the 3D-MOP experiments. This is explained by the previously performed landscape analysis: The landscape of the 4D-MOP provides about 50% more optimal solutions and is therefore of a higher optima density than the 3D-MOP. Consequently, the distance between the optimal solutions is reduced. Hence, the higher the selection probability for selecting solutions from the first front the lower are the ACV values and the lower are the diversity values.

An increase of the population size from 100 up to 200 (Fig. 25, 26, 27, 28) results in a stagnation of the ACV values independent of the selection parameter. The same holds for the diversity values and the selection parameter $p_0 = 0\%$. An increase of the selection parameter results in a slight increase of the means as well as an increase of the indicator spread for a population size of 150 and 200 (Fig. 28). The results of population 150 and 200 are depicted in one figure, as there is no visible difference within the indicator values.

The optimal configuration regarding the convergence or ACV values is achieved with a population size of 100. The optimal configuration for the diversity of the solutions is achieved with a population size of 70 followed directly by the configuration with a population size of 100. Concluding, the best configurations in general are expectable with a population size in range of 70 to 100 and a selection parameter of $p_0 = 30\%$.

The results of the multi-parent recombination is quite different in the case of the 4D-MOP: An increase of the parent number results in a continuous decrease of the ACV values and therefore in better performance regarding the convergence. In the case of the diversity values, the increase of the parent number from 2 to 3 reveals a clear decrease of the diversity values. A further increase of the parent number to 4 achieves slight better diversity values as the configuration with 2-parent recombination. Therefore, the best NSGA-II performance is expected with 4-parent recombination. This confirms the observations of Eiben as presented in the introduction that the optimal number of parents is problem depending.

Compared to the results of the 3D-MOP and the questions raised in the introduction, an increase of the population size does not results in an increase of the performance like in the case of 3D-MOP. The customized NSGA-II provides a good performance with regard to convergence and diversity within a limited range of population size for 3D-MOP as well as 4D-MOP. The selection parameter $p_0 = 30\%$ is a suitable choice for a good balance between at most low ACV values and at most high diversity values. Therefore, the optimal algorithm settings are equal for 3D-MOP as well as 4D-MOP.

D. Discussion of the results

The interdependence of the population size and the selection parameter in this customized NSGA-II as well as the influence of multi-parent recombination is exemplary examined on a generic three- and four-dimensional biochemical minimization problem and the results presented above are discussed according to the five questions raised in the introduction:

The first question is aimed at the influence of large populations on the convergence speed. Early convergence as a main goal of our moGA is defeated since an increase of the population size results in higher speed of convergence. The experiments show that the optimal population size regarding convergence and diversity is in a limited range from 70 to 100 for the three- as well as the four-dimensional optimization problem. An increase of the population size above 100 results in a stagnation of the convergence behavior and the diversity for the three-dimensional optimization problem. Furthermore, a population size lower than 50 does not provide a convincing diversity within the solutions. In the case of the four-dimensional optimization problem, an increase of the population size above 100 also results in a stagnation of the convergence behavior and no significant improvement of the diversity. A decrease of the population size below 70 results in worse convergence and diversity performance.

Our second question is focused on the impact of the population size and the selection parameter. A configuration rule for the selection parameter depending on the population size is necessary in the case of a large interdependence of both. However, the experiments of the three- dimensional optimization problem do not reveal an interdependence of the population size and the selection parameter. Though, the diversity of the configurations with a population size from 50 to 100 is remarkably improved with a selection parameter of 30% compared to $p_0 = 0\%$ (SUS). Further, higher values for p_0 are not advisable as the speed of convergence is reduced. In the case of the four-dimensional optimization problem, in increase of the selection probability above 30% is not advisable as this results in a significant decrease of the diversity independent of the population size.

The third question asks for a range of the population size providing the best performance: This range is fixed to a population size from 70 to 100 based on the evaluation of the experiments. More precisely, the optimal performance for the three- and the four-dimensional optimization problem is achieved within the same range of population size and the same parameter settings. This allows the hypothesis that the optimal algorithm settings are independent of at least three and four dimensions and the customized NSGA-II is an effective and robust tool for biochemical optimization.

The fourth question refers to the influence of the variation of the number of parents within the recombination on the algorithm performance. Three different numbers of parents for recombination are tested. The experiments on the three-dimensional optimization problem reveal no effect on the convergence behavior and the diversity for 2-, 3- and 4- parent recombination, whereas an increase of the parent number results in an improvement of the convergence behavior for the four-dimensional optimization problem. The highest diversity tendency is achieved for four parents. Therefore, an increase of the standard setting of 3 parents to 4 is advisable. A reason for these observations is challenging and the optimal number has to be empirically verified for each optimization problem.

The fifth question refers to the generalization of the results and algorithm settings on higher dimensional optimization problems. As mentioned above, the best performance is achieved for the same algorithm settings in the case of the three-dimensional and four-dimensional optimization problem. This confirms the hypothesis that the performance results are transferable on high-dimensional optimization problems. Only the number of parents within the recombination procedure is challenging for each optimization problem.

VII. Conclusion and Future Work

The presented customized NSGA-II provides a reliable good performance according to the convergence and diversity for a three- and four-dimensional biochemical minimization problem. This good performance is achieved with the same optimal settings, though the three-dimensional problem investigated here is more challenging than the investigated four-dimensional one due to the higher front diversity. This allows the hypothesis that this customized NSGA-II is an efficient and robust genetic algorithm, which potentially provides a high performance for a wider range of similar molecular problem classes with the property of early convergence.

For future work, we currently work on a selection strategy based on ACV indicator for ongoing improvements. Furthermore, an analysis concept is challenging to gain a deeper insight into molecular MOP or more general real-valued MOP and is in the focus of our research.

REFERENCES

[1] S. Rosenthal and M. Borschbach, "Impact of Population Size and Selection within a Customized NSGA-II for Biochemical Optimization Assessed on the Basis of the Average Cuboid Volume Indicator," BIOTECHNO 2014: The Sixth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, IARIA 2014, pp. 1-7.

[2] S. Rosenthal, N. El-Sourani, and M. Borschbach, "Introduction of a Mutation Specific Fast Non-dominated Sorting GA Evolved for Biochemical Optimization," SEAL 2012, LNCS 7673, 2012, pp. 158-167.

[3] D. J. Craik, D. P. Fairlie, S. Liras, and D. Price, "The Future of Peptide-based Drugs," Chemical Biology & Drug Design, 81(1), 2013, pp. 136-147.

[4] N. Röckendorf, M. Borschbach, and A. Frey, "Molecular Evolution of Peptide Ligands with Custom-tailored Characteristics," PLOS Comput Biol 8(12), 2012

[5] S. Rosenthal, N. El-Sourani, and M. Borschbach, "Impact of Different Recombination Methods in a Mutation-Specific MOEA for a Biochemical Application," L. Vanneschi, W. S. Bush, and M. Giacobini (Eds.): EvoBIO 2013, LNCS 7833, 2013, pp. 188-199.

[6] S. Rosenthal and M. Borschbach, "A Benchmark on the Interaction of Basic Variation Operators in Multi-Objective Peptide Design evaluated by a Three Dimensional Diversity Metric and a Minimized Hypervolume," M. Emmerich et al. (eds.): EVOLVE - A Bridge between Probability, Set Oriented Numerics and Evolutionary Computation IV, 2013, pp. 139-153.

[7] J. T. Alander, "On Optimal Population Size of Genetic Algorithms," in Proceedings of the IEEE Computer Systems and Software Engineering, 1992, pp. 65-69.

[8] V. K. Koumousis and C. P. Katsaras, "A Saw-Tooth Genetic Algorithm Combining the Effects of Variable Population Size and Reinitialization to Enhance Performance," IEEE Transactions on Evolutionary Computation, vol. 10, no. 1, 2006, pp. 19-28.

[9] T.-L. Yu, K. Sastry, D. E. Goldberg, and M. Pelikan, "Population sizing for entropy-based model building in genetic algorithms," Illinois Genetic Algorithms Laboratory, University of Illinois, Tech. Rep., 2006.

[10] T. Bäck, A. Eiben, and V. der. Vaart, "An empirical study on GAs without parameters," in Proceedings of the 6th International Conference on Parallel Problem Solving from Nature, 2000, pp. 315-324.

[11] Z. M. Jaroslaw Arabas and J. Mulawka, "GAVaPSa genetic algorithm with varying population size," in Proceedings of the IEEE International Conference on Evolutionary Computation, 1995, pp. 73-78.

[12] A. E. Eiben, M. C. Schut, and A. R. Wilde, "Is Self-Adaption of Selection Pressure and Population Size Possible? a Case Study," in Parallel Problem Solving from Nature - PPSN IX, vol. 4193, 2006, pp. 900-909.

[13] A.E. Eiben, P.-E. Raue, and Zs. Ruttkay, "Genetic Algorithms with Multi-parent Recombination," Proceedings of the Parallel Problem Solving from Nature III, 1994, pp.78-87.

[14] I. Ono and S. Kobayashi, "A Real-coded Genetic Algorithm for Function Optimization Using Unimodal Normal Distribution Crossover," Proceedings of the Seventh International Conference on Genetic Algorithm, 1997, pp. 246-253.

[15] S. Tsutsui and A. Ghosh, "A Study on the Effect of Multi-parent Recombination in Real Coded Genetic Algorithms," Proceedings of the 1998 IEEE ICEC, 1998, pp. 828-833.

[16] G. Grosan, M. Oltean, and D. Dumitrescu, "Performance Metrics for Multiobjective Evolutionary Algorithms," Proceedings of Conference on Applied and Industrial Mathematics (CAIM), 2003

[17] BioJava: CookBook, URL: http://www.biojava.org/wiki/BioJava/ [retrieved: December, 2013].

[18] S. Henikoff and J.G. Henikoff, "Amino Acid Substitution Matrices from Protein Blocks," Proc. Natl. Acad. Sci. USA, vol. 89(22), 1992, pp. 10915-10919.

[19] T. Sovany, K. Papos, P.Jr. Kasa, I. Ilic, S. Srcic, and K. Pintye-Hodi, "Application of Physiochemical Properties and Process Parameters in the Development of a Neural Network Model for Prediction of Tablet Characteristics," AAPS PharmSciTech, vol. 14(2), 2013, pp. 511-516.

[20] D. Heider, J. Appelmann, T. Bayro, W. Dreckmann, A. Held, and J.. Winkler, "A Computational Approach for the Identification of Small GTPases based on Preprocessed Amino Acid Sequences," in Technol. Cancer Res. Treat 8, 2009, pp. 333-341.

[21] T. P. Hopp and K. R. Woods, "A computer program for predicting protein antigenic determinants," Mol Immunol, 20(4), 1983, pp. 483-489.

[22] S. Needleman and C. Wunsch, "A General Method Application to the Research for Similarities in the Amino Acid Sequence of Two Proteins," Journal of Molecular Biology, vol. 48(3), 1970, pp. 443-453.

[23] Java API for Genetic Algorithm (JAGA), URL: www.jaga.org/ [retrieved: January, 2014].

[24] Metaheuristic Algorithms in Java (jMetal), URL: www.jmetal.sourceforge.net/ [retrieved: January, 2014].

[25] Java-based Evolutionary Computation Research System (ECJ), URL: www.cs.gmu.edu/∼edab/projects/ecj/ [retrieved: January, 2014]

[26] MOEA framework, URL: www.moeaframework.org/ [retrieved: August, 2014].

[27] Evolutionary Algorithms workbench (EvA2), URL: www.ra.cs.uni-tuebingen.de/software/EvA2/introduction.html/ [retrieved: January, 2014]

[28] Modular framework for Meta-Heuristic Optimization (OPT4J), URL: http://opt4j.sourceforge.net].

[29] P. Merz and B. Freisleben, "Fitness Landscape Analysis and MEmetic Algorithms for the Quadratic Assignment Problem," IEEE Transactions on Evolutionary Computation, vol. 4 (4), 2000, pp.337-352.

[30] P. Stadler, "Fitness Landscape," In Lecture Notes in Physics, vol. 585, 2002.

[31] G. Merkuryeva and V. Bolshakovs, "Benchmark Fitness Lands-cape Analysis," International Journal of Simulation Systems, Science and Technology 12, 2, 201, pp. 3845.

[32] E. Weinberg, "Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference," Biological Cybernetics 63, 1990, pp. 325336.

[33] T. Jones and S. Forrest, "Fitness Distance Correlation as a Measure of Problem Difficulty for Genetic Algorithm," In Proceedings of the 6th International Conference on Genetic Algorithms,1995, pp. 184192.

[34] V.K. Vassilev, T.C. Fogarty, and J.F. Miller, "Information Characteristics and the Structure of Landscapes," Evolutionary Computation, vol. 8(1), 2000, pp.31-60.

[35] A. Leier and W. Banzhaf, "Exploring the Search Space of Quantum Programs," In Proceedings of the 2003 Congress on Evolutionary Computation IEEE Press, vol. 1, 2003, pp.170-177.

[36] B.V. Lee, "Analysing Molecular Landscapes using Random Walk and Information Theory," Masterthesis, LIACS, University of Leiden, 2009.

[37] E. Zitzler and L. Thiele, "Multiobjective Optimization using Evolutionary Algorithms - a Comparative Case Study," in A. E. Eiben, T. Bäck, M. Schoenauer, and H. P. Schwefel (EDS.), Fifth International Conference on Parallel Problem Solving form Nature (PPSN-V), 1998, pp. 292-301.

[38] D. A. Van Veldhuizen and G. B. Lamont, "Multiobjective Evolutionary Algorithm Test," in Proceedings of the 1999 ACM Symposium on Applied Computing, San Antonio, Texas, 1999, pp. 351-357.

[39] K. Deb and S. Jain, "Running performance metrics for Evolutionary Multiobjective Optimization," Kan GAL Report No. 2002004, Kanpur Genetic Algorithms Laboratory, Indian Institute of Technology Kanpur, 2002.

[40] O. Schütze, X. Esquivel, A. Lara, and C. A. Coello Coello, "Using the Averaged Hausdorff Distance as a performance measure in evolutionary multiobjective optimization," IEEE Transactions on Evolutionary Computation, vol. 16(4), 2012, pp. 504-522.

[41] C. A. Coello Coello and N. Cruz Cortis, "Solving Multiobjective Optimization Problems using an Aritifical Immune System. Genetic," Programming Evolvable Mach., vol. 6 (2), 2005, pp. 163-190.

AMIDE – Automatic Molecular Inverse Docking Engine for Large-Scale Protein Targets Identification

R. Vasseur[1, 4], S. Baud[1, 2], L. A. Steffenel[2, 3], X. Vigouroux[4], L. Martiny[1], M. Krajecki[3], M. Dauchez[1, 2]
1-MEDyC, UMR CNRS 7369, University of Reims (URCA), Reims, FRANCE
2-P3M, Multi-scale Molecular Modelling Plateform, University of Reims (URCA), Reims, FRANCE
3-SYSCOM Team, CReSTIC EA3804, University of Reims (URCA), Reims, FRANCE
4-Center for Excellence in Parallel Programming, Bull SAS, Echirolles, FRANCE

romain.vasseur@etudiant.univ-reims.fr
stephanie.baud@univ-reims.fr
luiz-angelo.steffenel@univ-reims.fr
xavier.vigouroux@bull.net
laurent.martiny@univ-reims.fr
michael.krajecki@univ-reims.fr
manuel.dauchez@univ-reims.fr

Abstract **- Molecular docking are widely used computational technics that allows studying structure-based interactions complexes between biological objects at the molecular scale. We developed AMIDE, a framework that allows performing inverse virtual screening to carry out a large-scale chemical ligand docking over a large dataset of proteins. Molecular inverse docking has several applications in the field of drug research like identifying potential side effects of existing or new drugs, or to help choosing the less harmful treatment for a disease. We developed different strategies to distribute the docking procedure, as a way to efficiently exploit the computational performance of multi-core and multi-machine (cluster) environments. This tool has been validated on well-known experimental structures through 24 protein-ligand complexes extracted from the Kellenberger's set. Its ability to reproduce experimentally determined structures and binding affinities highlighted that AMIDE allows performing better exploration than existing blind docking methods.**

Keywords - Protein-Ligand docking; inverse docking; ranking methods; distributed computations; HPC experiments.

I. INTRODUCTION

In the field of drug discovery or drug design, molecular docking is focused on protein-ligand complexes to study how the chemical ligand that is a drug will bind to the target protein receptor. The prediction of the binding mode of a ligand into a protein target cavity, the structure of the complex and the estimation of the binding affinity between both partners is crucial to find new therapeutic compounds to cure life threatening diseases. For this, we develop a framework to perform accurate inverse docking and targets identification on multi-cores supercomputers [1]. Molecular docking represents a virtual alternative to costly and time-consuming systematic wet biological experiments such as High Throughput Screening (HTS) processes and/or Nuclear Magnetic Resonance (NMR)-based screening. Then, it is called virtual ligand screening (VLS) or *in silico* ligand screening and has become a method of choice for rational drug design, hits identification and hits to leads optimization [2][3][4]. At present, several applications are available for VLS, such as for instance PLANTS [5], DOCK Blaster [6], GOLD [7], AutoDock [8][9], FlexX [10], Glide HTVS [11], ICM [12] and LigMatch [13].

VLS tries to predict probable bindings of a huge number of ligands (to the order of millions) to a unique target receptor and is linked to multiple ligand dockings. Such methods require knowledge of the three dimensional structure of a receptor alone or associated with its experimental ligand. Many chemical databases and libraries provide millions of compounds, among which we can cite some public and free ones such as the PDBbind database [14] or the ZINC database [15], some with fees access as the Cambridge Structural Database [16] and several private pharmaceutical collections. Protein structures are obtained from the Research Collaboratory for Structural Biology (RCSB) Protein Data Bank (PDB) [17], an open source database that collects all public experimental data on tridimensional biological structures. For a large number of proteins, X-ray crystallography and NMR provide experimental structural data. In August 2014, the number of protein structures publicly available in the Protein Data Bank is over 95,000 ; the number of nucleic acids structures is about 2,700 and the number of structures of nucleic acids-protein complexes is about 5,000. The total number of structures available in the PDB increased on average by 6,500 structures per year during the last decade [17]. Yet, it is important to highlight that these statistics do not include the large number of proprietary structures as described above held by pharmaceutical companies that dispose of their own private structures databanks. To use non-resolved structures for a protein of interest, 3D prediction models can be built *de novo* [18] or based on partially known fragments by homology modelling [19][20].

The purpose of the current work is to develop an automatic virtual screening tool that allows performing large-scale structure-based inverse docking. The main idea of this approach is to perform molecular docking of a chemical ligand over a large dataset of proteins. In the fields of drug design and structural biology, inverse docking methodology would find several applications, such as searching for additional uses of new drugs, by searching for interactions with protein groups outside the usual research field. Inverse docking can also be used to identify potential side effects of new drugs or to help choosing the less harmful treatment for a disease. Several problems arise when performing inverse docking, as we are no longer targeting a single protein but thousands. One of the main concerns is the computation time, which represents a clear obstacle when dealing with a large number of different proteins. For instance, even with the use of multicore processing we shall not restrain the inverse docking to a single computer but rely on multiple computational environments such as clusters and grids. In order to effectively use wide computational resources, however, we cannot simply launch a batch of docking computations but we must rethink docking in terms of task distribution, of pipelining, as well as load balance and fault tolerance. Recently, the literature reports several implementations to performed massively parallel ligand screening. These implementations rely on Message Passing Interface (MPI or OpenMP) only [21] or combined with multi-threading programming [22], with cloud computing to treat Full Flexible Receptors (FFR) models [23][24] or even with FPGAs or GPUs accelerators [25]. AMIDE was preferentially developed for and with the AutoDock4.2 software [26] to keep a fine-grain control of algorithm parameters and to generate multiple grids. In fact, whereas AutoDock Vina [27] is a docking software that supports multithreading natively, only “black box” blind dockings are possible with Vina.

In this work, docking simulations were performed with AutoDock and we developed a set of Python scripts to automatize the docking process. We also developed a Python framework embedding different strategies to distribute the docking procedure, as a way to efficiently exploit the computational performance of multi-core and multi-machine (cluster) environments. Data presented in this paper result from the testing described hereafter. In addition, the experiments compare the docked poses obtained with our tool for a set of chemical ligands on their experimental target against the determined structure of the complex obtained by X-ray crystallography. The rest of this paper is structured as follows: Section II presents the methods employed and the different strategies we developed to decompose the docking computation. Section III presents the results of the computation performances and the monitoring of the AutoDock program over HPC architectures. Section IV presents the description of the test set and methods we used to generate and to rank the docking poses. In Section V, docking poses given by these strategies are compared to the native ones (X-ray structures). Finally, all results are afterward discussed in Section VI.

II. METHODS

A. Classical method

Molecular protein-ligand docking experiment consists in searching and finding a protein surface area able to host the ligand. For some particular protein targets, ligand active binding sites on the protein are widely known. In these cases, the user can limit the docking search around a protein location in order to decrease the computation time without affecting the quality of the docking results. In other cases, without any knowledge, the user must consider as exhaustively as possible the entire protein surface i.e., the whole protein “volume”. This way to do is called blind docking (BD). BD was introduced to detect possible binding sites and ligands binding modes by scanning the entire surface of protein targets. This represents the “naïve” approach to dock ligands on unknown targets but is barely parallelizable. In fact, for each complex the AutoDock software will launch only one infrangible docking task with the whole volume to explore. Depending on the shape of each receptor, a large number of runs/generations is required in order to systematically cover the entire protein surface and consequently obtain good docking results. Some success cases with AutoDock [28][29] or with others programs [30][31]were reported in the literature.

B. Parallel Decomposition

To obtain a better implication of the computational resources, we must imperatively improve task parallelism when conducting large-scale inverse docking. If decomposing a docking job in parallel task may trigger a better utilization of the computational resources through pipelining and load balance, it also contributes to the fault tolerance aspects since only a small segment of the execution is lost in the case of a computer crash or execution failure. For this, we developed two methods to decompose the docking computation and improve tasks distribution and fault tolerance.

The first strategy to distribute docking computations aims at the reduction of the exploring space through the multiplication of the number of small 3D boxes. For instance, the “single grid” used in a blind docking experiment and describing the whole protein volume is arbitrary split into several grids. Each grid is a sub-volume of points covering a piece of the protein. Assuming a regular decomposition, we define a geometrical Arbitrary Cutting method (AC) as a 12-part decomposition scheme, i.e., 3x2x2 (3 on the longest axis of the protein). We also tested multiple space cuttings of the whole-space to find a suitable decomposition ratio in prior experiment and the 12-part scheme showed better quality docking results than other geometrical cuttings into multiple subspaces as n-part schemes where n = 8 (2x2x2), 27 (3x3x3) or 64 (4x4x4) [32]. Indeed, a large number of 3D boxes may improve parallelism but the number of subspaces is also dependent on the size and shape of the protein. So, having too small 3D boxes may limit the movement of the ligand and impact the success of the ligand docking. Hence, the choice of decomposition must be carefully tuned and the number of generated chunks must be precisely balanced. Moreover, the several subspaces are overlapping each other to explore the entire protein surface and overcome the presence of the 3D boxes edges. Indeed, one of the constraints imposed by

AutoDock is that the ligand cannot bind outside of the box. The overlapping is inherently dependent on the ligand size, so in our experiments we set two ranges for the partial overlapping: a third of the juxtaposed boxes if the ligand size is inferior to it, or the size of the ligand if the ligand is larger than that.

This decomposition strategy is simple to implement and the subspace grids can be easily generated from the coordinates of the protein. By multiplying the number of 3D boxes we can deploy the docking over different processors in order to be computed in parallel. One drawback of this strategy, however, is that it does not check the protein surface for cavities (which are potential docking sites), and may therefore “cut” right in the middle of a potential cavity, making it less interesting. Another drawback of this method is that only ligands inside the grid can be evaluated. Indeed, any atom of the ligand outside the 3D box will not be treated and will eliminate the pose of the conformer during the sampling process, which may prevent the detection of potential bindings when part of the ligand crosses the boundaries of the 3D box. So, to overcome boundary problems, we also use a more rational knowledge-based method.

This second method to perform space cutting consists in predicting upstream pockets and cavities on the surface receptor with additional programs and carry out dockings only on these pockets [33][34]. For this Pocket Search method (PS), we used the Fpocket program [35] that screens pockets and cavities using a geometrical algorithm based on Voronoï tessellations. The second version of the software (Fpocket2) is compatible with a multiprocessing parallel use. Only pockets that show a long side superior to a third of the whole protein longest side and inferior to the half of the whole protein longest side are conserved in order to limit the number of generated jobs and to avoid multi-exploration of the same space. One advantage of the pocket strategy is that it refocuses the docking algorithm exploration zones only on predictive biological sites of interest (potential binding sites). As only these interesting zones are included in the docking procedure they can drastically improve the overall inverse docking performance. On the opposite side, the pocket search is a predictive method and as such it may exclude some potential zones, which should not be overpassed by the AC method described above.

C. *Proof of concept – biological relevance*

Thanks to our two decomposition strategies, we have been able to correctly replace experimental ligands in their crystallographic cavities with good free energy of binding (see IV.C). In fact, a n =12 and 50 runs of docking, allows to locate the ligand in the crystallographic cavity with an energy of -9.95 kcal/mol and a RMSD of 1,74 Å for the X23/XIAP ligand-protein complex (pdb_ID: 3CM2). In the same way, the pocket search strategy and 50 runs of docking gives the same experimental location for the ligand with the same order of values (ΔG = -10.11 kcal.mol-1; RMSD = 1.86 Å). Even they succeed to in replacing the ligand in the experimental cavity, the other cuttings n = 8, 27 or 64, do not give satisfying results in terms of binding energy and RMSD value. In fact, n = 27 and n = 64 seem to generate too small boxes whose the size is not adapted to a standard protein cavity. The 12-part decomposition scheme (n = 12) is a cutting method that may let the ligand rotate without constrains in the overlapped region and therefore lead to an exhaustive search of the binding site.

III. PERFORMANCES

As stated in the previous sections, we designed alternative decomposition methods to improve the potential parallelism and fault tolerance when performing an inverse docking. Indeed, when analyzing the performance of the different decomposition strategies, a number of parameters may impact the overall results, as for example the number of boxes or the overlapping of these boxes. As our main goal is to obtain a decomposition technique with a precision level at least as good as the precision of a "monolithic" blind docking, we chose to privilege precision and distribution, at the expense of raw performance.

Today, most computers CPUs are composed by several computing cores, an architecture that allows parallel executions inside a single node. Currently AutoDock does not explore multicore and therefore the execution of a blind docking on a multicore machine is far from being efficient. On the opposite, our decomposition methods allow a better resource usage, reducing the overall execution time, as indicated in Table I. In addition, the use of decomposition methods improve fault tolerance as a crash during the execution (for example, after 1 hour) may be resumed from the boxes that have not completed yet, contrarily to the blind docking that must restart from the beginning.

TABLE I. PERFORMANCES OF SAMPLING METHODS

Ligand	Target	Method	Cores	Tasks	Time
X23	3CM2	BD	1	1	4h13’
X23	3CM2	AC/PS	4	12+1	3h00’
X23	3CM2	AC/PS	2 x 4 (HT)	12+1	1h53’

In order to evaluate the scalability of the decomposing methods when performing in multicore machines, we performed additional experiments to evaluate the overhead/speedup on different families of processors. For instance, these experiments were run on different HPC architectures on Bull’s CEPP clusters. Binaries were run on bullx B510 blades nodes with Intel Xeon processors: E5-2680 v1 (Sandy Bridge) 2x8 2.7 GHz cores and E5-2695 v2 (Ivy Bridge) 2x12 2.4GHz cores interconnected by InfiniBand 4x QDR. All docking computations are performed with the AutoDock4.2 software. The Linux system binary and recompiled sources are available on the MG Lab website (http://autodock.scripps.edu/downloads). Bull’s Center of Excellence in Parallel Programming

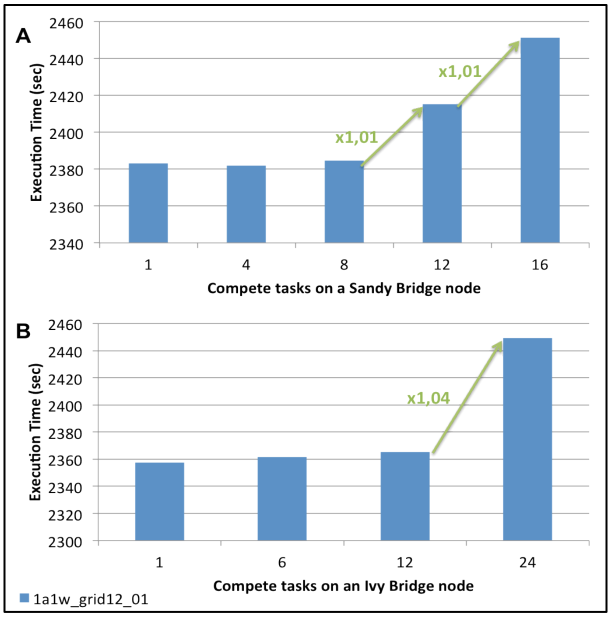


Figure 1. Performances of load tasks alternatively bound on each socket of a Sandy Bridge node (A) and an Ivy Bridge node (B).

expertise helped to optimize the parallel execution on multicores supercomputers. To obtain regular computation times to compare, we have to use a fixed single seed to initiate the conformational sampling to overcome the intrinsic stochastic effect of the AutoDock search algorithm. In the same manner, always the same sub-volume box AC-1 of n = 12 cutting is computed to obtain a common reference time value for all tests.

On Fig. 1 sockets were alternatively bound with docking tasks until one node was full. While the load occupy less than half core, there is no performance change but then we can observe a performance loss to reach performance degradation inferior to 3% on Sandy Bridge chips and inferior to 4% on Ivy Bridge chips. In spite of frequency difference between the two node types, times of execution are similar on both types. If we decrease core frequencies, we also decrease the memory bandwidth, so the execution time highly increases of one third at 1.8 GHz and is doubled at 1.4 GHz on Ivy Bridge cores. On hyper threaded Sandy Bridge cores, time of execution is highly degraded (x 1.8) for two threads on one core but allows gaining 10% on a machine throughput. The different experiments listed above demonstrate that a multicore code can achieve a good parallel performance with a good speedup on a single machine, as long as the number of tasks corresponds to the number of available cores.

Our final aim is to explore a much larger number of proteins; we need to associate resources from several machines, therefore deploying the docking tasks over a larger set of machines like a cluster or even a cloud environment. To coordinate the deployment of tasks over several machines, we developed a set of Python scripts to automate all the steps involved on inverse docking preparation and execution. Each docking experiment requires specific input files (e.g. grid coordinates files) and parameters. The generation of the required files constitutes pre-processing steps. Among these steps we can cite (a) downloading and acquisition of PDB files, (b) preparation of PDB files of the target structures, (c) extraction of coordinates for the grid creation, (d) grids decomposition. Pre-processing steps (a) to (d) involve data parsing and files manipulation. Depending on the decomposition strategy, step (d) creates one or several grids corresponding to the 3D boxes for each technique. In the case of the pocket approach, it creates 3D boxes only around the cavities identified by the Fpocket software. Because this step involves several computations (according to the number of 3D boxes), it represents the first parallel step in our implementation.

Distributed computation steps are (e) grid computation and (f) docking execution. Step (e) is managed by AutoGrid, a tool from the AutoDock suite that creates the affinity grids used on docking to evaluate binding energies. The parallel execution of step (e) is obtained through the use of a server-worker queue in a task-stealing strategy, where the master feeds the task IDs to the queue and the workers subsequently get a task from it. If no more tasks are available in the queue (they were all consumed and are being computed), a grid worker is authorized to become a docking worker and start the next step.

A similar queuing mechanism is set to the execution of step (f). However, for a better efficiency, the queue is not fed by the master but directly from the grid workers, i.e., as soon as a grid worker has prepared its task, it passes the task ID to the docking queue, which should be eventually consumed by the arriving docking workers. The Fig. 2 illustrates the processing flow of these two steps.

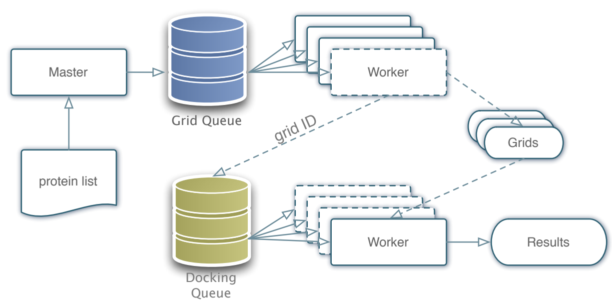


Figure 2. Queuing structure for parallel execution

IV. Comparison with Experimental Data

A. Preparation of the Test Set

The test set used in this study is constructed from the Rognan's group [36] set of 100 protein-ligand complexes. In order to be able to perform accurate High Definition (HD) docking only proteins structures with a long side inferior to 60 Angstroms are conserved.

Twenty-four complexes have passed this process and are included in the final test set (see TABLE II). Molecular weights of ligand molecules range from 114 to 659 Daltons, number of atoms in the ligand range from 10 to 52 and number of rotatable single bonds (rotors) in ligand molecules range from 0 to 23. All ligands molecules bind to their target protein non-covalently. Structures files and coordinates of all the complexes are downloaded from the Structural Chemogenomics Group website [36]. For the convenience of computation, each complex file was split into a protein molecule file in PDB format and a ligand molecule file, which was saved in Mol2 format. All preparation settings are available in the work from Kellenberger *et al.* [37]. The program automatically generates all docking parameters files and each complex is then subjected to an exhaustive conformational sampling procedure with AutoDock.

B. Conformationnal Sampling Procedure

The AutoDock program (version 4.2) is used to generate an ensemble of docked conformations for each ligand molecule. This program utilizes a Lamarckian Genetic Algorithm (LGA) for conformational sampling [38]. Each LGA run outputs a single docked conformation as a final result. For the AC method and the PS method 50 individual LGA runs are performed to generate 50 docked conformations for each ligand. All AutoDock docking experiments were performed with the default parameters of the Lamarckian algorithm for initial population size (*ga_pop_size = 150*), maximal number of energy evaluation (*ga_num_evals = 2500000*) and maximal number of generations (*ga_num_generations = 27000).* The protein structure is kept fixed during docking.

C. Ranking the Best Ligand Pose

AutoDock needs to compute an affinity grid for each atomic type to pre-evaluate the binding energy. The affinity grid is contained in a 3D box that frames the protein surface. The binding energy is evaluated with a tri-linear interpolation of the eight-grid points affinity value surrounding each atom of the ligand. For the scoring step, computation time will only depend of the number of atoms in the ligand and will be independent of the protein volume. The free energy of binding ΔG is computed with the AutoDock4 scoring function (AD4) [39]. The AD4 scoring function is composed by several energy terms of classical physics force fields. The free energy of binding (1) is expressed by the sum of molecular mechanics components such as a dispersion-repulsion term, a term for the hydrogen bonding, a term for the electrostatics contribution, a term describing the energy associated to bond lengths, bond angles and associated restriction entropy loss and a term for the desolvation energy.

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{tor} + \Delta G_{solv} \quad (1)$$

In the first experiment, the best ligand poses obtained by AC and PS methods are discriminated using the best energy of binding for each method with the AD4 function. In the second one, five best ligand poses are selected using the AD4 function. These ligand poses correspond to the lower binding energies structures whatever the method used to generate them (AC or PS). They are discriminated from the bulk of all generated structures. In the two experiments, the localization of best energy docked poses is compared to the experimental pose with the measurement of the Euclidian Distance (ED) between the two ligands geometrical mass centers. When ligands are in the same binding cavity as the experimental one and the ED is lower than 2.5 Angstroms, the ligand pose is considered similar to the crystallographic pose and is called X-pose. When ligands are partially docked in the experimental cavity or able to dock in a juxtaposed cavity and ED is included between 2.5 and 8.5 Angstroms, the ligand pose is called J-pose (for Juxtaposed-pose). Beyond this value, we checked that any ligand is localized in a binding area different from the experimental structure. In this case, the wrong ligand pose is called W-pose. (All of these ligands poses were checked by hand and visualized with VMD [40]).

Thus, ligand pair Root Mean Square Deviation (RMSD) computation evaluates the shift between the binding conformation of the best-docked ligands and the crystallographic conformation. The RMSD corresponds to the measure of the average distance between atomic positions of two structures expressed in Angstroms as it shows in (2).

$$RMSD(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{ix}-w_{ix})^2 + (v_{iy}-w_{iy})^2 + (v_{iz}-w_{iz})^2)} \quad (2)$$

V. RESULTS

As described above, our methodology was tested on 24 experimental protein-ligand complexes available in the PDB. Both AC and PS methods were used individually and in a combined procedure to evaluate their ability to re-dock an experimental ligand on its native protein target receptor. To evaluate the ability of our methodology to retrieve experimental ligands poses; a ranking protocol based on AC and PS results has been developed. Protocols parameters are explained in this section and are mentioned in section IV.C.

For the PS method, the experiment shows that for this size of proteins (see IV.A), the Fpocket algorithm found at the most five or six different well-sized pockets. TABLE III gives the volume of the three first pockets found for each experimental complex. For all proteins of the set (100%), one pocket at least is detected, for nineteen proteins in the set (19/24, 79%) two pockets are detected and for 14/24 (58%) three pockets are detected. If structures displaying at least 4 pockets are selected, the ratio of the set falls down to 9/24 (37.5%) and decreases even more when considering a higher number of pockets. Thus, it appears that for each

TABLE II. THE 24 EXPERIMENTAL PROTEIN-LIGAND COMPLEXES

PDB code	Res. (Å)	Protein	Ligand
1azm	2.0	Carbonic Anhydrase I	5-Acetamido-1,3,4-Thiadiazole-2-Sulfonamide
1cbs	1.8	Cellular Retinoic-Acid-Binding Protein Type II	Retinoic Acid
1ebp	2.1	Epididymal retinoic acid binding protein	Retinoic Acid
1fkg	2.0	Fk506 Binding Protein	(1R)-1,3-Diphenyl-1-Propyl(2S)-1-(3,3-Dimethyl-1,2-Dioxopentyl)-2-Piperidinecarboxylate (Rotamase Inhibitor)
1fki	2.2	Fk506 Binding Protein	(21S)-1-Aza-4,4-Dimethyl-6,19-Dioxa-2,3,7,20-Tetraoxobicyclo Pentacosane
1glp	1.9	Glutathione S-Transferase Yfyf	Glutathione Sulfonic Acid
1glq	1.8	Glutathione S-Transferase Yfyf	S-(P-Nitrobenzyl) Glutathione
1hfc	1.5	Fibroblast Collagenase	(N-(2-Hydroxymatemethylene-4-Methyl-Pentoyl)Phenylalanyl)Methyl Amine
1icn	1.7	Intestinal Fatty Acid Binding Protein	Oleate (Oleic Acid)
1lic	1.6	Adipocyte Lipid-Binding Protein	Hexadecanesulfonic Acid
1lmo	1.8	Mucopeptide N-Acetylmuramylhydrolase	Di-N-Acetylglucosamine
1mcr	2.7	Immunoglobulin delta Light Chain Dimer	N-Acetyl-L-His-D-Pro-Oh
1mmq	1.9	Matrilysin	Hydroxamate Inhibitor
1mup	2.4	Major Urinary Protein Complex	2-(Sec-Butyl) Thiazoline
1nco	1.8	Holo-Neocarzinostatin	Apo-Carzinostatin chromophore
1poc	2.0	Phospholipase A2	1-O-Octyl-2-Heptylphosphonyl-SN-Glycero-3-Phosphoenolamine
1rob	1.6	Ribonuclease A	Cytidylic Acid
1srj	1.8	Streptavidin	Naphthyl-Haba
1stp	2.6	Streptavidin	Biotin
1tng	1.8	Trypsin	Aminomethylcyclohexane
1tnl	1.9	Trypsin	Tranylcypromine
1ukz	1.9	Uridylate Kinase	Adenosine-5'-Diphosphate
3ptb	1.7	*beta*-Trypsin	Benzyldiamine
8gch	1.6	*gamma*-Chymotrypsin	Gly-Ala-Trp (peptide)

protein-ligand complex selecting only the first pocket found by the Fpocket algorithm is enough to consider the whole set; the results point that selecting at most the three first pockets should refine the search. In addition, the number of jobs launched partly depends on the number of pockets that will be explored. Thus, the number of jobs launched is precisely defined for each complex. A fixed number of jobs can be very interesting to monitor the speed-up and the scalability of the program over a variant number of available cores. In theory, the optimal load balance should be reached if the number of available cores is superior or equal to the number of launched jobs. So, to optimize the computation time we should set the best ratio jobs/cores and to do this a fixed number of jobs is necessary. For example, this set of complexes generates a pool of maximum 360 jobs (24 complexes x (12 AC method boxes + 3 pockets boxes from the PS method at the most)). So, the best energy structure of the ensemble of the twelve boxes is conserved for the AC method and the best energy structure of each of the first three pockets is conserved for the PS method. Finally, four docked poses at the most are obtained for each complex, which will be compared with the experimental ligand pose of the crystallographic ligand-protein complex. Previously, we defined that the re-docking is successful if an X-pose or a J-pose were obtained for the ligand (see IV.C).

A. *Comparison with the Blind Docking experiment*

Firstly, the results for AC and PS methods are compared with the corresponding blind docking experiment (BD). For twelve experiments out of the set (12/24, 50%) the best energy score was obtained by the PS method, for 10/24 (41.5%) it was obtained by AC method and only for 2/24 (8.5%), it was obtained by BD experiment (Fig. 3). Moreover, the combined results of AC method and PS method give a better energy of docking for 22/24 (91.5%) compared to BD. Furthermore, for 54% of the cases the combined methods gave a RMSD between the experimental structure and the best docking pose lower than 5 Angstroms and a RMSD lower than 10 Angstroms for 23/24 (96%)

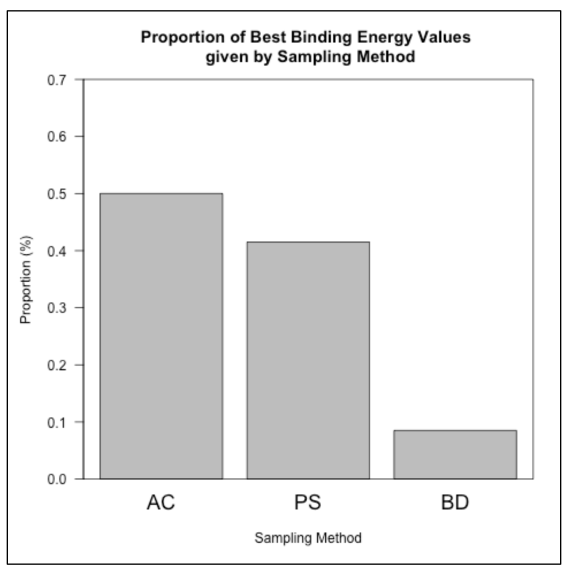


Figure 3. Proportion of Best Binding Energy Values given by the Sampling Method (AC: 62.5%, PS: 29%, BD: 8.5%).

TABLE III. NUMBER OF POCKETS DETECTED FOR EACH PROTEIN AND THEIR VOLUMES

	Pocket 1 (PS1)	Pocket 2 (PS2)	Pocket 3 (PS3)
PDB	*Volume (Å^3)*	*Volume (Å^3)*	*Volume (Å^3)*
1azm	833	786	244
1cbs	1626	378	557
1ebp	1262	370	616
1fkg	549	**N/A**	**N/A**
1fki	576	756	**N/A**
1glp	1307	370	640
1glq	607	637	686
1hfc	762	683	485
1icn	1655	**N/A**	**N/A**
1lic	978	927	**N/A**
1lmo	1306	143	561
1mcr	676	192	**N/A**
1mmq	409	276	548
1mup	479	583	756
1nco	350	**N/A**	**N/A**
1poc	1016	504	642
1rob	654	576	686
1srj	408	**N/A**	**N/A**
1stp	367	**N/A**	**N/A**
1tng	647	610	**N/A**
1tnl	602	466	512
1ukz	600	1072	**N/A**
3ptb	549	328	529
8gch	765	619	383

a. N/A: Non Applicable data –

versus only one for BD (4%) in both case (TABLE IV). These results highlight that our methods perform better exploration of the protein surface. Indeed, the ratio (volume/number of runs) explored in the case of our methodology is better optimized than in the case of BD. Both methods ensure a better conformational sampling and a better quality of docking than using the BD. The distribution of docked poses depending on the sampling method associated with the best energy is presented in Fig. 4. This figure also shows the real efficiency of the PS method that provides no wrong structures and the need to associate both methods (AC + PS) to obtain exhaustive search and good ratios in docking experiments.

B. Comparison between sampling methods

For 18/24 (75%) the sample methods that give the best free energy of binding give also the best docking poses (X-pose or J-pose) distributed as follows: 7/18 (39%) for AC method and 10/18 (55%) for the PS method and 1/18 (6%) for the BD experiment. Among these complexes, the combined method that gives the best free energy of binding gives also the best docking pose for 17 (94.5%) versus only one for BD (5.5%). From Table IV, we can extract the following correlation: comparing the docked poses at rank 1 of Euclidian distance and rank 1 for the lowest RMSD value, there is a match for 6/24 (25%) in the case where a J-pose is observed and for 14/24 (58.5%) in the case where an X-pose is observed. So, at rank 1 for the two previous criteria, the ligand docked poses (X-poses and J-poses) give the lowest RMSD value for 18/24 (75%). Comparing the docked poses at rank 1 of Euclidian distance and rank 1 and 2 for the lowest RMSD value the proportion reach 22/24 (91.5%). The match ratio is distributed by sampling method as follows: The AC method gives the X-pose for 3 complexes with a mean RMSD value equal to 2.32 Angstroms compared to the experimental structures (1glp, 1mup, 1tnl). The AC method gives also a J-pose for 3 complexes (1hfc, 1icn, 1rob) and an associated RMSD value equal to 7.82 Angstroms compared to the experimental structures. Nevertheless, it is important to mention that for 1hfc and 1icn poses are reverse poses that is to say the ligand acquires a head to tail conformation compared to the experimental one so the RMSD increases. The PS method gives the X-pose for 11 complexes (1azm, 1cbs, 1ebp, 1fkg, 1fki, 1mmq, 1nco, 1stp, 1tng, 3ptb, 8gch). In these cases, the mean RMSD with the experimental structure is 2.93 Angstroms. The PS method gives a J-pose for 4 complexes (1lic, 1mcr, 1poc, 1ukz) and an associated mean RMSD value with the experimental structure of 5.54 Angstroms (Fig. 5). The BD method gives an X-pose for 1srj with a RMSD value of 2.47 Angstroms. If the rank 2 for the Euclidian distance is also considered, the PS method is able to replace the ligand for 1srj in an X-pose with 2.35 Angstroms of RMSD. So, the combined method with these evaluation criterions gives the best pose for 22/24, 91.5% of the cases of the total set.

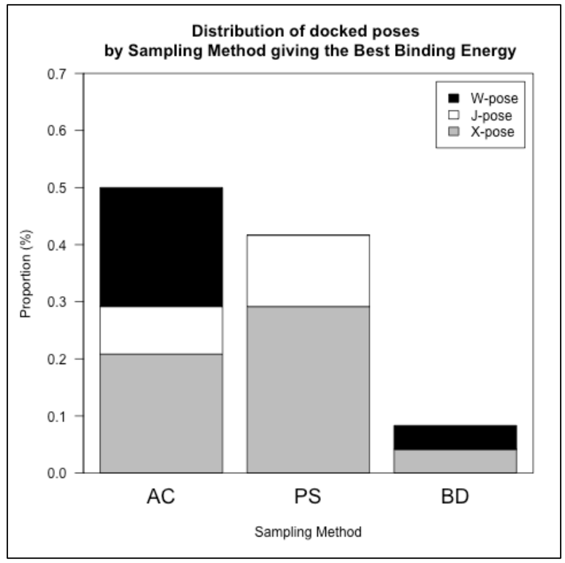


Figure 4. Distribution of docked poses (X-pose in grey, J-pose in white and W-pose in black) by Sampling Method giving the Best Binding Energy

TABLE IV. EVALUATION CRITERIONS OF THE SAMPLING METHODS

	Energy (kcal/mol)		RMSD (Angstroms)				Gravity Centers Euclidian Distance (Angstroms)					
	Rank 1		*Rank 1*		*Rank 2*		*Rank 1*			*Rank 2*		
PDB	*Method*	*Value*	*Method*	*Value*	*Method*	*Value*	*Method*	*Value*	*Pose*	*Method*	*Value*	*Pose*
1azm	AC	-5.15	PS1	1.95	**N/A**	**N/A**	PS1	1.12	X-pose	**N/A**	**N/A**	**N/A**
1cbs	PS2	-6.84	PS2	2.24	AC	8.86	PS2	0.96	X-pose	PS1	1.59	X-pose
1ebp	PS2	-8.68	PS2	2.00	AC	2.73	PS2	0.27	X-pose	PS1	1.23	X-pose
1fkg	PS1	-5.96	PS1	5.49	AC	8.22	PS1	1.24	X-pose	AC	3.98	J-pose
1fki	PS1	-10.49	PS1	0.60	PS2	1.75	PS1	0.59	X-pose	PS2	1.00	X-pose
1glp	AC	-4.46	AC	2.71	PS1	5.42	AC	0.60	X-pose	PS1	2.74	X-pose
1glq	BD	-3.66	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
1hfc	AC	-4.78	AC	8.75	**N/A**	**N/A**	AC	3.84	J-pose	**N/A**	**N/A**	**N/A**
1icn	AC	-3.97	AC	8.80	**N/A**	**N/A**	PS1	3.49	J-pose	AC	3.52	J-pose
1lic	PS1	-4.65	PS1	5.75	**N/A**	**N/A**	PS1	3.50	J-pose	AC	4.23	J-pose
1lmo	AC	-3.26	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
1mcr	AC	-4.03	PS2	4.41	**N/A**	**N/A**	PS2	2.81	J-pose	**N/A**	**N/A**	**N/A**
1mmq	AC	-6.31	AC	3.97	PS1	4.16	PS1	0.79	X-pose	AC	1.55	X-pose
1mup	AC	-4.23	AC	2.59	PS1	4.04	AC	1.56	X-pose	PS1	2.02	X-pose
1nco	PS1	-7.19	PS1	7.83	**N/A**	**N/A**	PS1	2.90	X-pose	AC	8.22	J-pose
1poc	PS1	-1.91	PS1	6.71	**N/A**	**N/A**	PS1	3.95	J-pose	**N/A**	**N/A**	**N/A**
1rob	PS2	-5.29	AC	5.91	PS1	9.89	AC	5.32	J-pose	PS2	8.45	J-pose
1srj	BD	-7.48	PS1	2.35	BD	2.47	BD	0.45	X-pose	PS1	1.13	X-pose
1stp	PS1	-6.10	PS1	1.34	AC	2.42	PS1	0.30	X-pose	AC	0.55	X-pose
1tng	PS1	-5.87	PS1	1.05	AC	1.53	PS1	0.60	X-pose	AC	0.83	X-pose
1tnl	AC	-5.96	AC	1.68	PS1	2.44	AC	0.35	X-pose	PS1	0.41	X-pose
1ukz	AC	-6.74	PS1	5.31	**N/A**	**N/A**	PS1	3.39	J-pose	**N/A**	**N/A**	**N/A**
3ptb	AC	-5.52	PS1	1.52	AC	2.07	PS1	0.19	X-pose	AC	0.19	X-pose
8gch	AC	-5.00	PS1	4.32	**N/A**	**N/A**	PS1	1.03	X-pose	**N/A**	**N/A**	**N/A**

a. N/A: Non Applicable data – Euclidian Distance > 10 Angstroms

Fig. 6 shows the results obtained with the AC method for the experimental complex 1stp. An X-pose with a Euclidian distance between ligands geometrical mass centers of 0.55 Angstroms (rank 1) with a RMSD value of 2.42 Angstroms (rank 1) is observed. As we can see on Fig. 6 and Fig. 7 with two different types of protein representations, the re-docked ligand reached successfully the experimental cavity of binding and adopts a similar conformation compared to X-ray structure. On Fig. 6, the New Cartoon style represents only the secondary structure of the backbone skeleton of the protein whereas on Fig. 7, all amino-acids side chains are included to build the protein surface thanks to the MSMS algorithm [41]. The local structure of side chains creates reliefs and since some of them display specific chemical properties, they can arrange themselves in binding cavities. The ligand pose and conformation in the binding site will be related to the cavity geometry. As we can see in Fig. 6, a good ligand pose implies a chemical conformation that precisely place the chemical groups implied in Hydrogen bonds in an appropriate range of distance (around 2.0 Angstroms). Hydrogen bonds are strong dipole-dipole interactions between electro-negative atoms, and according to local chemical composition they are partially in charge of ligand docking in a binding pocket. Indeed, according to the ligand pose in the binding cavity, hydrogen bonds can form or not and modify its contribution in the binding free energy value.

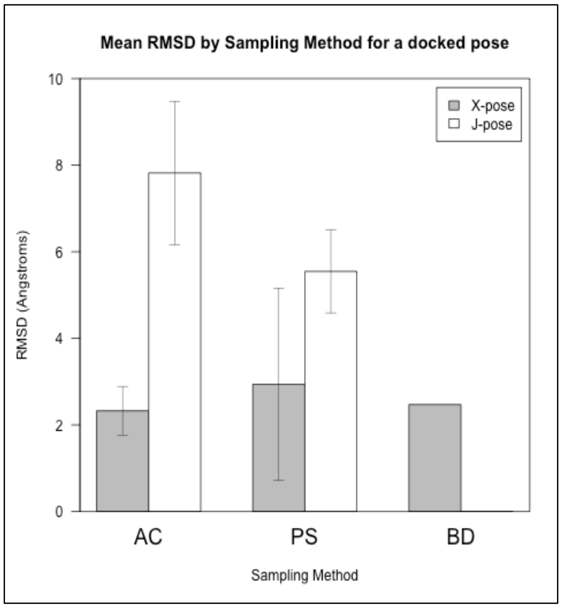


Figure 5. Mean RMSD (in Angstroms) for an X-pose (in grey) and a J-pose (in white) by Sampling Method at rank 1 of Euclidian distance and rank 1 and 2 of RMSD.

C. *Wrong cases study*

For ligands from seven complexes, there is a match between RMSD and mass centers distance but not between both and the best binding energy. In all cases the pose giving the best energy is localized in different cavities that the crystallographic ones. These results can be explained by several settings of decomposing method used (Fig. 7). For 1azm, the best energy is obtained with the box-11 of the AC method (-5.15 kcal/mol) whereas best RMSD with an X-pose is obtained by the PS method (PS1). The AC pose is localized in a different cavity from the crystallographic one. The box-11 dimensions do not allow to include the crystallographic area and they do not permit to refind the experimental pose. On the other side, the PS1 box dimensions do not allow to refind the AC pose cavity neither. The experimental cavity (S1) is included in an another AC box, box-7. The ligand pose obtained with this box is localized in the same cavity as the previous AC box (S2) and presents a better energy than PS1 pose. If we set the dimensions of a tuned box able to include the two binding sites S1 and S2, the ligand pose obtained binds into S1 with even better energy of -5.26 kcal/mol. Finally, to maximise the number of energy evaluations and the conformationnal sampling, we carried out a 256 runs on the previous tuned box and anew the crystallographic cavity is obtained with a poorer energy compared to S1 of -4.49 kcal/mol. So, just the box boundaries presence is not enough to conclude, 1azm complex may be wrong prepared or this case show the limits of the AutoDock force-field.

Crystallographic pose refinding may be precluded by boxes boundaries but it is also impacted by protein shape specifications. In fact, for 1ukz, the cavity is closer as a funnel with a long and slight pipe that sinks into the protein structure. The experimental ligand is housed at the bottom of the pipe in a burried area in the protein core. Fpocket detects the left large extremity as part as a full binding pocket (PS2) and the hidden area as an another binding pocket (PS1). The AC box (giving the best energy) only takes in the funnel cavity and does not include the burried site (like PS2 do) and inversely PS1 includes the crystallographic cavity but does not take in the large surface cavity. It explains why there is no match between the AC method that gives the better energy and the PS1 X-pose.

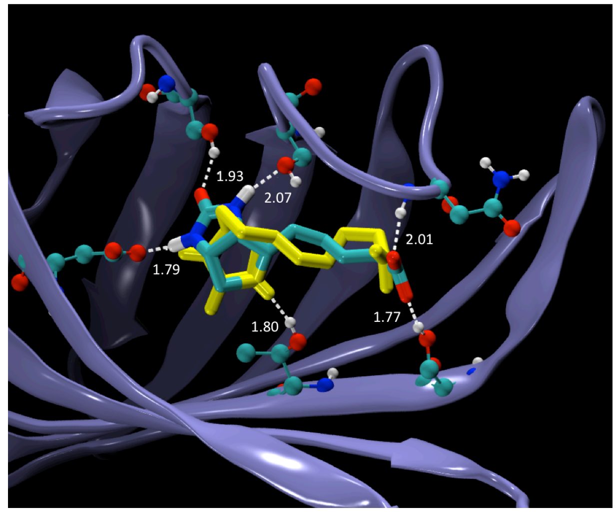


Figure 6. 1STP -- Streptavidin (New Cartoon, in purple)/Biotin (Licorice, X-ray in cyan, X-pose in yellow) protein-ligand complex stabilized by hydrogen bonds in the binding site.

Failed dockings can be explained by protein shape specifications but also by ligand chemical structure. Some ligands such as 1lmo or 1rob are very exposed in large valleys at the protein surface, which are correctly identified by the Fpocket program as a binding pocket but the docking program could fail to place correctly the ligand on a planar surface. Else, the chemical nature of ligands could increase the docking process weakness: 1lmo ligand is a big flexible di-saccharide and 1rob ligand is an ADN nucleoside both containing –ose residues hard to treat with the Autodock force field.

Only for 1glq in the test set, the best energy value is given by the blind docking experiment (-3.66 kcal/mol). The ligand pose is neither in the crystallographic cavity nor in any pocket cavity and binds on a relative open cavity. However, 1glq and 1glp are two crystallographic structures of the same protein with about the same degree of resolution complexed with two similar ligands (see Table II). For all this reasons, Fpocket is not able to find precisely the same pockets in 1glq so the boundaries are not exactly at the same place and do not allow to retrieve the experimental pose with the PS method. The AC method does but the energy of binding is worse than for the pose obtained by the blind docking. Nevertheless, if we launch multiple blind docking experiments, these artefact binding modes should not be

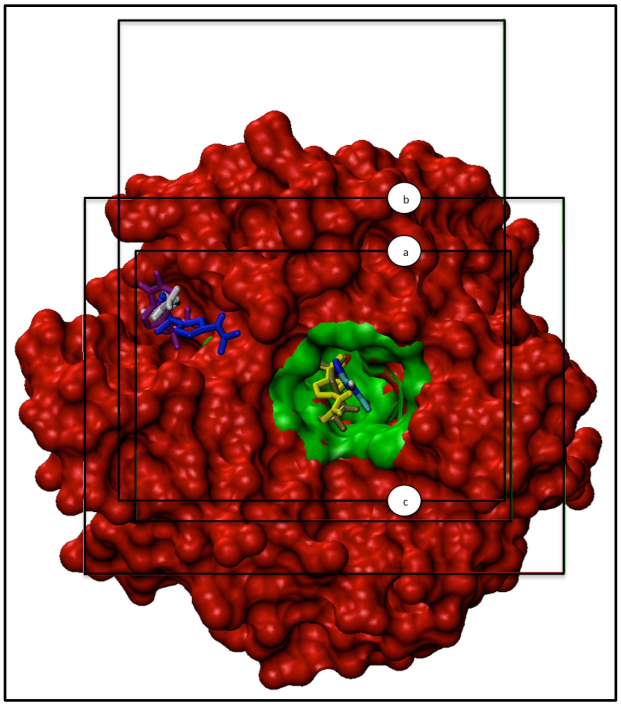


Figure 7. 1azm ligands in the crystallographic cavity (MSMS, in green): X-ray pose (in cyan), PS1 pose (in tan), Tuned box-256R pose (in yellow) and 1azm ligands in another cavity: AC box-11 pose (in blue), AC box-7 pose (in purple), Tuned box-50R pose (in purple) bound on the whole Carbonic Anhydrase I protein receptor with a: PS1 pocket box, b: Tuned box, c: AC box-7

retrieved several times. 1stp and 1srj are two crystallographic structures of the same protein (see Table II) with a large variation in resolution neatness. In fact, if a structure with a higher resolution than 2.0 Angstroms is available it is assumed that a structure with a lower resolution degree is a worse structure. In this case the two structures do not show remarkable difference of structure of the binding site. For the binding site in 1stp, the protein-ligand complex is the well known Streptavidin/Biotin complex in which the protein have a β-barrel secondary structure. This complex is one of a strongest non-covalent interactions known in nature. It is used extensively in molecular biology as a marker. The ligand fits perfectly in the binding site and the interaction are stabilized through a complex network of hygrogens bonds. For 1stp, the experimental ligand was well replaced by the PS1 method (0.37 Angstroms) and AC method (0.55 Angstroms) with the best binding energy equal to -6.10 kcal/mol for PS1.

For 1srj the ligand is Naphtyl-Haba docked in the same cavity as Biotin. It was well re-placed by the blind docking experiment (0.45 Angstroms) and PS1 method (1.23 Angstroms) with the best binding energy equal to -7.48 kcal/mol for BD. This results could be explained by the asymetric shape of the protein that confers a geometry less spherical than a regular globular protein. Consequently, the long axis of the protein takes a high value and imposes the same grid spacing as the other proteins. But in this case, the surface to explore included in the blind docking box is less important and the majority of the grid points are not on the protein surface. So, the ratio volume/runs is very high and the algorithm explore much more precisely the binding pocket and leads to the best energy pose with the maximum goodness.

D. A second protocol for ranking improvement

A second protocol briefly described in the IV. D. section was set up and the new results were compared to the ones obtained with the first protocol. In this experiment, the best ligand poses correspond to the five lowest binding energies structures (R1, R2, R3, R4, R5) from the bulk of all poses generated by AC method and PS methods. While in the first protocol only the first three pockets were selected, here the pockets were selected from a maximun of 9 pockets if they were ever detected by the Fpocket algorithm. The R1 value corresponds to the absolute lowest binding energy structure from the bulk of all docking runs. The R2, R3, R4, and R5 values are the four following ranked in ascending order. For each R value of binding energy (R_E) the associated Euclidian distance with the experimental ligand (R_D) is given in the Table V. The best five energy structures were compared with their experimental ligand with the same metric as previously (computation of the Euclidian distance between the two mass centers). Thus, for 11/24 (46%) of the test set, the use of this protocol gives structures that are found only in X-pose. For 16 percent of the complexes, the solutions belong to the J-Pose class only. If we combined the two kinds of class, one of the five structures at least gives a X-pose for 19/24 (79%) of the set. For all the set but one case (1mcr), the considered solutions are X-pose or J-Pose (23/24, 96%).

If we compare strictly the Euclidian distances between the two experiments we highlight that only a few values are present in both Table IV and V. Since the R1 value is the absolute lower energy from the AC and PS method, it is necessarily in both tables except if the $R1_E$ correspond to a structure provided by a pocket not taken into account during the first experiment but it has never happened with this test set. For the others R values, a few values of the Table IV are found in Table V. In the first protocol, each sampling method is considered separately : for each one of them the best binding energy conformation is selected among the 50 solutions given by Autodock. Thus, we get 15 conformations (three for the PS and one for the AC) among which we select the lowest one. Whereas we demonstrated that this protocol improves in a significtaive way the results obtained with blind docking [42], the fact that all the collected solutions are not treated with the same weight (solution 2 to 50 of a PS1 PS2 or PS3 are never compared to solution 2 to 50 of each AC box…) may introduce a bias. With the second protocol, all the solutions ((12+n)*50, n being the number of considered pockets) are considered together and compared on the same level. From this ensemble of soultions, the fisrt five best binding energies are selected. Following this scheme, it is possible to find in

TABLE V. EUCLIDIAN DISTANCE FOR R VALUES

	E (kcal/ mol)	Euclidian Distance (Angstroms)				
PDB	***R 1_E***	***R 1_D***	***R2_D***	***R3_D***	***R4_D***	***R5_D***
1azm	-5.15	**N/A**	**N/A**	**N/A**	1.08	**N/A**
1cbs	-6.84	0.96	2.37	**N/A**	1.79	2.47
1ebp	-8,68	0.27	1.30	0.14	0.05	0.43
1fkg	-5.96	1.24	2.85	3.15	4.69	2.77
1fki	-10.49	0.59	0.76	8.72	0.56	0.88
1glp	-4.46	0.60	1.74	**N/A**	5.94	2.69
1glq	-3.21	**N/A**	**N/A**	**N/A**	2.71	**N/A**
1hfc	-4.78	3.84	0.29	**N/A**	**N/A**	**N/A**
1icn	-3.97	3.52	4.51	2.93	2.27	5.21
1lic	-4.65	3.50	3.07	5.35	0.84	1.78
1lmo	-3.26	**N/A**	7.74	**N/A**	8.44	8.88
1mcr	-4.03	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
1mmq	-6.31	1.55	**N/A**	8.65	**N/A**	**N/A**
1mup	-4.23	1.03	1.90	1.87	1.56	**N/A**
1nco	-7.19	1.50	2.90	6.91	5.78	**N/A**
1poc	-1.91	3.77	**N/A**	3.95	7.11	5.20
1rob	-5.29	6.31	9.76	3.03	8.45	3.17
1srj	-6.18	**N/A**	**1.13**	6.16	7.50	6.06
1stp	-6.10	0.30	0.53	0.72	**N/A**	**N/A**
1tng	-5.87	0.60	0.27	**N/A**	9.97	1.28
1tnl	-5.96	0.35	**N/A**	0.24	1.28	**N/A**
1ukz	-6.74	8.17	3.02	0.94	8.97	8.39
3ptb	-5.52	0.19	0.09	1.01	0.91	2.16
8gch	-5.00	**N/A**	**N/A**	**N/A**	0.79	**N/A**

a. N/A: Non Applicable data – Euclidian Distance > 10 Angstroms

the final set of solutions ligand conformations that come from the same box or pocket. That could not be the case with the 4 solutions proposed in the first protocol. This way to proceed allows to retrieve a J-pose for 1glq and 1lmo and a X-pose for 1srj that increases de ratio of 4.5% to reach 96% of the test set. Only 1lmo and its di-saccharide ligand stay unruly and confirms the real force-field weakness with carbohydrate residues.

VI. DISCUSSION/CONCLUSION

In order to be able to treat many hundred proteins computations on High Performance Computing (HPC) architectures, we developed a set of methods to parallelize the treatment of each protein, as well as to distribute the tasks among a given set of machines as a way to speed up the overall execution of the inverse docking. For this, we developed a framework that can embed the AC and the PS method to explore as best as possible the protein surface and rationally dock the ligand into the binding cavity.

Our results show that the methods we are developing perform better volume exploration with a better ratio volume/runs than a classical blind docking experiment. In fact, to perform an accurate high definition docking we have to deal with coherent grid spacing. By default, AutoDock builds affinity grids with a spacing of 0.375 Angstroms that corresponds to a quarter of the bond between two atoms of Carbone. We defined a spacing interval between 0.375 and 0.450 in which we consider the accuracy of the simulation as a HD docking. The main drawback of this method is that AutoDock is able to build and also explore a 3D box of 126 x 126 x 126 points at the most. So, only a protein whose long axis is lower than 60 Angstroms can fit into the grid box.

Considering this kind of protein for a blind docking experiment, the AutoDock program is also limited in the number of simulations runs, that is to say in the number of times the initial LGA is reinterred (256 runs max.). So, AC method considers the BD box volume cut into 12 sub-boxes with a partial overlap. Each sub-box is explored by the LGA with 50 runs of simulations that roughly correspond to the half of the ratio volume/runs for the BD. Whereas the ratio is more difficult to precisely estimate, it is even better with PS method, which explains the effectiveness of the program to perform better exploration and to obtain better docking quality results than BD experiments.

As many docking programs [29][30], we have shown that our framework is a successful tool to re-place correctly the ligand into the active site of the target receptor in a non-covalent manner. Furthermore, it is also able to predict accurate ligands bindings independently of active site knowledge [33][34]. For this, we established two protocols in order to be able to distinguish in the best way the correct binding poses of the ligand. In the first protocol we evaluated a good docking pose using three criteria: free energy of binding, Euclidian distance between mass centers and RMSD of the re-docked ligand with respect to the crystallographic ligand. Combinations of these criteria are able to discriminate right docking poses from experimental data. The combination between the binding energy and the RMSD (rank1 and 2) is able to discriminate 66.5% of the test set and the one between the mass center distance (rank1 and 2) and the RMSD (rank1 and 2) is able to discriminate 91.5% of the test set. On the other side, the ratio is 75% for the combination of binding energy and center of mass distances (rank1 and 2) and 71% for the combination of the triad. This is explained by the nature of the evaluation criteria. RMSD and mass centers distances are implicitly correlated because they both describe a space position. Mass centers distances describe a space position for the entire ligand whereas RMSD describe a space position for each atom of the ligand, both always in respect to the experimental structure. In fact the RMSD reflects the ligand structure in a local environment, its capacity to adapt itself to the binding cavity. Consequently, taking into account the numbers of atoms implied both in the binding site and in the ligand structure and the number of torsions available for the ligand, the probability to obtain a low RMSD in a different cavity than the crystallographic binding site is close to zero. This is well shown in Table IV, for 8 cases out of 9 if the RMSD is higher than 10 Angstroms the corresponding mass center distance is higher than 10 Angstroms too (N/A data). That explains the good ratio for these criteria combination. On the other side, the space position adopted by the ligand in the binding site translated by the RMSD value impacts the chemical match between chemical groups able to make non-covalent interactions (Hydrogen bonds, van der Waals forces and electrostatics) with atoms in the binding cavity. These forces represent a major contribution into the energy function that is used to evaluate the free energy of binding (see IV.C). So, the ratio of the combination of RMSD and energy of binding can be explained partially by this relationship.

Nonetheless, in this experiment we have shown that we reproduce ligands experimental poses with our framework. As the references are experimental data, we dispose of comparison elements (RMSD and mass centers distances). The results obtained in this study (distances determining X-pose and J-pose and associated RMSD) validate the method for detecting workable binding sites. To identify already known binding sites or new ones the aim of this program is to perform predictive experiments on large sets of proteins for a given ligand of interest. For these, we will only dispose of the free energy of binding to discriminate good docking poses. For 7/24 there is no match between the binding energy and the geometric criterions. In some remarkable cases we have shown previously, only the free binding energy computation does not allow to retrieve similar poses to the crystallographic ones. That is demonstrating that the evaluation of the binding energy is not an absolute reference. To reduce the unsuccessful ratio we reinforced the ranking evaluation process in the second experiment with a better protocol. We showed that this protocol (a) allows removing the constraints of the first experiment that excluded some good results and (b) is able to compensate force field behavior failures.

In most of the cases we have seen that the PS method strongly performs to detect druggable cavities on a protein receptor. In fact some proteins present multiple binding sites well described in enzymology allosteric phenomena especially. The advantage of using multiple pockets search is to identify well differentiated multiple sites on the fly during a unique docking simulation. That allows us to consider ligand repositioning experiments and also second targets and off-targets hunting. In addition the AC method is able to overcome the PS method failures with adding search completeness and not excluding planar binding surfaces such as protein-protein binding area in particular. So, we demonstrate that the combination of the two methods is an accurate strategy to identify new protein targets for a given ligand. We developed an effective tool to perform large-scale inverse virtual screening works on both HPC hardware and personal computer able to identify proteins targets for a chemical ligand of interest.

ACKNOWLEDGMENT

The performance and scalability tests were performed at the Bull's CEPP and experiments presented on this paper were carried out on the ROMEO Computing Center (https://romeo.univ-reims.fr) and the Multi-scale Molecular Modelling Platform (https://p3m.univ-reims.fr). Romain Vasseur is founded by a Bull SAS CIFRE grant. We thank Romulo Gadelha de Moura Lima, for his great contribution to format the code and Soline Laforet from the Bull's CEPP for her precious expertise in HPC benchmarking. The AMIDE framework and its documentation will be available through the Multi-scale Molecular Modelling Platform website.

REFERENCES

[1] R.Vasseur *et al.*, "A framework for Inverse Virtual Screening," BioSciencesWorld 2014 – Proc. BIOTECHNO 2014: The Sixth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, IARIA, April 2014, pp. 20-29.

[2] R. Abagyan and M. Totrov, "High-throughput docking for lead generation," Curr. Opin. Chem. Biol., vol. 5, no 4, 2001, pp. 375-382.

[3] D. Giganti *et al.*, "Comparative Evaluation of 3D Virtual Ligand Screening Methods: Impact of the Molecular Alignment on Enrichment", J. Chem. Inf. Model., vol. 50, no 6, 2010, pp. 992-1004.

[4] G. Klebe, "Virtual ligand screening: strategies, perspectives and limitations," Drug Discov. Today, vol. 11, no 13-14, 2006, pp. 580-594.

[5] O. Korb, T. Stützle, and T. E. Exner, "Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS," J. Chem. Inf. Model., vol. 49, no 1, 2009, pp. 84-96.

[6] J. J. Irwin *et al.*, "Automated Docking Screens: A Feasibility Study," J. Med. Chem., vol. 52, no 18, 2009, pp. 5712-5720.

[7] G. Jones, P. Willett, and R. C. Glen, "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation," J. Mol. Biol., vol. 245, 1995, pp. 43-53.

[8] S. Cosconati *et al.*, "Virtual screening with AutoDock: theory and practice," Expert Opin. Drug Discov., vol. 5, no 6, 2010, pp. 597-607.

[9] S. Zhang, K. Kumar, X. Jiang, A. Wallqvist, and J. Reifman, "DOVIS: an implementation for high-throughput virtual screening using AutoDock," BMC Bioinformatics, vol. 9, no 1, 2008, pp. 126.

[10] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," J. Mol. Biol., vol. 261, no 3, 1996, pp. 470-489.

[11] R. A. Friesner *et al.*, "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes," J. Med. Chem., vol. 49, no 21, 2006, pp. 6177-6196.

[12] Y. Y. Li, J. An, and S. J. Jones, "A computational approach to finding novel targets for existing drugs," Plos Comput. Biol., vol. 7, no 9, 2011, pp. e1002139.

[13] S. L. Kinnings and R. M. Jackson, "LigMatch: A Multiple Structure-Based Ligand Matching Method for 3D Virtual Screening," J. Chem. Inf. Model., vol. 49, no 9, 2009, pp. 2056-2066.

[14] R. Wang, X. Fang, Y. Lu, and S. Wang, "The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures," J. Med. Chem., vol. 47, no 12, 2004, pp. 2977-2980.

[15] J. J. Irwin and B. K. Shoichet, "ZINC-a free database of commercially available compounds for virtual screening," J. Chem. Inf. Model., vol. 45, no 1, 2005, pp. 177-182.

[16] F. H. Allen, "The Cambridge Structural Database: a quarter of a million crystal structures and rising," Acta Crystallogr. B, vol. 58, no 3, 2002, pp. 380-388.

[17] RCSB Protein Data Bank - http://www.rcsb.org/pdb

[18] P. Bradley, K. M. Misura, and D. Baker, "Toward high-resolution *de novo* structure prediction for small proteins," Science, vol. 309(5742), 2005, pp. 1868-1871.

[19] J. Moult, "A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction," Curr. Opin. Struct. Biol., vol. 15, no 3, 2005, pp. 285-289.

[20] N. Eswar *et al*., "Comparative protein structure modeling using Modeller," Curr. Protoc. Bioinforma., 2006, pp. 5-6.

[21] A. P. Norgan, P. K. Coffman, JP. A. Kocher, D. J. Katzmann, and C. P. Sosa, "Multilevel parallelization of AutoDock 4.2," J. Cheminf., vol. 3(1), no 1, 2011, pp. 1-9.

[22] X. Zhang, S. E Wong, and F. C. Lightstone, "Message passing interface and multithreading hybrid for parallel molecular docking of large databases on petascale high performance computing machines", J. Comp. Chem., vol. 34, no 11, 2013, pp. 915-927.

[23] R. De Paris, F. A. Frantz, O. Norberto de Souza, and D. A. Ruiz, "wFReDoW: A Cloud-Based Web Environment to Handle Molecular Docking Simulations of a Fully Flexible Receptor Model", BioMed Res. Inter., 2013, pp. 1-12.

[24] T. Y. Tsai, K. W. Chang, and C. Y. Chen, "iScreen: world's first cloud computing web server for virtual screening and de novo drug design based on TCM database@Taiwan," J. Comput. Aided Mol. Des., vol 25, 2011, pp. 525-531.

[25] I. Pechan and B. Fehér, "Molecular docking on FPGA and GPU platforms", International Conference on Field Programmable Logic and Applications, 2011.

[26] G. M. Morris *et al*., "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," J. Comput. Chem., vol. 30, no 16, 2009, pp. 2785-2791.

[27] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy od docking with a new scoring function, efficient optimization, and multithreading," J. Comput. Chem., vol 31, 2010, pp. 455-461.

[28] B. Iorga, D. Herlem, E. Barré, and C. Guillou, "Acetylcholine nicotinic receptors : finding the putative binding site of allosteric modulators using the "blind docking" approach," J. Mol. Model., vol. 12, no 3, 2005, pp. 366-372.

[29] C. Hetényi and D. van der Spoel, "Blind docking of drug-sized compounds to proteins with up to a thousand residues," Febs Lett., vol. 580, no 5, 2006, pp. 1447-1450.

[30] A. Grosdidier, V. Zoete, and O. Michielin. "Blind docking of 260 protein- ligand complexes with EADock 2.0," J. Comp. Chem., vol. 30, no 13, 2009, pp. 2021-2030.

[31] I. W. Davis, K. Raha, M. S. Head, and D. Baker. "Blind docking of pharmaceutically relevant compounds using RosettaLigand," Protein Science, vol. 18, no 9, 2009, pp. 1998-2002.

[32] R. Vasseur *et al*., "Parallel strategies for an inverse docking method," Proc. PBio: International Workshop on Parallelism in Bioinformatics, EuroMPI User's Group Meeting (EuroMPI 2013), ACM, Sept. 2013, pp. 253-258.

[33] D. Ghersi and R. Sanchez, "Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites," Proteins Struct. Funct. Bioinforma., vol. 74, no 2, 2009, pp. 417-424.

[34] C. Hetényi and D. van der Spoel, "Toward prediction of functional protein pockets using blind docking and pocket search algorithms," Protein Sci., vol. 20, no 5, 2011, pp. 880-893.

[35] V. Le Guilloux, P. Schmidtke, and P. Tuffery, "Fpocket: An open source platform for ligand pocket detection," BMC Bioinformatics, vol. 10, no 1, 2009, pp. 168.

[36] Structural Chemogenomics Group - http://bioinfo-pharma.u-strasbg.fr

[37] E. Kellenberger, J. Rodrigo, P. Muller, and D. Rognan, "Comparative evaluation of eight docking tools for docking and virtual screening accuracy," PROTEINS: Struct. Funct. Bioinf., vol.57, 2004, pp. 225-242.

[38] G. M. Morris *et al*., "Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function," J. Comp. Chem., vol. 19, no 14, 1998, pp. 1639-1662.

[39] R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell, "A semiempirical free energy force field with charge-based desolvation," J. Comput. Chem., vol. 28, no 6, 2007, pp. 1145-1152.

[40] W. Humphrey, A. Dalke and K. Schulten, "VMD - Visual Molecular Dynamics," J. Molec. Graphics, vol. 14, 1996, pp. 33-38.

[41] M. Sanner, A. J. Olson and J. C. Spehner, "Reduced Surface: an Efficient Way to Compute Molecular Surfaces", Biopolymers, vol 38, no 3, 1996, pp. 305-320.

[42] R. Vasseur *et al*., "Parallel strategies for an inverse docking method," J. Parall. Comput., special issue on Parallelism in Bioinformatics, in press.

Towards an Interactive Web Tool That Supports Shared Decision Making in Dementia: Identifying User Requirements

Marijke Span, Carolien Smits, Leontine Groen-van de Ven, Jan Jukema
Research group Innovation of Care of Older Adults and Social Work, Windesheim University of Applied Sciences
Zwolle, The Netherlands
m.span@windesheim.nl, chm.smits@windesheim.nl, lm.groen-vander.ven@windesheim.nl, js.jukema@windesheim.nl

Marike Hettinga
Research group IT Innovation in Health Care
Windesheim University of Applied Sciences
Zwolle, The Netherlands
m.hettinga@windesheim.nl

Anita Cremers
TNO Human Factors, Utrecht University of Applied Sciences, department: Multi Modal User Interface Design,
Utrecht, The Netherlands
anita.cremers@tno.nl

Myrra Vernooij-Dassen
Department of Primary Care, IQ HealthCare, Radboud University Nijmegen; Kalorama Foundation, Radboud Alzheimer Centre
Nijmegen, The Netherlands
myrra.vernooij-dassen@radboudumc.nl

Jan Eefsting
Department of Nursing Home Medicine and EMGO Institute for Health ad Care Research, VU University Medical Centre
Amsterdam, The Netherlands
j.eefsting@zgijv.nl

***Abstract*— To support shared decision making in care networks of people with dementia, an interactive web tool for people with dementia, informal caregivers, case managers and other professional caregivers is being developed. This study aims to identify user requirements for such a tool. A multiple method study with an iterative, participatory design was conducted. Data collection involved 50 semi-structured interviews with end users (people with dementia, informal caregivers, case managers and other professional caregivers), eight focus group interviews with end users and experts, a dementia expert consultation, and two multidisciplinary workshops. Content analysis was applied to the data and resulted in two sets of user requirements for the interactive web tool. The first set of user requirements consists of care and well-being related topics addressing decision making in dementia. Most important topics are care, daily activities, mobility, safety, future, finances, living, and social contacts. The second set of user requirements addresses additional needs and preferences of end users such as: participation of the person with dementia in the decision-making, insight into the decision history, anticipation of possible future problems and decisions, and the degree of self-management and autonomy preservation of the person with dementia. The two sets of user requirements form a solid basis for the further development of a user-friendly, interactive web tool facilitating shared decision making in care networks of people with dementia.**

***Keywords*- dementia, decision making, user requirements, participatory design, assistive technology.**

I. INTRODUCTION

This study focuses on the development of an interactive web tool to facilitate shared decision making in care networks of people with dementia [1]. Dementia is a degenerative disease affecting increasing numbers of people worldwide [2][3]. Over a prolonged period of time people with dementia and their caregivers are faced with many problems and decisions [4][5][6]. Decision-making in dementia is complex; it involves multiple participants with different capacities and interests [7].

To better understand the needs and preferences of people with dementia, it is important to include them in the decision-making [8][9][10][11]. Research shows that involving people with dementia in decision-making increases well-being [9] and quality of life [12] for both the person with dementia and the informal caregiver. Informal caregivers also show lower levels of depression and lessened relationship strain [12]. However, professionals and informal caregivers do not include people with dementia in decision making as a matter of course [8][9].

Shared decision making (SDM) is a way to involve patients in decision making by emphasizing the collaboration between professionals and patients in making a shared decision [13]. There are seven consecutive steps in SDM: identifying problems, information exchange, clarifying preferences and values, presenting and exploring options, discussing pros and cons of the options, deciding together, and evaluating the decision [14][15][16][17]18]. SDM increases patient autonomy and empowers the patient [19]. This could also benefit people with dementia, who are well able to express their needs [8]. Moreover, they are capable of expressing their preferences, even in an advanced stage of dementia [20].

Case management in dementia is implemented differently. In Europe, Canada, and the United States, community-dwelling patients diagnosed with dementia and their caregivers may receive case management support [21]. In The Netherlands case management is offered to people with dementia. Case management in the Dutch context aims to support informal caregivers and people with dementia during the complex care trajectory. Moreover, it aims enabling people with dementia to live independently as long as possible [22]. As case management for dementia is a fairly recent phenomenon, case managers need support to address the complex needs and preferences of people with dementia and their caregivers, possibly in the form of tools [23].

Although research increasingly focuses on dementia, SDM or supportive tools separately, studies that combine these elements are lacking. Thus far, existing decision aids have supported SDM in various patient groups by means of supporting discussion of value-based choices and options about single decisions [24]. Most decision aids facilitate SDM in single medical decisions in the clinical area [18][25].

The new interactive web tool differs from existing decision aids in at least four aspects. First, people with dementia and their caregivers have to make many decisions over a prolonged period of time in the order of years, as opposed to single-issue decisions. Second, the web tool has to take into account an ongoing cognitive decline in contrast to existing decision aids that focus on cognitively able people. Third, the person with dementia is part of a network of informal and professional caregivers who may participate in decision-making. Regular decision aids focus mainly on the patient-clinician relation. Fourth, the decisions of people with dementia and their caregivers are not just about single medical decisions, as the existing decision aids are, but also about multiple care and well-being decisions. Besides, aimed end users of the interactive web tool are people with dementia, informal caregivers, case managers and other professionals.

To develop user-friendly and useful tools involvement of both end users [26][27] and other interested parties [28][29] in developing IT applications is important. We used the Center for eHealth Research and Disease Management (CeHRes) roadmap to increase a structural embedding of the new interactive web tool in practice [29]. This roadmap offers a holistic framework of five phases: contextual inquiry, value specification, design, operationalization, and summative evaluation. This paper describes the phase of value specification (analyzing the preferences of all those concerned). Therefore, this study aims to identify user requirements for an interactive web tool facilitating shared decision making in dementia of all participants involved (people with dementia, informal caregivers, case managers and experts).

The research questions are: 1) What topics can be identified for an interactive web tool facilitating shared decision-making in dementia? 2) What additional needs and preferences regarding an interactive web tool facilitating shared decision making in dementia can be identified?

This paper is organized as follows. Section II describes the design and methods of this study. Section III presents the results in detail. Section IV addresses the discussion of the findings of this study. The paper closes with the conclusions in section V and the acknowledgment.

II. Design and Methods

This multiple method study with an iterative participatory design is part of a larger longitudinal study on decision making in care networks op people with dementia [1]. To guarantee rigorous development, we used multiple data sources to answer the research questions [30]. For both research questions the same data set was used: 50 interviews and 8 focus group interviews. The focus group interviews were organized on the basis of the findings of the individual interviews to enhance data richness [31][32]. In addition, an expert consultation was undertaken for research question 1 and two multi-disciplinary workshops for research question 2. Table I provides an overview of the characteristics of methods used.

All participants gave their written informed consent. Special attention was paid to the informed consent of people with dementia in order to be sure of their voluntary participation. In line with Murphy and colleagues [33], we invested in ongoing consent. This started with time for social talk in order to get to know the person. After gaining initial consent we checked their consent during the interview. We ended the interview with a positive affirmation by emphasizing the importance of their contribution. The investigators were careful to notice any signs, non-verbal or otherwise, of discomfort or restlessness. In such a case, the participant was given ample opportunity to quit. People with dementia who were unable to give written consent because the form confused them were asked for their verbal consent (recorded on audio tape). The focus group participants at the day care center were first informed and then asked for their consent in a group meeting. Afterwards, they were explicitly and individually asked for their consent for participation.

A. *Analysis*

To answer research question 1, content analysis was used starting by reading and rereading the (focus group) interview transcripts [30]. For this research question paragraphs in the text that addressed problems and decisions were of interest. Therefore, two researchers coded the interviews independently focusing on experienced problems and decisions. They then clustered the identified problems and

TABLE I. CHARACTERISTICS OF METHODS USED ADDRESSING RESEARCH QUESTION 1 AND 2

Research Questions			
1. What topics can be identified for an interactive web tool facilitating shared decision-making in dementia? 2. What additional needs and preferences regarding an interactive web tool facilitating shared decision making in dementia can be identified?			
1,2	***1,2***	***1***	***2***
Methods			
Interviews	***Focus group interviews***	***Expert consultation***	***Workshops***
Respondents			
Fifty members of ten care networks of people with dementia (mild to moderate) were individually interviewed about decision making: - 10 persons with dementia - 20 informal caregivers - 10 case managers - 10 other professional caregivers	8 focus groups (n=34) were conducted: - 2 groups with people with dementia (n=9) - 2 groups with informal caregivers (n=11) - 2 groups with case managers (n=14) - 2 groups with dementia experts (n=13) We chose this approach of homogenous selection to enable respondents, and especially people with dementia, to speak for themselves.	7 national dementia experts were consulted via email. Member check was performed in a central meeting.	2 multi-disciplinary workshops were performed with de same participants (n=7): - Interpretation workshop - Affinity diagramming workshop
Recruitment			
The participants were recruited via case managers, Alzheimer cafés, the Dutch Alzheimers' Association, residential homes, and daycare centers.	People with mild to moderate dementia and informal caregivers were recruited from daycare centers, residential homes, and the Dutch Alzheimer Association. Case managers were recruited from regional case managers' networks.	Via the consortium of the SDM research program.	Via the consortium of the SDM research program.
Sampling/inclusion criteria			
The network inclusion was based on maximum diversity of sex of the person with dementia, stage of dementia, type of dementia, type of informal caregiver (e.g., spouse and child), type of formal caregiver (e.g., case manager and home care nurse), and the socio-economic status of the person with dementia. Participation of people with dementia required the ability to converse with a researcher. This corresponds with a score of 2-5 on Reisberg's Global Deterioration Scale and excluded people with severe dementia [34].	Participation of people with dementia required the ability to converse with similar people in a group and a researcher (corresponding with a score of 2-5 on Reisberg's Global Deterioration Scale). In order to create a safe environment, a daycare center was approached in order to recruit people with dementia who knew each other. Furthermore, the principal researcher (i.e., the focus group leader) attended the daycare center for 6 days, 5 before and 1 after the focus group interview.	Inclusion focused on diversity of organizations: e.g., Dutch Alzheimers' Association, University/University of Applied Sciences, and elderly care lobby.	Inclusion focused on diversity of the disciplines: e.g., older adults/dementia experts, technicians, and developers.
Topics			
The interviews were based on a topic list addressing decision-making elements such as: values, problems, decisions, personal considerations, options, and information need. The interviews were semi open and explorative, as we wanted to get insight into participants' personal experiences. The interview guide started with the questions: "How are you right now?", "What has changed for you lately?" and "What choices did you have to make because of these changes?". We then elaborated on respondents' answers. The face-to-face interviews took thirty to seventy five minutes.	Based on the findings of the interviews focus group members were asked to check the information of the interviews and to give additional information. Two researchers moderated all focus group interviews, that lasted 1 to 2 h each. The principal researcher led the focus group using an interview protocol to direct the interview. The second researcher assisted the principal researcher.	The consultation focused the refinement and prioritization of the dementia related topics identified in interviews and focus group interviews. Experts were invited to comment on the clustered experienced problems and decisions and what experienced problems and decisions were in line with the aim and scope of the interactive web tool.	The workshops aimed to align project members' views of the new IT application and identify user requirements emerging from the needs and preferences of the interviews and focus group interviews.
Both interviews and focus group interviews were audio taped, transcribed verbatim, and analyzed with Atlas.ti software.			

decisions in categories using Affinity diagramming [35] till consensus was reached. The clustered list of problems and decisions was then presented in focus groups for a check of this list and gathering supplementary problems and decisions and to enhance data richness [31][32]. Participants were asked whether they recognized the items on the list, and whether they had additional items for the list. Then, in an expert meeting with national dementia experts the results of the interviews and focus groups were discussed and a priority of decision-making topics was determined that addressed the aim and scope of the interactive web tool.

To address research question 2 the interviews were analyzed using content analysis for fragments containing implicit or explicit information regarding needs and preferences of end users about what a new interactive web tool should offer. Relevant fragments identified by the principal investigator were peer-reviewed by a second researcher.

The focus groups were used to check whether participants recognized the selected needs and preferences and gathering additional needs and preferences. The principal researcher made an initial clustering of these fragments using Affinity diagramming [35] with a second researcher. This resulted in a clustered list of fragments. The findings of the focus group interviews provided input for the consecutive workshops. Finally, the clustered list of needs and preferences was discussed in two consecutive multidisciplinary workshops; an "interpretation workshop" followed by an "affinity-diagramming workshop".

The workshops aimed to align project members' views of the new interactive web tool and identify user requirements emerging from the needs and preferences. Therefore, in the first workshop, the participants interpreted and translated the identified needs and preferences of end users in user requirements ("the tool facilitates/clarifies/supports/provides …"*)*. Subsequently, they defined initial categories of needs and preferences. Disagreements were discussed until consensus was reached [35]. In the second workshop, the definition of categories was reviewed, summarized, and reformulated. Moreover, views of the workshop participants were explored, the boundaries of the new interactive web tool were clarified, and the categories were clustered in domains using Affinity diagramming [35]. Determining boundaries of the interactive web tool was necessary. Not all needs, preferences and expectations of participants could be addressed in view of the aim of the interactive web tool. The two workshops, both consisting of the same seven respondents, were audio taped and field notes were taken.

B. Ethical considerations

The principal investigator contacted all interview participants, who gave their informed consent individually, mostly in written form. We paid special attention to the informed consent of people with dementia in order to be sure their participation was voluntarily [33]. The institutional review board of the regional medical ethics committee gave written approval for the study.

III. RESULTS

Analysis of the data resulted in two sets of user requirements addressing the two research questions. The first set of user requirements consisted of decision-making topics the new interactive web tool should address. The second set of additional user requirements consisted of needs and preferences the new interactive web tool should address.

A. Research question 1: decision making topics in dementia the interactive web tool should address

Eighteen topics of clustered experienced problems and eight topics of clustered decisions addressing decision making in dementia arose from the interviews (characteristics in Table IIa) and focus groups (characteristics in Table IIb). In the data of experienced problems and decisions eight similar topics emerged: social contacts, daily activities, mobility, safety, living, future, care, and finances (Table III).

Besides, other topics of clustered experienced problems were: decreasing autonomy of people with dementia, involvement of people with dementia, participants in decision making, communication, information, role of professionals, role of informal caregivers, options, and timing of decisions. These experienced problems show an overlap with the identified user requirements based on users' needs and preferences (research question 2).

More problems were identified than decisions. Moreover, the problems (e.g., loneliness, mobility, managing behavior of person with dementia, lack of safety, and overburdening of the informal caregiver) were more related to well-being than the decisions, that were more often related to care (e.g., decisions about household assistance, home care, resuscitation, and personal alarm devices). Informal caregivers and professional caregivers named more problems than people with dementia did. People with dementia reflected more on important values (e.g., autonomy and staying at home) than the informal and professional caregivers. Focus group participants (except people with dementia) were very well able to indicate problems in the decision-making. Expressing preferences for the new interactive web tool proved to be more difficult for the people with dementia.

B. Research question 2: needs and preferences regarding the new interactive web tool

Two hundred fragments were generated from the interviews (characteristics in Table IIa) and focus group interviews (characteristics in Table IIb). These fragments, containing implicit and explicit participants' needs and preferences about the new interactive web tool resulted in 36 additional user requirements that were clustered in eleven categories and three domains (Table IV) [29]. These three domains stemming from the data are: "involved people and their roles", "timeline", and "information and communication".

TABLE II. CHARACTERISTICS OF PARTICIPANTS

A. Characteristics of participating care networks in interviews										
Network	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Age of PWD[a)]	82	83	80	84	70	89	83	73	86	87
Gender of PWD Male=M/ Female=F	F	F	M	F	M	F	M	F	M	F
Education of PWD[b)] Low (L)/Medium (M)/High (H)	M	H	M	L	H	M	M	H	L	M
Type of dementia Alzheimer (AD)/Vascular (VD)/Lewy Body (LBD)	LBD	VD	AD	?	AD	AD	?	AD	?	AD
Marital status Married=M/Widowed=W/Single=S	W	W	W	W	M	W	M	S	W	M
Living situation pwd at T0 Nursing home=NH/Independent living=IL	NH	IL	IL	IL	IL	IL	IL	NH	IL	IL
Type of informal caregivers interviewed[a)]	D D	D F	D S	GS GDiL	Sp S	D SiL	S D	Si N	S S	Sp D
Type of formal caregivers interviewed[c)]	PA CM	E CM	HCN CM	HCN TLN	E HCN	HCN CM	E CM	PA CM	CC HCN	HCN TLN

a) PWD= person with dementia, D=daughter, F=friend, S=son, GS=grandson, GDiL=granddaughter in law, Sp=spouse, SiL=son in law, Si=sister, N=niece. b) Low: primary/secondary school graduate, medium: high school graduate, high: college graduate. c) PA=principal attendant, CM=case manager, E=employee day care center, HCN=home care nurse, TLN=team leader nurse, CC=care coordinator

B. Characteristics of focus groups' participants				
Characteristics people with dementia attending a daycare center (n=9)				
Gender	***Age***	***Education***[a)]	***Type of dementia***[b)]	***Marital status***
5 Male	68-86 (M=79,2)	3 low	4 AD	5 Married
4 Female		4 medium	1 VD	4 Widowed
		2 high	1 FTD	
			3 MCI/D	
Characteristics of informal caregivers (n=11)				
Gender	***Age*** *(2 did not fill in)*	***Education***[a)] *(2 did not fill in)*	***Relation pwd***	***Experienced caregiving***
3 Male	41-83 (M=64,6)	0 Low	7 Spouse	5 Heavy
8 Female		4 Medium	3 Daughter	6 Medium
		5 High	1 Friend	
Characteristics of case managers (n=14)				
Gender	***Age***	***Experience as case manager***		
1 Male	25-58 (M=42,1)	6: < 1year		
13 Female		7: 1 -5 year		
		1: > 5 year		

a) Low: primary/secondary school graduate, medium: high school graduate, high: college graduate. b) AD= Alzheimer's Disease; VD= Vasculair Dementia; FTD= Fronto Temporal Dementia; MCI/D= Mild Cognitive Impairment/Dementia

The first domain, "involved people and their roles", relates to the categorized user requirements: participation of the person with dementia in decision-making as a central user requirement named by all interested parties; roles of informal caregivers, case managers, and other professionals; self-management and autonomy; and organization of care around the person with dementia. The different parties involved have their specific roles in the network. They all focus on preserving the self-management and autonomy of the person with dementia for as long as possible. The degree of this preservation depends on the endurance capacity of the network. Furthermore, the degree of preserving self-management and autonomy changes over time as informal care givers and professional caregivers take over from persons with dementia when the disease progresses. Participation of people with dementia and self-management and autonomy were the only two categories of user requirements where all participants contributed to.

The second domain, "timeline", refers to the categorized user requirements: decision history and anticipation. One side of the timeline shows the decision history: information addressing decisions, values, and preferences of the network members in the past. Knowledge of decision history gives valuable information that can support network members in coping with current problems and decisions. This information is not always available. Reasons stated for this

TABLE III. FIRST SET OF USER REQUIREMENTS: IDENTIFIED TOPICS

Topics (experienced problems and decisions by participants) identified in decision-making in dementia		
Topics	***Problems***	***Decisions***
Social contacts	Decreasing social contacts of PWD[a)]. PWD Lonesomeness of PWD. PWD Incomprehension of the neighborhood. PWD, IC Gender differences in social skills. PWD Sensitiveness of the word 'dementia'. IC, CM, OFC PWD and IC avoid situations and people: shame about the disease dementia. CM, OFC	Church attendance of PWD. PWD, IC, CM No invitation of PWD at home. IC
Daily activities	Loneliness of PWD. IC, CM, OFC Lack of adequate daily activities of PWD. IC, CM, OFC	Starting day care. PWD, IC, CM, OFC Increasing day care. IC, CM, OFC
Mobility	(Limitation of) mobility of PWD with driving. PWD, IC, CM, OFC (Limitation of) mobility of PWD with cycling. IC, CM, OFC (Limitation of) mobility of PWD with walking. IC, CM, OFC	Driving. PWD, IC, CM, OFC Cycling. PWD, IC, CM, OFC Going outside alone/walking. IC, CM, OFC Scoot mobile. PWD, IC, CM, OFC Traveling with public transport. PWD, IC, CM, OFC
Living	Bureaucracy of indication institutions. IC, CM Stay at home. IC	Registration for nursing home/home for the aged. IC Location: where register/admission (which home and which village/town). IC Admission to an home for the aged/nursing home. IC Internal removals. IC, CM, OFC Timely admission to a nursing home of PWD during timely holiday/drop out of IC. PWD, IC, CM, OFC Forced admission to a psychiatric institution. IC
Safety	Wandering of PWD. PWD, IC, CM, OFC Unsafe behavior of PWD with gas-tap; smoking; electricity. IC, CM, OFC	Measures that limit PWD's liberty. IC, CM, OFC Enclosure of gas-tap; smoking; electricity. IC, CM, OFC Personal alarm device. PWD, IC, CM, OFC
Future	Life of IC after admission (to a nursing home) of the PWD. IC Discussing the future with the PWD is difficult. IC, CM Uncertainty regarding the future, for the PWD as well as the IC. IC, CM Settle things early to enable the PWD to co-decide: authorization, advanced directives/last will and testament. IC Delay of decisions. It's difficult to look in an early stage to the future. IC, CM Fear of PWD of early death of IC (and conversely). PWD, IC	Resuscitation. IC Timely setting things (e.g., authorization). IC Euthanasia wish/statement. IC Advance care planning. IC, CM
Care	Supply and demand of care do not match. IC Accessibility of care. IC Availability of care. IC Bureaucracy of care and welfare. IC, CM Supply of well-being and activities is not used. CM Call in CM in the dementia process is often too late. CM Too many organizations appeal to the PWD and the IC. This leads to commotion and conflicting advices. CM, OFC Timing: it is difficult to get PWD at the right moment in a good (and preferred) situation. CM Lack of relief centers in own region. Distance between family and PWD. Family does not want to travel too far. IC, CM Involved professionals (too many with conflicting opinions). IC Lack of realistic options; no real choice. IC Attention to financial part and consequences of care for PWD. IC	Household assistance. PWD, IC, CM, OFC (Increase of) home care (washing, showering, care, clothing, medication). IC, CM, OFC Housekeeping: checking refrigerator, medication, nutrition. IC, CM, OFC
Finances	Financial and economic consequences for IC influence decision-making. IC Availability of finances. IC	Settle/handling finances. PWD, IC Handling administration. PWD, IC Handling mail. PWD, IC
Autonomy	Encroachment of autonomy of PWD. PWD Increasing dependence of PWD. PWD, IC, CM	
Involvement of person with dementia	How to manage the PWD when he/she can't co-decide anymore? IC Involvement of the PWD in decision-making depends on the stage of dementia. IC, CM Single PWD don't want to burden their children. They don't want to appeal on them. PWD	
Participants in decision-making	Unclearness about who decides. IC, CM, OFC Involvement (yes/no) of PWD in decisions. IC, CM, OFC	
Communication	Lack of communication between members of the care network. CM, OFC	

	Diverging views of IC on situation PWD. CM Different views of involved ones regarding discussion point. CM The decision has unexpected consequences. PWD, IC, CM Not everything is expressed regarding the discussion point by involved ones. CM Promises that cannot be honored by IC. CM Uncooperative family systems. CM Family members are not on the same wavelength. CM, OFC The meaning of a decision has not been well expressed. IC	
Information	Lack of timely and available information (e.g., authorization). IC Lack of (the same) information. IC Insufficient information and/or conflicting information. IC	
Professionals	Balancing of the CM between interest PWD–IC. CM Burden of IC changes at some point in overburdening (no sliding scale). CM Lack of consultative/advisory experts. IC CM is available but in need of education. IC Relation with General Practitioner (not at hand and not well informed). PWD, IC Involvement of CM in the dementia process is too late. CM Lack of overview of the system of PWD/IC by CM (PWD or IC stop this). CM Lack of attention regarding the IC after the initial phase of the dementia process. IC) Formal caregivers are solution minded. PWD experience their help as invasive. PWD, IC	
Burden of informal caregiver	(Over)burdening of IC. IC, CM, OFC Overburdened IC have unrealistic expectations of relief work. CM Overburdening is not always visible and can lead to a crisis decision. CM Feelings of guilt and powerlessness of IC. IC, CM Mourning process of IC. IC, CM Feelings of shame of IC about the disease and therefore avoiding situations and people. CM Difficulties with decision-making of IC. IC, CM Coping with difficult behavior of the PWD. IC, CM	
Vulnerability of informal caregiver	Vulnerability of the IC/partner. What if IC is forced to stop caregiving? IC IC/spouse sometimes does not permit required care for PWD because breakthrough of daily routines. CM	
Options	There is no choice. IC Limited offer of options. IC, CM Formal caregivers offer familiar care that is quickly available and effective. CM	
Timing	Timing of admission (to a nursing home) is too quick/too slow. IC, CM What is the right moment to make a decision? IC, CM Sometimes people are not ready to face certain problems or to make certain decisions. CM, OFC Early diagnose, yes/no? An early diagnosis has negative aspects (no car driving; no change of insurance company; consequences for income). IC, CM	

[a)]PWD = people with dementia, IC = informal caregivers, CM = case managers, OFC = other formal caregivers

are lack of discussion within the network, replacement of professional caregivers, and a lack of information recording. The other side of the timeline shows anticipation: early discussion of possible problems in the future, and important values of network members, specifically of the person with dementia. This enables network members to support and decide in line with the wishes and needs of the person with dementia and the endurance of the informal caregivers.

The third domain, “information and communication”, including the categorized user requirements of information, communication, and support, seems to be more generic. These user requirements are lubricants for SDM in care networks of people with dementia and are therefore indispensable. Network members often do not have the same information starting point. There is a lack of information exchange between network members, and network members do not always have access to the same information. Furthermore, important information within the network is not (always) shared; interaction between all network members is not self-evident. Interaction between network members is characterized by a one-on-one contact, rather than interaction between all network members "around the table". Moreover, informal caregivers need the support of fellow sufferers.

IV. Discussion

In this multiple methods study with an iterative participatory design, we determined the user requirements for a new interactive web tool based on experienced problems and decisions (social contacts, daily activities, mobility, safety, living, future, care, and finances), and needs and preferences of participants (participation of the person with dementia in the decision-making, insight into the decision history, anticipation of possible future problems and decisions, and the degree of self-management and autonomy preservation of the person with dementia among others). The extensive and thorough research procedure for identifying user requirements resulted in divers and rich user requirements.

Most identified user requirements address aspects of well-being. User requirements addressing care, financial matters, and future are recognized by Livingston and colleagues [4]. Zwaanswijk and colleagues [5] emphasize experienced problems with social networks. In both studies only the informal caregivers were interviewed about their needs. Van der Roest and colleagues [6] interviewed both informal caregivers and people with dementia. Most important needs they experienced address daytime activities, company and information. Each of these studies affirms some of the topics that are identified in this study that involved not only the perspectives of informal caregivers and/or people with dementia but also the perspectives of case managers and other professionals.

In our study, people with dementia described fewer problems than informal and professional caregivers. This is in line with findings of van der Roest [6] and de Boer and colleagues [36]. They suggest that this discrepancy could be due to different perspectives; informal caregivers and professional caregivers experience and rate problems differently from people with dementia (e.g., behavioral problems). People with dementia seem to be very capable of describing what is important to them: their values (e.g., loss of control, autonomy, and independence). De Boer and colleagues [36] consider also, as in our study, the contribution of people with dementia as very valuable to improve care to the experience and wishes of people with dementia.

Results show that identified problems are more related to well-being, whereas decisions (in the sense of solving problems) are more related to care. This discrepancy may be due to the focus of professionals on care. Most professional caregivers and case managers in dementia care have a professional background in nursing. This might influence the options they provide. It can be argued that, if professional caregivers focus more on well-being, the options they offer will also likely be more focused on well-being. Improving person centeredness in dementia care may support such an attitude change of professional caregivers [37]. For the new interactive web tool it is of importance that we should take into account the character of problems and decisions in order to avoid a mismatch.

The user requirements show no conflicts. However, some user requirements are complementary or overlapping. Ten clustered topics of experienced problems with no decision counterpart (e.g., communication, information, role of professionals, and involvement of person with dementia) overlap with user requirements based on the needs and preferences. Furthermore, the contributions that the participants made to the user requirements differ. Only two items were stated by all participants: "participation of the person with dementia in decision-making" and "self-management and autonomy". The informal caregivers contributed broadly; they gave input to all user requirements. In this study, people with dementia participated in the development. They made varying contributions. The information they provided could not always be easily derived from their answers. This might be due to the abstraction level of our study object: decision making. Nevertheless, people with dementia contributed to the requirements: participation of the person with dementia in decision-making, self-management and autonomy, anticipation, social contacts, mobility, living, and daily activities. Comparison of the user requirements with national dementia care standards show that they are compatible with important domains of these care standards such as participation of people with dementia in decision-making as long as possible, monitoring the well-being of the person with dementia and the endurance capacity of informal caregivers, the case manager as a coordinator, the importance of information exchange, and communication with all those involved [38][39][40][41][42].

Participation of end users in the development of the interactive web tool is a key feature in our study. A previous review showed that participation of end users, and especially people with dementia, may contribute to the development of a user-friendly and usable interactive web tool [27]. We therefore conducted this study with maximum participation of the end users. To maximize inclusion of people with dementia researchers invested in the relationship with them by spending time with them [33]. This provided us with valuable user requirements. People with mild to moderate dementia were well able to participate in interviews and focus groups and could express their preferences. This affirms Whitlatch and colleagues' assertions [20]. The recruitment of people with dementia for the interviews was sometimes difficult. Informal caregivers and formal caregivers tended to shield them from participating. They were sometimes afraid that participation would be too intrusive for the person with dementia. This is in line with Savith & Zaphiris and Wilkinson [43][44]. Another reason for non-participation of people with dementia and informal caregivers was the risk of overburdening of the informal caregiver. The people with dementia who participated in the study enjoyed the conversations and stressed the importance of their participation. They hope that their contribution will benefit future dementia patients.

TABLE IV. SECOND SET OF ADDITIONAL USER REQUIREMENTS: NEEDS AND PREFERENCES

Additional domains and categories of user requirements addressing needs and preferences of participants		
Domains of requirements	***Categories of user requirements***	***User requirements addressing needs and preferences of end users***
Involved persons and their role	Participation of the person with dementia in decision-making	The tool: - facilitates involvement of the person with dementia when discussing issues and decisions. - facilitates acting in accordance with the uniqueness of the person with dementia. - gives insight into explicit choices about the participation of the person with dementia. - strengthens the position of the person with dementia in decision-making. - supports and strengthens the person with dementia as a (co-) decider.
	Self-management and autonomy	The tool: - supports the wishes of the person with dementia about self-control and independency. - facilitates a gradual takeover of tasks appropriate to the needs of the person with dementia
	Role informal caregivers	The tool: - facilitates the monitoring of the limits of informal caregivers regarding their burdening. - supports informal caregivers in deciding in accordance with the wishes of the person with dementia. - monitors the (possibly changing) need of informal caregivers in coordinating activities.
	Role case manager	The tool: - facilitates the role of the case manager - provides the case manager with supporting methods (e.g., network analysis, options). - monitors the activities and agreements made by the case manager. - facilitates the case manager to involve the person with dementia as co-decider. - supports the case manager in strengthening the communication in the network.
	Other professionals	The tool: - facilitates adding other professionals (e.g., GPs) to the network. - facilitates one medical contact for the person with dementia and the informal caregiver.
	Organization around the person with dementia	The tool: - facilitates efficient access to underlying (care) possibilities. - detects errors and delays in the settlement of processes and procedures. - facilitates the alignment of roles and self-management tasks within the care network.
Timeline	Anticipation	The tool: - facilitates timely information about possible future issues and decisions within the network. - facilitates timely discussion about possible future issues and decisions within the network.
	Decision history	The tool: - offers information about decisions made with regard to medical, care and welfare aspects. - supports providing insight into what was discussed by whom and from what perspective.
Information and communication	Information	The tool: - provides relevant and consistent information to network members about: dementia in general; the issues that may occur regarding the disease dementia; experience knowledge of network members; information about regional dementia provisions. - supports the accessibility to the same information for all network members. - facilitates the professional (case manager) as a signpost regarding information.
	Communication	The tool: - facilitates an open communication between all network members. - facilitates exchange of information between network members. - facilitates case managers in maintaining a regular contact with the family. - facilitates the network members to be informed about what is going on within the network.
	Support	The tool: - facilitates sharing experience knowledge of network members. - facilitates professionals in supporting network members proactively. - supports decision-making, timing of decisions, and the implementation of decisions. - supports the participation process of the person with dementia.

Aims of the interactive web tool facilitating SDM in dementia are open communication, transparency, deciding together step-by-step, and giving voice to the person with dementia. The user requirements determined in this study contribute to these aims and can be used in outcome metrics of the current pilot study: does the interactive web tool enable what it promises regarding topics of decision making and preferences of end users?

A. Strengths and limitations

This study has some methodological limitations. A first limitation concerns the fact that we did not quantify the occurrences of problems and decisions because we were interested in diversity. Several methods, in sequence, were used to determine the content of a new SDM interactive web tool. In order to gain as much information as possible, we gathered all the views of the respondents that related to decision making and their situation. This highlighted a wide range of needs and preferences, and problems and decisions. The frequency of occurrence was not a criterion for acceptance or refusal. Although there was some overlap, many single problems were inventoried.

The strength of this study lies in its extensive and thorough approach. Data triangulation, using multiple methods, in-depth and comprehensive data collection, thorough analyses, and the iterative participatory approach strengthen the results of this study. Furthermore, the findings complement each other; similar results came from several data sources. We paid special attention to thorough design and procedure of the study, especially with regard to the vulnerability of people with dementia. Attendance of the principal investigator at a day care center for six days made people with dementia more at ease in the focus group meeting. Nurse education can support this change by focusing increasingly on gerontology and well-being. This change nowadays is perceived in the Netherlands that are transforming from a welfare state to a participation society [45].

V. Conclusions

With the iterative, participatory, and sequential approach we identified needs and preferences of participants, as well as experienced problems and decisions. This resulted in user requirements for a user-friendly interactive web tool that facilitates SDM in care networks of people with mild to moderate dementia. People with dementia and their informal and professional caregivers made valuable contributions.

Decision making in dementia care networks addresses predominantly problems of the well-being of people with dementia and informal caregivers. Eight categories of problems and decisions addressing decision-making in dementia are identified: social contacts, daily activities, mobility, living, safety, future, care, and finances. Additional identified user requirements for the new interactive web tool concern: the participation of the person with dementia in decision-making, insight into decision history, anticipation of future problems and decisions, communication, information, and the degree of self-management, and autonomy preservation of the person with dementia and the informal caregivers over time. The next steps will be designing, developing, and improving the interactive web tool in collaboration with the end users, then testing it in experiments. In due course, a pilot study and its evaluation will follow.

Acknowledgment

The authors would like to thank all participants of the individual interviews and focus group interviews for their contribution.

This work is supported by The Dutch Foundation Innovation Alliance (SIA RAAK [Regional Attention and Knowledge Circulation] PRO), Zorgpalet Hoogeveen (Residential care organization for older adults) and Windesheim University of Applied Sciences.

References

[1] M. Span, C. Smits, L. M. Groen-van de Ven, A. Cremers, J. S. Jukema, M. J. F. Vernooij-Dassen, et al., "Developing eHealth technology for people with dementia: towards a supportive decision tool facilitating shared decision making in dementia," The Fifth International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED 2013) IARIA, Feb. 2013, pp. 69-72, ISBN: 978-1-61208-252-3

[2] M. Prince, R. Bryce, and C. Ferri, "World Alzheimer Report 2011," Available at: http://www.alz.co.uk/research/WorldAlzheimerReport2011.pdf. Last access date 20-10-2014.

[3] C. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, et al., "Global prevalence of dementia: a Delphi consensus study," Lancet, vol. 366, pp. 2112-2117, 2005.

[4] G. Livingston, G. Leavet, M. Manela, D. Livingston, G. Rait, E. Sampson, et al., "Making decisions for people with dementia who lack capacity: qualitative study of family carers in UK," BMJ, vol. 341, 2010.

[5] M. Zwaanswijk, A. P. A. v. Beek, J. Peeters, J. Meerveld, A. L. Francke, and S. Van Beek, "Problemen en wensen van mantelzorgers van mensen met dementie: een vergelijking tussen de beginfase en latere fasen van het ziekteproces [Problems and needs of informal caregivers of persons with dementia: a comparison between the initial stage and subsequent stages of the illness process]" Tijdschrift voor Gerontologie en Geriatrie, vol. 41, pp. 162-171, 2010.

[6] H. G. van der Roest, F. J. M. Meiland, H. C. Comijs, E. Derksen, A. P. D. Jansen, H. P. J. van Hout, et al., "What do community-dwelling people with dementia need? A survey of those who are known to care and welfare services," International Psychogeriatrics, vol. 21, pp. 949-965, 2009.

[7] R. M. Epstein and R. L. Street, "Shared Mind: Communication, Decision Making, and Autonomy in Serious Illness," The Annals of Family Medicine, vol. 9, pp. 454-461, September 1, 2011.

[8] M. von Kutzleben, W. Schmid, M. Halek, B. Holle, and S. Bartholomeyczik, "Community-dwelling persons with dementia: What do they need? What do they demand? What do they do? A systematic review on the subjective experiences of persons with dementia," Aging & Mental Health, vol. 16, pp. 378-390, April 2012.

[9] S. L. Dupuis, J. Gillies, J. Carson, and C. Whyte, "Moving beyond patient and client approaches: Mobilizing 'authentic partnerships' in dementia care, support and services," Dementia, vol. 11, pp. 427-452, 2011.

[10] K. Cubit, "Informed consent for research involving people with dementia: A grey area," Contemporary Nurse, vol. 34, pp. 230-236, 2010.

[11] L. Nygård, "How can we get access to the experiences of people with dementia? Suggestions and reflections.," Scandinavian Journal of Occupational Therapy, vol. 13, pp. 101-112, 2006.

[12] H. L. Menne, S. S. Tucke, C. J. Whitlatch, and L. F. Feinberg, "Decision-Making Involvement Scale for Individuals With Dementia and Family Caregivers," American Journal of Alzheimer's Disease & Other Dementias, vol. 23, pp. 23-29, 2008.

[13] G. Elwyn, A. Edwards, and P. Kinnersley, "Shared decision-making in primary care: the neglected second half of the consultation.," British Journal of General Practice, vol. 49, pp. 477-482, 1999.

[14] D. Stacey, F. Légaré, S. Pouliot, J. Kryworuchko, and S. Dunn, "Shared decision making models to inform an interprofessional perspective on decision making: A theory analyis," Patient Education and Counseling, vol. 80, pp. 164-172, 2010.

[15] F. Légaré, D. Stacey, S. Pouliot, F. P. Gauvin, S. Desroches, J. Kryworuchko, et al., "Interprofessionalism and shared decision-making in primary care: a stepwise approach towards a new model," Journal of Interprofessional Care, vol. 25, pp. 18-25, 2011.

[16] C. A. G. Wolfs, M. E. de Vugt, M. Verkaaik, M. Haufe, P. J. Verkade, F. R. J. Verhey, et al., "Rational decision-making about treatment and care in dementia: A contradiction in terms?," Patient Education and Counseling, vol. 87, pp. 43-48, 2012.

[17] G. Makoul and M. L. Clayman, "An integrative model of shared decision making in medical encounters," Patient Education and Counseling, vol. 60, pp. 301-312, 2006.

[18] P. D. A. Ottawa Hospital Research Institute. Available: http://decisionaid.ohri.ca/index.html. Last access date 24-11-2-14.

[19] E. Duncan, C. Best, and S. Hagen, "Shared decision making interventions for people with mental health conditions," Cochrane Database of Systematic Reviews, 2010.

[20] J. Whitlatch, Menne, H, "Don't forget about me. Decision making by people with dementia," Journal of the American Society on Aging, vol. 33, pp. 66-71, 2009.

[21] Alzheimer Europe, Alzheimer in Europe Yearbook 2008. Available at: http://ec.europa.eu/health/ph_information/reporting/docs/2008_dementiayearbook_en.pdf, Last access date 24-11-2014.

[22] J. Peeters, J. de Lange, I. van Asch, S. P., M. Veerbeek, A. M. Pot, et al., "Landelijke evaluatie van casemanagement dementie [National evaluation of case management in dementia]", NIVEL, Utrecht 2012.

[23] T. Koch, S. Iliffe, J. Manthorpe, B. Stephens, C. Fox, L. Robinson, et al., "The potential of case management for people with dementia: a commentary," International Journal of Geriatric Psychiatry, vol. 27, pp. 1305-1314, 2012.

[24] A. M. O'Connor, H. A. Llewellyn-Thomas, and A. B. Flood, "Modifying Unwarranted Variations In Health Care: Shared Decision Making Using Patient Decision Aids. A review of the evidence base for shared decision making.," Health Affairs Web exclusive, pp. 63-72, 2004.

[25] Med-Decs. De internationale verzamelplaats voor hulp bij medische keuzen [Med-Decs. The international meeting-place for medical choices support]. Available: http://www.med-decs.org/nl/. Last acces date 24-11-2014.

[26] F. Meiland, A. Reinersmann, S. Savenstedt, B. Bergvall-Kareborn, M. Hettinga, D. Craig, et al., "User-Participatory Development of Assistive Technology for People with Dementia - From Needs to Fucntional Requirements. First Results of the COGKNOW Project," Non-Pharmalogical Therapies in Dementia, vol. 1, pp. 71-91, 2010.

[27] M. Span, M. Hettinga, M. J. F. J. Vernooij-Dassen, J. Eefsting, and C. Smits, "Involving People with Dementia in the Development of Supportive IT Applications: a Systematic Review," Ageing Research Reviews vol. 12, pp. 535-551, 2013.

[28] S. Visser, T. Haaker, R. Menko, and M. Hettinga, "From Buzz to Business: eHealth Innovation Matrix". Abstract submitted and accepted for Med-e-Tel 2012, Luxembourg, April 18-20, 2012.

[29] J. E. W. C. Van Gemert-Pijnen, N. Nijland, M. A. H. Van Limburg, S. M. Kelders, B. J. Brandenburg, H. C. Ossebaard, et al., "Introducing a holistic framework for eHealth technologies," Journal of Medical Internet Research, vol. 13 (4):e111, 2011.

[30] A. Bryman, Social Research Methods, third edition ed. New York: Oxford University Press, 2008.

[31] S. D. Lambert and C. G. Loiselle, "Combining individual interviews and focus groups to enhance data richness," J Adv Nurs, vol. 62, pp. 228-37, Apr 2008.

[32] M. Bloor, J. Frankland, M. Thomas, and K. Robson, Focus Groups in Social Research. London: Sage Publiations Ltd, 2002.

[33] K. Murphy, F. Jordan, A. Hunter, A. Cooney, and D. Casey, "Articulating the strategies for maximising the inclusion of people with dementia in qualitative research studies," Dementia, January 8, 2014.

[34] B. Reisberg, S. H. Ferris, M. J. de Leon, and T. Crook, "The global deterioration scale for assessment of primary degenerative dementia.," American Journal of Psychiatry, vol. 139, pp. 1136-1139, 1982.

[35] H. Beyer and K. Holzblatt, Contextual Design. Defining Customer-Centered Systems, Morgan Kaufmann Publishers, 2010.

[36] M. E. de Boer, C. M. Hertogh, R. M. Dröes, I. I. Riphagen, C. Jonker, and J. A. Eefsting, "Suffering from dementia - the patient's perspective: a review of the literature," Int Psychogeriatr. , vol. 19, pp. 1021-1039, 2007.

[37] P. J. Doyle and R. L. Rubinstein, "Person-Centered Dementia Care and the Cultural Matrix of Othering," The Gerontologist, August 6, 2013.

[38] L. J. Boomsma, D. F. R. Boswijk, A. Bras, A. H. M. Kleemans, J. M. G. A. Schols, A. W. Wind, et al., "Landelijke Eerstelijns Samenwerkings Afspraak Dementie [National dementia collaboration agreement]", Huisarts & Wetenschap, vol. 52, pp. S1-S5, 2009.

[39] L. J. Boomsma, M. De Bont, C. Engelsman, J. Gussekloo, C. Hartman, A. Persoon, et al., "Landelijke Eerstelijns Samenwerkings Afspraak Dementie [National dementia collaboration agreement]", Huisarts & Wetenschap, vol. 48, pp. 124-126, 2005.

[40] Richtlijn Diagnostiek en medicamenteuze behandeling van dementie [Directive diagnostics and medical treatment of dementia], K. v. d. G. CBO, 2005. Available at: http://www.platformouderenzorg.nl/uploads/files/guidelines/dementie1.pdf. Last access date 20-11-2014.

[41] E. Moll van Charante, M. Perry, M. J. F. Vernooij-Dassen, D. F. R. Boswijk, J. Stoffels, L. Achthoven, et al., "NHG-Standaard Dementie (derde herziening) [General

practitionars' dementia standard]" Huisarts & Wetenschap, vol. 55, pp. 306-317, 2012.

[42] AlzheimerNederland and Vilans, "Zorgstandaard Dementie," ed [Dementia care standard], 2012. Available at: http://www.alzheimer-nederland.nl/media/19342/Alzheimer%20Nederland_Brochure_Goede_Zorg_Dementie_LR.pdf. Last access date 24-11-2-14.

[43] N. Savitch and P. Zaphiris, "Accessible websites for People with Dementia: a Preliminary Investigation into Information Architecture," in ICCHP, 2006, p. 8.

[44] H. Wilkinson, "Including people with dementia in research: methods and motivations". In: The perspectives of people with dementia, research methods and motivation. London: Jessica Kingsley, 2002.

[45] C. H. M. Smits, H. K. van den Beld, M. J. Aartsen, and J. J. F. Schroots, "Aging in The Netherlands: State of the Art and Science," The Gerontologist, September 2, 2013.

On Benefits of Interactive Online Learning in Higher Distance Education

Repeating a Learning Analytics Project in the Context of Programming Education

Klaus Gotthardt, Bernd J. Krämer
Faculty of Mathematics and Computer Science
FernUniversität in Hagen
Hagen, Germany
klaus.gotthardt@fernuni-hagen.de
bernd.kraemer@fernuni-hagen.de

Johannes Magenheim, Jonas Neugebauer
Department of Didactics of Informatics
University of Paderborn
Paderborn, Germany
jsm@uni-paderborn.de
jonas.neugebauer@uni-paderborn.de

***Abstract* - All generations of the Web have given rise to new technologies and services for teaching and learning. Ample research projects have investigated their impact on educational processes and students' learning outcomes. This is especially true for the comparison of online and face-to-face learning or distance and classroom instruction. There is, however, very little original research and evaluation evidence about the impact of incorporating web-based learning technology and multimedia content in higher distance education. This article presents the results of two studies with distance students of FernUniversität. These results address four strands: the students' competence gain on key course topics; an analysis of the learning behavior of a subset of online students; a performance comparison between online students and a control group using classical correspondence learning materials; and online student satisfaction with the online version of the course and the e-learning environment. Authentic comments from two students who were involved in all evaluation strands supplement the formal findings.**

Keywords - distance learning, Web-based learning, learning analytics, competence analysis, learning performance, technology assessment.

I. Introduction

Distance education describes an organization of educational processes in which teachers and students are separated in space and time and communicate with each other mainly asynchronously. Traditional teaching media were and still are prepackaged self-instructional correspondence courses that allow distance students to study at the time and location of their choice.

Starting out in the early '90s, this teaching model was challenged by the advent of the Worldwide Web. The initial Web of Information opened up the possibility to exploit interactive multimedia content and deliver learning materials electronically. It also allowed their adaptation to different learning styles by drawing appropriate learning objects from the Web or content repositories. Hypermedia structures in online courses could be used to lead students on different paths at the students' own paces through the course content, depending on their individual performances in intermediate online tests and quizzes.

The Web of Services then brought about the possibility to include all kinds of collaboration and communication services in online learning activities. More recently, the Social Web offered a range of social software tools that allowed students to participate actively and in real-time in educational processes, independent of any physical proximity.

Numerous studies have compared online and face-to-face learning to understand the impact of web-based learning technology on higher and extended education. There is, however, very little original research and evaluation evidence about the possible advantages and disadvantages of incorporating web-based learning technology and multimedia content in higher distance education. To fill this gap, an interdisciplinary group of researchers from FernUniversität in Hagen and the University of Paderborn performed a quantitative study and experimental research on online learning. Subjects were volunteer students from FernUniversität in Hagen, the only German language distance teaching university. The project set out to find answers to the following research questions:

1. To what extent can the learning objectives of a distance-learning course be achieved with traditional custom textbooks and asynchronous tutoring activities using email, text forums, and phone?
2. Is there a significant difference in competence gains and learning outcomes between students relying on traditional distance learning settings and online distance students?
3. Do students accept or even prefer online learning technologies to traditional correspondence courseware?

During the summer semester 2012, we invited all 693 students enrolled in the distance-learning course on "Object-oriented Programming" (OOP) to participate in a pre- and post-test evaluating the students' modeling and comprehension competencies in the topic area. Further, we asked for volunteers who would agree to study a new interactive version of a selected course module online and allow us to log and evaluate their online behavior anonymously. The group of online students was also asked to perform a technology assessment at the end of the course.

Study settings, evaluation methods and tools, and preliminary evaluation results of this project were published in the proceedings of the Sixth International Conference on Mobile, Hybrid, and On-line Learning [1]. Due to the small number of subjects participating in all phases of the study, these results were, however, statistically not reliable. Therefore we decided to repeat the study in the winter semester 2013/14 with the same course but a different group of students.

In the following section, we summarize the evolution of higher distance education over the last two decades. We also provide a brief review of significant research close to the work reported here. Section III sketches the learning setting in which the study was performed, including course content, learning objectives, and the newly designed digital learning environment. Section IV presents the quantitative study design and sketches the research methods and tools used in the analysis of the data gathered in the form of log data and surveys. This section also describes the modifications applied to the competence test used in the second survey. Section V summarizes the results of the competence analysis. The interpretation of the log data collected from online students is described in Section VI. Key results of the technology acceptance test are reported in Section VII. Then, Section VIII relates the test scores of those students who attended the final examination to their online retention times and interaction with learning objects. Here, we also compare the examination results of the online students with those of a control group of offline students. The latter used traditional correspondence materials for learning. Section IX summarizes the feedback from two online students who were interviewed 6 months after the end of the course. Section X concludes the paper and presents the lessons learned from the two projects.

II. Background and Related Research

The new technological and communicative options of the Web allowed educators at distance-teaching universities to tailor course design and instruction to interactive online learning [2]. As an intermediate step, existing textbooks were supplemented with interactive multimedia systems. They were delivered electronically through the Web and allowed students to experiment with simulated or animated virtual worlds, access remote labs, and thereby get meaningful feedback on student-system interactions in real-time [3]. Asynchronous e-mail, chat tools, or text forums and synchronous webinars allowed educators to narrow the distance between fellow students and teachers concerning social interaction possibilities.

Ever since distance education emerged, researchers and educators tried to answer the question whether distance learning can be as effective as learning in face-to-face settings. T. Russel has summarized a rich body of literature addressing this question in an annotated bibliography entitled "The No Significant Difference Phenomenon" [4]. The essence of this work is not the claim that distance education and classroom education are equally effective for all students, courses, and instructors but that individual variations and extraneous variables may render many of the findings in related literature inconclusive. A huge meta-analysis of empirical literature on the comparison between distance education and classroom instruction basically supports Russel's conclusion [5].

More than thousand empirical studies can be found in research literature comparing online to face-to-face learning, measuring student learning outcomes, or proposing innovative learning designs. The meta-analysis in [6] found that, on average, online students performed slightly better than students who attended face-to-face instruction. In detail, opinion is divided, however. Ferguson and Tryjanowski can show that students in their face-to-face class scored significantly higher than online students [7]. In contrast, Cooper's study on a "Fundamentals of Computer Applications" course with 94 in-class and 37 online students, could not observe any significant difference in scoring between the two groups [8].

Today, many people equate distance learning and online or e-learning, but, in fact, these concepts do not refer to the same thing [9]. Physical distance is not a defining feature for e-learning, but it is for distance learning. The target population of distance universities is students who are unable to regularly attend on-campus classes due to family or work commitments or because of personal limitations. These students depend on the institution "distance-teaching university" because it offers:

- Complete distance learning programs,
- Learning materials especially designed for self-study,
- Effective asynchronous communication media,
- Flexible examination conditions,
- Well-trained tutors, and
- A network of study centers.

Against this background, the key question is not: "online or face-to-face learning". Rather, the issue is how online learning solutions can be effectively used in higher distance education.

III. Setting the Scene

In this section we introduce the learning setting both for A) traditional higher distance learning with correspondence material and established distance learning support tools and B) a newly designed online version of a selected course module.

A. Course Objectives, Structure and Learning Support

Since 2004, the course OOP is taught on the basis of correspondence materials consisting of 41 chapters summing up to a 500 pages textbook. The course objectives claim that by the end of the course, successful students will be able to:

- Identify, explain, and properly apply at least 10 core concepts of object-oriented programming. They include: object, class, class relationships, single and multiple inheritance, interface, constructor, method and method invocation, variable, array, data and object type, program flow statements, and package.
- Design appropriate data structures and algorithms solving a given problem.
- Govern the essential sequential programming concepts of Java.
- Develop the data structures and algorithms of a design solution into a correct Java program.
- Understand and modify a given Java program.

- Distinguish between and implement iterative and recursive algorithms.
- Read and interpret simple UML diagrams.

The course content is divided into 7 course modules that are processed every other week. Besides course-related learning objectives, more detailed learning objectives are set for each module. Questions and self-assessment exercises with sample solutions and software-based self-tests providing real-time feedback aim to provide stimuli and learning guidance.

In distance education, regular task blocks, course-related newsgroups and forums replace exercise hours of on-campus instruction. Students edit the optional task blocks in a 14-day cycle. To this end, they use FernUniversität's online exercise system to submit their solutions and get automated responses. Student solutions can be IMS QTI-type interactions, but in this course they mostly consist of program snippets or small programs the students have to develop themselves. Once a program is uploaded, it is automatically compiled on the remote university server. If the compilation succeeds, the resulting code is run through a style checker and a sequence of pre-defined test cases. The style checker and test results provide a quick feedback and aim to motivate the student to improve his or her solution if some test cases failed or the programming style is weak. This edit-compile-test cycle can be repeated as often as needed until the final delivery schedule is reached. A tutor then corrects the submitted final version of each student program manually to provide individual feedback.

B. *Digital Learning Environment*

Distance students possess above average experience with self-directed learning. They are used to organize their learning freely but have limited time to participate in synchronous learning events. This requires the barriers for group learning actions to be kept low.

To challenge students in the online group to perform more demanding learning activities, we redesigned both the instructional design and the content of one module of the course substantially. The selected course module deals with program exceptions, testing, program documentation, and error handling. The learning objectives of this module require both conceptual knowledge and practical skills. According to Bloom's taxonomy of learning objectives [10], these are, in particular, application, analysis, and synthesis skills. Concept knowledge is needed when students must:

- Recognize the benefits of meaningful program documentations.
- Understand the difficulties and limitations of program testing.
- Identify and properly apply error-handling techniques.
- Avoid typical programming errors.

Practical skills are needed when students must:

- Document their own and foreign Java programs in a meaningful way using Javadoc.
- Write Java code that raises meaningful exceptions and properly intercepts them or passes them on to a higher level of exception handling.
- Plan test cases, perform program tests, and interpret test results.

The online course module, which can be access at [11], addresses exactly the same course topics and learning objectives as the textbook version. Besides short reading passages in the form of HTML pages capturing basic factual knowledge students need to learn, the online module includes ten interactive learning objects. They are implemented in Adobe Flash and allow the students to experiment with alternative solutions of program designs, explore, modify and explain the behavior of given program solutions, evaluate their own solutions, or grasp the semantics of certain programming concepts in a trial-and-error fashion. To encourage teamwork between physically remote students, we also designed a few new learning tasks that involve 2-4 students playing different roles in the programming task, such as a programmer or tester. Furthermore, we proposed homework assignments whose solution was composed of several program modules to be contributed collaboratively by several students.

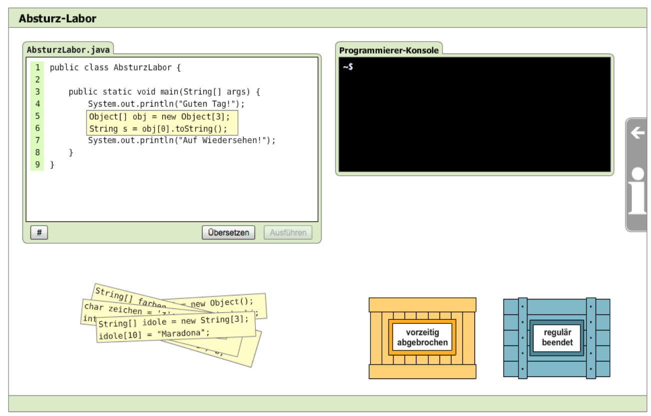


Figure 1. Crash lab providing explorative programming experiences.

Fig. 1 depicts one of the interactive learning objects, called crash lab. It allows students to explore the behavior of programming exceptions. They drag one of the program statements in the pile underneath the editor window on the left into a code frame provided in the editor window. Then, they push the button "Übersetzen" to compile the resulting code and run it if the compiler succeeds (in which case the button on the right is highlighted). Students are supposed to predict whether the different programs they build this way will fail or terminate successfully. The results of the test runs are collected in the two boxes labeled "vorzeitig abgebrochen" (aborted) and "regulär beendet" (regularly completed) at the bottom right.

An example of a team problem is a simple game with a treasure being hidden in an area of 24 cells. A player has to find a treasure by moving strategically in this area. One student has to develop the program component controlling the game; the partner has to implement the behavior of the player. Some constraints are imposed on the behavior of both components, which may lead to program exceptions.

The online course was delivered through the learning management system Moodle [12]. It was supervised and tutored by the same faculty who also taught the textbook version. The installation of Moodle we used in the study was seamlessly connected to the learning object repository edu-sharing [13][14]. This repository enabled the online students to share and collaboratively work on their contributions in a protected space. The repository functionality allowed them to control which access rights to their personal workspace they wanted to grant to selected peers.

IV. Quantitative Study Design

To acquire data about the students' perception of the new online learning environment and their competence gain on the subject of this course, two online questionnaires were developed.

Our first experimental study was conducted during the summer semester 2012 [1]. It consisted of a:

- Pre- and post-test to determine the students' competence gain in core topics of the course.
- Learning behavior analysis of a subgroup of online students based on various log data.
- Technology assessment of the online student group.
- Analysis of homework submissions.
- Written examination at the end of the semester covering all learning objectives for the course.

Unfortunately, a disappointingly low number of students participated in the post-test performed in July 2012. As a consequence, we were unable to determine the competence gain of the anonymized subjects. In addition, we observed hardly any overlap between students who participated in the competence tests and in the final examination. This fact prevented us from correlating competence gains and examination scores. For similar reasons, a statistically valid comparison of the final grades of online and offline students was impossible.

These weak results encouraged us to repeat the study in the winter semester 2013/14. To reduce the problems of the first study, we adapted some study parameters that had caused problems before. These modifications are discussed in Section C.

A. Learner Satisfaction Analysis

Besides acquiring a detailed insight into the students' online behavior, we also wanted to collect their satisfaction with the online course material and the learning environment provided.

Based on Davis' technology acceptance model (TAM) [15], Venkatesh's extension TAM2 [16], and Brooke's System Usability Scale (SUS) [17], we developed a technology acceptance questionnaire tailored to our specific learning environment. Both TAM and SUS have been successfully used in the study of e-learning environments [18][19]. We also considered the most recent model TAM3 [20] but did not find extensions like "Perceptions of External Control", "Computer Anxiety", "Perceived Enjoyment" and others relevant for our learning analysis purpose. The online questionnaire we gave our students includes 28 Likert-scaled items and covers the following analytical dimensions:

- Preferred computer equipment.
- Individual experiences with e-learning.
- Usability of the online course and tool environment.
- Subjective evaluation of prior knowledge on the course topics.
- Communication and cooperation competencies gained in the online course.
- Comparison of the effectiveness of the study process with traditional distance learning textbooks and the online study material, respectively.

At the end of the semester, the questionnaire was delivered online via LimeSurvey [21]. Those students who attended the online-course and also agreed to participate in the behavior analysis were asked to work on it.

B. Competence Measurement

For the competence analysis, a new competence measurement instrument was applied. The instrument was developed in several steps in the project "Measurement Procedure for Informatics in Secondary Education (MoKoM)" by researchers in the fields of didactics of informatics and psychology. The German Research Foundation (DFG) funded the MoKoM project. The goal of the project was the development of an empirically sound competence model and corresponding measurement instrument for the domains of informatics modeling and system comprehension.

In the first step, several relevant national and international curricula and syllabi like the ACM/IEEE Computing Curriculum 2011 [22] and the ACM Model Curriculum for K-12 Computer Science [23] were analyzed and used to derive a theory based competence framework. It consisted of three cognitive and one non-cognitive competence dimension.

To refine this framework, the Critical Incident Technique was used to conduct 30 expert interviews with experienced persons in the field of computer science education. Each interviewee was presented four problem scenarios and was asked how (s)he would solve the given task [24]. The scenarios were derived from the theoretical competence framework. The transcribed interviews were analyzed by means of the qualitative content analysis [25]. The results were used to refine and restructure the competence framework. This process led to an empirically refined competence model with six dimensions: K1 System Application, K2 System Comprehension, K3 System Development, K4 dealing with system complexity and K5 Non-Cognitive Skills [26].

In the next step, for each competence a test item was developed to compile a measurement instrument for the considered domains. The items were developed following the principals of Situational Judgment Tests and the experiences made in other competence measurement studies like TIMMS and PISA [27]. Based on detailed competence descriptions, tasks for every single competence item were created. To ensure an objective and coherent evaluation, a comprehensive grading manual was created alongside the test items. Due to the resulting large amount of test items, the instrument was split into six blocks and then compiled into six booklets with three blocks each. Additionally, each booklet contained a block for the non-cognitive dimensions of the competence model. These booklets could be completed in 90 minutes and were used in a test with 800 students in upper secondary education computer science classes.

To use the instrument in the study at hand, all items were transformed into respective online versions. The resulting questionnaire was delivered through the online survey system LimeSurvey. Due to the nature of the questions, some items could not be transformed properly, e.g., the drawing of diagram. For each of these items, a decision had to be made whether the associated competence was appropriate for the field of OOP. If the answer was positive, a new item had to be developed. As we wanted to exploit the breadth of the competence model and expected a high number of participants in the test, the partition into six booklets was kept for the first online survey conducted in 2012.

To allow for an anonymous survey, an additional item was added to ask for a unique code. This code was generated individually for each student based on personal information. This allowed for the association of pre- and post-tests without revealing the students' identities.

To assess the competence gain of the students, they were asked to complete the online survey created from the MoKoM measurement instrument twice: Once at the start of the term and once at the end. The students were randomly partitioned into six groups. Each group had access to one of the six test booklets provided in LimeSurvey.

C. Revision of the Competence Analysis Questionnaire

The underwhelming results of the first survey (see Table V and Section V.A) led us to redesign the test instrument for a second survey that was carried out during the winter semester 2013/14. To get more students to finish the test, the target was to enable its completion within 60 minutes. For this reason, instead of testing the complete range of competences of the MoKoM model, a subset of items especially tailored to the requirements of the course at hand was selected. To choose the appropriate test items, we matched the learning objectives of the course with the competence descriptions of the MoKoM model. This way, we could accumulate a proper selection of relevant test items and combine them into a custom-designed competence test. This tailoring also eliminated the need for several test booklets and let us assume that more students than before would now work on the same set of items.

D. Study Groups in the First Study

Before the course started in summer semester 2012, we invited all 693 students enrolled to evaluate the online competence test. To raise their interest in the study, we announced to give away three books in a raffle based on voluntarily provided email addresses. 146 students followed this request and worked on the test with different degrees of completion. They formed the *competence study group*.

Then we asked for students who would be willing to study the online course unit and allow us to track their online behavior. We found 12 volunteers who formed the *online student group*. This group was also invited to participate in a technology assessment survey conducted after the online course module was passed.

At the end of the course, we asked the whole student population again to evaluate the competence (post-) test and mark it with the same code the students had used in the pre-test.

Finally, we wanted to differentiate the test scores of final examination participants among online and offline (textbook) students to search for significant differences between both groups. The written examination test typically consists of 6 or 7 problems addressing a) major levels of Bloom's cognitive taxonomy and b) core course topics. A typical test includes:

1. A block of 5-7 questions testing the understanding of core OOP concepts and their relationships.
2. One or two problems for which the meaning of a given program must be interpreted.
3. Four constructive tasks including the design of a class structure and the implementation of selected methods, the design of an iterative and a recursive algorithm for a given computational problem, and the design of simple data structures and algorithms operating on them.

The test lasts two hours and is conducted simultaneously in more than 10 different locations throughout Germany, Switzerland, and Austria (usually in lecture halls of universities) under the supervision of FernUniversität academic personnel.

E. Study Groups in the Second Study

980 students were enrolled in the course OOP in the winter semester 2013/14. Following the communication process of the first study, all students were informed about the study before the course started. Simultaneously, they were asked to work on the competency pre-test and volunteer as online students and thereby allow us to track their online behavior.

To inspire the participation in these studies and balance out the extra workload, we offered all students 5 bonus point if they completely answered both competence tests and up to 5 bonus points for online students, depending on their degree of online learning activity.

165 students volunteered for the online study. As we considered these students as particularly motivated, we decided to accept only 98 subjects as *online students* and form a control group of motivated students, called *offline students*, from the other 67 volunteers.

At the end of the course, all students enrolled were asked to work on the post-test using the same personal code they had used in the pre-test. The online students were asked to evaluate their subjective impression of the learning environment by answering the TAM questionnaire. The students who had registered for the final examination were requested to provide their personal code on the examination paper and thus allow us to correlate competence test results with the examination test scores.

V. Results of the Competence Analyses

We separate our discussion of the analysis results into two subsections, one for each study.

A. First Study performed 2012

Unfortunately, the participation in the survey performed during the summer semester 2012 was disappointing. Of the 693 students registered for the course, only 57 started their pre-test booklet and just 19 students in total finished more than 75% of all items. These numbers got even worse for the post-test with 30 subjects who started to work on the booklets

and only 5 finished it. IN both tests intervals around 150 users visited the survey page but only one third and one fifth, respectively, started to answer the first item at all. Since the participants who finished their booklet were spread across the six booklet variants, the useable data for each test item was too small to get any meaningful results regarding the competence gains of the students. Only one student completed both tests. Even tendencies supporting or falsifying any of the hypotheses we started from are hard to state.
To at least get a notion what to change for subsequent tests, the answering habits of the participants were examined. By counting the number of students who tried to solve an item over all 87 datasets, it is easy to see that the completion rate drops to 60% after only four items (see Fig. 2). The eighth item was completed by only 50%.

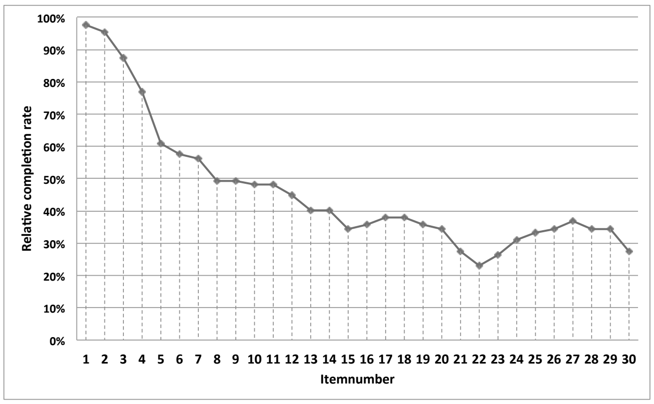


Figure 2. Relative completion rate of survey items.

A reason for these results might have been the length of the competence test. Tailored to be conducted in a German classroom setting, where two successive lessons equals 90 minutes, it seems the time a distance-learning student is willing to spend on his own in answering the survey is considerably shorter. Examination of the response rates for online surveys even showed that the ideal length for an online survey is thirteen minutes or less [28]. This ideal timeframe is, however, too short for a competence test. As discussed in Section IV.C, the average workload for the revised competence test was reduced to 60 minutes.

B. Second Study Conducted 2013/14

This time, the pre-test was called 162 times. After cleanup of duplicates and removal of data sets not completely answered, 126 evaluable records remained. The post-test was edited only by half as many students and just produced 49 complete data sets. Based on the personal code used on both tests, 40 data sets could be paired with each other to compare the results on an individual though anonymous basis.

In each test, a maximum of 77 points could be obtained from 21 tasks. The blue bars in Fig. 3 depict the points students achieved in the pre-test, while the red bars on top indicate the gain in points in the post-test. Four students even showed a negative gain ranging between -3 and -28 points less in the second test.

In the pre-test, an average of 43.54 (57%) points was achieved, in the post-test 55.15 (72%). Based on the personal code, 40 data sets of both tests could be paired to see the competence gain of individual students (see Fig. 3).

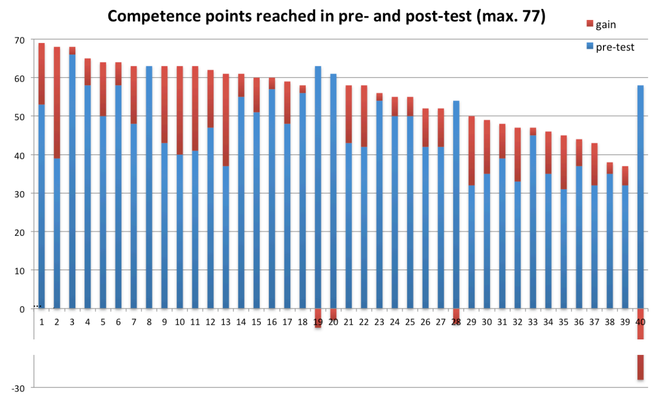


Figure 3. Competence gain of anonymized students.

Based on this reduced sample, the mean value increases in the pre-test to 46.38 points (60%), while it nearly stagnates in the post-test with 55.25 points (72%). This result represents a significant improvement in skills (t (39) = 5.68, p <0.001). The boxplot in Fig. 4 visualizes the distribution of points.

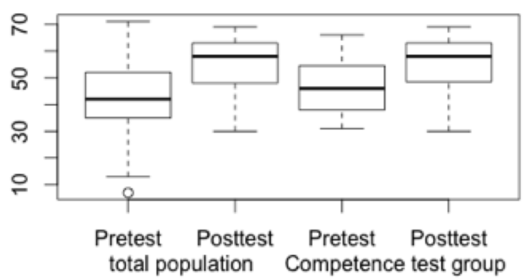


Figure 4. Boxplot of the competence points achieved.

Out of the 40 pairs of competence analysis sets, 16 were delivered from online students and 24 from students in the control group (offline students). With regard to the total number of points, both groups indicate significant increases. However, the online group improved by an average of 10.7 points compared with 7.6 in the control group (see Fig. 5).

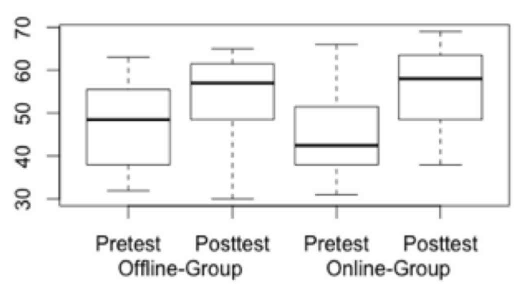


Figure 5. Boxplot of the competence points achieved.

But this difference turns out to be insignificant. At the item level, the post-test reveals significant differences in two tasks: The online group explains the concept “inheritance” more reliably and can better explain the function of a stack. Fig. 6 depicts the increase of competences on the topic “ordering of test sequences” as a bar chart.

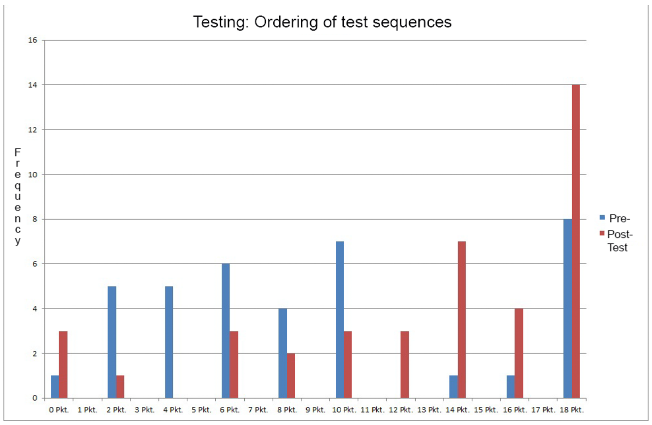


Figure 6. Pre- and post-test results of competence test for topic “ordering of test sequences.

The shift from lesser to more points from the first to the second test is obvious.

VI. Behavior Analysis of Online Students

The number of online students involved in the first study (12) is too small to provide statistically valid results. Moreover, as only 6 of the online students took the examination test at the end of the semester, these figures are not sufficient to compare the outcomes of online and offline students. Therefore, we focus our discussion about the online behavior of distance students enrolled in the course OOP on the results of the second study. Here, 57 of the 98 volunteer online students we had accepted were finally active in the learning environment. But first, we sketch the data sources and analysis techniques employed.

A. Data Sources and Analysis Techniques

The database for the behavior analysis of the online students was compiled from the log data provided by Moodle and edu-sharing and the log data captured by the Flash-based interactive learning objects. The latter allowed us to see how successful a student was in the interaction with a learning object, which errors he or she made, and how often a student tried to solve a given task implemented by an object.

All log data were time-stamped. These time-stamps helped to integrate the data coming from different sources. The raw data were cleaned and integrated to a single database that was then analyzed with the help of the business analytics software SAS [29]. SAS was particularly used for structure and usage mining. Structure mining relies on the links between information pages and links from within course pages to self-assessment examples, homework assignments, forum entries, and objects maintained in the repository and workspaces.

The objective of structure mining is to identify recurring patterns of behavior, e.g., in the form of paths through the learning materials or repeated experiments with exercises and programming problems. These paths form a network that visualizes how students navigate through the course material and the learning environment. Particular indexes of the network analysis are the weighted in- and out-degrees of course elements, which indicate the frequencies of visits.

Usage mining provides useful descriptive statistics. This includes:

- Information about the number of page visits.
- Retention time, i.e., the cumulative elapsed time between page views; an average value was assumed for exit pages.
- Typical entry and exit pages.

B. Results of the Second Behavior Study

Among the 98 students who had been admitted to the online study group, a total of 55 subjects logged-in at least once in Moodle. 52 students accessed at least once the online content. Of these 52, twelve students spent less than 15 minutes online, 19 were less than an hour online. As these overly short online times seem unreasonable, we chose a retention time of one hour as the lower threshold for the data analysis. This decision left 33 online students to evaluate.

The study conditions with remote subjects who learn autonomously render it impossible to know whether a student’s online time was fully dedicated study time. However, the retention times neither vary conspicuously by clock time, nor did we detect meaningless values. A realistic workload for the online learning module should, however, be much higher than the measured average retention time. It is therefore likely that some online students had also used other study materials for learning, e.g., the course textbook.

The number of pages viewed varies from 25 to 696 with an average of 169 and a median of 142 page views. The average time spent on a text page varies from1:00 to 4:51 minutes (median: 02:21 minutes). The maximum time spent on a page varies between 8:32 and 107:38 minutes (median: 58:41 minutes). Although we cannot be sure whether students spent this time on effective learning, the high values appear plausible, since they were observed for self-assessment pages. The highest values occurred for pages with solutions to self-assessment tests on the topics “exceptions” and “testing”. A quiz on topic “program documentation” had also very high page view figures.

TABLE I. Retention Times Shared of Page Types

Chapter	Documentation	Exceptions	Testing	Error handling
Reading & Understanding	51.68%	54,64%	55,52%	75,34%
Self-assessment	30.11%	42,99%	32,46%	18,98%
Forum	18.21%	2,37%	12,02%	5,68%

Table 1 shows the retention times students spent on the average per chapter on the different page types. We distinguish between knowledge acquisition, practice, and forum activities. The former occur when students need to read into and understand new course topics. In the online course, information pages are intertwined with self-assessment pages. On the latter, students find interactive learning objects and programming tasks to test their understanding and give them hands-on experience with smaller programs. Each chapter has its own forum, which takes a prominent place at the beginning of each chapter.

Students spent between 52% and 75% of the time on reading and understanding new topics, 20-43% for self-assessment tests, and 6-18% in the forum. The differences may be explained by the fact that Chapter Error Handling includes only one self-assessment exercise, while Chapter Documentation provides 6, Chapter Exceptions 9, and Chapter Testing 13. Although students visited the forums quite frequently, only a few students contributed by posing and answering questions or exhibiting an own problem solution. The largest proportion of practice time was spent on the topic Exceptions. The corresponding chapter offers a few challenging programming tasks and interactive learning objects for self-assessment.

The usage data collected through the learning objects record very detailed observations. They show that about half the online students interacted with them quite intensively, whereas the other online students largely ignored them. In the discussion of the examination scores in Section VIII we will fall back on these observations.

Interestingly, 30% of the volunteers for online study were female students although their share among the students enrolled in the course is only 25%.

C. *Reading into the Log Data of Interactive Learning Objects and Programming Tasks*

The log data captured from the students' interactions with Flash objects and a built-in compiler back-end to compile and run own solutions to programming tasks provide some interesting insights. In the following discussion we focus on learning activities related to Chapter Exceptions because this theme was particularly tested in the written examination.

For instance, 17 online students worked with the learning object Crash Lab depicted in Fig. 1. Together they spent 2:40 hours working with this object. 16 students placed all 6 statements properly, 9 students created 1-3 own statements. Only five students evaluated the predicted behavior of the test programs, which was correct in four cases. Students #7, #27, #40 and #44 were successful in the first attempt. We explicitly indicate anonymous student identities here because we will refer to them later in the discussion of grades in Section VIII.

Another learning object tests the students' ability to understand the passing of exceptions. It is composed of 4 sub-tests in which the last sub-test can be varied and explored several times. 13 students interacted with this learning object, summing up 3:11 hours total interaction time. 11 students successfully solved the first sub-test and 9 students also completed the second sub-test. Finally, 5 students worked successfully on the other two sub-tests as well, including students #14, #27, #40, and #44. We observe a decreased activity rate with each sub-test, which supports the author's intention to increase the difficulty of the sub-tests step by step.

An investigation of the log data of a learning object in which students had to write code reveals that one student performed significantly worse than his or her peers who worked on the same object. This behavior seems to exhibit a trial-and-error strategy as opposed to the structured progress of the other students.

VII. TECHNOLOGY ACCEPTANCE

In the second study, 34 online students completed the TAM questionnaire. The results show that the majority of students has at least two devices in the selection "notebook, tablet, smartphone, or desktop" to access the online study material. Only 16% of all respondents used the services of the learning environment less than once per week, 61% used it two or three times, or more. In particular, the self-assessment tasks and reading materials were retrieved and used frequently.

The online students exhibit a high degree of technical affinity: 85% work frequently with web applications and 67% are online most of the day. Most students found the ergonomics of the e-learning environment convincing (see Fig. 7). It consists of the online course and the learning tools Moodle, edu-sharing, news, forum, and chat.

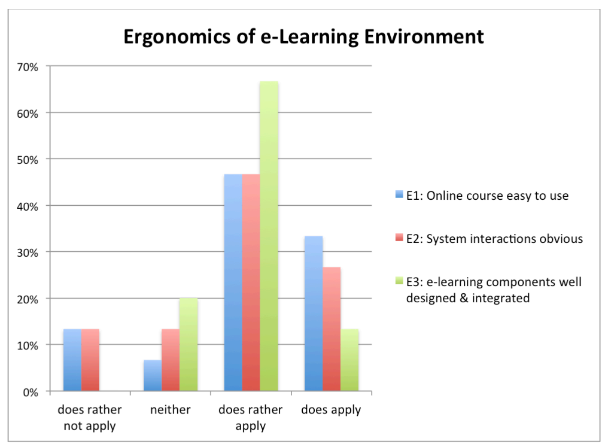


Figure 7. Student perception of the e-learning environment

Fig. 8 illustrates the students' evaluation of the value of the online course module to mediate methodological knowledge about program documentation. The high values of positive answers (84% and 78%, respectively) to the items:

Q1: The online module helped me to understand the importance of program documentation and quality assurance.

Q5: I'm able to produce a program documentation using Javadoc.

confirm that only a minority of online students was unable to assess the learning support function of the online version of the course positively.

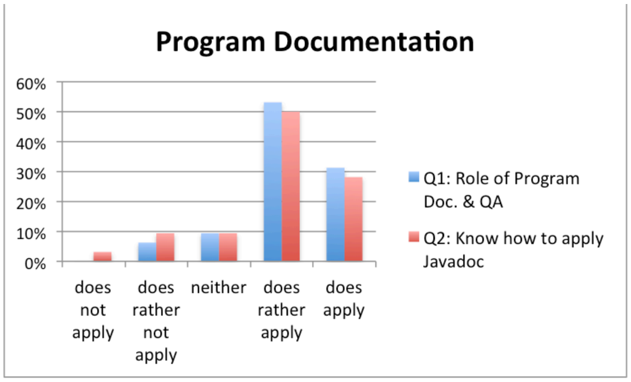


Figure 8. Mediation of methodlogical knowledge on program documentation. Q1: The online module helped me to understand the importance of program documentation and quality assurance. Q5: I'm able to produce a program documentation using Javadoc.

Similar questions were raised for the other key topics of this course module, including program testing, exception handling, and error recovery. Table II summarizes the students' perception of the e-learning environment's ability to convey methodological knowledge on program testing. Again, the positive answers predominate with the exception of constructive knowledge and practical skills needed to work with JUnit. Note that such skills were also lower rated in the domain program testing.

TABLE II. IMPARTING OF METHODOLOGICAL KNOWLEDGE ABOUT

Question	Rather applies	Applies
Q6: I understand the possibilities and limits of testing	53%	25%
Q7: I know about test levels and test methods	67%	15%
Q8: I know how to identify and apply test cases	48%	12%
Q9: I'm able to implement test cases in JUnit	24%	18%

Fig. 9 shows again a high confidence in factual and conceptual but a weaker result in procedural knowledge about exception handling.

Most respondents also certify that the online learning environment exhibits a slightly higher motivating factor. This increased motivation does, however, not carry over to cooperative tasks. The vast majority (82%) did not feel encouraged to interact with other students and to solve team tasks cooperatively. An aversion to synchronous person-to-person interaction over a distance is also reflected in the preference of asynchronous communication methods such as email and forums over synchronous forms such as instant messaging and Skype. This trend towards independent study habits is confirmed by the low interest in cooperative problem solving reported in Section VI.B and previous experience of the course tutors. The main reason quoted for the low interest in cooperation with peer students is a job-related lack of time. A few respondents also saw no reason to contact other students to help them get over hurdle because they had solved their study problems themselves, e.g., by researching the Internet.

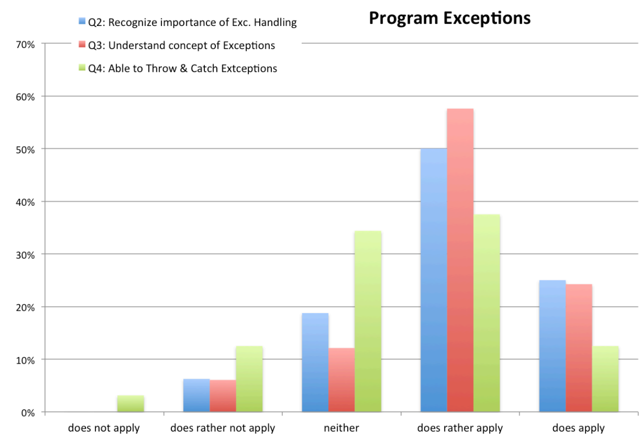


Figure 9. Mediation of methodological knowledge on exception handling. Q2: This course module animated me to reflect on exception handling during program development. Q3: Now, I understand the concept of exceptions. Q4: I'm able to throw and catch exceptions in Java

At the beginning of the semester, the students had not much prior experience with program testing and exception handling. Only about a quarter was knowledgeable in these areas. This changed after completion of the course unit addressing these topics. About 60% of respondents now believed to be able to identify proper test cases and test levels. More than 70% were convinced to know the possibilities and limits of testing. This self-evaluation is also confirmed by the competence test (see Section V.B). Here, the tasks related to testing showed the statistically most significant improvements in competences.

Several items in the TAM questionnaire addressed a comparison of the classical correspondence material and the online version of the course module. Table III summarizes the students' judgment.

TABLE III. STUDENT PREFERENCES: ONLINE VS. DISTANCE STUDY TEXTBOOK

Question	NA	RNA	NN	RA	A
The online module illustrates complicated facts better than the traditional textbook	0%	16%	25%	47%	13%
Due to the online module I feel easier to work on a typical problem from the topic area than with the traditional textbook.	7%	13%	40%	33%	7%
I prefer the online course to the classical textbook	0%	27%	33%	20%	20%
I prefer to study with the textbook	9%	25%	19%	34%	13%
I would like to see more course content be offered in a similar online version	0%	7%	20%	47%	27%
I totally dislike e-learning	64%	6%	18%	9%	3%

NA: does Not Apply; RNA: does Rather Not Apply; NN: Neither; RA: Rather Applies; A: Applies

The responses to these questions suggest that the students are rather indifferent towards the use of the online course material compared to the traditional textbook. Though most students are open-minded regarding the new learning opportunities (74% would like more courses to offer online material) and a majority of 60% acknowledge, that the online version can help to illustrate complicated matter in better ways, only 40% of the students actually prefer the online course to the classical textbook and 47% prefer to study with the classical version exclusively. These answers suggest, that the students think of the online course as a supplement to the traditional textbooks. An explanation might be the new experience the online material offers and the lack of experience with it. Thus, the students don't want to rely on an e-learning approach alone. When online courses become more common, the willingness to utilize them might increase.

VIII. Examination Results Put in Context

The maximum test score students can reach in the final examination of this course is 100 points. At least 50 points are needed to pass the test. The grade acquired for the course is derived from a student's score in the final examination. Table IV illustrates the mapping of test scores to grades. Grade 5.0 means failed.

TABLE IV. Scores mapped to Grades

Score Points	Grade
100 - 95	1.0
94 - 90	1.3
89 - 84	1.7
83 - 80	2.0
75 - 79	2.3
74 - 69	2.7
68 - 65	3.0
64 - 60	3.3
55 - 59	3.7
50 - 54	4.0
49 - 0	5.0

A. Examination in August 2012

Out of the 693 students who were enrolled in the course during the first study 199 took the final examination in September 2012. This group included 6 of the 12 online students. 142 students passed, including 4 online students, and 57 failed, including 2 online students.

B. Examination in February 2014

141 students participated in the final examination test conducted in February 2014. (In our experience, the relative number of examination participants is lower in the February than in the August examination test. For instance, in the summer semester 2014 only 636 students were enrolled in the course OOP but 250 took the examination test.)

89 students passed, producing a typical success rate in the range 60-67%. The examination group included 22 students from the online and 28 from the control group. Fig. 10 shows the distribution of scores for the group of online students (N = 22) and the control group (N = 31) for the examination element program exceptions. The distribution curves are very similar. Slight differences can be seen only in the failure rates (grade 5.0 or score points < 50). They were: 45.5% in the online and 51.6% in the control group. In both groups, seven students achieved a grade than better 2.3. Given this small number of subjects, no effect of the study material, online course unit versus textbook, can be reliably derived from these variances.

The failure rate of the whole examination test, which included four additional problems, was only 39.6% (N=141). 9 students in the online group and 7 students in the control group, who would have failed on test element "exceptions", finally passed the examination test because they scored better on other test problems. This can be explained by the fact that other test problems have a higher success rate over the whole examination group. Test element "exceptions", which ranges at level "Analysis" in Bloom's cognitive taxonomy [10], achieved a rate of 49,5%, while two other test problems ended up with 67.9 and 72.1. In Bloom's taxonomy they range on levels "Comprehension" and "Application", respectively. This observation about the difficulty level of test problems coincides with remarks that students communicated to the tutors independently of the study: They found the examination topic "exceptions" very difficult because it has not yet been tested this intensively in previous examinations. (Note here that examination problems and master solutions are made available after a test has been concluded to help future students to prepare for their own examination.)

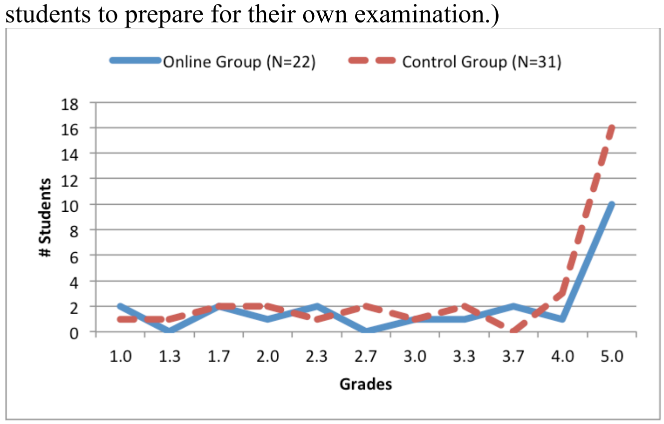


Figure 10. Distribution of examination grades among students in online and offline group.

When overlaying the examination scores of the online students for the test element "exceptions" (line diagram) and their retention times spent on information and exercise pages of this chapter, the result in Fig. 11 emerges. If we would draw a trend line, it would slightly fall from left to right. But let us discuss a few more striking issues of the bar charts and score line in Fig. 11.

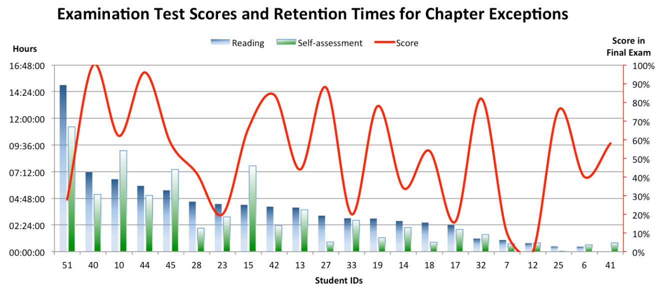


Figure 11. Retention times and examination results of individual students for test item “exceptions”.

In the discussion of Fig. 11, we also consider both the log data collected from within the online learning environment (see Section VI.C) and the online exercise system for homework assignments, which was mentioned in Section III.A, and the competence tests.

Let us pick a few extreme examples, starting with the positive ones. Students #40 and #44 achieved 100 and 96% of the maximum sore. Both accessed the information pages between 7 and 5 hours and worked on the exercise pages roughly 5 hours. We also know from Section VI.C that both were higly active and successful with the self-assessment elements of this course unit. In their homework, they acquired 100 and 95%, respectively. Student #40 also achieved the highest score in the competence post-test (69 score point out of 77) with a gain of 16 points. Student #44 scored a bit worse with a final score of 62 points and a gain of 15 points.

With 82%, student #32 scored relatively close to these two students but spent only 2.5 hours in total on exception topics, and a bit more on assessment than reading activities. (S)he ranges second in the competence post-test with 68 points and showed the higehst gain with 29 points. Student #25 exhibits the lowest time period spent on Chapter Exceptions but still ended up with a score of 76% in exception problems. All these students worked quickly on the excercises and largely error-free. Most likely, some of them acquired the necessary topic knowledge from other sources than the online course.

On the other hand, the student with the longest retention time both on reading and exercise pages (#51) missed the 50% mark necessary to pass the examination. This student achieved 95% of possible scores in the homework but did not access any of the self-assessment elements on topic exception. The latter is also true for Student #23 who also failed in the examination. Student #51 had a competence score of 59 and a gain of 11 points, while #23 collected 60 in the pre-test and gained 9 points compared to the pre-test.

The result of the correlation of examination scores and retention times indicates that we have no evidence to draw any conclusion on examination success from retention times in the online learning system. However, the targeted use of interactive learning objects and self-assessment programming tasks seems to have a positive impact on the examination test scores.

IX. Verbal Student Feedback

After all survey and log data had been evaluated, we felt the need to add a qualitative data analysis to clarify some results of the quantitative research. Questions of interest include:

- How do online students learn?
- Had the online students used other sources than just the online materials provided?
- What motivated them to volunteer for online study?
- Were their expectations satisfied?
- Did they find the online materials sufficient for reaching the learning objectives?
- How do they compare learning with traditional correspondence materials and online media?

The plan was to perform semi-structured telephone interviews with a representative sample of online students and analyze the interview content to find possible answers to these questions and detect unexpected relationships between them.

An interview guideline with 32 questions was designed, and in July 2014 a request to volunteer for a telephone interview was sent to all online students of the winter semester 2013/14 course. Unfortunately, only two students were able to take an interview. Many respondents regretted not to be able to follow the request because they were preparing for the September 2014 examination in other courses. Some students did not react at all. For reliable qualitative research, this number of subjects is by far too small. Nevertheless, we decided to organize at least the two interviews offered in search of authentic responses that would support or question our assumptions about online learner behavior.

The typical method of learning of both students is to read the course material cursorily first to judge the learning effort. Then, in a second run, they study the content in detail and thereby take notes, mark and annotate the text, perform self-assessment tests, and solve homework assignments. Some students seem to produce their own summary of course topics; our two interview partners did not. Shortly before the examination test takes place, they go through previous examination tests to get a better feel for what a typical examination test looks like and how they operate under time constraints. In addition, they re-evaluate homework assignments. To overcome understanding problems or acquire more detailed information on a topic, both students preferably use a search engine to find resources in the web that help clarify their questions. Depending on the given subject, they also buy additional textbooks recommended in the reading list. Both students felt no need to contact peer students or tutors to discuss unclear course topics. They also did not work on the collaborative tasks because they consider it difficult to synchronize with peers. This confirms our earlier experiences that distance students largely learn autonomously, mostly due to time problems.

Both interview partners state that their performances in this course match those of other courses or are slightly better. They believe that this course has a difficulty level comparable to other courses they have been enrolled in. Both students felt that they learned a lot. One student was inspired to apply and extend his/her programming skills by adding dynamic web pages to his/her web presence using PHP [30].

The motivation to volunteer for online studies were: a) a general interest in e-education paired with e-learning experiences in the work context; b) a personal interest in the question whether online learning is advantageous over the study with classical distance learning textbooks. One student said that his expectations were partly satisfied. On one side, the interactive learning objects and programming tasks were considered helpful in assessing the learning progress. The lack of possibilities to highlight and annotate passages in the online materials was seen negative. In summary (s)he felt that "Basic knowledge about object-oriented programming concepts and basic programming skills definitely came across." The other student said: "I liked the general layout, the content, and the mix of reading and exercise pages of the online course. I would like to have more of this." Both students solved all online tasks but the collaborate task completely.

Asked whether the students perceived a difference in learning effort for both media types, one student thought that the learning objectives could be achieved faster with the online version. The other student was undecided.

Both students stated that they also studied the textbook version after having studied the online version to be on the safe side because they had noticed that both versions differ in structure and presentation. This may explain why the online times of most students were relatively short and largely dedicated to the examination of online tests. When comparing both versions, one student replied: "Studying the online material was full of variety, while the textbook is drier to go through". The other said that (s)he is better motivated by working with the online material, while studying the course text on paper leads earlier to a loss of concentration. The students did not feel exhausted when learning for a longer time in front of a computer or tablet screen. Both students said that the communication means offered were useful. In particular the topic-specific online forum is considered more effective than the course newsgroup, which requires special newsreader and addresses all course topics.

In an independent student evaluation, which is carried out every semester for every course of the curriculum, this course typically receives a rating above average. This is also true for the evaluation performed at the end of the winter semester 2013/14 (1.6 on a scale 1-3). The students also find the workload in the usual range. Some were asking for additional learning video and interactive online media. Many students consider the supervised forum essential for an effective learning process.

X. Conclusion and Lessons Learned

With this study we wanted to learn more about the influence of e-learning elements in higher distance education. To this end, we used an introductory programming module in the undergraduate degree program on Business Informatics as a case to compare learning attitudes and outcomes of students using the classical correspondence study material with students learning along an interactive online version of a selected learning module. The study was performed twice in two different years on the same course but with different student populations because the first run produced statistically unreliable results. Table V depicts relevant figures of the two investigations undertaken. The questionnaires used in the second study and the raw data collected thereby are accessible online [31].

In summary, the research questions we started from can be answered as follows:

1. The competence analysis shows that the students achieved significant growth in competence, particularly on the course topics "testing" and "program exceptions". However, no significant difference can be observed between online students and the control group (see Fig. 5).
2. Even if we chose the students' final grades acquired in a written examination that took place at the end of the course, no significant difference in learning outcomes can be seen (see Fig. 10).
3. The retention times in the online environment are no useful indicator for test success or failure, a result that has been confirmed before in a workload study [32].
4. The behavior analysis depicts a correlation between high and targeted use of interactive learning objects.
5. The results of the technology acceptance analysis and the online behavior analysis suggest that the students appreciate technology-supported learning. At the same time, both studies clearly show that students, on the one hand, have little tendency to adopt cooperative learning and, on the other hand, do not feel capable because time constraints.

TABLE V. TEST PARAMETERS FOR BOTH STUDIES

Activity	SS2012	WS 2013/14
Students enrolled in course	693	980
Competence pre-test	57	162
Pre-test: complete sets	19	126
Competence post-test	12	78
Post-test: clean sets	5	49
Post-test paired with pre-test	1	40
Online study volunteers*	12	186
Volunteers selected	12	98
Volunteers actively studying	10	45
Technology acceptance test		34
Final examination	141	200
Participating online students	6	21

*Online students have experience with traditional correspondence materials from other course units and courses

As a result of this investigation, we decided to enrich other topics of the course OOP with further interactive learning objects for self-assessment and practice. For the near future, we will keep the textbook version as the main learning medium for our students.

To support the autonomous preparation for an examination test, we plan to produce short video clips explaining how to understand a given problem and constructively approach its solution.

In general, we find that technology-assisted learning elements for distance students increasingly arouse interest, to the extent that they are efficient and supportive for the learning process. We believe that e-learning still offers many degrees of freedom for pedagogic innovations of distance learning, while the pedagogy of the traditional correspondence-based distance education model is largely exhausted. Research in the field of online distance education is, however, still largely disconnected and disparate. In their new book, for the first time, Olaf Zawacki-Richter and Terry Anderson have tried to integrate major research trends in online distance education into a systematic research agenda [33].

ACKNOWLEDGMENT

We are particularly grateful to all students who volunteered for the online study and the three surveys. Winfried Hering owes our thanks for performing the log data and behavior analysis of the first study as part of his Bachelor thesis research. We thank Silvia Schreier for being overly supportive in the organization and execution of the first study. Helga Huppertz owes our gratitude for co-designing and implementing the online sections of the course. In the second study, she also helped to analyze the log data of the online student group. Christoph Wöldecke helped us setting up the competence measurement instrument and evaluating the results of the competence measurements.

REFERENCES

[1] W. Hering, H. Huppertz, B. J. Krämer, J. Magenheim, J. Neugebauer, and S. Schreier, “On Benefits of Interactive Online Learning in Higher Distance Education – Case Study in the Context of Programming education,” eLmL 2014: The Sixth International Conference on Mobile, Hybrid & On-line Learning, 2014. Also available from http://www.thinkmind.org/index.php?view= article&articleid=elml_2014_3_20_50048, last access: Dec. 10, 2014.

[2] B. J. Krämer, “Web-based Learning Environments: Tools and Engineering Issues,” in Handbook of Software Engineering and Knowledge Engineering, vol 1: Fundamentals, pp. 697-718, World Scientific, 2001.

[3] B. J. Krämer and L. Wegner, “From Custom Text Books to Interactive Distance Teaching,” ED-MEDIA/ED-TELECOM 98, Prentice Hall, 1998.

[4] T. Russel, The No Significant Difference Phenomenon, 5th edition Raleigh, NC, North Carolina State University, 1999.

[5] R. M. Bernard, P. C. Abrami, Y. Lou, E.Borokhovski, A. Wade, L. Wozney, P. A. Wallet, M. Fiset, and B. Huan, “How does distance education compare with classroom instruction? A meta-analysis of the emirical literature,” Review of Educational Research, pp. 379-439, vol. 74, 2004.

[6] B. Means, Y. Toyama, R. Murphy, M. Bakia, and K. Jones, Evaluation of evidence-based practices in online learning: a meta-analysis and review of online learning studies, U.S. Department of Education, 2010, https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf, last access: Dec. 10, 2014.

[7] J. Ferguson and A. M. Tryjankowski, “Online versus face-to-face learning: looking at modes of instruction in Master’s-level courses,” Journal of Further and Higher Education, vol. 33, no. 3, pp. 219-228, 2009.

[8] L. M. Cooper, “A comparison of online and traditional computer applications classes,” T.H.E. Journal, vol. 28 no. 8, March 2001.

[9] S. Guri-Rosenblit, “‘Distance education’ and ‘e-learning’: Not the same thing”, Higher Education, vol. 49, no.4, pp. 467-493, Springer, June 2005.

[10] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, Taxonomy of educational objectives: The classification of educational goals - Handbook I: Cognitive domain, New York: David McKay Company, 1956.

[11] B. J. Krämer, *Object-oriented Programming – Module 5: Exceptions and Testing*. [Online,in German]. Available from http://fernuni-hagen.edu-sharing.net/OOP_M5 (user name: elearning, password: TryIt), last access Nov. 22, 2014.

[12] Moodle, *Open Source Learning Platform*. [Online]. Available from http://moodle.com/, last access: Nov 15, 2014.

[13] edu-sharing, *Open source repository network for learning content and educational knowledge*. [Online]. Available from http://edu-sharing.net/, 2009, last access: Dec. 10, 2014.

[14] M. Klebl, B. J. Krämer, and A. Zobel, “From content to practice: sharing educational practice in edu-sharing,” British Journal of Educational Technology, vol. 41, no. 6, pp. 936-951, Nov. 2010.

[15] F. Davis, “A technology acceptance model for empirically testing new end-user information systems - theory and results,” PhD thesis, Massachusetts Institute of Technology, 1985.

[16] V. Venkatesh and F. D. Davis, “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” Management Science, 46, pp. 186-204, 2000.

[17] J. Brooke, “SUS - A quick and dirty usability scale.” [Online]. http://www.usabilitynet.org/trump/documents/Suschapt.doc, last access: Dec. 10, 2014.

[18] C. Ong, J. Lai, and Y. Wang. “Factors affecting engineers‘ acceptance of asynchronous e-learning systems in high-tech companies,” Information & Management, vol. 41. No. 6, pp. 795-804, 2004.

[19] M. Zviran, C. Glezer, and I. Avni, “User satisfaction from commercial web sites: the effect of design and use,” Information and Management, vol. 43, no. 2, pp. 157-178, 2006.

[20] V. Venkatesh and H. Bala, “Technology Acceptance Model 3 and a Research Agenda on Interventions,” Decision Sciences, vol. 39, no. 2, pp. 273–315, 2008.

[21] LimeSurvey, *Open source online survey system, 2003*, [Online]. Available from http://www.limesurvey.org, last access: Dec. 10, 2014

[22] The Joint Task Force on Computing Curricula. (2001). Computing Curricula 2001 Computer Science. IEEE-CS, ACM. Final Report.

[23] Tucker, Allen (ed.), “A Model Curriculum for K-12 Computer Science: Final Report of the ACM K-12 Task Force Curriculum Committee,” 2nd Edition, New York: Association for Computing Machinery (ACM), 2006. [Online]. Available from.http://www.acm.org/education/education/curric_vols/k12 final1022.pdf, last access: Dec. 10, 2014.

[24] B. Linck et al., “Empirical refinement of a theoretically derived competence model for informatics modelling and system comprehension,” Proc. IFIP Conference Addressing educational challenges: the role of ICT (AECRICT), 2-5 July 2012

[25] Lehner, L. et al., “Informatics Systems and Modelling–Case Studies of Expert Interviews”, Reynolds, N. and Turcsányi-Szabó, M. (eds.) Key Competencies in the Knowledge Society IFIP AICT. pp. 222–233, Heidelberg: Springer, 2010..

[26] B. Linck et al., “Competence model for informatics modelling and system comprehension,” IEEE Global Engineering Education Conference (EDUCON), pp. 85–93, Berlin 2013.

[27] T. Rhode, “Development and testing of an instrument for measuring informatics modelling competences in a didactic context,” dissertation, Universität Paderborn, 2012 (in German).

[28] W. Fan and Z. Yan, “Factors affecting response rates of the web survey: a systematic review,” Computers in Human Behavior, vol. 26, no. 2, pp. 132–139, Feb. 2010.

[29] SAS, *Business Analytics Software*. [Online]. Available from http://www.sas.com, last access: Dec. 10, 2014.

[30] PHP, *Hypertext Preprocessor*. [Online]. Available from http://php.net/, last access: Dec. 10, 2014

[31] Bernd J. Krämer and Johannes Magenheim: Datasets Acquired in a Comparative Study on the Impact of Higher Distance Education with Traditional Correspondence and Interactive Online Media, FernUniversität 2014. [Online]. Available from http://deposit.fernuni-hagen.de/2990/ last access: Dec. 10, 2014

[32] R. Schulmeister and C. Metzger (Eds.), “The Workload in Bachelor Programs: Time Budget and Study Behavior,” Waxmann, 2011 (in German).

[33] O. Zawacki-Richter and T. Anderson (Eds), “Online Distance Education: A Research Agenda,” Athabasca University Press, 2014.

International Journal on Advances in Life Sciences, vol 6 no 3 & 4, year 2014, http://www.iariajournals.org/life_sciences/

A Virtual Patient Simulator Based on Human Connectome and 7 T MRI for Deep Brain Stimulation

Giorgio Bonmassar
Athinoula A. Martinos Center for Biomedical Imaging
Department of Radiology
Massachusetts General Hospital, Harvard Medical School, Charlestown, MA
Giorgio.bonmassar@mgh.harvard.edu

Leonardo M. Angelone
Office of Science and Engineering Laboratories,
Center for Devices and Radiological Health, U.S. Food and Drug Administration,
Silver Spring, MD
Leonardo.angelone@fda.hhs.gov

Nikos Makris
Department of Psychiatry and Neurology
Massachusetts General Hospital, Harvard Medical School, Charlestown, MA
nikos@cma.mgh.harvard.edu

***Abstract* — This paper presents a virtual model of patients with Deep Brain Stimulation implants. The model is based on Human Connectome and 7 Tesla Magnetic Resonance Imaging (MRI) data. We envision that the proposed virtual patient simulator will enable radio frequency power dosimetry on patients with deep brain stimulation implants undergoing MRI. Results from the proposed virtual patient study may facilitate the use of clinical MRI instead of computed tomography scans. The virtual patient will be flexible and morphable to relate to patient-specific neurological and psychiatric conditions such as Obsessive Compulsive Disorder, which benefit from deep brain stimulation.**

Keywords - VPS; DBS; OCD; MRI; CT; safety; specific absorption rate; SAR; heating.

I. INTRODUCTION

Many deep brain stimulation (DBS) patients may require regular MRI examinations throughout the course of their lives since MRI is often the diagnostic tool of choice for monitoring structural changes in the brain [1]. Whole-body MRI examination is used in many common injuries following accidents and comorbidities. Moreover, functional MRI could be potentially useful to assess the effects of electrical stimulation of the basal ganglia [2]. Because of safety concerns more than 300,000 per year patients with implanted leads (e.g., pacemakers, DBS, and catheters) are denied MRI [3]. One of the main concerns regarding use of MRI for DBS implants is related to potential radio frequency (RF) – induced heating [4][7]. The RF waves used in MRI to elicit signal from the tissue interact with the conductive leads generating potentially high induced currents along the leads ("antenna effect") and increased RF power deposition near the distal tip of the leads [8]. Unfortunately two cases of serious, permanent neurological injury related to the antenna effect of DBS leads during MRI were reported [9][10]. In one case, one patient with two bilateral implants underwent a routine MRI of the lumbar spine and reported a hemiplegia. The authors of the report associate the injury to the RF-induced currents generated by the body coil on the DBS implants, which produced an edema near one of the implants with the consequent paralysis [10]. Additionally, a case of neurological deficit (i.e., double-vision and severe right-sided motor contractions) in a patient with bilateral DBS leads who

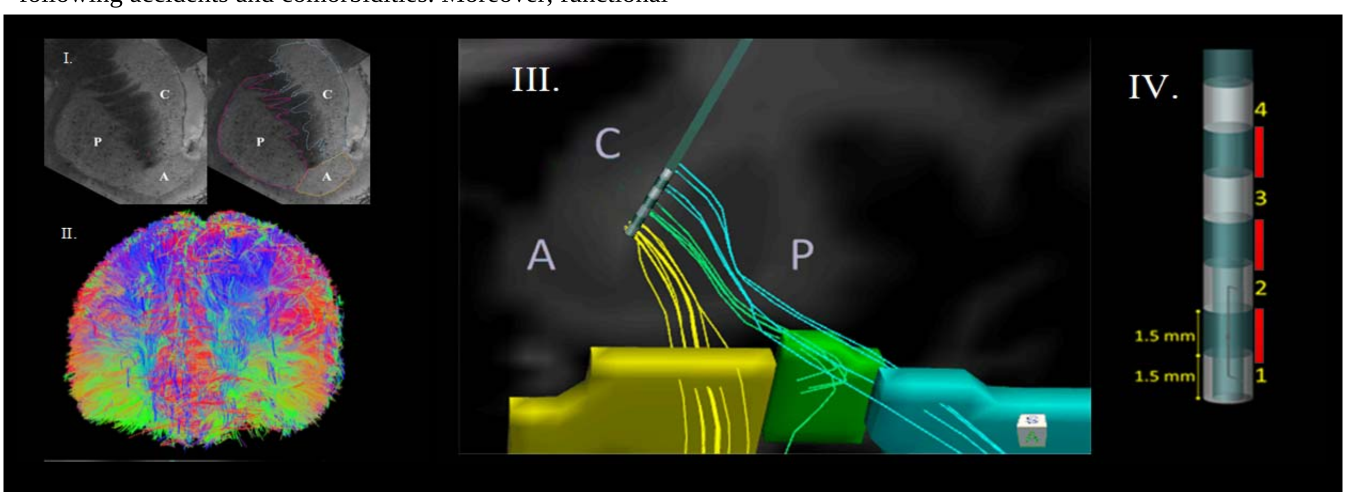


Figure 1: (I) Coronal section at the C.A.P. (Caudate-Accumbens-Putamen) level of a 7T high-resolution (100 μm^3 isotropic) ex-vivo human dataset showing detailed anatomy of the VC/VS target area (cross-hair in red) within the anterior limb of the internal capsule. (II) DSI connectome data set. The Connectome MRI data (voxel size: $2 \times 2 \times 2$ mm^3) could be reliable in identifying the fiber tracks for connecting OCD targets, (III) electrode placement in the VC/VS target area traversed by fibers (yellow fibers) of the ventromedial prefrontal-basal ganglia tract at the C.A.P. coronal slice. The latter fibers stem from the ventromedial prefrontal area (yellow block). (A=nucleus accumbens, C=caudate nucleus, P=putamen), (IV) Model of the implanted DBS electrode.

underwent diathermy was also reported [11]. Recently Shrivastava et al. reported temperature increases up to 30°C in the tissue near DBS electrodes implanted in suine undergoing high field MRI (i.e., 9.4 Tesla) [12]. Based on guidelines approved by the U.S. Food and Drug Administration (FDA), in the US MRI can be used in DBS patients with very specific conditions [13]. For example, it is possible to scan patients with DBS implants only with a transmit head coil. Also, the use of 3T MRI systems, including the state-of-the-art MRI multichannel transmit coils, is contraindicated.

Computational models have been used as a supporting tool to evaluate RF-dosimetry in MRI. Numerical DBS models are based on a wire or set of wires, which represents the virtual DBS implant, superimposed to healthy human brain models [8][14]. Several numerical models for electromagnetic analysis of DBS have been proposed [15][20]. Recently, anatomically precise head models with implanted DBS leads with a multiscale resolution up to $0.2mm^3$ isotropic were also developed [8][21]. The models were based on Finite Difference Time Domain (FDTD) simulations. FDTD models have been validated in the past against temperature measurements, showing an accuracy of 20% as predicted by the bioheat equation [22] with respect to whole-body specific absorption rate (SAR). The DBS models described above allowed for accurate geometrical modeling of implanted leads as well as precise computation of 1g- and 10g-averaged SAR. Nevertheless, the models were characterized by a limited precision with respect to anatomical modeling of the structures involved in the stimulation, anisotropic dielectric constants, information about head perfusion, and the tissue scar from the surgery.

The rest of this paper is organized as follows. Section II describes the significance of the different tools used for the modeling. Section III describes the methods used. Section IV presents the results. Section V includes the discussion. The acknowledgement and conclusions close the article.

II. Background And Significance

A. *Importance of structural MRI-based modeling*

Neurocircuitry models of Obsessive-Compulsive Disorder (OCD) pathophysiology are currently supported by neuroimaging studies indicating the critical role of abnormal frontal-basal ganglia–thalamic circuits (cortico-striato-thalamo-cortical or CSTC circuitry). The most relevant anatomical structures involved are the orbital frontal cortex (OFC), anterior cingulate cortex (ACC), the ventral striatum and medial thalamus. It has been reported that in OCD fiber tracts connecting these centers may be abnormal [23] and that ablating or modulating these white matter pathways (e.g., in anterior cinglulotomy or anterior capsulotomy) as well as DBS on the Ventral Capsule/Ventral Striatum (VC/VS) may be associated with therapeutic improvement [24]. The VC/VS in the Anterior Limb of the Internal Capsule (ALIC, which is approximately 20 mm long in its dorsoventral extension and 2-5 mm wide mediolaterally at the coronal sections where the nucleus accumbens is present) is currently a target for DBS in OCD (Medtronic, Minneapolis, MN) (Fig.1). More specifically, a small fiber tract connecting the ventromedial prefrontal cortex (vmPFC) with the medial thalamus, basal ganglia and brainstem (vmPFC-BG) is thought to be a central to OCD pathophysiology and treatment response. The latter is a relatively small fiber bundle, of about 2 mm in diameter, which courses approximately 4-5 mm above the nucleus accumbens septi interconnecting ventral prefrontal cortex with basal ganglia, thalamus and the brainstem [24]. Given the increased clinical practice and efficacy of DBS in OCD, the accurate anatomical characterization of this surgical target in terms of its precise location, relative topography with surrounding structures (in particular with the nucleus accumbens septi and the anterior commissure) and size represents an important goal for numerical models – such as the Virtual Patient Simulator described herein– for DBS modeling.

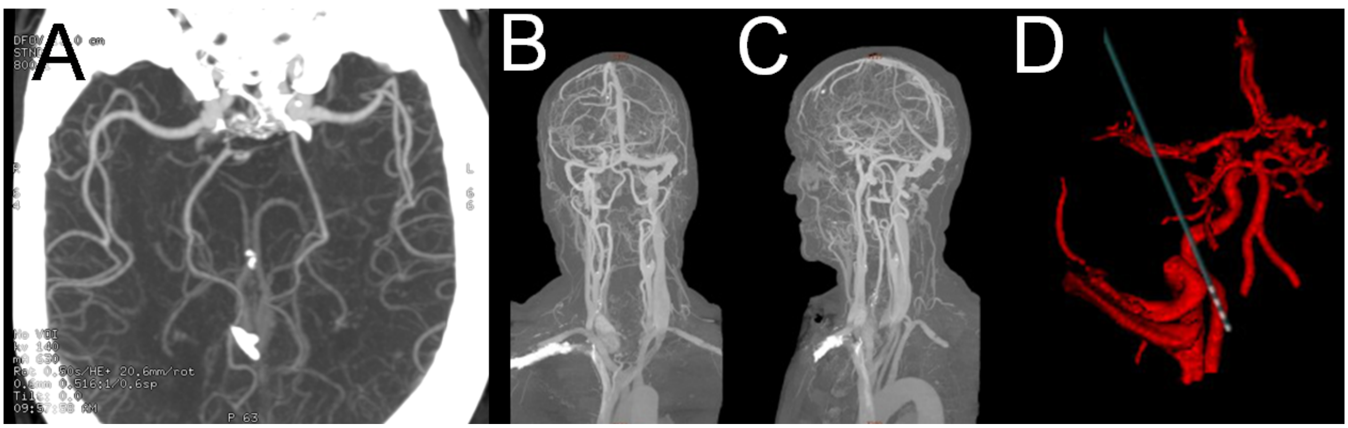


Figure 2: (A-C) Images of CTA acquired at the Radiology Department at MGH with 600µm in plane resolution, (B,C) 3D reconstruction with bone subtraction using dual energy CTA. (D) Illustration of how vasculature may wrap around an electrode.

B. Importance of Connectome-based modeling

Anisotropy of electrical properties of tissue arises in nerve and muscle fibers which consist of bundles of long, parallel myelinated elements. The VPS model described here will include Diffusion Tensor Imaging (DTI) data, to model tissue anisotropic conductivity [25]. To generate the model, the acquired diffusion tensor **D** will be converted into a complex relative permittivity tensor $\boldsymbol{\varepsilon^*}$ using a simple linear transform $\boldsymbol{\varepsilon^*} = \frac{\varepsilon^*}{d} \cdot \mathbf{D}$, where ε^* is the tissue complex relative permittivity [26] and d is the diffusivity.

Preliminary data (Fig. 1.B) show that Connectome MRI data (voxel size: 2 × 2 × 2 mm^3) may be reliable in identifying the fiber tracks connecting vmPFC and BG. The gradient strength determines the sensitivity, accuracy, and resolution of diffusion imaging. Data were acquired with a new Connectome scanner, installed at the A. A. Martinos Center, Massachusetts General Hospital, purpose-built for diffusion MRI with ultra-high gradients of 300 mT m^{-1}. As shown by a recent study [27] this technology allows for enhanced sensitivity and resolution of white matter imaging and MRI tractography of 5-10 fold over any other human DTI or High-Angular Resolution Diffusion Imaging (HARDI) such as the Diffusion Spectrum Imaging (DSI) technology. Fiber tracks were generated by using spatial information derived from high spatial resolution ex-vivo diffusion imaging (voxel size: 100 µm isotropic). .

C. Importance of Computed Tomography Angiography (CTA)-based modeling

Heat transport in biological tissues, which is usually expressed by the Pennes bio-heat equation, is a complex process that involves thermal conduction in tissues, convection and perfusion of blood (delivery of the arterial blood to a capillary bed in tissues). The underlying assumption of lack of perfusion used for many in vitro studies is often considered as a worst-case scenario for temperature changes at the distal tip and related tissue injury. However, a more accurate solution may be computed by solving the partial differential bioheat equation [28]:

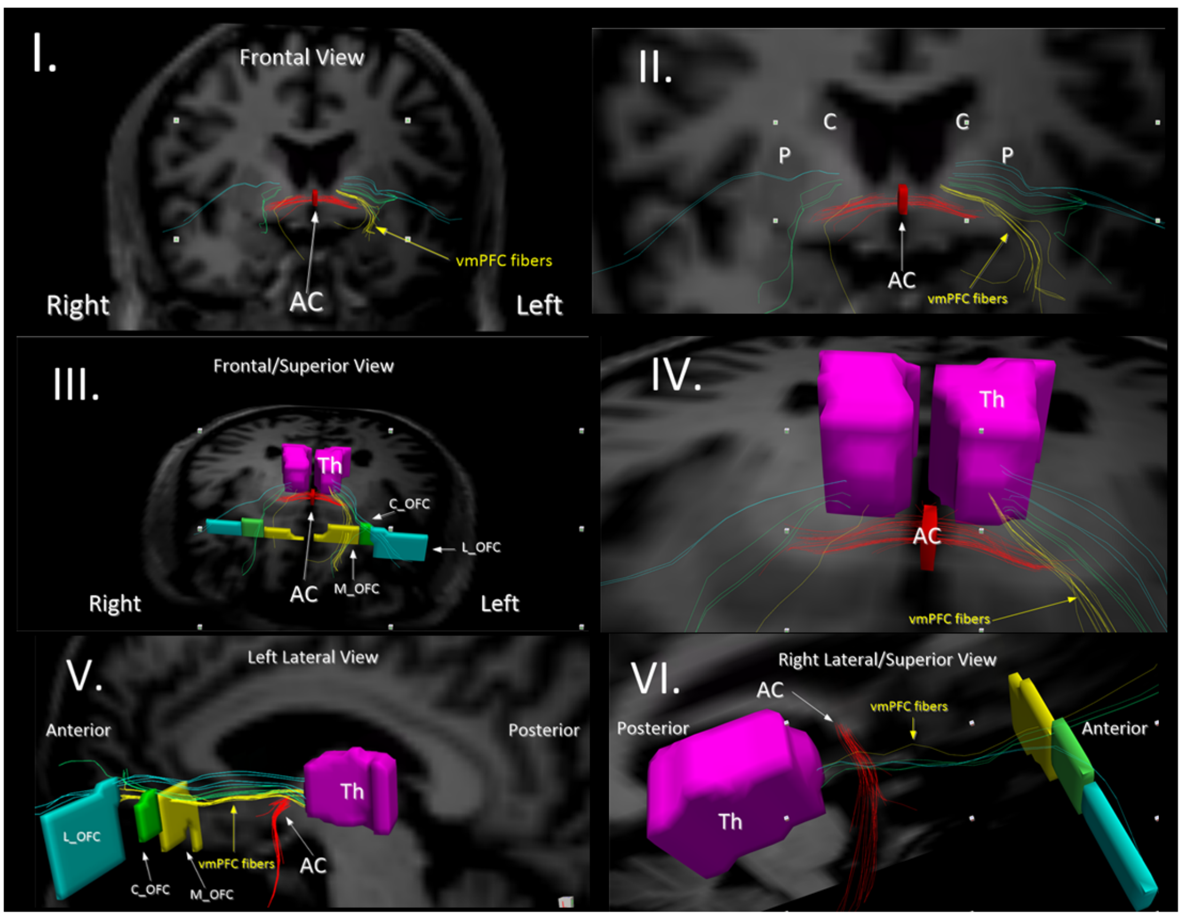


Figure 3: 3D reconstructions of ventromedial prefrontal (vmPFC) fibers and the anterior commissure (AC) from DSI connectome data of an individual subject (same as Fig. 1). Frontal (I., II, III, and IV) and lateral (IV and V) views showing vmPFC fibers connecting different regions of the orbital frontal cortex (lateral L_OFC, medial M_OFC and central C_OFC orbital frontal cortex) with the thalamus (Th). The vmPFC fibers are located dorsal to AC as shown in zoomed views (II and IV). Other abbreviations: C = caudate nucleus and P=putamen.

$$\boldsymbol{\rho C_p}\frac{\partial T}{\partial t} - \nabla \cdot (\boldsymbol{k}\nabla T) - \boldsymbol{M}(T) + \boldsymbol{P}(T)(T - T_B) + \boldsymbol{R_L}(T) = 2\rho\ \boldsymbol{SAR} \quad (2)$$

Where **ρ** is the mass density [kg/m^3], $\boldsymbol{C_p}$ is the heat capacity [J/(kg °C)], ∇T the gradient of temperature [°C], **k** is the thermal conductivity [W/(m °C)], **M** is the metabolic heat production [J/(s m^3)], **P** is perfusion parameter [J/(s °C m^3)] which becomes the heat loss when multiplied by the difference between the tissue temperature (T) and the blood temperature (T_B), $\mathbf{R_L}$ is the respiratory heat loss [J/(s m^3)], and $\boldsymbol{SAR}$ is the local specific absorption rate [W/kg] in the tissue generating the Joule heating. Bold symbols indicate vector fields or space dependent quantities (e.g., mass density is spatially variant, as it depends on the particular tissue [29]. The anatomical head model can be improved by including a realistic vascular structure and studying its thermal effect of perfusion and estimation of perfusion parameter (**P**) in the bioheat equation (eq.2). The realistic vasculature can be extracted from CTA images (Fig. 2).

D. Importance of Tissue Scaring around the DBS Electrode

Chronically implanted recording electrodes provoke an immune reaction against them. The histopathological finding is that of gliosis and spongiosis around the electrode track, which forms an encapsulation layer referred to as the “glial scar”. This reactive glial tissue which surrounds the implanted electrodes is approximately 200 μm thick and progressively isolated the electrode from surrounding neurons modifying the electric field and acts as a natural neuro protector against heating of the electrode tips. The conductivity and permittivity of the scar tissue can be modelled at the frequency of interest (e.g., 128 MHz for a 3T MRI system) as a ring around the electrodes. The scar tissue will affect the bio-heat modeling behaving as a thermal insulating shield, thus improving the accuracy of the model.

E. Impact of the new RF pulse

The issue of heating of implants during MRI has been studied for several years [30] and analytical solutions (e.g., Green’s function) to the problem have been proposed for simple geometries. Modifications of the implant leads and wires for reducing the RF-induced heating have been proposed introducing chokes or special geometrical paths of the wire. However, design modifications may be not appropriate for patients who already have DBS implants, since replacing the original leads with new leads requires major brain surgery. Because RF heating depends on the MRI RF transmit field, recent studies have focused on the modification of the transmit field in order to minimize the electric field in and around the implant and thus reducing the RF-induced heating near the implant.

III. METHODS

The study is based on *ex-vivo* analysis of collected 7 Tesla (7T) T2* and Connectome diffusion MRI data.

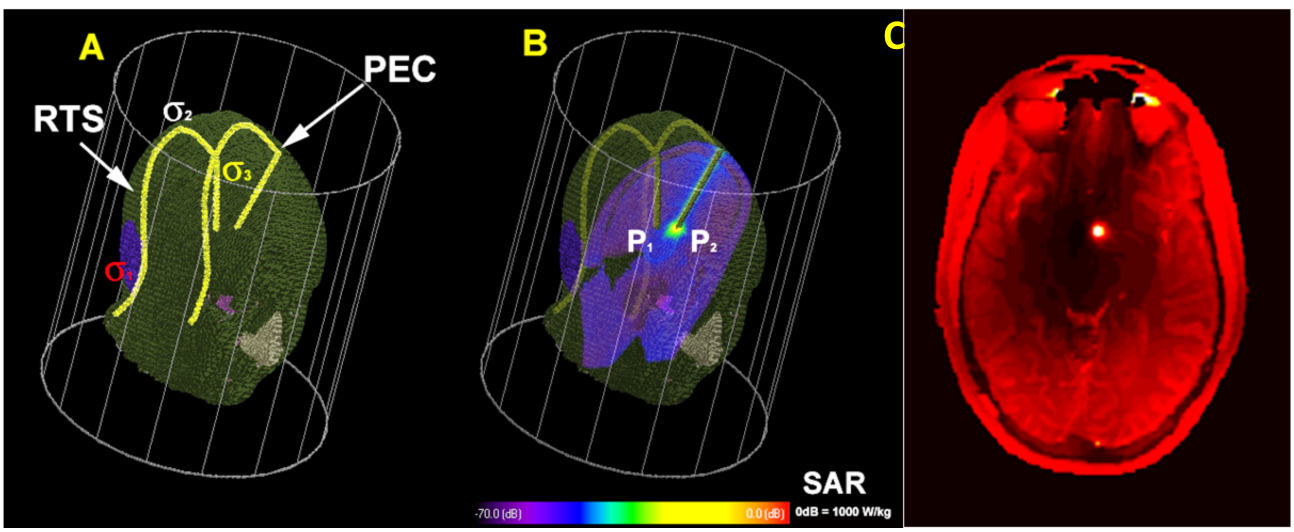


Figure 4: (A) Human head model with prototype of RTS and PEC implant (B) Computed SAR for the head model and the two implants. The 1mm^3 SAR at the tip of the RTS lead (P_1) was five –order of magnitude lower compared to SAR at the tip of PEC lead (P_2) (C) Results of electromagnetic simulations showing a high energy deposition near the distal tip of a monolateral metallic DBS lead.

A. Overview of study design, data acquisition and analysis.

(a) *Structural T2* 7 Tesla MRI data* The model is generated by segmenting 21 different brain structural entities on the *ex-vivo* structural MRI, as well as 28 non-brain structural entities following the approach described in Makris and colleagues [29]. The new head model now contains both the PD and the OCD targets found using tractographic methods. This *ex-vivo* brain consisted of MRI data of a hemisphere fixed in Periodate-Lysine-Paraformaldehyde (PLP) using the following parameters: T2*-W, 100 μm^3 isotropic resolution,

TR/TE/flip=40ms/20ms/20°, 1600×1100×896 matrix. Segmentation of OCD-DBS target-related cerebral structures – specifically the putamen, caudate nucleus, nucleus accumbens and anterior limb of the internal capsule – were manually outlined (on a re-sampled dataset at 1 mm isotropic spatial resolution of the original *ex-vivo* high-resolution – 100 μm^3 isotropic resolution – dataset acquired at 7 Tesla) using the segmentation methods by Filipek [31] and Makris [32], which have been developed and validated at the MGH Center for Morphometric Analysis, and have been implemented in several clinical studies.

The VC/VS in the ALIC is currently a target for DBS for OCD. It was approved in 2009 by the FDA with a Humanitarian Device Exemption (HDE) [33]. Importantly, a structure within this large territory, namely the vmPFC-BG tract is thought to be a more specific target for this procedure in OCD. Even though this is a relatively small fiber bundle, the ALIC/VS was segmented as follows by reaping the benefits of the high-resolution dataset.

(a) The anterior limb of the internal capsule is delimited medially by the head of the caudate nucleus, laterally by the putamen and ventrally by the nucleus accumbens. These structures were delineated by direct visualization using intensity-based contours in the Cardviews software system.

(b) *Connectome DSI data* were acquired on the Connectome 3T scanner using Gmax = 300 mT/m. This data acquisition was done on a 64-channel coil array using a single refocused spin echo sequence with the following features: TR = 2800 ms, TE = 56 ms, 63 axial slices, FOV = 220 mm, R = 3 GRAPPA, 2.0 mm isotropic resolution (BW = 2390 Hz/px), b = 15,000 s/mm^2. These data were

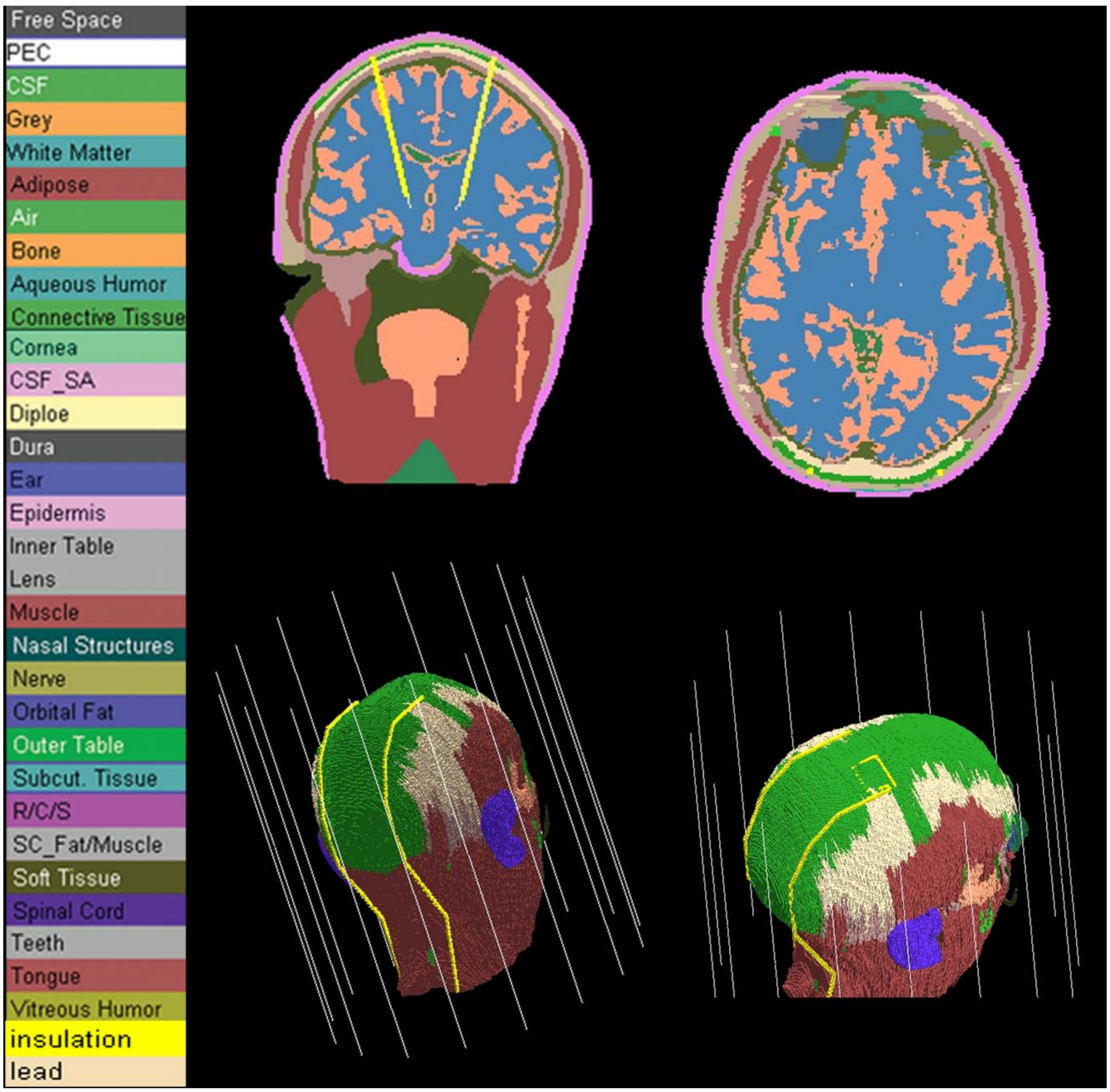


Figure 5: Anatomically fine-grained head model with a co-registered intracranial implant (white) and insulation (yellow). The coronal view show the 29 tissues manually segmented. The corresponding color code for each tissue is to the left of the image. (Bottom Left) 3D view of the head model and DBS implant, where external tissues (epidermis and subcutaneous tissue) have been removed for illustrative purposes.

transformed into DTI data to estimate complex relative permittivity tensor introduced in eq. (1). The vmPFC-BG tract was delineated using diffusion Connectome data as shown in Fig. 1.II. DTI/DSI data were visually validated by comparing the computed fiber tracks with anatomical atlases giving particular emphasis to the basal ganglia region as shown in Fig. 1.I. Furthermore, DTI/DSI data provided detailed information on the fiber tract connectivity between the ventromedial prefrontal cortex, basal ganglia, and thalamus. The latter is useful for DBS programming [35] and basic neuroscience research.

B. Numerical Model of Deep Brain Stimulation implant

A high-resolution model was generated based on the MRI data by distinguishing twenty-one brain [31] and twenty-eight non-brain anatomical structural entities. Each of the anatomical structural entities was associated to an electrical structural entity with electrical properties defined according to the literature [36][37]. The overall head dimensions were 170 mm in width, 217 mm in depth, and 238 mm in height and the total volume of the head model was 4,642,730 mm^3. The model was used for RF dosimetry studies of clinical significance [8], including preliminary studies with DBS implants. Two bilateral implants were modeled as insulated wires connected to the left and/or right targets in the head [10]. The wires were modeled as a perfect electric conductor (PEC) and the dielectric was modeled as Teflon. A four-electrode connection [35] was also modeled in full detail reaping the benefits of the high spatial resolution (Fig. 1).

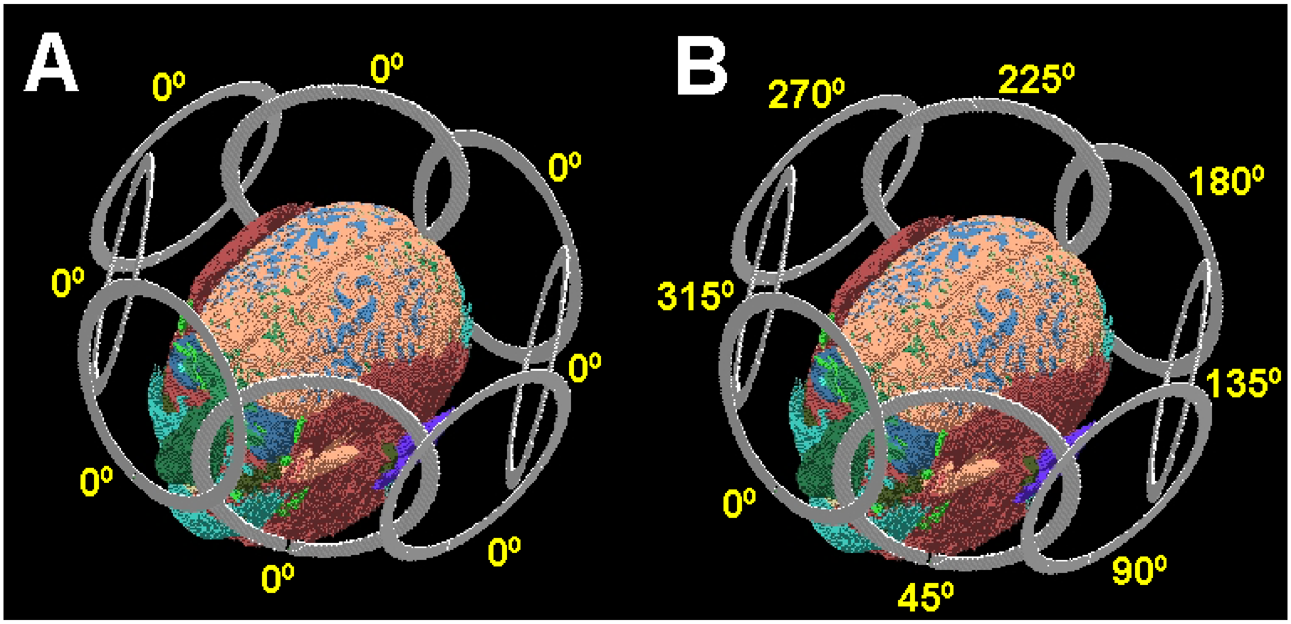


Figure 6: Geometrical model of the coil array and our model [29] with examples of null (A) and uniform (B) eight-channel transmit phase array vectors [34]. .

IV. Results

In the model preliminary developed, the OCD target structures for DBS were identified and segmented using high-resolution T2*-W, diffusion spectrum MRI data, and electromagnetic simulations for MRI safety.

Fig. 3.I and 3.II show data from a manual segmentation procedure to outline the different anatomical structures related to the VC/VS target area derived from a high-resolution 7 Tesla dataset and DSI-based tractography. Furthermore, the fiber connections between the orbital frontal cortex and the thalamus were reconstructed tractographically, visualizing them in relation to the anterior commissure (AC) to elucidate the spatial relationships between these fiber tracts. Fig. 3 shows that the vmPFC fibers – connecting different regions of the orbital frontal cortex (lateral L_OFC, medial M_OFC and central C_OFC orbital frontal cortex) with the thalamus (Th) – follow a pattern of spatial organization. Specifically, in the orbital frontal region within the anterior limb of the internal capsule they verticalize from a lateral to medial horizontal deployment showing the following spatial distribution above the nucleus accumbens. The L_OFC fibers are more dorsal, the M_OFC ones are more ventral and the C_OFC fibers are in-between. With respect to the AC, vmPFC fibers are located dorsal to AC as shown in zoomed views (Fig 3.II and 3.IV).

Fig. 4 shows an example of the numerical head model with implanted DBS leads modeled as resistive tapered stripline (RTS) [38] or metallic (PEC) wires. The model allowed simulating the high local SAR near the PEC lead as well as the lower local SAR near the tip of the RTS lead.

Values were normalized to obtain a whole-head averaged SAR of 3.2 W/kg [10][39][40].

Additionally, dual energy CTA images with bone subtraction (Fig. 2.B and 2.C) were performed using standard contrast agents. Segmentation was obtained by standard thresholding [41] from the images containing enhanced vascular information validated by comparison with vascular atlases [42][44]. These data may be used to inform more precise thermal analysis and prediction of thermal response of target nuclei surrounding the DBS electrode exposed to RF energy (Fig. 2.D).

V. Discussion

Clinical work in OCD indicates that several compulsive behaviors in this disorder are related to avoidance of putative dangerous situations. The neural system that mediates avoidance is the same one that also underlies reward-seeking and comprises anterior cingulate/orbital frontal connections. Currently, neurosurgical procedures such as anterior cingulotomy, subcaudate tractotomy (SCT) and DBS are offered as an option to reduce symptom severity and increase quality of life in cases of medically intractable OCD. Given that 40-60% of patients with OCD do not have satisfactory response to optimal management [45], these procedures have been used for therapeutic benefit with very positive outcomes [46][49]. At Massachusetts General Hospital it has been recently shown that in 73% of these patients symptomatology was improved [50][51]. DBS is a therapeutic procedure aiming to change activity in fronto-basal ganglia circuits by injecting electrical stimulation in the Ventral Anterior Internal Capsule (VA-ALIC) and adjacent Ventral Striatum (VS), (i.e., nucleus accumbens septi). Given the efficacy of DBS and CST in cases of medically intractable OCD, an understanding of the disrupted connectivity is important to elucidate OCD psychopathology and provide guidance in making the SCT lesion or the DBS target more effective. Current Connectome and 7 Tesla MRI technology allows for more precise characterization of these targets for therapeutic intervention and of their relative topography.

There have been reports of serious accidents associated with MRI-related heating [10]. However, OCD patients with DBS would benefit from regular MRI examinations, as MRI is often the diagnostic tool of choice to diagnose injury due to trauma or evaluate comorbidities. There are over 75,000 patients with DBS implants worldwide and approximately only one patient over twenty is assessed with MRI. In this pilot study the use of VPS in patients with DBS implants is investigated gathering preliminary results on the anatomical characterization of the OCD target structures for DBS. Within the study high-resolution 7 Tesla and Connectome MRI data were used and electromagnetic simulations for MRI safety were performed. The surgical procedure may introduce structural changes that maybe seen with the MRI once it is found to be safe using the precautionary principle or precautionary approach to risk management states, which is adopted by all safety organizations including the FDA. Using VPN offers an avenue for showing that MRI may be performed conditionally safe, since VPN allows the estimation of safety parameters such as the Specific absorption rate, etc. Furthermore, the proposed new OCD VPS may allow for improved RF power dosimetry on patients with DBS implants undergoing MRI. For example, the proposed VPS could help evaluating the effect of different lead pathway (e.g., use of loops) (Fig. 5). Note that DBS electrodes/leads are very poor receiving coils, since are not tuned coils connected through coaxial cables [52]. Additionally, the VPS could help evaluate safety of patients with DBS implants undergoing MRI when using novel technologies, such as transmit array [34] (Fig. 6). Future OCD patients with DBS implants may profit from the proposed VPS by allowing for a MRI investigation instead of CT.

Acknowledgment

Funding provided by the National Institute of Health, National Institute of Biomedical Imaging and Bioengineering 1R21EB016449-01A1. he mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

References

[1] G. Bonmassar and N. Makris, "The Virtual Patient Simulator of Deep Brain Stimulation in the Obsessive Compulsive Disorder Based on Connectome and 7 Tesla MRI Data" in - *COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive Technologies and Applications*, Venice, Italy, vol.10, pp.235-238, 2014.

[2] P. R. Arantes, E. F. Cardoso, M. A. Barreiros, M. J. Teixeira, M. R. Goncalves, E. R. Barbosa, *et al.*, "Performing functional magnetic resonance imaging in patients with Parkinson's disease treated with deep brain stimulation," *Mov Disord,* vol. 21, pp. 1154-62, Aug 2006.

[3] F. G. Shellock and A. Spinazzi, "MRI safety update 2008: part 2, screening patients for MRI," *AJR Am J Roentgenol,* vol. 191, pp. 1140-9, Oct 2008.

[4] B. Stevenson, W. Dabney, and C. Frysz, "Issues and design solutions associated with performing MRI scans on patients with active implantable medical devices," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2007, pp. 6167-70, 2007.

[5] K. B. Baker, J. A. Tkach, M. D. Phillips, and A. R. Rezai, "Variability in RF-induced heating of a deep brain stimulation implant across MR systems," *J Magn Reson Imaging,* vol. 24, pp. 1236-42, Dec 2006.

[6] R. Buchli, P. Boesiger, and D. Meier, "Heating effects of metallic implants by MRI examinations," *Magn Reson Med,* vol. 7, pp. 255-61, Jul 1988.

[7] C. K. Chou, J. A. McDougall, and K. W. Chan, "RF heating of implanted spinal fusion stimulator during magnetic resonance imaging," *IEEE Trans Biomed Eng,* vol. 44, pp. 367-73, May 1997.

[8] L. M. Angelone, J. Ahveninen, J. W. Belliveau, and G. Bonmassar, "Analysis of the role of lead resistivity in specific absorption rate for deep brain stimulator leads at 3T MRI," *IEEE Trans Med Imaging,* vol. 29, pp. 1029-38, Apr 2010.

[9] J. Spiegel, G. Fuss, M. Backens, W. Reith, T. Magnus, G. Becker, *et al.*, "Transient dystonia following magnetic resonance imaging in a patient with deep brain stimulation electrodes for the treatment of Parkinson disease. Case report," *J Neurosurg,* vol. 99, pp. 772-4, Oct 2003.

[10] J. M. Henderson, J. Tkach, M. Phillips, K. Baker, F. G. Shellock, and A. R. Rezai, "Permanent neurological deficit related to magnetic resonance imaging in a patient with implanted deep brain stimulation electrodes for Parkinson's disease: case report," *Neurosurgery,* vol. 57, p. E1063; discussion E1063, Nov 2005.

[11] C. Roark, S. Whicher, and A. Abosch, "Reversible neurological symptoms caused by diathermy in a patient with deep brain stimulators: case report," *Neurosurgery,* vol. 62, p. E256; discussion E256, Jan 2008.

[12] D. Shrivastava, A. Abosch, T. Hanson, J. Tian, A. Gupte, P. A. Iaizzo, *et al.*, "Effect of the extracranial deep brain stimulation lead on radiofrequency heating at 9.4 Tesla (400.2 MHz)," *J Magn Reson Imaging,* vol. 32, pp. 600-7, Sep 2010.

[13] Medtronic, "MRI guidelines for Medtronic Deep Brain Stimulation Systems.," 2008.

[14] E. Cabot, T. Lloyd, A. Christ, W. Kainz, M. Douglas, G. Stenzel*, et al.*, "Evaluation of the RF heating of a generic deep brain stimulator exposed in 1.5 T magnetic resonance scanners," *Bioelectromagnetics,* vol. 34, pp. 104-13, Feb 2013.

[15] M. Oliveri, G. Koch, S. Torriero, and C. Caltagirone, "Increased facilitation of the primary motor cortex following 1 Hz repetitive transcranial magnetic stimulation of the contralateral cerebellum in normal humans," *Neurosci Lett,* vol. 376, pp. 188-93, Mar 16 2005.

[16] C. R. Merritt, H. T. Nagle, and E. Grant, "Fabric-Based Active Electrode Design and Fabrication for Health Monitoring Clothing," *Information Technology in Biomedicine, IEEE Transactions on,* vol. 13, pp. 274-280, 2009.

[17] C. R. Butson, S. E. Cooper, J. M. Henderson, and C. C. McIntyre, "Patient-specific analysis of the volume of tissue activated during deep brain stimulation," *Neuroimage,* vol. 34, pp. 661-70, Jan 15 2007.

[18] M. Seyal, A. J. Shatzel, and S. P. Richardson, "Crossed inhibition of sensory cortex by 0.3 Hz transcranial magnetic stimulation of motor cortex," *J Clin Neurophysiol,* vol. 22, pp. 418-21, Dec 2005.

[19] F. Tyc, A. Boyadjian, and H. Devanne, "Motor cortex plasticity induced by extensive training revealed by transcranial magnetic stimulation in human," *Eur J Neurosci,* vol. 21, pp. 259-66, Jan 2005.

[20] M. M. Elwassif, A. Datta, A. Rahman, and M. Bikson, "Temperature control at DBS electrodes using a heat sink: experimentally validated FEM model of DBS lead architecture," *J Neural Eng,* vol. 9, p. 046009, Aug 2012.

[21] M. I. Iacono, N. Makris, L. Mainardi, L. M. Angelone, and G. Bonmassar, "MRI-based multiscale model for electromagnetic analysis in the human head with implanted DBS," *Comput Math Methods Med,* vol. 2013, p. 694171, 2013.

[22] P. Gajsek, T. J. Walters, W. D. Hurt, J. M. Ziriax, D. A. Nelson, and P. A. Mason, "Empirical validation of SAR values predicted by FDTD modeling," *Bioelectromagnetics,* vol. 23, pp. 37-48, Jan 2002.

[23] P. A. Cannistraro, N. Makris, J. D. Howard, M. M. Wedig, S. M. Hodge, S. Wilhelm*, et al.*, "A diffusion tensor imaging study of white matter in obsessive-compulsive disorder," *Depress Anxiety,* vol. 24, pp. 440-6, 2007.

[24] J. F. Lehman, B. D. Greenberg, C. C. McIntyre, S. A. Rasmussen, and S. N. Haber, "Rules ventral prefrontal cortical axons use to reach their targets: implications for diffusion tensor imaging tractography and deep brain stimulation for psychiatric illness," *J Neurosci,* vol. 31, pp. 10392-402, Jul 13 2011.

[25] D. S. Tuch, V. J. Wedeen, A. M. Dale, J. S. George, and J. W. Belliveau, "Conductivity mapping of biological tissue using diffusion MRI," *Ann NY Acad Sci,* vol. 888, pp. 314-316, 1999.

[26] C. Gabriel, S. Gabriel, and E. Corthout, "The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues," *Phys. Med. Biol.,* vol. 41, pp. 2271–2293, 1996.

[27] J. A. McNab, B. L. Edlow, T. Witzel, S. Y. Huang, H. Bhat, K. Heberlein*, et al.*, "The Human Connectome Project and beyond: initial applications of 300 mT/m gradients," *Neuroimage,* vol. 80, pp. 234-45, Oct 15 2013.

[28] P. Bernardi, M. Cavagnaro, S. Pisa, and E. Piuzzi, "Specific absorption rate and temperature elevation in a subject exposed in the far-field of radio-frequency sources operating in the 10-900-MHz range," *IEEE Trans Biomed Eng,* vol. 50, pp. 295-304, Mar 2003.

[29] N. Makris, L. Angelone, S. Tulloch, S. Sorg, J. Kaiser, D. Kennedy*, et al.*, "MRI-based anatomical model of the human head for specific absorption rate mapping," *Med Biol Eng Comput,* vol. 46, pp. 1239-51, Dec 2008.

[30] C. K. Chou, H. Bassen, J. Osepchuk, Q. Balzano, R. Petersen, M. Meltz*, et al.*, "Radio frequency electromagnetic exposure: tutorial review on experimental dosimetry," *Bioelectromagnetics,* vol. 17, pp. 195-208, 1996.

[31] P. A. Filipek, C. Richelme, D. N. Kennedy, and V. S. Caviness, Jr., "The young adult human brain: an MRI-based morphometric analysis," *Cereb Cortex,* vol. 4, pp. 344-60, Jul-Aug 1994.

[32] N. Makris, J. W. Meyer, J. F. Bates, E. H. Yeterian, D. N. Kennedy, and V. S. Caviness, "MRI-Based topographic parcellation of human cerebral white matter and nuclei II. Rationale and applications with systematics of cerebral connectivity," *Neuroimage,* vol. 9, pp. 18-45, 1999.

[33] Available: http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm125520.htm

[34] L. M. Angelone, N. Makris, C. E. Vasios, L. Wald, and G. Bonmassar, "Effect of transmit array phase relationship on local Specific Absorption Rate (SAR)." in *ISMRM Fourteenth Scientific Meeting*, Seattle, USA, 2006.

[35] T. Eichele, S. Debener, V. D. Calhoun, K. Specht, A. K. Engel, K. Hugdahl*, et al.*, "Prediction of human errors by maladaptive changes in event-related brain networks," *Proc Natl Acad Sci U S A,* vol. 105, pp. 6173-8, Apr 22 2008.

[36] C. Gabriel, S. Gabriel, and E. Corthout, "The dielectric properties of biological tissues: I. Literature survey," *Phys. Med. Biol.,* vol. 41, pp. 2231–2249, 1996.

[37] C. Gabriel, S. Gabriel, and E. Corthout, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz," *Phys. Med. Biol.,* vol. 41, pp. 2251–2269, 1996.

[38] L. M. Angelone, E. Eskandar, and G. Bonmassar, "Resistive tapered stripline for deep brain stimulation (DBS) leads at 7 T MRI: Specific absorption rate analysis with high-resolution head model.," in *The 24th Progress in Electromagnetics Research Symposium (PIERS)*, Cambridge, MA, USA, 2008.

[39] W. R. Nitz, G. Brinker, D. Diehl, and G. Frese, "Specific absorption rate as a poor indicator of magnetic resonance-related implant heating," *Invest Radiol,* vol. 40, pp. 773-6, Dec 2005.

[40] IEC, "International Standard, medical equipment - 60601-2-33: Particular requirements for the safety of the magnetic resonance equipment for medical diagnosis, 2nd revision," Geneva2010.

[41] H. M. Duvernoy, *Human brain stem vessels : including the pineal gland and information on brain stem infarction*, 2nd completely rev. and expanded ed. Berlin ; New York: Springer, 1999.

[42] H. M. Duvernoy, *The human brain stem and cerebellum. surface, structure, vascularization, and three-dimensional sectional anatomy with MRI*. New York, 1995.

[43] H. M. Duvernoy, S. Delon, and J. L. Vannson, "Cortical blood vessels of the human brain," *Brain Res Bull,* vol. 7, pp. 519-79, Nov 1981.

[44] G. Schaltenbrand and P. Bailey, *Einführung in die stereotaktischen Operationen, mit einem Atlas des menschlichen Gehirns. Introduction to stereotaxis, with an atlas of the human brain*. Stuttgart,: G. Thieme, 1959.

[45] S. Pallanti and L. Quercioli, "Treatment-refractory obsessive-compulsive disorder: methodological issues, operational definitions and therapeutic lines," *Prog Neuropsychopharmacol Biol Psychiatry,* vol. 30, pp. 400-12, May 2006.

[46] D. D. Dougherty, L. Baer, G. R. Cosgrove, E. H. Cassem, B. H. Price, A. A. Nierenberg*, et al.*, "Prospective long-term follow-up of 44 patients who received cingulotomy for treatment-refractory obsessive-compulsive disorder," *Am J Psychiatry,* vol. 159, pp. 269-75, Feb 2002.

[47] H. H. Jung, C. H. Kim, J. H. Chang, Y. G. Park, S. S. Chung, and J. W. Chang, "Bilateral anterior cingulotomy for refractory obsessive-

compulsive disorder: Long-term follow-up results," *Stereotact Funct Neurosurg,* vol. 84, pp. 184-9, 2006.

[48] P. K. Bridges, J. R. Bartlett, A. S. Hale, A. M. Poynton, A. L. Malizia, and A. D. Hodgkiss, "Psychosurgery: stereotactic subcaudate tractomy. An indispensable treatment," *Br J Psychiatry,* vol. 165, pp. 599-611; discussion 612-3, Nov 1994.

[49] A. D. Hodgkiss, A. L. Malizia, J. R. Bartlett, and P. K. Bridges, "Outcome after the psychosurgical operation of stereotactic subcaudate tractotomy, 1979-1991," *J Neuropsychiatry Clin Neurosci,* vol. 7, pp. 230-4, Spring 1995.

[50] S. K. Bourne, A. Conrad, J. S. Neimat, and T. L. Davis, "Linear measurements of the cerebral ventricles are correlated with adult ventricular volume," *J Clin Neurosci,* vol. 20, pp. 763-4, May 2013.

[51] L. Yang, A. V. Sycheva, D. M. Black, and R. Eastell, "Site-specific differential effects of once-yearly zoledronic acid on the hip assessed with quantitative computed tomography: results from the HORIZON Pivotal Fracture Trial," *Osteoporos Int,* vol. 24, pp. 329-38, Jan 2013.

[52] E. Atalar, "Catheter Coils," in *eMagRes*, ed: John Wiley & Sons, Ltd, 2007.

Using Neural Networks and Feature Selection Algorithms in the Identification of Protein Signatures for the Prediction of Alzheimer's Disease

Lara Dantas and Mêuser Valença
Polytechnic School of Pernambuco
University of Pernambuco
Recife, Brazil
Email: {ldc, meuser}@ecomp.poli.br

***Abstract*—Alzheimer's Disease is now considered the most common type of dementia in the population. Although, it is a degenerative and irreversible disease, if diagnosed early, medications may be administered to slow the progression of symptoms and provide a better quality of life for the patient. Ray et al., and Gòmez and Moscato conducted studies with classifiers contained in the software Weka using a database with values of 120 blood proteins, and they noticed that they could classify the patient may or may not be diagnosed with AD with an accuracy rate of 93% and 65%, respectively. Thus, this study aims to use neural networks such as Multi-layer Perceptron, Extreme-learning Machine and Reservoir Computing to perform early diagnosis of a patient with or without AD. This article also envisions to utilize the Random Forest Algorithm to select proteins from the original set and, therefore, create a new protein signature. Through experiments it can be concluded that the best performance was obtained with the Multi-layer Perceptron and the new signatures created achieved better results than those available in the literature.**

Keywords–Neural Networks, Alzheirmer's Disease, Feature Selection Algorithms.

I. Introduction

Most developed countries are undergoing a major demographic shift. The oldest segments of the population are growing at a faster rate, and therefore, there is a constant increase in age-related diseases, especially progressive dementia disorders. First described by psychiatrist Alois Alzheimer in 1907, Alzheimer's Disease (AD) is, today, the most common cause of dementia in the elder population [1].

According to the Brazilian Institute of Geography and Statistics (IBGE) and the World Health Organization (WHO), there is 1.2 million people with AD in Brazil. It is believed that only 5% of the patients developed the disease at an early stage, i.e., before 65 years of age. In patients where the AD started after 65 years old, it is estimated that between 10% and 30% of these cases started after 85 their years old [2].

AD is a degenerative disease that causes irreversible death of several brain cells, the neurons. The patient suffering from this disease has a brain with microscopic pathologic lesions, known as neuritic plaques, and neurofibrillary tangles [3]. In addition, the brain of a person with Alzheimer's is much smaller than the brain of a healthy person, as it is shown in Fig. 1.

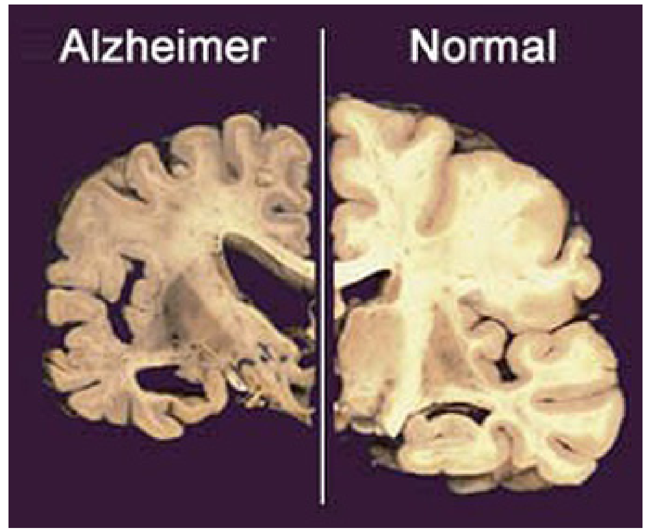


Figure 1. Brains illustrations. In the left side the one of a person with AD and in the right side a brain of a healthy person.

This disease develops in each patient in a unique way; however, there are several symptoms common to all of them, e.g., loss of memory, language disorders, depression, aggression, among others. Initially, the patient loses episodic memory, i.e., memory that holds information of events and their spatio-temporal relations. Thus, the old facts and the facts that just happened are easily forgotten. With the progress of the disease, semantic memory is also lost, i.e., lexical knowledge, rules, symbols are forgotten and the patient begins to lose its cultural identity [4].

The diagnosis of AD is often performed late since this disease can be confused with several other types of dementia and even the normal symptoms of aging. Although it is an irreversible disease, if it is discovered in its early stage, medications may be administered to slow the progression of symptoms and prolong the patient's welfare [5]. Thus, it is extremely important that mechanisms are developed for an earlier prediction of AD in the whole population.

Another common type of disease is the Mild Cognitive Impairment (MCI). The MCI causes cognitive changes that are noticed by the individuals experiencing them or to other people that live with the patient. However, this changes are not severe enough to interfere with daily life or independent function. People who are diagnosed with MCI have an increased risk of eventually developing AD.

In the literature, Ray et al. [6] conducted a study using a database of 120 samples of proteins contained in plasma of several patients. He concluded in his research that a combination of 18 out of the 120 available proteins enabled the realization of early diagnosis of AD with an classification rate of 91% using a set of tests with data from 92 patients who were diagnosed with or without AD.

In addition, he also used another set of tests containing data from 47 patients diagnosed with Mild Cognitive Impairment (MCI). For this set, the classification rate was 81%. These values were calculated from the average success rates found for all classifiers used in clinical trials for both sets [7].

Afterwards, Gòmez and Moscato conducted a study using 20 different classifiers available in the software Weka and defined various signatures with 18, 10, 6 and 5 proteins. These proteins were all contained in the set described by Ray et al. The 10 proteins signature reached a classification rate of 89% using the AD test set and 66% for the MCI test set. Furthermore, the 5 proteins signature reached a classification rate of 93% using the AD test set and 65% for the MCI test set. These success rates were also calculated from the average values of the 20 classifiers used [7].

Recently, Dantas and Valença [1] made a significant contribution. In the work mentioned, it was used the Random Forest Algorithm to create a new signature with 10 proteins and test its accuracy with 2 topologies of neural networks: Multi-Layer Perceptron and Reservoir Computing. After all the experiments, it was statistically proven that the new signature has a higher classification rate for the diagnosis of AD and MCI.

This paper aims to use another neural network topology to calculate the classification rate with the same signatures that were described in the work performed by Gòmez and Moscato and Dantas [7][1]. To meet this goal, the Extreme-learning Machine was the chosen technique.

Beyond that, this work also aims to use the Random Forest Algorithm, intending to create new signatures with 5 proteins. After that, this new signature will be tested using the Multi-layer Perceptron [8], Extreme-learning Machine [9] and the Reservoir Computing [10] and all results will be compared with the one available in the literature [1][6][7].

This article is organized into several sections. The first contains information about the neural network topologies that will be used. The next section describes the methodology used throughout this work, i.e., what database is used and how it is organized, the experiments and statistical analysis that was performed. Finally, there is a section that displays the results and the last one contains the conclusions obtained in this work.

II. Artificial Neural Networks

The Artificial Neural Networks (ANN) are models that intend to simulate the behavior of the human brain. The ANN contains simple processing units that are interconnected in order to process information. The knowledge of this model is stored in the weight contained in each of the connections between the artificial neurons.

The ANNs have been extensively used in many fields due to their ability to approximate complex nonlinear functions. These models have some advantages such as generalization, adaptability, ability to learn from examples.

Over the years, many studies have emerged in this field and several ANN topologies were created. In this paper, some of them will be discussed in the next sessions.

A. Multi-Layer Perceptron

One of neural networks topologies used in this paper was the Muti-Layer Perceptron (MLP). This network has the advantage of having intermediate layers. Thus, this characteristic guarantees that this neural network can approximate any continuous function as long as it has at least one intermediate layer, or any mathematic function if the number of intermediate layers is more than one.

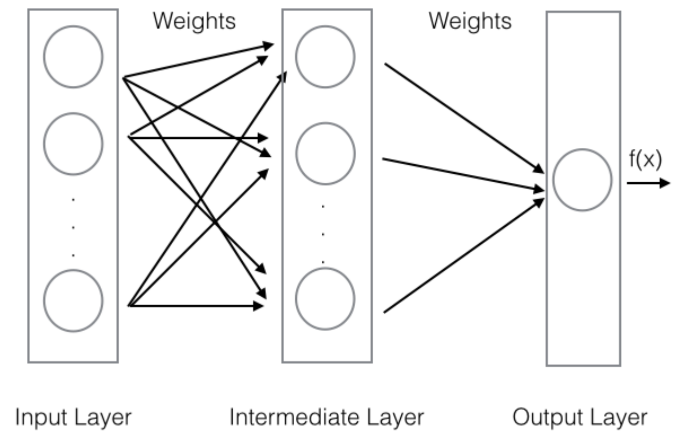


Figure 2. MLP with three layers

All the connections between the neurons of the MLP have a weight that, initially, have a random value but that during the training will be optimized. These neurons are disposed in three kinds of layers that is listed below:

- **Input layer:** Represents the input variables of the problem;
- **Intermediate or hidden layer:** This layer is responsible for the capacity of the MLP in solving non linear problems;
- **Output layer:** Represents the output variables of the problem

Fig. 2 shows an example of this network.

As well as other topologies of neural networks, the MLP also needs a training algorithm to optimize the weights of the MLP. In this paper, it was used the Back-propagation, a gradient-based algorithm [11].

1) Back-propagation algorithm: This algorithm is divided in two phases. In the first one, the forward phase, it happens the progressive signal propagation, i.e., it goes from the input layer to the output layer. In this phase, the weights are not changed and the output of the neural network and the error is calculated.

At this moment, the second phase, backward, of the algorithm is initiated. During this stage, the back-propagation error occurs, that is, based on the calculated output error, the weights connecting the hidden layer to the output layer are adjusted. This is accomplished using Equation (1) [11].

$$W_{i,j}^{m}(t+1) = W_{i,j}^{m}(t) + \alpha\delta_i^m f^{m-1}(net_j^{m-1}) + \beta\Delta W_{i,j}^{m}(t-1) \quad (1)$$

where

- α: represents the learning rate of the algorithm;
- β: represents the momentum rate;

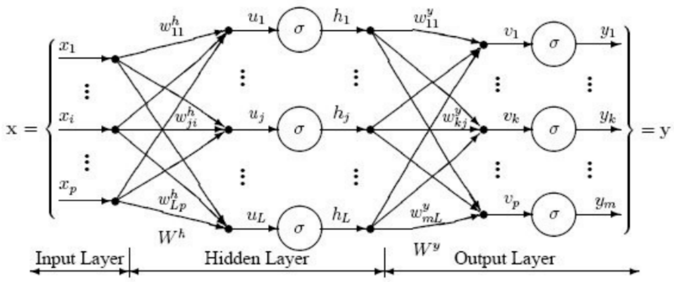


Figure 3. Single Layer Feedforward Network

- δ: represents the sensibility that can be calculated following Equation (2) to the neurons in the output layer and using Equation (3) to the neurons in the hidden layers [8].

$$\delta_i^m = (d_i - y_i) f'(net_i) \tag{2}$$

$$\delta_j^{m-1} = f'^{m-1}(net_j^{m-1}) \sum_{i=l}^{N} w_{ij}^m, \delta_i^m \tag{3}$$

More detailed information about the execution of this algorithm can be found in the literature [11][12][8].

B. Extreme-learning Machine

With the increase in the use of neural networks in many applications, it was noted that the learning speed of feedforward networks, such as MLP, usually was very slow. This problem also makes it difficult to use these neural networks for problems that require real-time response .

The main reason for this speed problem was that almost all ANNs were trained using gradient based algorithms. These kind of algorithms have several issues, such as the slower speed in the learning process, overfitting and local minimums.

In order to try to solve this problem, Huang et. al. created in 2004 the Extreme Learning Machine (ELM), a new training algorithm for Single-hidden Layer Feedforward Neural Networks (SLFNs) [9]. This algorithm randomly chooses the input weights and analytically determines the output weights of a SLFN [13]. The architecture of a SLFN is shown on Fig. 3.

ELM can be modeled following the Equations (4) and (5) [9]:

$$\sum_{i=l}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i.x_j + b_i) = o_j, j = 1, ..., N \tag{4}$$

$$\sum_{i=l}^{\tilde{N}} \beta_i g(w_i.x_j + b_i) = t_j, j = 1, ..., N \tag{5}$$

where

- (x_j, t_j): N input patterns;
- w_i: Weight vector of the neuron i from the hidden layer;
- b_i: bias of the neuron i from the hidden layer;
- $\tilde{N}$: number of neurons in the hidden layer;
- β_i: Weight vector between the hidden neuron i and the output layer.

The ELM also can be demonstrated in a matrix form as it described in Equation (6) [13].

$$H\beta = T \tag{6}$$

where

- H is expressed on Equation (7) [9];
- β is expressed on Equation (8) [9];
- T is expressed on Equation (9) [9];

$$H(w_1, ..., w_{\tilde{N}}, b_1, ..., b_{\tilde{N}}, x_1, ..., x_{\tilde{N}}) = \begin{bmatrix} g(w_1.x_1 + b_1) & \cdots & g(w_{\tilde{N}}.x_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(w_1.x_N + b_1) & \cdots & g(w_{\tilde{N}}.x_N + b_{\tilde{N}}) \end{bmatrix} \tag{7}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N}xm} \tag{8}$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_{\tilde{N}}^T \end{bmatrix}_{Nxm} \tag{9}$$

The ELM pseudo code training algorithm is available in Fig. 4.

Algorithm 1: ELM Pseudo code

```
BEGINNING
Randomly select values for the weights w_i and bias b_i, i = 1, ..., N;
Calculate the output matrix H of the hidden layer;
Calculate the output weights β = H†T;
END
```

Figure 4. Pseudo-code of the ELM algorithm [13]

where $H^\dagger$ is the Moore-Penrose Pseudo Inverse [9] and needs to satisfy the following properties:

- $HH^\dagger H = H$;
- $H^\dagger H H^\dagger = H^\dagger$;
- $(HH^\dagger)^T = HH^\dagger$;
- $(H^\dagger H)^T = H^\dagger H$.

This pseudo inverse can be efficiently calculated using the Single Value Decomposition (SVD) [13].

Furthermore, the ELM offers significant advantages such as ease of use, better generalization performance with much faster learning speed and it can be used in real-time applications.

A remarkable point to be considered is that the gradient based algorithms can be used for training neural networks with multiple hidden layers. In contrast, the ELM can only be used in SLFNs. However, as previously mentioned, with only a single hidden layer, the ANN can approximate any continuous function.

Thus, the use of this algorithm in this paper is valid, since it is a problem that can be solved with SLFNs networks.

C. Reservoir Computing

Recurrent Neural Networks (RNN) were created to enable the solution of dynamic problems. This is accomplished through a feedback of a neuron in a layer i to that found in some previous layer, $i-j$. This neural network topology has a better resemblance to the operation and behavior of the human brain [14].

In 2001, a new approach for the design of the training of a RNN was proposed by Wolfgang Mass called Liquid State Machine (LSM) [15]. At the same time, but independently, the same approach was described by Herbert Jaeger and called Echo State Machine (ESN) [16].

Both ESN and LSM networks have the Echo State Property (ESP) [17], i.e., due to the recurrent network connections, information from previous entries are stored. However, these data are not stored for an infinite period of time, and as well as the human brain, old information must be forgotten over time. Thus, the neural network has a rich set of information from the past and present therefore enhancing its applicability to dynamic systems [18].

In 2007, Verstraeten coined the term Reservoir Computing (RC) that unified the concepts described in ESN and LSM. Since then, this term is used in literature to illustrate learning systems which are represented by a dynamic recurrent neural network [10].

The RC is composed of three parts: an input layer, which as the MLP, represents the input variables of the problem, a reservoir, which can be seen as a large distributed and dynamic RNN with fixed weights, and a linear output layer called readout.

Fig. 5 represents the RC topology with two neurons in the input layer, three in the reservoir and one neuron in the output layer.

Although the problem addressed in this work is not dynamic, it was decided to test this neural network topology in order to verify the processing time and the memory effect of the RC and if it would have some positive impact on the diagnosis of AD.

The RC used in this work was based on the ESN approach and was developed in the Java programming language to make use of the object-oriented paradigm. This framework was created in order to solve classification and prediction problems and it was validated by three Benchmarks: Iris species, Thyreoide, Cancer and Diabetes.

The training algorithm for the RC was the same as the ELM, i.e., the outputs are calculated using the Moore-Penrose generalized inverse and it is more detailed on the next section.

1) Construction and Simulation of RC: The first step in order to prepare the RC and perform the data set classification is configure its architecture. Thus, it is necessary to define the number of neurons that will be used in the input, output and in the reservoir layers.

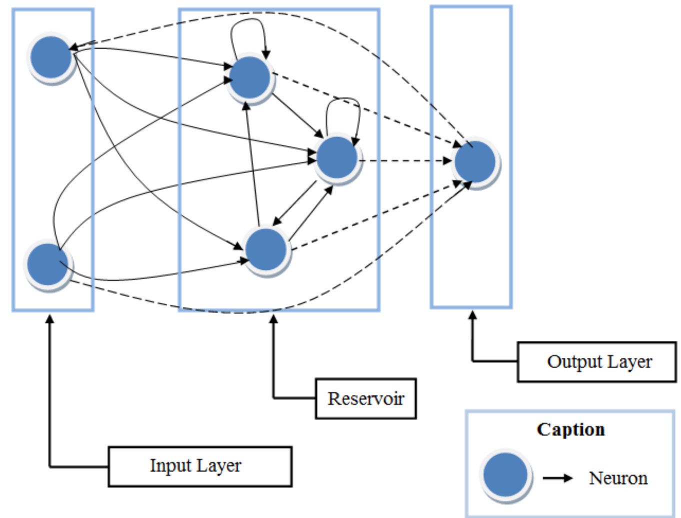


Figure 5. RC architecture. The dashed lines represent the weights that should be adjusted during the training of the network.

Furthermore, several RC parameters should also be determined. Being a recent methodology, there are no studies that prove how many neurons in the reservoir are necessary so that the neural network has better performance, or the rate of connectivity between these neurons. Therefore, for this work, values were defined based on some empirical tests performed.

Once the architecture of the neural network is determined, the next step is to generate the weight matrices connecting the input layer to the reservoir, W_{in}, and the matrix with the weights between neurons in the reservoir, W_{res}. Both matrices are generated with random values between -1 and 1.

Studies claim that the matrix W_{res} must have a spectral radius equal to 1 to provide a more numerical stability [19], i.e., when W_{res} is initialized, it must have its values changed as follows:

- Initially it must be decomposed into singular values;
- Then, W_{res} should have its values changed until the maximum value of the main diagonal of the eigenvalues matrix is less than or equal to 1.

To perform the simulation of the RC, the database is divided into three sets: training, used to perform the update of the states of the neurons of the reservoir, cross-validation, used to stop the training of the neural network, and test set, used to calculate the RC classification rate [20].

The states of the neurons in the reservoir must be initialized to zero. Since this is a recurrent network and RC stores its states (M_{est}) in a matrix, it is necessary that the final values found by the network are not so influenced by this initialization. Therefore, the literature suggests that before start training, a set of cycles called warm up is executed in order to perform updates in the states of the neurons in the reservoir and overlook the influence of the initial value [10]. The states are updated according to (10) [10]:

$$x[k+1] = f(W_{res}x[k] + W_{in}u[k]) \tag{10}$$

where $W_{in}u[k]$ represents the matrix containing the result of the product of the values derived from the input layer by the weights connecting these neurons to the reservoir at a time k and $W_{res}x[k]$ is the matrix with the states of the neurons from the reservoir at the same time k. The result will be assigned to $x[k+1]$, i.e., the state of the neuron RC in an instant forward will be the result of calculating the activation function of the neuron from the sum of the two parcels described above. In this work, the activation function used was the hyperbolic tangent according to equation 11[8].

$$f(net_i) = \frac{e^{net_i} - e^{-net_i}}{e^{net_i} + e^{-net_i}} \quad (11)$$

Where, y is the output value and net_i is the weighted average of the weights with the entries of the ith neuron.

Once the period of warm up is over, the training of the RC can be initialized. The first step should be to load the training set and perform the update of the states of the reservoir, noting that the matrices W_{in} and W_{res} should not be changed. They are randomly generated during construction of the RC, as described in the previous section, and should not be adjusted.

Still during training, the weights matrix that connects the neurons of the input layer to the output (W_{inout}) and the one that connects the reservoir to the output must be calculated by the pseudo-inverse of Moore-Penrose. As they are non-square matrices and their determinants can approach zero, it is necessary to calculate the pseudo-inverse.

At the end of each training cycle, a cross validation cycle should be initiated. This process should be repeated until the stopping criteria is reached and the training set is finalized. During the process of cross-validation, the matrices W_{inout} and W_{out} should remain being readjusted.

When the process of training is finished, the testing process begins. The set of tests is presented to the RC and at this time, all the weights matrices, W_{in}, W_{res}, W_{inout} and W_{out}, should remain unchanged, as the matrix M_{est}. At this point, the classification error is calculated. These values will be used in the future to make the necessary comparisons.

The behavior of the RC can be best viewed through the algorithm described in Fig. 6[10].

Algorithm 1: Pseudocode of RC

```
Set the number of neurons in the input layer ;
Set the number of neurons in the reservoir layer ;
Set the number of neurons in the output layer ;
Randomly generate the weights of Win matrix between
-1 e 1;
Randomly generate the weights of Wres matrix between
-1 e 1;
Normalize the weights of Wres matrix so that the
spectral radius of the matrix is smaller than or equal to
1;
while until the end of the number of warm up cycles do
  | updates the states of the neurons of the RC;
end
while until the stopping criterion is reached do
  for each value of the input set do
    | updates the states of the neurons of the RC;
  end
  Calculates the Moore-Penrose inverse matrix to find
  the weights connecting the RC to the output layer;
  Calculates the Moore-Penrose inverse matrix to find
  the weights connecting the input layer to the output
  layer;
  for each value of the cross-validation set do
    | updates the states of the neurons of the RC;
  end
  Calculates the output values of the RC;
  Calculates the RMSE;
  Checks if the stopping criterion has been reached;
end
for each value in the set of tests do
  | updates the states of the neurons of the RC;
end
Calculates the output values of the RC;
Calculate the accuracy rate;
```

Figure 6. Reservoir Computing pseudo-code

III. RANDOM FOREST ALGORITHM

Although, the Random Forest (RF) Algorithm is an excellent classifier, this technique can be used to rank the importance of variables in a classification problem [21].

To measure the importance of each variable in a data set D_n, it has to fit a random forest to the data. The data set can be expressed as it shown in Equation (12) [22].

$$D_n = (X_i, Y_i)_{i=1}^{n} \quad (12)$$

where

- X is the training set;
- Y is the responses set;
- n is the number of the examples in the data set.

Fig. 7 shows the pseudo code used to calculate the Variable Importance (VI) with the RF. It is important to mention that this algorithm uses the Out-of-bag (OOB) error estimation in the formula to measure the VI [21].

Algorithm 2: Pseudo code of the calculus of VI using RF

```
for each tree t do
  | Consider the associated OOB_t sample; Denote by
  | errOOB_t the error of a single tree t on this OOB_t
  | sample; Randomly permute the values of X^j in
  | OOB_t to get a perturbed samples denoted by
  | OOB_t^j, the error of predictor t on the perturbed
  | sample.
end
```

Figure 7. Pseudo code of the calculus of VI using RF

With all errors calculated with the pseudo code described in Fig. 7, the VI coefficient must be calculated using the Equation (13) [22].

Algorithm 3: Pseudo code of the variable selection with RF

```
while Preliminary elimination and rankin doesn't finish
do
    Sort the variables in decreasing order of RF scores
    of importance; Cancel the variables of small
    importance. Denote by m the number of remaining
    variables.
end
while Variable selection doesn't finish do
    Construct the nested collection of RF models
    involving the first k variables, for k = 1 to m, and
    select the variables involved in the model leading to
    the smallest OOB error.
end
```

Figure 8. Pseudo code of the calculus of VI using RF

$$VI(X^j) = \frac{1}{ntree}\sum_{t=0}^{j}(err\tilde{O}\tilde{O}B_t^j - errOOB_t) \qquad (13)$$

Since all VI have been calculated, now it is possible to select the most important variables from the original data set. To perform this action the steps described in Fig. 8 must be followed [21].

In the case of this paper, the RF algorithm was applied in the database and Fig. 9 shows the graph with all proteins ordered by the VI coefficient.

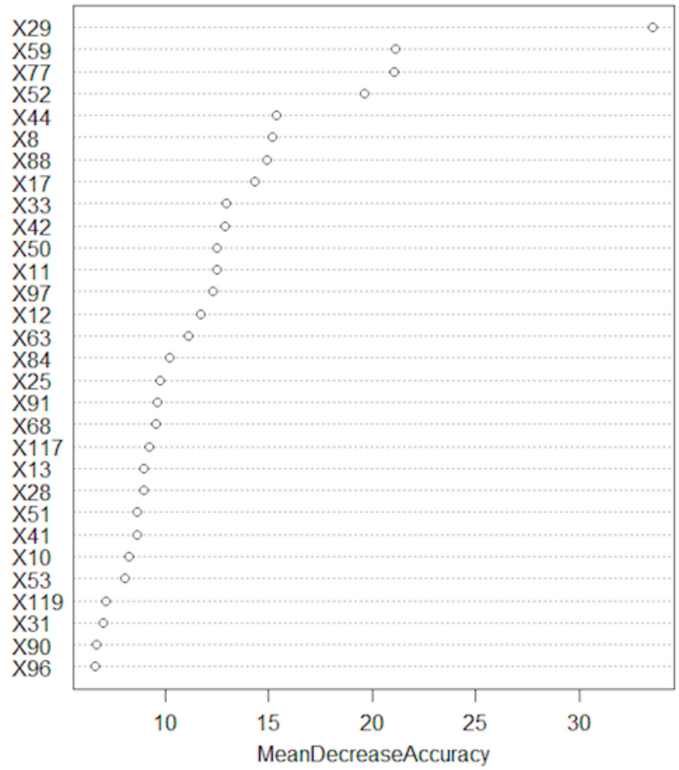


Figure 9. Application of the RF algorithm in the database used in this work

IV. Experiments

A. Database

The database used in the development of this work was the same used by Gòmez and Moscato et al. and Dantas in their publication. It has values of 120 proteins found by analysis of blood samples from different patients. The ultimate goal of the database is to classify whether a patient can be diagnosed or not with AD or MCI [7] [1].

In his work, Gòmez and Moscato et al. subdivided the database in 2 sets. The first set contained the results of blood samples of 83 patients. Of these 83 patients, 68 were allocated to the training process of the chosen classifier. The data for the remaining 15 patients were used in the process of cross-validation of the classifier, i.e., a process that determines the optimal point to stop its training [8].

The second set, used in the testing process of the classifier, has two options. It could be used to diagnosis AD and in this case, this set will contain the samples related to the 92 patients that could be diagnosed with AD. The second option is use this set to perform diagnosis of MCI. In this other case, the test set will contain blood samples related to 47 patients with a possible diagnosis of MCI.

Recently, Dantas defined two new signatures with 10 proteins with the objective to compare the results with the one obtained by Gòmez and Moscato et al. In this work those two signatures will be used to verify the accuracy of the ELM algorithm.

Besides that, aiming a comparison with the signatures previously defined by Gòmez and Moscato [7], two new signatures of 5 proteins were proposed. One of them used to perform the diagnosis of Alzheimer's Disease and the other for MCI.

The Random Forest algorithm was executed 30 times and in each of the simulations, a signature of 5 proteins was defined. After all these 30 simulations were over, the best signature was chosen according to the gini metric.

It is important to mention that Gòmez and Moscato [7] used the same signature in both cases, that is, the signature composed by 10 and 5 proteins is used for the AD and MCI testing sets.

Table I shows the signatures of proteins that are contained in Gòmez and Moscato et al. work and which are used in this study.

In this work, 8 databases were prepared in order to reproduce the experiments described by Gòmez et al. [7] and Dantas [1], using the MLP, RC and the ELM. They were:

- 1 database for testing the signature of 10 proteins defined by Gòmez et al. with the AD set of tests, called now by **Database 1**;
- 1 database for testing the signature of 10 proteins defined by Gòmez et al. with the MCI set of tests, called now by **Database 2**;
- 1 database for testing the signature of 10 proteins defined by Dantas with the AD set of tests, called now by **Database 3**;

TABLE I. REPRESENTATION OF THE PROTEINS CONTAINED IN EACH ONE OF THE SIGNATURES USED.

Abbreviation	Signature	Proteins
S1	**10 proteins signature defined by Gòmez et al. [7]**	CCL7/MCP-3, CCL15/MIP-1d, EGF, G-CSF, IL-1a, IL-3, IL-6, IL-11, PDGF-BB, TNF-a
S2	**10 proteins signature defined by Dantas [1] for AD test set**	IL-1a, TNF-a, G-CSF ,PDGF-BB, IGFBP-6, M-CSF, EGF, IL-3, GDNF, Eotaxin-3
S3	**10 proteins signature defined by Dantas [1] for MCI test set**	IL-1a, PDGF-BB, EGF, TNF-a, RANTES, FAS, GCSF, MIP-1d, FGF-6, IL-11
S4	**5 proteins signature defined by Gòmez et al. [7]**	EGF, G-CSF, IL-1a, IL-3, TNF-a
S5	**5 proteins signature defined by the Random Forest for AD test set**	IL-1a, TNF-a, G-CSF ,PDGF-BB, M-CSF
S6	**5 proteins signature defined by the Random Forest for MCI test set**	IL-1a, PDGF-BB, EGF, TNF-a, RANTES

- 1 database for testing the signature of 10 proteins defined by Dantas with the MCI set of tests, called now by **Database 4**;
- 1 database for testing the signature of 5 proteins defined by Gòmez et al. with the AD set of tests, called now by **Database 5**;
- 1 database for testing the signature of 5 proteins defined by Gòmez et al. with the MCI set of tests, called now by **Database 6**;
- 1 database for testing the signature of 5 proteins defined by Random Forest Algorithm with the AD set of tests, called now by **Database 7**;
- 1 database for testing the signature of 5 proteins defined by Random Forest Algorithm with the MCI set of tests, called now by **Database 8**;

All databases described above maintained the organization used by Gòmez et al. regarding the division of values for the training, cross validation and testing set.

1) Pre-processing of data: To properly execute the training of the neural network it is necessary that your data is normalized, i.e., the input values of the neural network must be contained in the same numerical range. This is important since very different values can influence the training and generate a loss in the generalization ability of the neural network [8].

One of the most commonly used normalization techniques in literature is the linear transformation and it was the one chosen for this work. Equation (14) is the formula used to normalize the values of the database.

$$y = ((b - a) \times (\frac{x - x_{min}}{x_{max} - x_{min}})) + a \tag{14}$$

In (14), a and b represent the maximum and minimum values that the data should take. In this work, it was used the value of -0.85 for a and 0.85 for b, as the activation function chosen for this neural network is the Hyperbolic Tangent. Therefore, the values contained in the database must be between -1 and 1.

B. Simulations

1) MLP Parameters: Although it is widely used in many researches, several parameters of the MLP must to be configured. The choice of each one directly influences the final outcome of the prediction.

Below are the main parameters of the MLP and Backpropagation algorithm:

- Number of neurons in the input layer;
- Number of neurons in the hidden layer (only one hidden layer);
- Number of neurons in the output layer;
- Activation function of the hidden layer;
- Activation function of the output layer;
- Learning Rate;
- Momentum.

During the experiments, the number of neurons in the input layer was varied according to the amount of proteins in the signature of each experiment. These values can be either 5 or 10.

This work aimed to perform early diagnosis of a patient with or without AD or MCI, thus the number of neurons in the output layer is 2.

The algorithm used to adjust the weights is the Backpropagation and its formula can be found in section II-A.

The activation function chosen of the hidden layer is the hyperbolic tangent. This function returns values in the interval [-1, 1] and is given by 11

Several tests were performed to define the learning rate, momentum and number of neurons in the hidden layer. The best results correspond to the values of 0.3 for the learning rate, 0.6 for the momentum and 20 neurons in the hidden layer.

The MLP used has been implemented in the JAVA programming language and in the Eclipse development environment [23].

2) RC Parameters: As the MLP, the Reservoir Computing technique has several parameters that require configuration. Taking into account that it is a recent area of research, the choice of these settings can not be considered ideal and it is often performed randomly. One way to do this is to evaluate each chosen parameter value and determine if it was better or worse for the network performance. This process is repeated until a value is considered optimal, which does not necessarily means the best.

In this work, these parameter values were changed and all the classification rates at the end of the trainings were stored and then compared. The configuration which presented the best parameter values for the network performance was the one chosen.

Below are the parameters whose settings were required to be defined during this project:

- Number of neurons in the input layer;
- Number of neurons in the output layer;
- Number of neurons in the reservoir;
- Activation function of the reservoir;
- Activation function of the output layer;

- Initialization of weights;
- Connection rate of the reservoir;
- Number of warm up cycles;

The number of inputs and outputs remain the same as used in the MLP, since the goal is to compare the ANN's techniques using the same classification scenario. It means the number of neurons in the input layer can be 5 or 10, depending on the signature used. The number of neurons in output layer is 2.

The number of neurons in the reservoir is one of the parameters for which there is no fixed criterion that defines it. It was chosen randomly after checking the classification rate at the end of each training. It was observed that the ideal number of neurons in the reservoir was 4.

As mentioned in Section II-C, the weights of the input layer to the reservoir and the weights of the reservoir are randomly generated with values between [-1, 1].

The reservoir states are initialized to zero (0). Due to this, as also mentioned in Section II-C, it was decided to add to the network a phase called warm up. During the warm up, it is not necessary to calculate the weights of the output layer, or to calculate an output value. This warm up phase is done just to update the states of the reservoir and remove the dependence on the initial state. The number of cycles chosen for warm up was 100.

The connection rate of the reservoir neurons was 10%. That is, only 10 % of the connections have weight values different from zero associated to them.

The activation function chosen in the reservoir was the hyperbolic tangent. In the output layer, the selected function was linear one.

During this work, we implemented a Neural Network with the technique of RC in the programming language Java and the Eclipse development environment.

3) ELM Parameters: Finally, the ELM also needs to have its parameters adjusted. Below are the parameters whose settings were required to be defined during this project:

- Number of neurons in the input layer;
- Number of neurons in the hidden layer (only one hidden layer);
- Number of neurons in the output layer;
- Activation function of the hidden layer;
- Activation function of the output layer;

The same number of neurons in the input and output layer used in the two other techniques is maintained for the ELM. That is, 5 or 10 inputs and 2 outputs.

The number of neurons in the hidden layer remains the same used with the MLP which means 20 neurons.

The activation function of the hidden layer is the hyperbolic tangent and of the output layer is the linear.

Table II summarizes the parameters used in all experiments with the neural networks topologies mentioned above.

After defining the settings of the MLP, ELM and RC, 30 simulations were performed with each of the databases in each of the chosen neural network topologies in this work. This number is considered ideal to perform more meaningful statistical comparisons [24].

TABLE II. REPRESENTATION OF THE PARAMETERS USED FOR THE SIMULATIONS WITH THE RC, MLP AND ELM.

Parameters	RC value	MLP value	ELM value
RC connectivity	10%	Not applicable	Not applicable
Number of neurons in the input layer	Depends of the amount of proteins in the signature	Depends of the amount of proteins in the signature	Depends of the amount of proteins in the signature
Number of neurons in the RC	4	Not applicable.	Not applicable
Number of neurons in the hidden layer	Not applicable	20	20
Number of neurons in the output layer	2	2	2
Number of warm up cycles	100	Not applicable.	Not applicable
Activation function of neurons in the reservoir or hidden layer	Hyperbolic Tangent	Hyperbolic Tangent	Hyperbolic Tangent
Activation function of neurons in the output layer	Linear	Linear	Linear
Learning rate	Not applicable.	0.3	Not applicable
Momentum	Not applicable.	0.6	Not applicable

TABLE III. NULL AND ALTERNATIVE HYPOTHESIS FOR THE SHAPIRO-WILK TEST.

Hypothesis	Description
Null Hypothesis	The sample is normally distributed
Alternative Hypothesis	The sample isn't normally distributed

TABLE IV. NULL AND ALTERNATIVE HYPOTHESIS FOR THE F TEST.

Hypothesis	Description
Null Hypothesis	The samples have variances statistically equal
Alternative Hypothesis	The samples don't have variances statistically equal

C. *Statistical Analysis*

When all the simulations were completed, it was necessary to perform a sequence of statistical tests in order to scientifically validate the results. For this, it was used the R mathematical software, since it contains all the implementations of the tests used. This software uses as default a level of significance (α) previously defined with the value of 0.05.

Before using a parametric test on a data set it is necessary to check whether the samples are normally distributed and if they have statistically equal variances. If these two assumptions are validated, one can apply a parametric test, otherwise it must be used a non-parametric test.

In order to verify the first assumption, that is, check if the samples were normally distributed, the Shapiro-Wilk test was applied with the hypothesis described in Table III.

After that, it is necessary to verify whether or not the samples were drawn from the same population, i.e., if their variances were statistically equal. The hypothesis for this test are available on Table IV.

If these two premises were true, a parametric test can be used. In this case, it was chosen the Student's T-test. The hypothesis are described in Table V.

In case of the first two premises aren't true at the same time, it is necessary to use a non-parametric test, i.e., one that

TABLE V. NULL AND ALTERNATIVE HYPOTHESIS FOR THE STUDENT'S T-TEST

Hypothesis	Description
Null Hypothesis	The average of the analyzed samples are statistically equal
Alternative Hypothesis	The average of the analyzed samples aren't statistically equal

TABLE VI. NULL AND ALTERNATIVE HYPOTHESIS FOR THE WILCOXON RANK-SUM TEST

Hypothesis	Description
Null Hypothesis	The median of the analyzed samples are statistically equal
Alternative Hypothesis	The median of the analyzed samples aren't statistically equal

TABLE VII. REPRESENTATION OF THE AVERAGE ACCURACY RATES AFTER THE 30 EXPERIMENTS.

Database	Average classification rate with RC / Standard Deviation	Average classification rate with MLP / Standard Deviation	Average classification rate with ELM / Standard Deviation
Database 1	86.62% / 0.026	93.44% / 0.017	87.78% / 0.025
Database 2	69.29% / 0.024	68.15% / 0.018	68.45% / 0.027
Database 3	90.57% / 0.022	94.31% / 0.008	91.05% / 0.023
Database 4	76.59% / 0.047	78.86% / 0.031	74.32% / 0.050
Database 5	93.22% / 0.026	95.61% / 0.017	91.12% / 0.021
Database 6	65.44% / 0.024	69.14% / 0.018	66.81% / 0.019
Database 7	92.97% / 0.022	93.26% / 0.008	91.62% / 0.028
Database 8	72.90% / 0.047	73.33% / 0.031	73.88% / 0.040

makes no assumptions about the probability distribution of the samples. It was chosen the Wilcoxon Rank-Sum Test with the hypothesis in Table VI.

For all those cases, the p-value must be compared with the significance level adopted in the R software (α) . In case of this value is less then α the null hypothesis should be rejected implying that the alternative hypothesis is automatically accepted.

V. RESULTS

After all simulations were performed with the databases, it was calculated the arithmetic mean for each set of simulations and Table VII displays those values found.

As none of the eight datasets met the two premises necessary for the application of a parametric test at the same time, it was not possible to perform the Student's T-test. And so, the Wilcoxon Rank-Sum Test was chosen, since it is a non-parametric test, i.e., it makes no assumptions about the probability distribution of the samples.

When applied the Wilcoxon Rank-Sum Test for each of the eight cases, the results found were that the MLP has a statistically better performance than the RC and ELM.

In order to verify the performance of the new proposed signatures, simulations with RC, MLP and ELM topologies were performed for both the diagnosis of AD and MCI. The results were compared with those found by the same neural networks when the signatures used were the ones defined by Gòmez and Moscato et al.

In all cases, the classification rate showed improvement when the new signatures were used for all architectures. After the Wilcoxon Rank-Sum test, this statement was confirmed statistically. The maximum and average values are also bigger than those described by Gòmez and Moscato et al.

TABLE VIII. COMPARISON OF THE RESULTS WITH THE NEW PROTEIN SIGNATURE PROPOSAL OBTAINED WITH RC, MLP, ELM AND THE ONES AVAILABLE IN LITERATURE.

Protein Signature	RC - Maximum value	MLP - Maximum value	ELM - Maximum value	Results found by Gòmez and Moscato et al.
New 10-protein signature for AD proposed by Dantas[1]	96.73%	95.65%	96.81%	93%
New 10-protein signature for MCI proposed by Dantas [1]	89.36%	82.97%	89.20%	66%
New 5-protein signature for AD	98.15%	97.33%	95.29%	93%
New 5-protein signature for MCI	73.50%	74.21%	74.49%	66%

Table VIII summarizes the maximum values found in the simulations of the RC, MLP and ELM. Those results were found using all new signatures proposed. Table VIII also display the results obtained in the work of Gòmez and Moscato et al. using their own signature [7].

From Table VIII, it can be concluded that all neural network topologies used obtained results consistent with those described by Gòmez and Moscato et al. and succeeded in reaching a maximum value greater than the average found in the literature.

VI. CONCLUSION

Nowadays, Alzheimer's disease is one of the most common diseases in the elderly population. In recent years, the number of patients has grown significantly since the life expectancy in most developed countries has increased.

AD is a degenerative disease, i.e., brain cells will deteriorate and there is no way to reverse the disease. However, the earlier the drugs are administered, the better the quality of life of the patient since the medication will slow the progression of the symptoms.

Thus, this study aimed to verify the performance of MLP, RC and ELM to early classify if a patient can be diagnosed with AD or not. Moreover, another goal was to make a comparison of the performance of the MLP with the RC and ELM neural networks, and also with the results available in the literature.

From the statistical tests and simulations, it can be concluded that the MLP presented a superior performance in all cases. It is also possible to conclude that the four new signatures proposed achieved better results when compared to those shown by Gòmez and Moscato et al. Furthermore, they also had better performance when compared to the results obtained from the same neural network topologies when the signatures used were the ones proposed by Gòmez and Moscato et al.

As future work, it is intended to use other neural networks topologies to make a comparison with those already used in this work. Besides that, it is intended to invest in more variable selection techniques in order to further optimize the results and to reduce the number of proteins in the signatures used to perform early diagnosis of Alzheimer and MCI.

References

[1] L. Dantas and M. Valença, "A comparative study of neural network techniques to perform early diagnosis of alzheimer's disease," The Sixth International Conference on Advanced Cognitive Technologies and Applications, 2014, pp. 178–183.

[2] H. Portal. Portal of the brazilian health. [retrieved: July, 2014]. [Online]. Available: http://portal.saude.gov.br/saude/ (2014)

[3] R. Green, Diagnosis and Treatment of Alzheimer's Disease and Other Dementias. Editora de Publicações Científicas Ltda., 2001

[4] P. Marques, Dementia of Alzheimer type: diagnostic, treatment and social aspects. Editora de Publicações Científicas Ltda., 1997.

[5] E. Giusti and V. Surdo, Alzheimer's care and family counseling: psychological needs and treatment of dementia. Gryphus, 2010.

[6] S. Ray, M. Britschgi, C. Herbert and A. Boxer, "Classification and prediction of clinical alzheimer's diagnosis based on plasma signaling proteins." Nat. Med., vol. 13, 2007.

[7] M. Gómez and P. Moscato, "Identification of a 5-protein biomarker molecular signature for predicting alzheimer's disease," PLos One, vol. 3, 2008, p. 12p.

[8] M. Valença, Fundamentals of Neural Networks: Examples in Java, 2nd ed. Livro Rápido, 2009.

[9] G. Huang, Q. Zhu and C. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," International Joint Conference on Neural Networks, 2004.

[10] D. Verstraeten, "Reservoir computing : computation with dynamical systems," Ph.D. dissertation, Ghent University. Faculty of Engineering, Ghent, Belgium, 2009.

[11] P. Werbos, "Beyond regression: New tools for prediction and analysis in behavioral sciences." Ph.D. dissertation, Harvard University, 1974.

[12] S. Haykin, Neural Networks: Principles and Practices. Bookman, 2007.

[13] R. Rajesh and J. Prakash, "Extreme learning machines - a review and state-of-the-art," International Journal of Wisdom Based Computing, vol. 1, 2011, pp. 35–49.

[14] M. Valença, Applying Neural Networks: A Complete Guide, 1st ed. Livro Rápido, 2005.

[15] W. Maass, Motivation, theory, and applications of liquid state machines. Imperial College Press, 2011.

[16] H. Jaeger, "The echo state approach to analysing and trainning recurrent neural networks," German National Resource Center for Information Technology, Tech. Rep., 2010.

[17] M. Massar and S. Massar, "Mean-field theory of echo state networks," PHYSICAL REVIEW E, vol. 87, 2013.

[18] M. Lukoševičius, A practical guide to applying echo state networks. Springer Berlin Heidelberg, 2012, pp. 659–686.

[19] A. Araújo, "A method for design and training reservoir computing applied to prediction of time series," Ph.D. dissertation, Universidade de Federal de Pernambuco, 2001.

[20] M. Kulkarni and C. Teuscher, "Memristor-based reservoir computing," in Proceedings of the IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH'12), 2012.

[21] A. Liaw and M. Wiener, "Classification and regression by randomforest," R news, vol. 2, no. 3, 2002, pp. 18–22.

[22] R. Genuer, J. Poggi and C. Malot, "Variable selection using random forests," Pattern Recognitions Letters 31, 2010.

[23] S. Santos, "An application for prediction of water resources based on neural networks and particle swarm," Master's thesis, Universidade de Pernambuco, 2012.

[24] N. Juristo and A. Moreno, Basics of Software Engineering Experimentation. Kluwer Academic Publishers, 2001.

www.iariajournals.org

International Journal On Advances in Intelligent Systems
issn: 1942-2679

International Journal On Advances in Internet Technology
issn: 1942-2652

International Journal On Advances in Life Sciences
issn: 1942-2660

International Journal On Advances in Networks and Services
issn: 1942-2644

International Journal On Advances in Security
issn: 1942-2636

International Journal On Advances in Software
issn: 1942-2628

International Journal On Advances in Systems and Measurements
issn: 1942-261x

International Journal On Advances in Telecommunications
issn: 1942-2601