



DELIVERABLE REPORT

D9.1.1

“Evaluation Plan”

Collaborative project

MASELTOV

Mobile Assistance for Social Inclusion and Empowerment of Immigrants with Persuasive Learning Technologies and Social Network Services

Grant Agreement No. 288587 / ICT for Inclusion

project co-funded by the
European Commission

Information Society and Media Directorate-General
Information and Communication Technologies
Seventh Framework Programme (2007-2013)















Due date of deliverable:	October 31, 2012 (month 10)
Actual submission date:	September 15, 2013 (updated after YR1 review)
Start date of project:	January 1, 2012
Duration:	36 months

Work package	WP9 – Field trials and Evaluation
Task	T9.1 – Preparation of Field Trials and Evaluations
Lead contractor for this deliverable	CURE
Editor	Jan Bobeth (CURE)
Authors	Jan Bobeth, Susanne Schmehl (CURE)
Quality reviewer	Lucas Paletta (JR)

Project co-funded by the European Commission within the Seventh Framework Programme (2007–2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

COPYRIGHT & DISCLAIMER

This report is in the public domain and may be used and reprinted without permission, provided that the source is acknowledged. The European Union is not responsible for any use that may be made of the information contained herein. The sole responsibility of the views, opinions, and findings expressed in this publication lies with the author.

partner	Organisation	ctry
01	 JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH	AT
02	 CURE – CENTER FOR USABILITY RESEARCH AND ENGINEERING	AT
03	 RESEARCH AND EDUCATION LABORATORY IN INFORMATION TECHNOLOGIES	EL
04	 FUNDACIO PER A LA UNIVERSITAT OBERTA DE CATALUNYA	ES
05	 THE OPEN UNIVERSITY	UK
06	 COVENTRY UNIVERSITY	UK
07	 CESKE VYSOKE UCENI TECHNICKE V PRAZE	CZ
08	 FH JOANNEUM GESELLSCHAFT M.B.H.	AT
09	 TELECOM ITALIA S.p.A	IT
10	 FLUIDTIME DATA SERVICES GMBH	AT
11	 BUSUU ONLINE S.L	ES
12	 FUNDACION DESARROLLO SOSTENIDO	ES
13	 VEREIN DANAIDA	AT
14	 THE MIGRANTS' RESOURCE CENTRE	UK

CONTENT

1. Introduction	4
2. User-Centred Design	5
2.1 The User-Centred Design process in MASELTOV	5
2.2 Justification of the user-centred design process.....	6
3. Planning of the Iterative Evaluation of User Interfaces (T9.2)	9
3.1 Expert Review	9
3.2 First and second user tests	9
3.3 Design and Evaluation Roadmap	10
4. Summary and Outlook.....	12
5. References	13

1. INTRODUCTION

The evaluation plan presented in this deliverable describes how the MASELTOV services will be evaluated at the various stages of the project. Early and continuous feedback is very important to avoid the development of undesired functionalities and not useable services. Especially when designing for a vulnerable target group like immigrants that just arrived in a new country and have a different cultural background (see MASELTOV target group definition in D2.3.1). For this reason, immigrants will be involved at any stages of the design process of the MASELTOV services. To better understand this user-centred approach it is described in chapter 2. presenting an overview and the status of the accordant tasks in Figure 1.

This deliverable will be updated twice at later stages of the project. In this version the iterative evaluation of user interfaces is explained (chapter 3.). The second version will replace the accordant chapter 3. with the detailed setup of the first field trials, while the third version will describe the planning for the assessment of the final integrated prototype of the MASELTOV project. Furthermore, both upcoming versions of this document will contain an updated version of Figure 1 showing the status within the user-centred design process in the project.

2. USER-CENTRED DESIGN

In this chapter we describe the general approach of involving users in the design process and the service development in the MASELTOV project.

2.1 THE USER-CENTRED DESIGN PROCESS IN MASELTOV

Figure 1 presents an overview and the current status of the tasks forming the user-centred design process in MASELTOV. The user involvement was started with the requirements analysis in which we have collected specific service requirements of immigrants in order to profoundly understand their specific service needs (see D2.3.1). The next step is to design the services and discuss them with immigrants. To benefit from their ideas we conducted two participatory design sessions (see D2.4). The goal of this serious user involvement was to identify potential differences in design and solution approaches and to discuss them with end users.

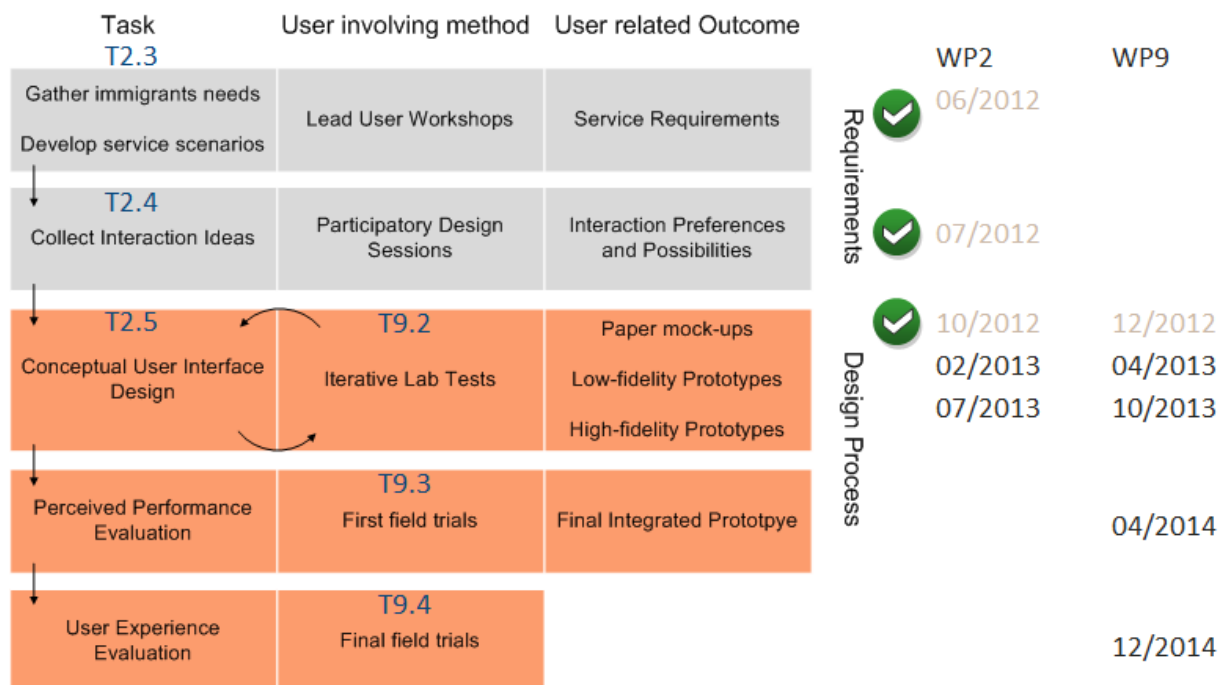


Figure 1: Overview and status of tasks within the user-centred design process of MASELTOV. While the planning of T9.2 is described in this version of the document (D9.1.1), T9.3 and T9.4 will be planned in the two upcoming versions (D9.1.2, D9.1.3).

The MASELTOV user interface is designed in an iterative manner (see T2.5) based on the outcomes of the participatory design sessions and the identified user needs. This means various elementary user interface concepts will be elaborated and the resulting mock-ups will be presented first to usability experts and then to the users in form of usability tests (see T9.2). So, the design of the user interface is directly linked with the iterative usability testing that takes place in Task 9.2, in which the interface concepts will be evaluated frequently from an early stage onwards to ensure user validation. For this reason, the concepts of the first iteration are visualised in a simple wireframe format. They can be created rather fast and easy as necessary changes and updates are expectable after the first evaluation. As shown in Figure 1 the feedback flows directly into the next iteration step. There the concepts will be refined to low-fidelity and within the third iteration to high-fidelity click-dummies. Graphics and a more sophisticated visual representation will be added at this later phase of the design process when

the interaction flow has finally been specified. The high-fidelity prototypes shall represent the majority of the services and functions that MASELTOV offers so that the immigrants get a real impression of how MASELTOV will work.

At the end of the creative design phase the user interfaces will be implemented and connected with the accordant services (WP6, WP7 and WP8). The several service components remain separately and thus can be evaluated individually. They will be tested in the first field trials taking place in the three cities (London, Madrid/Barcelona, Graz) of the NGOs with the accordant immigrants group (see T9.3). In every city six immigrants will be recruited to use the various MASELTOV services for one week under real conditions in order to detect bugs and report usability problems.

After the final iteration and revision the several service components will be integrated into the final prototype of the MASELTOV system. This final prototype shall cover all specified use cases and scenarios and thus provide service support in various situations. We conduct the final field trials in order to evaluate the developed MASELTOV services under real world conditions (see Task 9.4). Therefore, the methods “Remote Usability Testing” (e.g. Andreasen et al., 2007) and “Experiences Sampling Methods” (e.g. Larson et al., 1983) will be applied in order to gather real time information about the usage of MASELTOV that can be analysed to determine the user experience.

2.2 JUSTIFICATION OF THE USER-CENTRED DESIGN PROCESS

Firstly brought up by Donald Norman in the 1980s, user-centred design is an approach to enhance the output quality in software development. The goal of user-centred design is to make a product easy to use for its intended users (Vredenburg et al. 2001). The process has emerged from traditional software development approaches to align the strategic focus to the user’s needs and not to get lost in solving technical challenges first (Lowdermilk 2013). The main characteristic is to involve users (i) from the start of a system development cycle and (ii) in all stages of a software project in an iterative manner: moving from a technology-driven approach to a user-driven approach (Vredenburg et al. 2001). User-centred design is not only about deepening the understanding of the users but further imposes the conceptual understanding of (i) the tasks users are supposed to perform with the system and (ii) the environment in which the system is supposed to be used (Stone et al. 2005). This process is based on four main principles about how interactive systems should be developed as certified in the ISO standard 9241-210 (2010):

- The active involvement of users
- An appropriate allocation of function between user and system
- The iteration of design solutions
- The engagement of a multidisciplinary design team

The procedure includes to ask or observe users for their needs, to present design ideas frequently to users for their feedback and to update the design iteratively, to evaluate the functional prototypes under real conditions (Petrie and Bevan 2009). These authors provide a practical overview on the important concepts of human-computer interaction and user-centred design, mainly usability, accessibility and user experience (Petrie and Bevan 2009). Based on these concepts and the aforementioned principles the user-centred design process is set up in MASELTOV as described in the previous section 2.1.

2.2.1 NARROWING DOWN THE TARGET GROUP

An important precondition for applying a user-centred design process is to know who the users of a system will be in order to involve participants as close as possible to the intended target group (Grudin and Pruitt 2002). However, the identification of users can be particularly demanding in case of dealing with large numbers of heterogeneous users (Kujala and Kauppinen 2004). In the special case of MASELTOV where we work with immigrants representing such a large and very heterogeneous group in Europe, we needed to specify the target group carefully. Moreover immigrants may form attitudes of mistrust towards the researchers, which hinders their recruitment as study participants and might bias study results, due to inaccurate answering behaviour (Hynes 2003). For these reasons, the target group definition needed to be narrowed down from the whole population of immigrants to a subset of people with similar attributes. The MASELTOV target group has been defined and updated according to recommendations of the reviews (Szwachertowska et al. 2013) in the two versions of the MASELTOV deliverable D2.3 Use Cases and Service Scenarios.

2.2.2 SAMPLE SIZES

When it comes to concrete user involvement in ICT projects the questions raises how many users are needed for the current task which has been discussed extensively in the literature (e.g. Nielsen 1993, Spool and Schroeder 2001, Hwang and Salvendy 2010). In user-centred design, sampling can be based on groups that are identified by the main user characteristics (Kujala and Kauppinen 2004).

For the elicitation of cultural-specific user requirements Aykin et al. (2006) recommended the application of qualitative methods. Vulnerable target groups like newly arrived immigrants may be reluctant to take part in formal research studies (Atkinson & Flint 2001). Building up trust to those groups can be difficult and might require more than the usual guaranteeing of anonymity, confidentiality, and the application of ethical principles. One possibility might be to establish credentials by working voluntarily with a refugee community organization (Hynes 2003). Potential participants of research studies (e.g. for interviews) should be supported in their freedom of decision whether to participate or not and whether to stop collaboration or not (Lammers 2005). In qualitative research, it is practiced to pick up representative users from each group based on the main user characteristics which allows that all necessary users to be represented (Kujala and Kauppinen 2004). Researcher and developer should focus on a smaller amount of participants that can be studied more intensively (e.g. observing and interviewing the concrete target group) than in quantitative research to better understand behavioural patterns. Beyer and Holtzblatt (1998) recommend that between six and twenty users should be involved, depending on the scope of the study. Thus we conducted interviews and focus groups within the requirements analysis of MASELTOV with selected representatives of target groups.

To assess the design ideas the most important aspect is to work in iterations and to present updated design frequently to users (Stone et al. 2005). Therefore even a small amount of users per iteration is appropriate. A precise number of necessary users cannot be defined per default as this decision always depends on the concrete case (Kujala and Kauppinen 2004). Nielsen (1993) states that the majority of usability issues of a system (around 80 %) are identified by the first five users. However, in other studies evaluating more complex applications up to 15 participants were needed to identify 80% of the usability issues (Spool and Schroeder 2001). Later on, Hwang and Salvendy (2010) proposed 10 ± 2 participants for usability studies. Schmettow (2012) argues that usability studies differ so much in terms of complexity and context that a “magic number” of involved users can never be defined in a reliable way. In

MASELTOV we stick to usual practice and conduct three formative evaluation iterations with 5-10 users per iteration during the design phase in the lab (T2.5 and T9.2) as experts also suggested recently (Petrie and Bevan 2009, Hwang and Salvendy (2010), Lowdermilk 2013).

For the summative testing of the working system to take place in the field, Petrie and Bevan (2009) suggest to involve between 8 and 30 users. In MASELTOV, we plan to involve 36-72 users (see DOW Part B Table 11). A detailed evaluation and assessment plan for the field studies' conduction will be reported in the subsequent versions of this document (D9.1.2, D9.1.3).

2.2.3 EXTERNAL ETHICAL APPROVAL

Applying the user-centered design process in MASELTOV reveals a clear difference between this project and other ICT projects involving vulnerable target groups: in MASELTOV an external ethical approval is conducted for all user involving tasks. An independent expert from the ICMPD (International Centre for Migration Policy Development) reviews all guidelines for user studies before these studies take place. More details can be found in the Ethical Manual (D1.4).

3. PLANNING OF THE ITERATIVE EVALUATION OF USER INTERFACES (T9.2)

The iterative evaluation of the user interfaces consists of three individual assessments that go along with the design process: (i) an expert review, (ii) first user tests and (iii) second user tests.

3.1 EXPERT REVIEW

Within the first iteration the first design concepts for the MASELTOV services are evaluated by usability experts. The expert review (i.e. heuristic evaluation) is an evaluation against usability principles and heuristics introduced by Nielsen and Molich (1990). In a user-centred design process expert reviews are usually conducted to get first feedback for new user interfaces before presenting them to end users. Expert reviews don't replace usability tests with end users, they take place before. In this early stage of the design process expert reviews have various important advantages in comparison with usability tests with end users. First of all, it is enough to provide simple mock-ups to visualise the interaction concept to the experts. End users might be irritated when looking at these obviously unfinished designs. Another advantage is that there are no constraints through a small amount of tasks that are usually created for end users in form of testing scenarios. Thus, in this early stage expert reviews are more thorough than user testing as they can inspect all interactions.

The heuristic evaluation within MASELTOV will be based on the heuristics of Bettini et al. (2008). These heuristics have been elaborated by reviewing and adapting existing heuristics to the mobile context. One of their resources was the refined set of heuristics by Nielsen (1994) which apply to all user interfaces albeit in a very general way. The final set of mobile usability heuristics we will use is (Bettini et al. 2008):

- Visibility of system status and losability/findability of the mobile device
- Match between system and the real world
- Consistency and mapping
- Good ergonomics and minimalistic design
- Ease of input, screen readability, and glanceability
- Flexibility, efficiency of use, and personalisation
- Aesthetic, privacy, and social conventions
- Realistic error management

It has also been recommended to involve multiple evaluators as they find more problems and can also identify overlapping problems. For this reason, two usability experts from CURE will conduct the expert reviews separately from each other. After their reviews the detected usability problems will be consolidated and summarized in a report (D9.2.1).

3.2 FIRST AND SECOND USER TESTS

For the 2nd stage testing (first user tests) of the UIs of the MASELTOV services, the updated concepts of the first iteration (results from heuristic evaluation) are visualised in a simple wireframe format. For the 3rd stage testing high fidelity prototypes that simulate the final look and feel of the services as they already include more detailed graphical design, will be provided. The aim of the studies is to identify and describe existing usability and user experience issues and to further provide recommendations for improvement to the according partners.

The method used for the evaluations is a task-based empirical investigation. A pre- and a post-interview will be conducted to cover users' expectations (before the test) and to understand their experience (after the test). The services will be presented to the participants on a smart phone in form of click-dummies. This is an established and helpful tool/process to collect early user feedback efficiently. The functionality is faked but the user interface is available and reacts to the input of users. The detected usability issues of the evaluated services will be prioritized in a severeness ranking to assure that the major problems are processed first. The usability will be evaluated with the help of the System Usability Scale (SUS), which is a simple ten item Likert-scale giving a global view of subjective assessments of usability (Sauro 2011). It has the advantage that results of the different services can be compared in an objective way. As this questionnaire is used widely in research and industry, international benchmarks can be used to estimate the overall usability quality of the services in relation to international usability studies.

A detailed guideline for the testing will be provided by CURE. This test protocol will be reviewed for ethical issues by ICMPD before the user tests take place. The MASELTOV NGOs will conduct the testing in their according country with the according test users in the according language (favourable in the mother tongue of the participants). Per country 4-5 participants will evaluate the MASELTOV services in 04/2013 and 10/2013. Herewith, each service will be evaluated by at least 12 immigrants which is an amount above average (see section 2.2.2 Sample Sizes). The recruitment will be conducted by the NGOs as well. As mentioned above the results will be reported (D9.2.2/D9.2.3) and presented to the partners to update the MASELTOV services and UIs.

3.3 DESIGN AND EVALUATION ROADMAP

There are a number of services that will be designed and developed within the MASELTOV project. As some of them are based on existing services and thus benefit from previous work, the design process differs partly for some services. Besides there are other services that run in the background and will not have a user interface. For this reason, the roadmap depicted in Table 1 gives an overview about which service will be designed by which partner and when the according lab test will take place.

Services with user interfaces	T9.2: Iterative lab tests: Expert Review	T9.2: Iterative lab tests: First Usability Testing		T9.2: Iterative lab tests: Second Usability Testing	
Text Lens	December 2012	April 2013	12 users	September 2013	12 users
Navigation	December 2012	April 2013	12 users	September 2013	12 users
Information Service	No → No mock-ups as this is based on existing service	April 2013	12 users	September 2013	12 users
Serious game	No → the game has independent interactions, still in concept creation state	April 2013 (already something playable)	12 users	September 2013	12 users
Language lessons	No → No mock-ups as this is based on existing service	April 2013	12 users	September 2013	12 users
GeoRadar	December 2012	April 2013	12 users	September 2013	12 users
Social Network Service	December 2012	April 2013	12 users	September 2013	12 users

Table 1: Designing partners and evaluation roadmap of the services which will have a user interface.

4. SUMMARY AND OUTLOOK

The evaluation plan of MASELTOV and the status of the user-centred design process we follow within this project are documented in this deliverable. This is the first of three versions and will be updated twice before the first field trials to be conducted in April 2014 and before the final field trials to be conducted in December 2014.

In this version of the deliverable we described how the evaluation that accompanies the design process will be organised. Both according tasks (T2.5 and T9.2) include three iterations in which various stages of designs starting from elementary mock-ups (first iteration), creating low-fidelity prototypes (second iteration) and ending with high-fidelity prototypes (third iteration). Those tasks are tied strongly together as the accordant work packages WP2 and WP9 form the user-centred design process. The results of the evaluations will be documented in the upcoming deliverables D9.2.1, D9.2.2 and D9.2.3. Based on the detected problems and elicited recommendations the user interface concepts will be updated within the following iterations.

5. REFERENCES

- Aykin, N., Honold Quaet-Faslem, P., Milewski, A. E. (2006): Cultural Ergonomics. In: Salvendy, G. (ed.) Handbook of human factors and ergonomics. Hoboken: John Wiley, pp. 1418-1458
- Andreasen, M. S., Nielsen, H. V., Schröder, S. O., & Stage, J. (2007). What happened to remote usability testing? An empirical study of three methods. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '07. ACM, New York, NY, 1405-1414.
- Atkinson, R., & Flint, J. (2001): Accessing hidden and hard-to-reach populations: snowball research strategies. Technical Report. Issue 33., Department of Sociology, University of Surrey, Guildford, England.
- Bertini, E., Catarci, T., Dix, A., Gabrielli, S., Kimani, S., Santucci, G. (2008): Appropriating Heuristic Evaluation Methods for Mobile. In: Lumsden, J. (Edt.): Handbook of research on User Interface Design and Evaluation for Mobile Technology. Information Science Reference, Hershey PA, 780-801.
- Beyer, H. and Holtzblatt, K. (1998): Contextual Design: Defining Customer-Centered Systems. San Francisco: Morgan Kaufmann Publishers.
- Grudin, J. and Pruitt, J. (2002): Personas, participatory design and product development: An infrastructure for engagement. Proc. PDC 2002, 144-161.
- Hwang, W. and Salvendy, G. (2010): Number of people required for usability evaluation: the 10±2 rule. Communications of the ACM 53(5), 130–133.
- Hynes, P. (2003): The issue of “trust” or “mistrust” in research with refugees. New Issues in Refugee Research, Working Paper 98, Geneva: UNHCR Evaluation and Policy Analysis.
- ISO (2010): Ergonomics of human- system interaction. Part 210: Human-centred design process for interactive systems. Geneva, International Standards Organization
- Kujala, S., & Kauppinen, M. (2004). Identifying and selecting users for user-centered design. In Proceedings of the third Nordic conference on Human-computer interaction - NordiCHI '04 (pp. 297–303).
- Lammers, E. (2005): Refugees, asylum seekers and anthropologists: the taboo on giving. Global Migration Perspectives 29.
- Larson, R., & Csikszentmihalyi, M. (1983). "The experience sampling method". New Directions for Methodology of Social and Behavioral Science, 15, 41-56.
- Lowdermilk, T. (2013): User-Centered Design. O'Reilly.
- Nielsen, J., Molich, R. (1990): Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people (CHI '90), ACM, 249-256.
- Nielsen, Jakob and Landauer, Thomas K. A. (1993): Mathematical Model of the Finding of Usability Problems. INTERCHI'93, 206-213.
- Nielsen, J. (1994): Heuristic evaluation. In: Nielsen, J., and Mack, R. L. (Eds.): Usability Inspection Methods, John Wiley & Sons, New York, NY, 25-62.
- Norman, D. (1988): The Psychology of Everyday Things. Basic Books.

Petrie, H., & Bevan, N. (2009). The evaluation of accessibility, usability and user experience. In Constantine Stephanidis (Ed.), *The Universal Access Handbook*. CRC Press.

Sauro, J. (2011): *A Practical Guide to the System Usability Scale: Background, Benchmarks and Best Practices*. Measuring Usability LLC, Denver, USA.

Schmettow, M. (2012). Sample size in usability studies. *Communications of the ACM*, 55(4), 64–70.

Spool, J., & Schroeder, W. (2001). Testing web sites: Five users is nowhere near enough. In CHI'01 extended abstracts (pp. 285–286).

Stone, D., Jarret, C., Woodroffe, M., Minocha, S. (2005): *User Interface Design and Evaluation*. Morgan Kaufmann Publishers

Szwochertowska, M., González Enríquez, C., Johnson, J., Sabour, M. (2013): *MASELOTV YR1, Technical Review Report*.

Vredenburg, K., Isensee, S., Righi, C. (2001): *User-Centered Design: An Integrated Approach*. Prentice Hall