Usability Analysis of Archetyped Interfaces for the Electronic Health Record: a Comparative Study

Danielle Santos Alves, Valéria Cesário Times,
André Magno Costa de Araújo, Marcus Urbano da Silva
Center for Informatics (CIN) at Federal University of Pernambuco (UFPE)
Recife, Brazil
{e-mail: dsa3, vct, amca, mus}@cin.ufpe.br

Amadeu Sá Campos Filho, Magdala de Araújo Novaes
TeleHealth Center (NUTES), Clinical Hospital, Federal University of Pernambuco (UFPE)
Recife, Brazil
{e-mail: amadeu.campos, magdala.novaes}@nutes.ufpe.br

Abstract— Few studies about OpenEHR standards assess usability aspects. This paper aims to evaluate archetyped interfaces, built by a user interface building tool, with respect to usability requirements of health care professionals. Such an assessment is carried out by comparing two user interface building tools. We carried out experimental tests with Health professionals to evaluate the generated graphical user interfaces by a standard openEHR tool and a framework proposed by researchers to build Health applications dynamically using archetypes. Quality in Use Integrated Map (QUIM) and Questionnaire for User Interface Satisfaction (QUIS) questionnaires were used to evaluate the usability aspects. The Likert Scale was adopted to evaluate the interface concepts of the tools, such as efficiency, effectiveness and satisfaction. A T-student test was performed to compare the results, which showed that the second tool achieved better ratings in all the analyzed concepts when compared to the first tool, all being statistically significant. The usability characteristics raised by the users are listed. The conclusion is that the interface generated by the second tool brought more user satisfaction in comparison to the first tool.

Keywords- Archetypes; Electronic Health Records; usability tests; archetyped interfaces; user interface building tool.

I. INTRODUCTION

Nowadays Electronic Health Records (EHR) are the subject of many studies, especially those related to achieving interoperability between systems for the future of international health services [1][2]. The main global initiatives focused on EHR interoperability are International Organization for Standardization (ISO), HL7 and the openEHR foundation. To standardize the content semantics, the global academic community has suggested the use of archetypes and terminology [3].

Archetypes may be conceptualized as a set of specifications that define a reference model of Health information; i.e., a language for constructing "clinical models" [3]. For this paper, we define an 'archetyped interface' as a User Interface (UI), which is generated from the specifications, restrictions and terminologies defined within archetypes. Archetyped EHR systems are those that use archetypes in their content definition.

Many authors have studied the human-computer relationship and reflected on this interaction, which relates especially to the interface component "with emphasis on the human side, the relevance, the utility, among others" [4].

Research on the use and reuse of archetypes to build the EHR achieved progress in the development of archetyped clinical systems, but many of these studies have not taken into consideration the usability aspects [5]. It is known that the usability aspects are very important for the successful implementation and compliance of an EHR system by the users[4].

There are several possibilities to promote the inclusion of evaluation and usability tests with potential users of EHR products during the system development. Ideally, the users are not engaged in the system design phase (formative tests), but are regularly consulted during business process validation steps and interface design prior to the approval of the final product (summative tests) [6].

Furthermore, the interface is always mentioned in studies as a relevant aspect to be improved by reducing the total number of stages of the process and the percentage of mental effort required while performing tasks, so users feel more confident when they are using the system [7][8]. This paper aims to evaluate the archetyped interfaces built by a user interface building tool with respect to usability requirements of the health care professional through a comparison between two user interface building tools.

The remainder of this paper is organized as follows: Section II describes the background and the related work. Section III presents the usability concepts adopted in the study and describes the influence factors linked to each usability concept used. In Section IV, the methodology is explained, spanning experiment design, data analysis, support instruments and the controlled experiment users. Section V shows the results and includes some discussion about the findings. Finally, Section VI presents the authors’ conclusion.

II. BACKGROUND AND RELATED WORK

Some recent studies have described factors that negatively and positively influenced the process of implementing a computerized system in large hospitals in the
world. All these studies mentioned the importance of identifying the views of users in the context of implementation. The target population of the studies was the nursing team, which was considered the least resistant team regarding the use of computers during the computerization process [9]–[11].

A study performed at The Johns Hopkins Hospital, (Baltimore, Maryland) in 2003 and published in 2008 by the Association of periOperative Registered Nurses (AORN) Journal, shows the importance of including the nursing staff and members of the information technology staff in the pre-selection of the requirements process, as well as in the final analysis system and implementation. The researchers noticed that for the successful development and implementation of a hospital information system, a technological development is necessary (albeit not enough), alongside a scientific and methodological deepening. According to Saletnik, Niedlinger and Wilson [10], “nurses played key roles in the planning and implementation phase of the system or of information management at Johns Hopkins”. To conclude, the authors related that “implementation of information technology is not a purely technical project and, as such, cannot be left to the information technology (IT) department. Health is characterized by a level of complexity that defies predictability required in many IT implementations. The development team, therefore, must be prepared to learn and adapt to the problems” [10].

In a study with 48 nursing leaders, researchers aimed to evaluate usability aspects of the management system used in the hospitals through a focus group. The researchers summarized the following results: the nurses said that there is a reduction in efficiency in the work process after the application of not fully tested computerized systems. However, they consider electronic systems essential in daily work. Nurses also reported strategic management problems and a lack of coordination during the implementation process of electronic systems [12].

Another study that aimed to analyze and evaluate the perception of the users towards EHR was based on the evaluation of 113 nurses from different shifts of a primary health care institution in Catalonia, Spain, dedicated to adult and pediatric outpatient visits. The sample evaluated nursing users (men and women) with an average age of 44.27 years. The results showed that there is no statistically significant relationship between the nurses’ opinions of EHR and the ages of the subjects. However, there are significant differences in results concerning how long they have been using EHRs, as the longer the nurses work with EHRs, the greater is their degree of satisfaction. Nurses considered the contribution of EHRs positive to their nursing care. This work concluded that the usability of the EHR was satisfactory, but also emphasized that, when assessing usability, one must also take into account the training and the need for technical support during the process of implementing the EHR [9].

Many applications in the Health area failed because their interfaces were difficult to use and imposed a heavy cognitive load on its users to navigate through the system. Unfortunately, the aspect of usability is poorly analyzed by software developers, which has a negative effect on the acceptance of electronic solutions by Health professionals.

III. Usability Concepts vs. Influencing Factors

Many authors over the years try to conceptualize the usability aspects and its metrics. International regulations also specify metrics and parameters to be considered when assessing the usability of a product [13]–[18].

In the context of information technology applied to Health, the challenge is to provide professionals (final users) with computer products that have been developed considering their applicability in contextualized practice, respecting the following usability requirements: easy to learn, use, re-using, among others. Therefore, attention is drawn to the parameters and usability metrics that are used from the moment of planning, to implementation and final evaluation, always involving users in the design (formative tests) and final analysis (summative tests) [19].

Of the 14 concepts of usability found in the literature, four were taken into consideration at this time as being the most relevant in this context that we intend to analyze. These concepts were based mainly on ISO 9241, Nilsen, and Preece, which define as important the concepts of effectiveness, efficiency, satisfaction and usage context. Besides these, the aspect “system screen or interface” was also considered for evaluation [10].

According to [10], for each concept of usability to be evaluated, some factors hereby referred to as "influencing factors" must be taken into consideration as they have a direct impact on results. For each concept, two or more influencing factors may be associated. In the present work, we selected 31 influencing factors for the main concepts. These were weighted and evaluated as part of the issues included in the usability evaluation questionnaire for archetyped clinical systems.

The influencing factors analyzed were: behavioral time; use of resources; attractiveness; pleasantness; flexibility / customization; synthesis; user orientation; consistency; generalization; familiarity; self-description; feedback; observability; compliance task; accuracy; system compatibility with the real world; migration tasks; user control and freedom; recoverability; error recovery; readability; navigability; simplicity; charging time; and help and documentation.

IV. Methodology

This project followed the ethical and legal principles of research involving human subjects, being submitted for review and approved under the number 33667214.3.0000.5208, according to the resolution number 466/12 of the National Health Council in Brazil.

To assure the validity of the tests in this section, we will describe the methods used to evaluate the usability aspects of archetyped interfaces generated by two different user interface building tools. In order to achieve this, the experiment observed some parameters: 1) Time to fill in the form: the total time spent by each user in the experiment to
complete a specific task; 2) Value assigned by users to the application form: considers the scores awarded by each user to the tool regarding the usability aspects evaluated; 3) Requested usability requirements: the usability requirements listed by Health-domain users.

By carrying out this experiment, one expects to answer the following research questions: [Question 1 (Q1)]: How long each user takes to complete a task on each user interface building tool? [Question 2 (Q2)]: What is the score given for each user interface building tool and for each group? [Question 3 (Q3)]: What are the usability requirements for archetyped interfaces that will be listed by Health-domain users?

A. Experiment Design

For this experiment, we tried to analyze the user interface aspects that influence Health professionals. This was done by testing user interaction with two user interfaces: a) an interface dynamically generated by an archetyped interfaces tool recommended by the openEHR Foundation and b) another tool designed by the researchers.

This first experiment did not take into account aspects of database features and other aspects of security; the choice was made considering exclusively tools that carried the functionality of building interfaces, so that users involved in the experiments could analyze the main usability concepts assessed in this study, i.e.: effectiveness, efficiency, satisfaction, usage context, and system screen or interface.

The aim was to identify the strengths and weaknesses with respect to the usability concepts of the context of use, system screen (interface), effectiveness, efficiency and satisfaction. In addition, the influencing factors mentioned in Section III were also considered.

For the sake of avoiding the influence of biases, the tools were named Tool 1 (T1) and Tool 2 (T2). Tool 1 (T1) is the standard tool of the openEHR Foundation, used by researchers worldwide for archetype specification [20]. Tool 2 (T2) is a tool designed by [21] that aims to build dynamic interfaces from specified archetypes available in the database of the Clinical Knowledge Manager (CKM) [22]. Figures 1 and 2 show the interface of T1 and T2, respectively.

Both user interface building tools were evaluated by 14 users (Health professionals with expertise in information systems). From the findings of this assessment, a list of the main positive and negative aspects of the tools was produced. From the negative aspects, proposals for improvements were listed.

The subjects who agreed to participate in the experiment were placed in a computer lab where they sat randomly on machines with the same configuration, in which tools T1 and T2 were already open. In front of their machines, they had the documents indicating to which group they would be allocated (randomization). They were divided into two groups (Group 1 - G1 and Group 2 - G2). This was done to prevent researchers from knowing to which group an individual belonged, rendering the experiment double-blind. All the computer screens were filmed from the beginning of the test with the tools on each machine. The collected data was stored by researchers to complement the purposes of qualitative analysis.

![Figure 1. Example of an interface of the “Blood Pressure” archetype in Tool 1 (T1)](image1)

![Figure 2. Example of an interface of the “Blood Pressure” archetype in Tool 2 (T2)](image2)

To eliminate the influence of previous experience with the tools and ensure that the answers are influenced only by the dependent variables detailed in the methodology of this paper, we used the technique of drawing study known as the Latin Square experiment 2x2 [21].
Latin Square is done by dividing volunteers into two groups (G1 and G2), who then evaluate T1 and T2 randomly. Participants in G1 commenced by using Tool 1 (T1), filling in the usability questionnaire for T1 and then using and evaluating Tool 2 (T2). Similarly, participants in G2 began using Tool 2 (T2) and then evaluating it. After that, they used Tool 1 (T1) and filled in the questionnaire for Tool 1.

For the experiment to represent the use of a health system faithfully, a clinical case of a fictitious patient with a maximum wealth of detail was designed and provided to the volunteers as the basis of the interface interaction exercise.

Four archetypes from the international repository CKM of the openEHR Foundation [22] were selected and used as the basic blocks for both tools. These were: patient admission; family history; problem diagnosis; and blood pressure. These four archetypes were chosen for minimally representing an admission and physical examination of the patient. They were captured and inserted as input for the interface design in both tools. It is worth noting that, during the assessment, the terminology and aspects of validation information of archetypes were not taken into account.

B. Data analysis

The usability questionnaire for the evaluation of archetyped interfaces was designed with the concepts (effectiveness, efficiency, satisfaction, use of context and system screen) and influencing factors of usability as explained in the section above. The questions in this survey were based on international validated questionnaires, such as Quality in Use Integrated Map (QUIM) and Questionnaire for User Interface Satisfaction (QUIS) [17][19].

For each of the concepts and influencing factors, a number of questions concerning the parameters analyzed were answered. Respondents were to analyze each question based on a Likert scale of 1 to 5, where 1 was considered the worst rating and 5 the best.

From these results, the arithmetic mean was calculated within each parameter. For this study, each concept was attributed an equal value. Tools with a final rating of 3.5 or more for each concept were considered satisfactory. In order to compare results numerically, the T-student test was applied, adopting a 95% confidence level.

C. Support Instruments

To perform the tests, each participant received the following support instruments; 1) Definition of archetypes and their Health application; 2) Instructions on Tool 1; 3) Instructions on Tool 2; 4) Free and clarified Consent Term; 5) map with information about the experiment; 6) Online access to usability evaluation questionnaire for each tool.

D. Controlled Experiment Users

According to Nielsen [15], 5 evaluators are sufficient to identify about 75% of the total usability issues at the interface and 10 are sufficient to identify 100%. For this experiment, 14 volunteers were invited - 7 for each experimental group.

Participants in the experiment were Health professionals who work in direct health care and Professors at the Federal University of Pernambuco (UFPE). These professionals were chosen on the basis of having at least a 2-year experience in health care and in-patient care in different specialties, as well as being active in various levels of complexity of health care. The full experiment took a week to complete.

V. RESULTS AND DISCUSSION

This section shows the results of the study, comparing with the literature review. The health professionals came to the lab and start responding a questionary about their sociodemographic information. Then, they test both systems. The sociodemographic profile and the user experience are shown ahead.

A. Users Profile

An invitation was sent to 20 Health professionals among doctors and nurses, with 14 Nursing Professionals attending the day of the experiment. All who attended were female who work in education and/or patient care. There were four Ph.D. researchers, three Ph.D. students, three experts, two Masters students and 2 undergraduates. The mean age was 34 years [25-55 years.] The training ranged from 1.5 years to 33 years with a mean of 10.5 years and the service time in the current workplace was 5.5 years [2 months - 32 years].

With regard to computer skills, all participants reported using web tools (send and receive email, do research on the Internet); 12 type and format text and prepare slides; 8 use statistical software; 12 type and format text and prepare slides; 8 use the EHR and 3 some other Health application.

Regarding the computational skills, four users were familiar with the use of the EHR, with long experience ranging from 1 year and 2 months to 9 years. The time spent operating an EHR per week ranged from 10 to 40 hours, averaging 21.25 hours. These users were characterized based on their experience with EHR and classified as inexperienced (2 users), intermediate (1 user) and only one was considered very experienced.

B. Quantitative results

As it can be seen in Table I, the analysis of Tool 1 and Tool 2 was carried out and divided by study groups (G1 and G2); then, we reached a final result based on the largest quantity of positive responses for each tool. The columns show the mean value of each group for each tool.

<table>
<thead>
<tr>
<th>TOOL 1</th>
<th>TOOL 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>G2</td>
</tr>
<tr>
<td>Fill Time</td>
<td>22,00</td>
</tr>
<tr>
<td>[C1] System Screen</td>
<td>3.02</td>
</tr>
<tr>
<td>[C2] Effectiveness</td>
<td>2.71</td>
</tr>
<tr>
<td>[C3] Efficiency</td>
<td>2.89</td>
</tr>
<tr>
<td>[C4] Satisfaction</td>
<td>2.71</td>
</tr>
<tr>
<td>[FI] Influencing Factors (31)</td>
<td>2.62</td>
</tr>
<tr>
<td>Overall average tool</td>
<td>3</td>
</tr>
</tbody>
</table>

G1: Group 1 assessment; G2: Group 2 assessment; T1: Tool 1; T2: Tool 2
Table I shows the time spent filling in the form, the final average of all questions for each concept of usability, the final average for the influencing factors and the overall mean rating for each Tool. The analyses were performed comparing the groups and tools.

1) Time to fill in form

As it can be seen in Figure 1, the mean value of filling time of each web form of group 1 was lower for Tool 2. For Group 2, the shortest time was for the Tool 1, i.e., when the user began the experiment by filling the form for Tool 1, they spent less time filling in Tool 2 and vice-versa. We may infer that the user was already familiar with the information (terminology) used in the archetype and with the reported case, making it easier to fill in the second form. Thus, for each group beginning of the second tool used had a lower filling time.

![Figure 3. Average filling time for both evaluated Tools](image)

In the statistical analysis, the filling time difference between the groups for the tool 1 and 2 was statistically significant (p = 0.005), and tool 2 had a higher average filling time.

2) System Screen

Regarding the concept 1 [C1] that evaluated the screen used, Tool 2 achieved a greater usability score for the both groups in all questions of the questionnaire. However, this result was not statistically significant (p = 0.077).

3) Effectiveness

Concerning Effectiveness [C2], Tool 2 achieved better scores than Tool 1. As compared per group, G2 preferred Tool 1, but with a minimum percentage difference with statistically significant difference (p = 0.052).

In this category, the users said that they could not complete the required tasks easily, correctly and completely. Additionally, errors that appeared were impediments to finishing the task.

4) Efficiency

Concerning Efficiency [C3], the results obtained were similar to those found in the previous concept. For Group 1, Tool 2 had a higher score than the grades assessed by the Group 2. To Group 2, Tool 1 was better regarded, with a statistically significant minimum percentage difference (p = 0.047).

The users said that while using Tool 1, they had difficulty to accomplish the tasks set; these have not been completed more quickly and completely when compared to Tool 2. In order to complete a task, users declared it was necessary to go through many steps, i.e., using a lot of graphics and mental resources.

5) Satisfaction

Satisfaction [C4] for Tool 2 achieved higher scores than Tool 1. Group 1 displayed a clear preference, whereas Group 2 attributed score average difference was only 0.01 for Tool 1 or 2. This difference was statistically significant (p = 0.055), leading to the conclusion that Tool 2 brought more satisfaction to its users.

In terms of satisfaction, Tool 1 was considered frustrating, tedious, difficult, poor in resources, and easily attached to negative feelings. Finally, Tool 1 did not promote the feeling of satisfaction, did not meet expectations, and was not attractive aesthetically.

6) Influencing factors

We evaluated 31 influencing usability factors. Group 1 considered the Tool 2 better than Tool 1. Since Group 2 found Tool 1 better, but with a little statistically significant difference (p = 0.038), the final overall average score of the tools with respect to evaluation of all concepts of usability shows that Tool 2 was preferred in comparison to Tool 1 for Group 1 and obtained the same average score (4) when compared to Group 2.

The highest average presented by Tool 2 shows that there were improvements in usability aspects presented by it, but it still had some aspects that should be improved.

A critical influencing factor was customization. Neither Tool 1 nor 2 allow for customization of the graphics and logical sequence of information in the interface.

The users said it was not possible to identify the recognizable elements of the tool interface from previous interactions with the same tool or other elements of the real world, such as a printer image on the screen identifying an impression.

Regarding the aspects of errors, the tool was not able to offer the user a clearly identified emergency exit; also, it did not allow to easily recover from unexpected situations. It was not able to help the user to recognize, diagnose and recover from errors. Finally, the software did not use simple language to present the errors and show how to circumvent them.

A positive point that was cited by the volunteers was that all the information for each Health aspect evaluated was listed in the same screen, requiring users to simply scroll down the screen to gain access to other information.

C. Final evaluation

The final results can be seen in Table II. For concept 1 [C1], which evaluated the screen, Tool 2 was preferred by both groups (G1 and G2), obtaining a mean score of 87.5, 93% and 75% approval, respectively.

Regarding Effectiveness [C2], Tool 2 was preferred by Group 1, but Tool 1 received higher marks from Group 2.
Both tools achieved the same degree of acceptance in this aspect. (80% for each specified group)

Regarding Efficiency [C3], Tool 2 was approved by 100% of the volunteers that were part of Group 1, and Tool 1 was preferred by Group 2, achieving a 75% approval rate.

On the issue of satisfaction [C4], participants in Group 1 were 100% satisfied with Tool 2 and 60% of participants in Group 2 preferred Tool 1.

### TABLE II. FINAL RESULTS OF THE USABILITY TESTS FOR TOOL 1 AND 2

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>G2</th>
</tr>
</thead>
<tbody>
<tr>
<td>[C1] System Screen</td>
<td>T2 (87.5%)</td>
<td>T2 (93.75%)</td>
</tr>
<tr>
<td>[C2] Effectiveness</td>
<td>T2 (80%)</td>
<td>T1 (80%)</td>
</tr>
<tr>
<td>[C3] Efficiency</td>
<td>T2 (100%)</td>
<td>T1 (75%)</td>
</tr>
<tr>
<td>[C4] Satisfaction</td>
<td>T2 (100%)</td>
<td>T1 (60%)</td>
</tr>
<tr>
<td>[FI] Influencing Factors (31)</td>
<td>T2 (87.1%)</td>
<td>T1 (71%)</td>
</tr>
<tr>
<td>Overall average Tool</td>
<td>T2</td>
<td>T1=T2</td>
</tr>
</tbody>
</table>

Of the 31 influencing factors of usability, Tool 2 was approved in 87.1% of the items by Group 1, and Tool 1 was preferred in 71% of the items by Group 2.

The final overall average for each tool appointed Tool 2 as being better according to group 1, whereas Group 2 evaluated both tools similarly. This difference was statistically significant (p 0.042).

Regarding overall satisfaction, as it can be seen in Figure 4, an assessment of the degree of satisfaction has been made for each evaluated aspect.

![Satisfaction of the evaluated Tools](image)

In general, Tool 1 achieved a degree of satisfaction of 63.8%, while Tool 2 achieved a higher level of satisfaction equal to 75.8%. This result was statistically significant (p 0.055), confirming that Tool 2 has brought greater satisfaction to its users.

### VI. CONCLUSION

The contribution of this study was to make usability tests with archetyped clinical tools, which are being used worldwide as a new approach for modeling content in Health information systems, albeit having never been tested on usability aspects. From these findings, the researchers will have parameters to design future systems. The usability of interfaces archetyped had not yet been explored in the literature with much detail.

We highlight the fact that end users ought to be asked to expose their preferences early in the prototyping process of future systems. This practice has been observed in some papers on software specification. User participation since the design phase is crucial in determining the success of a computer system.

More attention should be given to the customization feature. This feature was highly requested by end users because, with customization possibilities, interface usability problems such as font size, background color, and location, can be easily addressed without dependence on computer programmers. A system with this feature is required by Health professionals, and new research on development and validation of such functionality is a possible future path.

The results reported by users will be the basis for the development of a visual library prototype of functionalities for archetyped systems. All the aspects presented will serve as input for the development of future systems, including the specificities demanded by users. The requested adjustments will be specified in a future system that is being developed by the researchers.

Another aspect observed in the study that directly impacts usability was the terminology. Although it was not the focus of this validation study, the researchers were called upon many times during testing by various professionals to answer questions about terms, even after the researchers informed that they could not provide information that would influence the test. This situation led us to believe that the terminology can increase the probability of errors while providing required information. A suggestion then would be to carry out archetype validation studies.

Even though semantics is not the focus of this study, the importance of content validation of the data found in the archetypes became apparent. In addition, coherence in the workflow sequence is of great importance for Health professionals, having a significant impact on the acceptance, use and satisfaction of the future Electronic Health Record. This validation has already been initiated in another study with some users and is part of the doctoral thesis of one of the researchers.

Knowledge of usability requirements will allow computer professionals to analyze the clinical needs expressed by customers (healthcare professionals and/or leaders) and apply this knowledge in the systems they will develop and implement.

### ACKNOWLEDGMENT

This work was partially supported by the National Counsel of Technological and Scientific Development in Brazil (CNPQ)

### REFERENCES

Barriers, approaches and research priorities for semantic interoperability in support of clinical care delivery,” in SemanticHEALTH project - Semantic Interoperability Deployment and Research Roadmap, Radboud University Nijmegen Medical Centre; Department of Medical Informatics, 2008, pp. 1–33.


