# Evaluating the Global Public Inclusive Infrastructure: Cloud4all Evaluation Framework

Eleni Chalkia[1], Juan Bautista Montalva Colomer[2],
Silvia de los Rios[2], and Ivan Carmona Rojo[3]

[1] Centre of Research and Technology Hellas/Hellenic Institute of Transport (CERTH/HIT), Thessaloniki, Greece
hchalkia@certh.gr
[2] Universidad Politécnica de Madrid (UPM), Madrid, Spain
{jmontalva,srios}@lst.tfo.upm.es
[3] Technosite - Fundación ONCE, Madrid, Spain
icarmona@technosite.es

**Abstract.** Moving rapidly into digital economy expands the need for accessibility coming from the growing number of people with disabilities, in various contexts. Additionally, ubiquitous computing has amplified the need for interactive systems to be able to adapt to their context of use, enhancing their utility while preserving usability. Cloud4all project [0] aims to develop a complete new paradigm in accessibility, by replacing adaptation of individual products and services, with auto-configuration of any mainstream product or service, using cloud technologies to activate and augment any natural accessibility the product or service has, based upon a set of the user's Needs & Preferences (N&Ps). In order to assess this goal, Cloud4all has developed an evaluation framework, as part of the User Centred Design (UCD) iterative process. This paper provides an overview of the 1st pilots' evaluation framework, together with ideas and plans about the general framework of the pilot test.

**Keywords:** Accessibility, evaluation framework, auto-configuration, scenario, usability, user experience, Cloud4all.

## 1 Introduction

In order to be useful, ubiquitous systems need to be designed following user centered design, so that the users' Needs & Preferences (N&Ps) are taken into account in the entire design and development process [2]. The goal of this user centered design (UCD) is to create tools and products that satisfy the user who is willing to use them due to their increased utility, ease of use, and pleasure provided by the interaction with them. However, the evaluation of pervasive computing systems and their influences on users is quite difficult because it requires analysis of real users in a real context [3].

User centered design is enhanced by the involvement of real users in a real context, doing multiple evaluations of the products under development during the development

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part IV, LNCS 8516, pp. 358–367, 2014.

cycle. User's evaluation planning itself though, has been a very critical point in the development process, since it includes various aspects to be examined, which are yet not clearly defined by the literature. Terms like usability, accessibility and user experience are heavily involved in the evaluation of systems under development.

Usability is an overall term that covers aspects of a system as user friendliness and ease of use, and has been nominated with various definitions over time [4, 5, 6, 7, 8, 9] which are rather complementary than contradictory. ISO standards for software quality refer to broad view of usability as quality in use, as it is the user's overall experience of the quality of the product [10]. Thus, usability is related to the users, the goal and the contexts of use that are appropriate to the particular circumstances.

Nevertheless, while using new technologies, users are not necessarily seeking to achieve a task, but also to amuse and entertain themselves. Therefore the term user experience, initially popularized by Norman [11], covers the components of users' interactions with systems that go beyond usability studies. User experience, the newest term in the set of criteria against which a system should be evaluated, is a multi-dimensional concept and a commonly accepted definition is still lacking. According to Hassenzahl and Tractinsky [12], user experience attempts to go beyond the task-oriented approach of traditional Human Computer Interaction (HCI) by bringing out aspects such as beauty, fun, pleasure, and personal growth that satisfy general human needs but have little instrumental value. Therefore, when compared to basic usability, enjoyability and the hedonic quality [12, 14, 15] play an essential role.

The features of ubiquitous systems and products and the context of use affect the human's experiences and preferences about their use, utility and usability. Thus, user experience in user-product interaction, usability of the product and its utility, are terms closely linked to each other, which have to be evaluated. In Cloud4all, a European co-funded project, there will be 3 consecutive test iterations, to enhance the user centered design, as the project and its developments evolve. This paper presents the framework that has been developed for the 1st iteration of Cloud4all tests that focuses on the usability of the tools under development, and also provides some ideas, based upon the lessons learned from the 1st iteration for the next iteration phases.

## 2 Evaluation Framework Applied within Cloud4all

Cloud4all project aims to develop a complete new paradigm in the domain of HCI accessibility, by replacing adaptation of individual products and services for a person, with auto-configuration of any mainstream product or service users' encounter, using cloud technologies to activate and augment any natural (built-in) accessibility the product or service has, based on a set of the user's Needs & Preferences (N&Ps). The scope of Cloud4all is to provide to the users a seamless experience when changing platforms; having their settings transferred and transformed from one platform to the other in such a transparent to the user way, that he/she will not have to interact at all with the specific device settings.

In the core of this auto-configuration, the user has to assess this scope and evaluate the achievements of Cloud4all towards its objectives. In order to assess this goal, Cloud4all has developed an evaluation framework, as part of the User Centred Design

(UCD) iterative process that is twofold. On the one hand, the scope is to evaluate the usability and user experience of tools that are developed in Cloud4all and require human machine interaction to achieve their goal. And on the other hand, to validate the usability and the utility of the auto-configuration, that actually does not create UIs but triggers the procedure of already existing UIs to auto-configure themselves based on the user that uses them each time.

In this context, we addressed the following questions:

1. How will the auto-configuration affect the user with disabilities experience executing common tasks every day, in familiar devices and platforms?
2. Will users feel more confident using technology and devices and platforms that they are not familiar with when the auto-configuration exists?
3. Will the developers be willing to include their application to the Cloud4all concept and enrich the solutions repository?

In Cloud4all we have developed a common framework for evaluating the project developments addressing the above research questions. It is obvious from the questions above that the target groups of users of Cloud4all address both users with disabilities and the developers of Assistive technologies and mainstream solutions that want their products to be used by people with (or without) disabilities. The scope of the tests with each group is totally different, but they are both included under the umbrella of Cloud4all/GPII [17] concept. More details on the users involved and the scope of the evaluation are available in the following sections.

Due to the complexity of the study and the need for gathering feedback from the users in a very early stage of the project, as part of the UCD, preliminary tests have been planned in early phases of product development. Thus, 3 consecutive pilot iteration phases have been planned, starting from an early stage of the project. In each of the 3 iteration phases a different experimental plan will be developed, since the tools under evaluation will be evolving and the scope of the testing will also evolve. All 3 experimental plans though will be designed using the same patterns and including the same components of the evaluation framework. Thus in all 3 experimental plans the following evaluation framework will be defined:

- Scope of the evaluation (what to evaluate, goals of the study, research questions).
- Participants' details and recruitment (disability groups, participants per site).
- Research hypothesis.
- Pilot scenarios.
- Techniques and tools (what to measure and why, how to measure).
- Indicators and metrics per scenario.
- Test protocol (exact agenda, tasks).
- Results analysis tools and communication of the results with the developers.

At this point we have to highlight that Cloud4all testing goes beyond strict usability testing, since in its context both tools that are actually used by the users in order to achieve a goal, but also tools that are backend and don't have interface for the users, have to be evaluated. Thus, we have used various applications (desktop, mobile, tablet) as the means for the evaluation of tools that don't have an interface for the user and not as the target of the evaluation. This means that the actual target of the testing is not to evaluate the applications that are actually used, but use them in order to have an interface to evaluate the auto-configuration features of Cloud4all. Facilitators' role is very important in this perspective, since confusion to users could and has been quite obvious and could lead to biased results.

# 3 Specificities of Cloud4all 1st Evaluation Phase

## 3.1 Scope of Tests

The conducted experiment was planned in order to address these research questions. The first 2 questions refer to the users with disabilities, while the last question refers to developers. Starting from the first 2 questions, we can see that the core of the testing with the users with disabilities is the auto-configuration. The auto-configuration is consisted of various components which are connected to each other in such a way to create a complex architecture.

The scope of Cloud4all 1st pilots' iteration is three fold. Firstly, the goal was to introduce the concept of Cloud4all to users and get their general reaction feedback on this. At this phase, since not all the architectural components were ready to work with each other and a live demonstration of the Cloud4all vision is not possible, the users were presented with a video of a case study of Cloud4all and were asked to provide some general information about their opinions on this vision.

Additionally, the users were presented with scenarios that demonstrated the ability of the basic infrastructure to automatically launch and set up access solutions for users according to their preferences. Specific applications that have been connected to the Cloud4all architecture using specific user N&P sets (personas) were used to demonstrate this and prove to the users the extendability of Cloud4all vision.

Finally, very early testing of the auto-configuration scenario took place, using specific tools developed for the pilots. The goal of this testing was the evaluation of some components of the architecture, as well as the identification of the usability and the utility of the auto-configuration scenario from different types of users.

Regarding the part of the evaluation with the developers, the scope was to introduce them to the Cloud4all concept and the tool they can use in order to join Cloud4all/GPII and gather their initial feedback about acceptance and usability aspects.

## 3.2 Participants' Details and Recruitment

Three different types of users are included in the pilots of Cloud4all: developers, stakeholders and end users. Each user group tested different scenarios which were addressing their needs.

Cloud4all has established three pilot sites where all the evaluation activities will take place: Spain, Greece and Germany; covering, this way, geographically Europe, which will allow for cross-relation to cultural and socioeconomic issues.

In each one three iteration phases and for three pilot sites, the user groups above will actively participate. For that reason, the active recruitment of users for the activities that involve them directly (i.e. pilot tests, online user forum, demonstrators) has to be supported by other dissemination activities included in the project. Therefore, the attendance to workshops, conferences and other forums, as well as the social networks, can also be invited to participate in the pilot activities. In this way, new end-users, people with disabilities, experts and developers interested in the aim of the project can participate in the evaluation phases by entering in the pool of candidates to participate in the sites.

It is crucial to follow a proper procedure in the liaison with and involvement of users groups. Recruiting and retaining volunteers is an essential process that will guarantee a reliable outcome regarding users input in the concept validation, co-design and design stages as well as in all the evaluations and testing activities that need to be carried out during the project.

To that end, in each evaluation phase, some quotas for participation and selection criteria for these activities are established previously. In the first evaluation phase, the selection criteria for the end users with disabilities were the following:

- Type of disability;
- Gender;
- Age;
- Educational background;
- Operating system used

The selection criterion for experts (stakeholders) was the type of disability in which each one has experience, which is directly linked to the beneficiaries of each solution, and for developers, it was agreed to involve different profiles according to their expertise (mobile, desktop, web, AT).

Regarding the users with disabilities, the project initially did a classification of the profiles that will be involved in the different phases of the Project in order to try to represent as much as possible the wide range of problems, needs and preferences of the users in the different environments where Cloud4all will work. According to this classification, those are the profiles of users involved in this first iteration:

- Blind users
- Low vision users
- Users with dyslexia
- Cognitive impaired users
- Low literacy Users
- Elderly users

Specifically in the 1st iteration phase in the three sites 140 beneficiaries participated: 93 users with disabilities, 30 developers and AT providers, and 17 key stakeholders.

### 3.3 Pilot Scenarios

The pilot scenarios have been the first part of the pilot framework that was drafted in Cloud4all, and have been used as a basis on knowing what we have ready to test, and why and how we should test it. The pilot scenarios are actually stories of tasks the users have to perform in order to achieve a goal. The achievement of the goal and the execution of the tasks, as well as the users' experience while executing the tasks are the feedback that we gathered at this phase. The methodology of the pilot scenarios is the continuation of the methodology used for the identification of the use cases of Cloud4all, which is based on Rosson & Carroll's "Scenario based Design" [16]. According to Rosson & Carroll, 2001b and also in Cloud4all, the most important issue in the pilots' evaluation is the usability parameters. The feedback received from the 1st evaluation phase will help to re-define the "problem" and the "problem scenarios" and, successively, the use cases application interaction and pilot scenarios. This feedback will help us to guide the 2nd phase development and evaluation and so on [16].

Two are the main scenarios tested in this iteration phase. The first one is the auto-configuration scenario, which is depicted at the figure below. Before the auto-configuration scenario started, the specific preferences of the user where gathered from the pilots facilitators, for a specific platform (Platform A), and a N&P set of the users was created, together with a user token that includes all the information. The user was asked to perform a specific task in platform A which was configured with the settings that the user defined. Then the user was asked to go to another platform (Platform B), log in with his/her token and perform the exact same task, while the UI has been auto-configured. Afterwards, the user was asked to define his/her settings from Platform B, from scratch, log out and log in again with the new N&Ps set to Platform A.

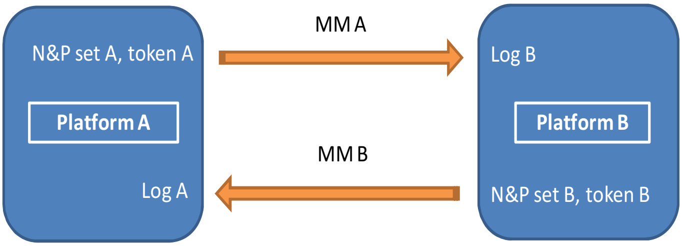

**Fig. 1.** Auto-configuration scenario

Moving forward to the tests with developers, in this pilot phase the scenario for the developers was focusing on getting feedback for the usability of the tool that the developers will use to add their application in Cloud4all/GPII and the developers where explicitly asked to perform the following tasks:

1. Add a new solution or common term: Adding a new solution/setting to the solutions ontology (without previous experience in cloud4all) /proposing a new term in the common terms registry.
2. Edit an existing solution: Changing the preferences in a solution that already exists on the ontological framework and the registry.
3. Search the ontology: Searching (by a free text and /or by categorization lists) for solutions, settings, etc., stored in the solutions ontology.

### 3.4 Experimental Plan

The techniques used in the Cloud4all pilots are divided in two categories based on the number of users and the interactions involved (i.e. between, among). Individual techniques are usually associated with in depth data gathering and group techniques have to do with breadth and diversity in data gathering.

Thus, the following individual participation techniques were used:

- Structured and semi-structured interviews.
- Contextual interviews.
- Usability and accessibility testing.
- Observations, including Think-Aloud Protocol Analysis.
- Paper mock up.
- Click demo.
- Wizard of Oz.

In addition, focus groups with experts and creativity sessions took place as group participation techniques.

Objective and subjective evaluation tools were used to accomplish and allow carrying out the evaluation techniques designed to assess each solution to be tested in Cloud4all pilots. To this end, subjective tools, such us, questionnaires, forms or service diaries, together with objective tools like timed tasks, video/sound/screen recording, whenever possible, were applied for the validation of Cloud4all applications with end-users.

The relations between the tools evaluated and the research hypothesis, tasks to be performed per tool, measures to be taken, plus the indicators and metrics used per tasks and sub-tasks are a critical point of research. The indicators and metrics used during this first iteration are standard metrics and procedures in usability. The general subjective measures are: perceived ease of use, usefulness, satisfaction, learnability and trust. Each pilot scenario has been related to target audience, research hypothesis, scope, usability metrics and thresholds and has been evaluated so.

The following table shows an example of the indicators and thresholds for a task of the scenario for the developers; “The user opens the semantic alignment tool to search the ontology”.

**Table 1.** Indicators and thresholds for task “The user opens the semantic alignment tool to search the ontology”

| Indicators | Metrics | Success thresholds |
|---|---|---|
| Timing measure | Minutes & Seconds | < 1’ 30’’ sec |
| Task completion | Binary (Yes, No) | Yes |
| Errors | Number | [< 2] |
| Help requests | Number | [< 2] |

Since the environment has a great role in the pilots’ realization and in Cloud4all, as mention above, we have pilot sites in 3 different countries in Europe (Greece, Spain and Germany), we tried to be as detailed as possible in order to keep a standard baseline among the different sites. Thus, a common agenda has been created with detailed steps of the procedure to be followed by the pilots' facilitators. Even if some small details have changed from one site to the other, this would not create important bias in our study. Nevertheless, the variety of the pilot sites and the complexity of the study in relation to the deferent environments and cultural issues may bias the trials, but this was an issue that was not taken into account in this first evaluation phase. Results of this phase allowed us to gain some conclusions for the next phases.

### 3.5 Results Analysis Tools and Communication of the Results with the Developers

The purpose of using data collection instruments (e.g. questionnaires, facilitator's diary) is to expedite the collection of necessary data for meeting the objective of each iteration phase. Good collection tools are simple, concise, and reliable. Similarly, good analysis tools assist analysis and reporting in all evaluation phases. According to Rubin & Chisnell [17], data analysis falls into two different types. The immediate (preliminary) reporting of problems aims to quickly ascertain any arising issues. Main patterns and trends identified within each pilot site were reported in users.gpii.net. Recommendations based on identified problems and bugs should be swiftly reported, in order for members of the development teams to be able to implement any changes on time. These issues should include only obvious problems (i.e. evaluators should be cautious and sometimes report less than more). In addition, any consequent changes at the tools to be tested should be in agreement with pilot site leaders to avoid using- and therefore testing -different versions in different sites.

The second level of analysis is the overall evaluation for each iteration cycle. It is a thorough analysis which aims to provide a comprehensive and in depth analysis leading to a concise report of main results and findings, and their diffusion mechanisms to interested parties within the project in the form of specific recommendations for the next stage of development.

As the level of complexity within the Cloud4all evaluation framework is rather high and potentially increases in the next iterations, two perspectives for data analysis were adopted and presented: data source-wise (driven by metric type) and tool-wise (driven by tool tested) with consideration and account for the inevitable overlap between the two perspectives.

Statistical analysis entails three separate steps and is associated with two reporting stages:

- Data compilation;
- Creating summaries;
- Data analysis;
- Creating a detailed account of recommendations for the second iteration phase based on a common template;
- Produce the first iteration evaluation report.

Chosen analysis path depends on data types. Subjective data from questionnaires were quantified based on a unified coding scheme and will be treated as ordinal. Objective data are easily quantifiable and mostly treated as continuous.

# 4 Conclusions

This paper describes the pilot evaluation framework for the 1st evaluation phase of Cloud4all project. It depicts the main highlights of the testing, providing to the reader an overview of the research questions to be addressed in Cloud4all in general and continues with specific details of the plans for the 1st iteration tests. The main scenarios performed by the users, the participants details, the experimental plan and the results analysis tools follow to give to the reader a holistic idea of the Cloud4all pilots testing.

This 1st iteration phase of pilots of Cloud4all was an important lesson for the facilitators and the developers of the project. The results of this evaluation where gathered and distributed to the developers, following a feedback loop that was very helpful for the updated of the various tools. The developers received a document with issues from the different pilot site, that was specifying the severity of the issue, the tool and the specific task this issue was met and the recommendation from the pilots.

During this phase, some issues were raised and important lessons were learned for the preparation and the realization of the pilots. These lessons will be taken into account when planning the next iteration phases. The evaluation of the auto-configuration provided a lot of valuable feedback, for the next iteration of the design process. The concept was perceived to be very promising and interesting by the participants. The evaluation also revealed that we still have a long way to go with the developers' tools and the way of involving them in the evaluation procedure. Also, there are a lot still to be done with regard to the development of methodologies for evaluation and the design of the measurement instruments.

**Acknowledgements.** Results presented in this paper have been researched within the Cloud4all project. Cloud4all is an R&D project that receives funding from the European Commission under the Seventh Framework Program (FP7/2007-2013) under grant agreement n° 289016.

# References

1. Cloud4all project, CN. 289016, 7th Framework Programme, ICT-2011.5.5 ICT for smart and personalised inclusion (November 2011), `http://cloud4all.info/`
2. Consolvo, S., Arnstein, L., Franza, B.R.: User Study Techniques in the Design and Evaluation of a Ubicomp Environment. In: Borriello, G., Holmquist, L.E. (eds.) UbiComp 2002. LNCS, vol. 2498, pp. 73–90. Springer, Heidelberg (2002)
3. Belotti, F., Berta, R., DeGloria, A., Margarone, M.: User Testing a Hypermedia Tour Guide. IEEE Pervasive Computing, 33–41 (2002)

4. International Standards Organization. ISO 9241-11: Ergonomic equirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability. Geneva: International Standards Organization (1998)
5. Gould, J.D., Lewis, C.: Designing for usability: key principles and what designers think. Communications of the ACM 28(3), 300–311 (1985)
6. Shackel, B.: Human factors and usability. In: Preece, J., Keller, L. (eds.) Human-Computer Interaction: Selected Readings. Prentice Hall, Hemel Hempstead (1990)
7. Shackel, B.: Usability – context, framework, definition, design and evaluation. In: Shackel, B., Richardson, S. (eds.) Human Factors for Informatics Usability, pp. 21–37. Cambridge University Press, Cambridge (1991)
8. Sharp, H., Rogers, Y., Preece, J.: Interaction design: beyond human-computer interaction. John Wiley, London (2007)
9. Stone, D., Jarrett, C., Woodroffe, M., Minocha, S.: User interface design and evaluation. Morgan Kaufmann, San Francisco (2005)
10. Bevan, N.: Quality in use for all. In: Stephanidis, C. (ed.) User Interfaces for All: Methods, Concepts and Tools, pp. 353–368. Lawrence Erlbaum, Mahwah (2001)
11. Norman, D.A.: The invisible computer. MIT Press, Cambridge (1998)
12. Hassenzahl, M., Tractinksy, N.: User experience: a research agenda. Behaviour and Information Technology 25(2), 91–97 (2006)
13. Hassenzahl, M.: Hedonic, emotional and experiental perspective on product quality. In: Ghaoui, C. (ed.) Encyclopedia of Human Computer Interaction, pp. 266–272. Idea Group (2006)
14. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität (AttrakDiff: A questionnaire for the measurement of perceived hedonic and pragmatic quality). In: Ziegler, J., Szwillus, G. (eds.) Mensch and Computer 2003: Interaktion in Bewegung, pp. 187–196. B.G. Teubner, Stuttgart (2003)
15. Hassenzahl, M., Law, E.L.-C., Hvannberg, E.T.: User experience: towards a unified view. In: Law, E.L.-C., Hvannberg, E.T., Hassenzahl, M. (eds.) Proceedings of the 2nd COST294-MAUSE International Open Workshop (2006)
16. Rosson, M.B., Carroll, J.M.: Usability Engineering: Scenario-Based Development of Human-Computer Interaction. Morgan Kaufmann, San Francisco (2001b)
17. Rubin, J., Chisnell, D.: Handbook of Usability Testing: Howto Plan, Design, and Conduct Effective Tests. Wiley (2008)
18. GPII, Global Public Infrastructure, http://gpii.net/