Mobile Living Labs 09:
Methods and Tools for Evaluation in the Wild

http://mll09.novay.nl

Table of Contents

Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild

http://mll09.novay.nl

Henri ter Hofte
Novay
P.O. Box 589,
7500 AN Enschede
The Netherlands
Henri.terHofte@novay.nl

Kasper Løvborg Jensen
Dept. of Electronic Systems
Aalborg University
Niels Jernes Vej 12,
9220 Aalborg
Denmark
klj@es.aau.dk

Petteri Nurmi
Helsinki Institute for Information Technology HIIT
University of Helsinki
P.O. Box 68
FI-00014, Helsinki, Finland
petteri.nurmi@cs.helsinki.fi

Jon Froehlich
DUB Institute
Computer Science and Engineering
University of Washington
Seattle, WA 98195 USA
jfroehli@cs.washington.edu

ABSTRACT

In a Mobile Living Lab, mobile devices are used to evaluate concepts and prototypes in real-life settings. In other words, the lab is brought to the people. This workshop provides a forum for researchers and practitioners to share experiences and issues with methods and tools for Mobile Living Labs. In particular, we seek to bring together people who have applied methods for Mobile Living Labs and people who build tools for those methods.

The aim of the workshop is twofold. First, to make an up-to-date overview of current methods and tools for conducting user studies in Mobile Living Labs – highlighting their individual strengths and weaknesses. Second, to uncover challenges that are not adequately addressed by current methods and tools and to come up with ideas and requirements that could fill this gap thus serving as beacons for further research and development in this area.

Categories and Subject Descriptors

H.5.2 [**Information Systems and Presentation**]: User Interfaces — Evaluation/methodology, Graphical user interfaces (GUI), Prototyping, Theory and methods.

General Terms

Measurement, Performance, Experimentation, Human Factors.

Keywords

Living Labs, mobile, user experience, field study, in-situ evaluation, methods, tools.

1. TOPIC

Mobile devices have become truly ubiquitous computing platforms, offering a wide range of functionality from traditional voice/text communication to GPS navigation and mobile friend finding. Much of their value stems from their ability to be taken anywhere offering quick, convenient access to applications and services in a variety of contexts. These "contexts of use" are even more important now as context-aware applications (such as location-aware services) are becoming a reality in many contemporary smartphones. This context of use, however, makes it difficult for researchers to evaluate mobile device applications and interactions using traditional human-computer interaction methods. Whereas some aspects of the user experience can be evaluated with lab experiments, interviews, focus groups and/or surveys, many other aspects are harder to investigate if taken out of the natural context of use (e.g., out in the wild).

Instead of focusing solely on bringing people to the lab, researchers who want to evaluate mobile devices and services are increasingly doing the opposite: *bringing the lab to the people* [1]. This is a key ingredient of the "Living Labs" approach [7], which employs a range of methods, including: self-report methods (e.g., experience sampling, diaries, and day reconstruction), measurement (e.g., application usage logging, context logging), as well as observation (e.g., ethnography).

For *Mobile* Living Labs, smartphones are often used, not only as a platform to deliver new mobile services, but also as a platform for data collection tools. Contemporary smartphones can be utilized to collect all kinds of data (audio, video, user input), according to a wide range of methods. These methods range from automatic data collection without user intervention (logging) to asking users small questions at random moments (experience sampling). Some tools used in Mobile Living Labs recently include ContextPhone [6], MyExperience [1], Xensor [3], RECON [4] and BeTelGeuse [5].

Of course, data gathered by such tools is different from observations by researchers (e.g., well-trained ethnographers), but using smartphones enables studying more persons, longer and at times and locations where observations by researchers would have been difficult.

This workshop provides a forum for researchers and practitioners interested in sharing experiences and issues with methods and tools for Mobile Living Labs. In particular, both people who have applied methods in Mobile Living Labs as well as those who built tools for Mobile Living Labs participate in the workshop.

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

Topics and questions addressed in the workshop include:

- When to choose *taking the lab to the people* instead of taking people to the lab?
- What are the merits and limitations of Mobile Living Lab methods in general?
- For which kind of applications are Mobile Living Labs beneficial? (e.g., only applications that involve social networks and context-aware applications, or are there other categories as well?)
- What are the merits and limitations of Mobile Living Lab evaluation methods in general? What are the merits and limitation using it for formative evaluation (i.e., to inform the design of new applications) and for summative evaluation (i.e., to assess the (user interface of) an application)?
- What are the relative merits and limitations of self-report methods, measurement methods and observation methods for studying various aspects of mobile user experience?
- How should one deal with long-term and large-scale Mobile Living Labs?
- Which tool support is needed for configuring, deploying, data collection and analysis of studies in Mobile Living Labs?
- What are the merits and limitations of using people's own mobile devices versus handing out new devices for a study?
- To what extent can we study mobile user experience with tools embedded in stationary infrastructure?
- What is known about the reliability and validity of these methods and tools? What are the open issues?
- How to deal with informed consent, privacy, data/device theft, loss and corruption?
- What have we learned so far? What have been the big outcomes from previous Mobile Living Lab workshops?

2. WORKSHOP AGENDA AND RESULTS

The agenda of the workshop, position papers and a summary of the workshop results are published on the workshop website, http://mll09.novay.nl.

3. WORKSHOP ORGANIZERS

- Henri ter Hofte, Novay, The Netherlands
- Kasper Løvborg Jensen, Aalborg University, Denmark
- Petteri Nurmi, Helsinki Institute for Information Technology, Finland.
- Jon Froehlich, University of Washington, USA

4. PROGRAMME COMMITTEE

- Tom Broens, Novay, The Netherlands
- Sunny Consolvo, Intel Research, USA
- Alexandre Fleuri, Aalborg University, Denmark
- Anne Marie Kanstrup, Aalborg University, Denmark
- Joke Kort, TNO Information & Communication Technology, The Netherlands
- James Landay, University of Washington, USA
- Yelena Nakhimovsky, Google, USA
- Ingrid Mulder, Delft University of Technology & Rotterdam University, The Netherlands

5. ACKNOWLEDGEMENTS

This work is supported by the Amsterdam Living Lab project (http://www.amsterdamlivinglab.nl/).

6. REFERENCES

[1] Connelly, K.; Siek, K.A.; Mulder, I.; Neely, S.; Stevenson, G.; Kray, C. (2008). Evaluating Pervasive and Ubiquitous Systems , *IEEE Pervasive Computing, 7*(3), 85-88.

[2] Froehlich, J., Chen, M., Consolvo, S., Harrison, B., & Landay, J. (2007). MyExperience: A System for In Situ Tracing and Capturing of User Feedback on Mobile Phones. In *Proceedings of MobiSys 2007, San Juan, Puerto Rico*.

[3] Hofte, G.H. ter (2007). What's that Hot Thing in my Pocket? SocioXensor, a smartphone data collector. In: *Proceedings of e-Social Science 2007, the Third International Conference on e-Social Science, October 7-9, 2007, Ann Arbor, MI, USA*.

[4] Jensen, K.L. (2009). RECON: Capturing Mobile and Ubiquitous Interaction in Real Contexts. In: *Proceedings of MobileHCI 2009, Bonn, Germany*.

[5] Kukkonen, J., Lagerspetz, E., Nurmi, P., Andersson, M. (2009). BeTelGeuse: A Platform for Gathering and Processing Situational Data. *IEEE Pervasive Computing, 8*(2), 49-56, 2009.

[6] Raento, M., Oulasvirta, A., Petit, R., Toivonen, H. (2005). ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications. *IEEE Pervasive Computing, 4*, 51-59.

[7] Schumacher, J, & Niitamo, V.P. (Eds). (2008). *European Living Labs: A new approach for human-centric regional innovation*. Wissenschaftliger Verlag, Berlin, Germany.

Innovation: A question of Fit – The Living Labs approach

Esteve Almirall, Jonathan Wareham
ESADE Business School
Av. Predalbes 60-62
08034 Barcelona (Spain)
{esteve.almirall,jonathan.wareham}@esade.edu

ABSTRACT
In the recent years Living Labs have manage to draw a significant amount of attention to both their methodologies and organizations. Because of that, a significant amount of effort has been diverted to its understanding. However, very little in assessing its contribution and comparing it to existing methodologies.

This work aims to cover that gap by summarizing the most common European Living Labs methodologies and positioning them in the user-contributed innovation methodology landscape. And by doing that, assess its merits and appropriateness together with policy implications.

Categories and Subject Descriptors
H.1 [**Information Systems**]: Models and Principles.

General Terms
Management, Economics, Experimentation, Human Factors.

Keywords
Living Labs, Living Labs Methodologies, Innovation, Open Innovation

1. INTRODUCTION
When in December 25, 2006 Time magazine [1] select the user as the person of the year for its front page it was doing nothing more than publicly acknowledging the increasing importance of user involvement and participation in generating contents and ultimately in innovation.

Even if users are the final recipients of the innovation process, their participation in the process itself has been precluded by the inability to reach and use the technologies needed to innovate. However, during the past decades and specially since the emergence of the personal computer technology has suffered a process of democratization [2] that translated into two streams: access and the virtualization capacities of information technologies [3]. This process of democratization together with the connectivity and coordination capacities of the Internet [4] have been driving and fueling the raise of user involvement.

However, even if we narrow our focus to innovation, we can witness how this democratization process led to a multiple and diverse practices that will certainly benefit of structures that could facilitate understanding.

In the case of user involvement in innovation, its level of contribution is the obvious classification dimension. Applying this criteria, we can differentiate users as creators, such as in the case of lead users [5] or Open Source; co-creators in practices such as Design Thinking [6], participatory or user-centered design or simply being treated as passive subjects whose insights are captured and introduced in the innovation process, such as in the case of applied ethnography, usability, human interaction or market validation exercises.

Living Labs trials and organizations are situated in this fertile middle ground of considering users as co-partners in the process of innovation and actively involving them in materializing their own needs, aspirations and wishes in their real-life context.

This research aims to examine some of the leading methodologies in the Living Labs community trying to find out through its comparison where are their strengths situated, what spaces of inquiry are they addressing that by capturing the imagination and insights of users could foster innovation. Thus, in our study we address the following research questions,

1. Where can Living Labs methodologies be situated in comparison with other innovation practices?
2. What is the new contribution of Living Labs methodologies that differentiate them from the existing ones?
3. Where are Living Labs methodologies more appropriate in terms of the innovation problem being addressed?

The understanding of these questions is highly relevant, not only for the agents directly involved in innovation, such as companies or researchers, that must select methodologies to address innovation problems, but also to policy makers because of the Open nature of Living Labs, their capacity in developing the Information Society and the importance of the public sector in their development.

The paper is organized as follows. First we briefly review the concept of Living Labs and present the research methodology. Second we describe four leading methodologies coming from CDT (Luleä, Sweden), IBBT iLabo (Belgium), CKIR (Finland) and i2Cat (Barcelona, Spain). Third, we map Living Labs methodologies against others that also seek user involvement/contribution and analyze their unique contribution. Finally we discuss where and when their use could be more appropriate and policy implications.

2. What are Living Labs?
Living Labs are commonly characterized as both a methodology that stresses user involvement in innovation projects and the organizations that focus on its use.

Living Labs are driven by two main ideas: a) involving users as co-creators on equal grounds with the rest of participants and b) experimentation in real world settings. Living Labs therefore provide structure and governance to user participation in the innovation process [7].

There is nothing that prevents the use of Living Labs methodologies in private companies or closed settings. In fact,

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

some well known companies have largely explored its use. However, Living Labs organizations are possibly more interesting because of its open nature and its role as intermediaries in an Open Innovation environment [7].

Living Labs organizations, despite of the difficulties and thanks in part to the support of the EU, in the last two years have grown fast and a network comprising 129 members from Europe, Brazil, South Africa, Mozambique, China and Taiwan has been established.

Our research took this network as the point of departure and examined the most established methodologies, drawing from a combination of secondary sources and field research derived from the active participation in the network and in Living Labs projects during the last three years.

3. Living Labs Methodologies

3.1 CDT. Luleä, Sweden

FormIT [8] is the last iteration of the most used Living Lab methodology in CDT, Luleä (Sweden) (http://www.cdt.ltu.se), one of the oldest and more developed Living Labs.

FormIT tries to put users at the center of the process by involving them through different methods and tools, mostly qualitative. In FormIT three states of product/service development are differentiated: the design of concepts, the design of prototypes and the design of the final system. The methodology evolves in spiral through these three stages.

In each stage we can find a three step process that begins with the appreciation of existing opportunities in applying a new technology, process or product. Once the opportunities are clearly established the process continues with a collaborative design of concepts, prototypes and the final system, depending on the stage. Real life environment validation is maintained through the whole process as much as possible. This three step process is repeated until the results is considered satisfactory.

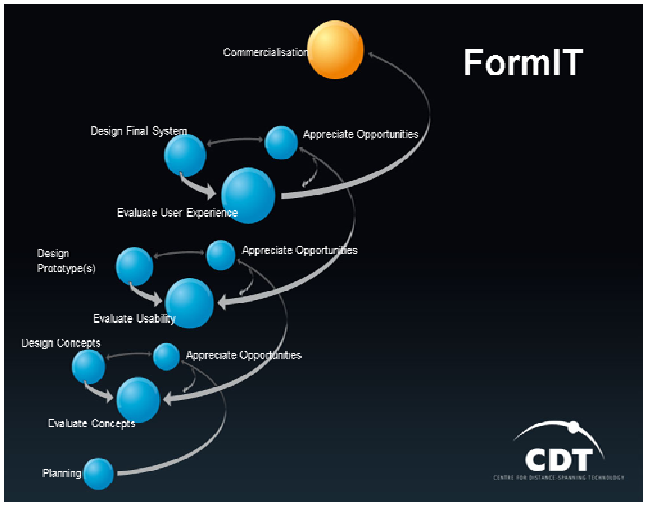


Fig 1. FormIT methodology.

3.1 iLabo – IBBT, Belgium.

iLabo in Belgium (http://www.ibbt.be/ilabo) also uses a three step methodology plus a feedback phase [9]. In this case, probably the most salient aspect is the importance given to the context. This is in a way similar to the appreciation of opportunities step that we encounter in the previous methodology, but here a special focus is devoted to the technological and socioeconomic context.

The first phase is contextutalization that after appreciating the technological and socioeconomic context evolves to user selection, finding groups of users whose insights could be relevant in this context.

The second phase is concretization, where departing from an initial measurement, the concept is developed.

The third phase corresponds to its implementation and testing in real life environments using a combination of logging analysis and traditional qualitative methods.

Finally, an ex-post measurement is conducted and on the basis of the final report a new evolution of the project could be carried out if appropriate.

Similar to the previous case, each phase can be conducted iteratively, but in this case each phase can lead not only to the previous one but to contextualization.

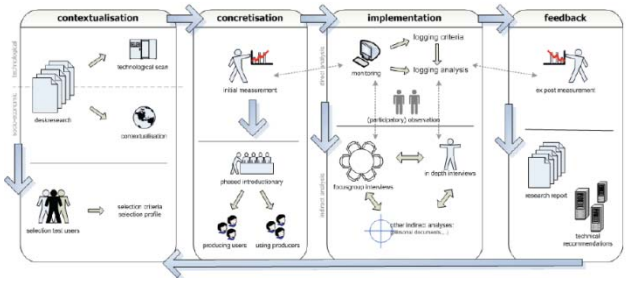


Fig 2. IBBT, iLabo methodology.

3.2 CKIR, Finish Living Labs

To our accounts, a developed conceptualization of Living Lab methodologies in Finland is not published yet. Therefore we will rely on initial versions presented in conferences that try to summarize current practices [10].

In this case, the methodology is guided by pre-defined scenarios that lead the focus of the project. It is again a three phase methodology that evolves in spiral.

In the first phase, called the grounding phase, a similar process than in the previous contextualization one, is conducted, identifying stakeholders and selecting the group of users.

The second phase, interactive and iterative co-design, covers the definition of concepts and the design of prototypes in a co-creative manner.

Finally, the third phase, appropriation and implementation is where public trials occur and feedback is gathered.

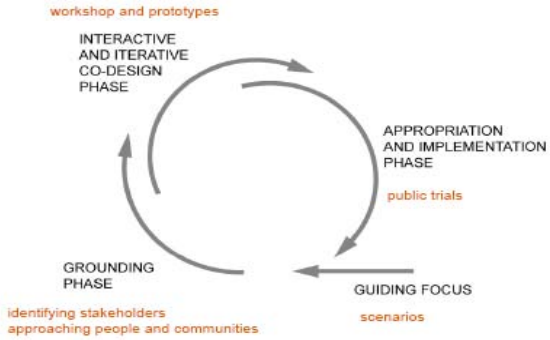


Fig 3. CKIR, Finish Living Labs methodology.

3.3 Catalan Living Labs

In Catalan Living Labs, even if there is not a formalized methodlogy, we can rely on documented cases [7] and presentations given in conferences and workshops.

Catalan Living Labs rely again in a three pahse methodology conducted in spiral, but with an important shift in focus towards implementations in real life environments that serve not only as a prove of concept but as a starting point for a public or commercial venture.

The first phase is devoted to group selection and here users are considered on equal basis with respect to the rest of the team (researchers, companies, etc…). However, the majority of projects are in B2B, where users are nurses in hospitals, patients, IT technicians, etc… Great care is taken in involving the relevant set of users, not only because their insights could contribute to develop a better product or service but because they could help in achieving a successful implementation in the market.

The second phase is devoted to the creation of an innovation arena where the project can develop free from hierarchical structures of the institutions participating. Also, many times, this involves the construction or the use of some kind of infrastructure such as high speed networks.

The final phase corresponds to the actual experimentation in real life environments, paying special attention in experimenting and developing business models that could make the project sustainable.

Maybe the distinctive characteristic of this methodology is the development of an innovation arena with the objective to reduce the uncertainty and therefore the associated risk, while creating an initial demand by involving the relevant actors and showing its viability in real life environments.

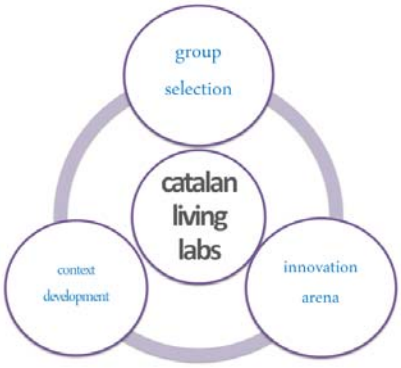


Fig 4. Catalan Living Labs.

4. Mapping User Involvement in Innovation

Graphically mapping methodologies is a way not only of positioning them in relation to each other but also of relating them towards dimensions of interest.

If our aim is to map innovation practices that seek to involve users, the first dimension seems pretty obvious: the level of user involvement in the innovation process. Therefore at one end we will find methodologies led by experts, where users are seen as subjects of investigation while on the other we will have methodologies where users are the ones in charge of the innovation process.

The second dimension that seems relevant in terms of mapping user-centered innovation is the degree of openness. In fact, a user driven innovation process will benefit more from involving a larger and more diverse user base than a homogenous and small one. Examples of that are Open platforms and Open Source.

Besides selecting these two dimensions, we divided methodologies in five different categories,

1) Traditional. Considering innovation as a process similar to engineering, led by experts.
2) User-Centered. Where users are mostly passive subjects of study. This is the case of usability testing, Human factors and Applied Ethnography.
3) User-Driven. Where is the user the one who drives the innovation process. Such is the case of Open Source, Lead Users or Open Platforms.
4) Participatory. That seeks to involve users on equal grounds to the rest of partners in a co-creative process. Here we have Design Thinking, Participatory Design and Living Labs.
5) Collaborative. Where the collaboration between different actors is stressed, ranging from closed networks to open marketplaces. There we find Join ventures, Collaborative Projects and MarketPlaces.

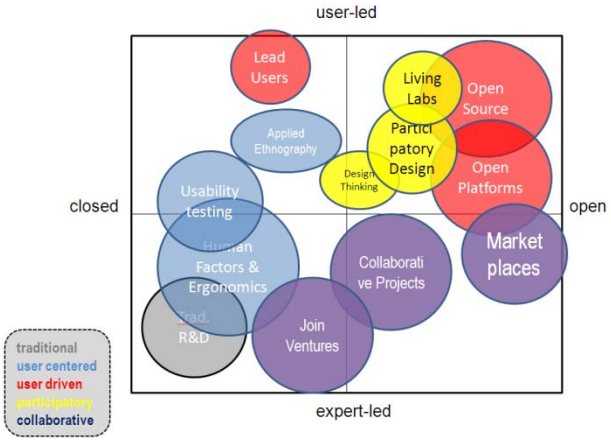


Fig 5. Mapping Innovation Methodologies

5. Analysis

Living Labs methodologies can be and are used in closed environments, but they probably make more sense when applied to open ones. However, because they are instrumented as projects instead of platforms, once the project starts, the group of participants is defined and remains mostly stable. Therefore Living Labs are positioned in the open side of the map, but not as open as open source or open platforms.

We find a similar situation when dealing with user involvement. Even if Living Labs aim to involve users as co-creators, their participation is mediated and they are not the ones leading the process. This contrast with, for example, lead users who effectively drive the innovation process with an entrepreneurial spirit. Therefore we situated Living Labs again in the upper side but in the middle instead of in the edge.

Looking at the map we can observe how Living Labs methodologies contrast with user centered ones in the way that they capture the insights of users. While in user centered methodologies the insights of users are captured and interpreted by experts, in Living Labs are the users themselves the ones that shape the innovation in their own real life environments.

On the other hand, revisiting Living Labs methodologies we can observe that even if each one stresses different aspects, all of them cover the innovation process at three levels,

1) Technological. Ensuring that the technological solution is viable and fills a space of opportunity.
2) Social. Assessing the social an user acceptance in terms of needs and interaction.
3) Economic. Assessing its viability in terms of business model.

Therefore, we can portray the process that takes place in Living Labs as a process of fitting high level knowledge to mid-low level solutions in particular contexts at these three levels: technological, social and business model.

This is done with the help of two mechanisms. First by involving the constituency that possesses the tacit non articulated knowledge about the context where the success of failure of the solution is going to be determined. And second by continuously assessing the validity of the hypothesis formulated with the introduction of this knowledge, in real life environments.

This process of fit is important and relevant because we know that most of the innovation occur in this mid-low knowledge level [11] and is in this level where most of the value can be captured, because globalization made science both global an easily available, excluding it to a great extent from being a source of competitive advantage.

6. Conclusions

Our first and most obvious conclusion of portraying the process carried out in Living Labs as a process of fit is that they will be more relevant where the fit of a particular technology or set of technologies to a precise context is more significant. Therefore, products and services that depend more on their soft characteristics for user acceptance and economic viability seem to be more appropriate.

The second conclusion is that Living Labs will be more appropriate where the fit is less trivial. Indeed, if the fit is trivial, it can be possibly inferred from observing users without having to involve them. However, in situations with multiple stakeholders, conflicting interests and large space of solutions, the innovation problem can possibly only be addressed by involving all constituencies and through its active participation, aiming to trap into their tacit knowledge that will be incorporated in solutions which will be validated in real life environments.

Both conclusions can be easily translated into policy by portraying Living Labs as a resource that allows this exploration exercise in situations where the solution of the innovation problems is hidden behind a complex web of stakeholders and possible solutions.

Historically, this type of infrastructures have been portrayed as a public good and defended because the competitive advantage that they could provide. This is probably even truer in the times of Open Innovation [12] where the presence of three factors: large solutions spaces, global availability of knowledge and technical platforms that make possible the coordination of a large and distributed number of actors, seem to demand the creation and development of this type of infrastructures.

7. REFERENCES

[1] Time Magazine, http://www.time.com/time/magazine/article/0,9171,1569514,00.html, accessed on the 4 of May, 2009.

[2] von Hippel,E. 2005. *Democratizing Innovation*, edn. Cambridge: The MIT Press.

[3] Dodgson, M., Gann, D. and Salter, A. (2005) 'Think, Play, Do: Technology, Innovation and Organization.' Oxford: Oxford University Press.

[4] Shirky, C. 2008. *Here Comes Everybody: The power of organizing without organizations.*The Penguin Press HC.

[5] von Hippel, E.1986. Lead users: a source of novel product concepts. *Management Science*, 32, 7, 791-805.

[6] Brown, T. 2008. Design Thinking. *Harvard Business Review*, June 2008.

[7] Almirall E. and Wareham J.2008 . Living Labs and Open Innovation: Roles and Applicability. *The Electronic Journal for Virtual Organizations and Networks*, vol 10, August 2008.

[8] Bergvall-Kåreborn, B, Ståhlbröst, A, Holst, M, and Mirijamdotter, A. (2006). FormIT - A Methodology for Forming Human Centric Innovative Information Technology OASIS, Presented at 10 December 2006, at Milwaukee.

[9] Ballon, P., Pierson, J. & Delaere, S (2007) Fostering Innovation in Networked Communications: Test and Experimentation Platforms for Broadband Systems. In: Heilesen, S.B. & Siggaard Jensen, S. (eds.) Designing for Networked Communications – Strategies and development. Hershey (PA)/London, IDEA, 2007, pp.137-166

[10] Mikkelä, K. 2008. Finnish Living Lab and User Driven Open Innovation Situation. Workshop on LL, Halmstad, May 29-30, 2008.

[11]] Bhidé, A.2008. The Venturesome Economy – How Innovation Sustains Prosperity in a More Connected World, edn. Princeton University Press,2008: New Jersey.

[12] Chesbrough, H. 2003.*Open Innovation: The new Imperative for Creating and Profiting from Technology*, edn. Boston 2003, Massachussetts : Harvard Business School Press.

User Evaluations in the Wild – Experiences from Mobile Living Labs

Anna Ståhlbröst
Luleå University of Technology
Social Informatics
971 87 Luleå
+46920492091
Anna.stahlbrost@ltu.se

Annika Sällström
Luleå University of Technology
Centre for Distance-spanning Technology
971 87 Luleå
+46920492947
Annika.sallstrom@cdt.ltu.se

Marita Holst
Luleå University of Technology
Centre for Distance-spanning Technology
971 87 Luleå
+46920492453
Marita.holst@cdt.ltu.se

ABSTRACT
Involving users has become a prerequisite these days in IT-product and service development processes; hence, the question is not so much about why we should involve users, but rather how they should be involved. Embracing a Living Lab approach means to strive to involve users throughout the innovation process and to make users engaged co-creators of the innovation. In this paper, we present different degrees of user involvement in design and evaluation processes and relate these to our experiences of involving users in Mobile Living Lab situations. We identify aspects we have grappled with in these process and issues that needs to be elaborated on further since the area of Mobile Living Lab is growing and concepts such as ubiquitous computing and context awareness is emerging. This in turn, sets new demands on methods for user involvement in the wild.

Categories and Subject Descriptors
D.2.10 [**Design Methodologies**]

General Terms
Experimentation, Human Factors

Keywords
User Involvement, Design, Mobile Service, Experiences, Evaluation

1. INTRODUCTION
Involving users in the development of new mobile technologies means giving users the opportunity to participate in the innovation process as representatives of a target user group with the aim to improve the chances of successful innovation [1]. Nowadays, there are many approaches to user involvement which range from developers making assumptions about users' needs without actually involving users, to users' developing the final innovation themselves. To clarify our perspective on user involvement, we refer to Barki and Hartwick [2], who state that the concepts of user participation and user involvement needs to be detached. They mean that the concept participation represents the actions a user performs during the development process, while involvement relate to a psychological state in which the users are more concerned about the system. Adding to that, Olsson [3] declares that the participation concept is imprecise, and techniques claiming to be participatory treat users as sources of information instead of equal partners. In our projects, we have involved users with the aim to give them the opportunity to influence the development and design of future innovative mobile systems where their view is as important as all the others. Hence, the users have been involved, not merely participated, in the development process [4].

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

Users can also be involved to different degrees such as *for*, *with*, and *by* categorization [5-7]. This refers to both users' degree of involvement and their responsibility in processes in which users are involved in evaluation and innovation processes in different ways.

- The first type, design *for* users, means that the system is developed on behalf of the user. Data about the users, general theories, and models of users' behavior are used as a base for the design. This approach often includes specific studies of users, such as interviews or focus groups. Here, users are typically involved in the evaluation of the usability of the system and they give their feedback on predetermined features.

- The second type, design *with* users, denotes a product development approach, focusing on the user, utilizing data on user preferences, needs, and requirements as in a design *for* approach, but, in addition, includes a demonstration of different solutions/concepts for the users, so they can react to the differing design solutions [7, 8]. Hence, the users are involved as co-designers of the system where users are involved and evaluate different design suggestions.

- The third type, design *by* users, apply a product development approach, in which the users are involved actively and partake in the design of their own product [7, 8]. Within this perspective, users are active and continuously give their feedback on the developed

system without being controlled and asked to. This evaluation perspective builds on spontaneous input stemming from real world experiences.

In figure 1 below, we clarify our view of the different perspectives of degrees of users' involvement. The top illustration represents design *for* users. Here, the designer is represented by the driver, who has full control of the situation; the user, represented by the car in the back of the trailer, is following passively from behind, being mostly a source of information. Inherent in this approach, and illustrated by the users' car being on the trailer, is the designer's responsibility to lead and know where to go. In this perspective, the users are involved relatively late in the development process, with the focus on verifying requirement specifications and user interface designs.

The middle illustration represents the perspective design *with* users. Here, the users are involved throughout the process and are on equal terms in co-creation of future solutions based on their needs and experiences. This is represented by the two persons sitting next to each other in the car. In this perspective, the designer is active and in charge of design and development activities (driving the car) and the user is active in design and evaluation activities related to their context (reading the map and giving the directions).

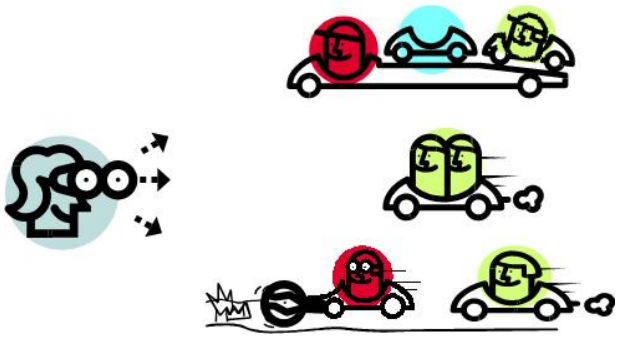


Figure 1: Design for, with, and by users

The bottom illustration is the design *by* users' perspective. Here, users are involved in the role of process initiators; hence, they drive the process. This is illustrated by the car in the back, where the driver has full control and can determine the speed and if s/he wants to follow. In this perspective, users contribute with inspiration and ideas; they produce content and they develop products or parts of products. The role of the designer is to be the facilitator, represented in the picture by the car in front paving the way for the user driving the car in the back. This means that the designer still has influence over what is possible to do or where to go, but the user decides how, when, and if s/he wants to follow.

2. User Evaluations in the Wild – Tales from the Battlefield

SMART - The SMART project is one of many EU-projects aimed to increase citizen involvement in different matters by using information and communication technology (ICT). In SMART this was done by exploring the concept of "reaction media", which allows individuals to engage and take active part in the development and improvement of their municipality. More specifically, we developed a mobile and context aware services that facilitated communication between users and providers. In addition, this service aimed to stimulate the users to give suggestions and opinions for products or services they wanted the company to develop.

The development of these services was carried out in an interactive manner in cooperation between citizens, companies, and official authorities. To facilitate a participatory approach the project was set in a Living Lab context. The foundation of Living Labs is the involvement of four different stakeholders in innovation processes; government, companies, researchers and end-user representatives. Since a Living Lab approach was applied, the aim of the project was to, in close cooperation between involved stakeholders, facilitate innovation, and develop products and services that users really need and that are designed to fit their life pattern and preferences. During this development process, the products and services were tested by end-users in their real-world environments. Since the Living Lab activities can go on 24/7 this means that users can test a product or service in their private context in real usage situations and from the perspective of the different roles they shift between during a day: citizen, parent, sports fan, patient, student, or employee. Hence, the users gain a thorough understanding of how a new product or service function, and fits into their usage context [9-11].

PredicTool - In this project, the focus was to develop and evaluate a mobile marketing service where the retailers could create selling campaigns and offers that were distributed via the mobile channel was developed. The Living Lab approach in this project was to involve the stakeholders; hence, the development was carried out in close co-operation with the retailer organization (the content providers) as well as their customers (the content consumers). The aim was also to develop the mobile service in close relation to the end-users needs, accordingly they were involved early on in the process as well as in the evaluation of the services once it was ready and usable. The purpose of the mobile marketing service was to enrich present customer loyalty-clubs, and, in so doing, give consumers offers that were related to their personal profile [11]. The result from this project revealed quit a diverse view from the users and their expectations on a mobile marketing service which then was used as valuable input for the redesign of the system.

Skygd – a Moblie Security Alarm - In this project, the focus was on evaluation of a mobile security alarm. The service was based on the fact that many people experience that the level of insecurity is increasing in our society. Hence, security businesses are working hard to develop products and services which offer a feeling of security. One way of doing that is to use the mobile phone since most people bring it with them everywhere.

Skygd was a service under development which aimed to increase the feeling of security. With this service, the user can easily send an alarm as soon as they get a feeling of being anxious or in an unsafe situation. The mobile phone is connected to GPS satellites so the users' exact position can be shown on a detailed map. Hence, when they send an alarm the receiver can follow the users' position in real time on a web-based map at the same time as an

audio connection is created and pictures from the mobile phone are transferred to the receiver.

In the project, 20 young girls were involved and they used the service in their real life for three weeks. They were given phones with the GPS function and were asked to use the service. During and after the test period interviews and questionnaires were used as follow up and evaluation methods. The results from this test was used as feedback to the developers and thereafter the service has been launched on the open market.

2.1 Experiences

When performing user involvement situations in their natural setting, we have identified some issues that need to be considered. One such issue is to recognize that the users' response related to a product or service are dependent on how well the tested mobile service fits into its intended context [12].

Hence, to consider the milieu in which the service or product is intended to be implemented is important when designing tests and evaluation of mobile innovations. This indicates that if the context of the test of the mobile system is wrong, the users' attitudes towards it can be affected negatively.

2.1.1 Observation of Usage

As a means to understand what happens in the context during a test, observations can be used as a method for data-collection. The usage of observations makes it possible to observe actual usage in an authentic environment; hence, it becomes possible for the observer to understand factors occurring in the context which might influence the evaluation.

Even though observation is a suitable method for gaining understanding of what has occurred in the context, during the test, this is not always possible in evaluations of mobile systems when a Living Lab approach is applied due to the aim of letting users use the innovation in real world contexts in real world use situations. If the users then are observed, the data will not reveal real use situations when an observer watches their every step [11].

2.1.2 Influence from the Context

When evaluating mobile services in real world settings, it is also important to recognise the risk that users' attitudes can be influenced by things that are not directly related to the service as such, and that these things can influence their experience of the use of the service to become overly negative or positive. For example, the quality of what is being communicated to users affects their attitudes towards the mobile service, as such. This means that if the users are testing, for example, the mobile marketing channel and the offers they get on their mobile are unattractive, they might just as well dismiss the mobile marketing service as such, even though it might the content they dislike. Another example of user's difficulties to detach the mobile service under development, from the context has been viewed in how influential the persons they meet related to the service are. For example, a bad behavior of a shop assistant can make the users stop using the service even though this is more related to them as customers, not as users of the mobile service.

Consequently, recognising issues in the mobile services surroundings that might influence users' attitudes, when performing tests and evaluations in an uncontrollable context, are important [11].

2.1.3 Create an Authentic Usage Situation

The aim of an innovation is to change a reality into a more desirable state and this can not be experienced until the system being evaluated is used, hence, evaluating innovations in an authentic environment is essential to gain as valid user input as possible. In these situations, it is important to understand how the mobile system would be used if it was a real usage situation. Hence, an issue such as timing of the test and evaluation becomes important to give the users a chance of experience a real world use situation. This means that if the mobile service is focused on security, it is best to test the service in situations where the users feel insecure, for example, when it is dark outside.

In addition, we have also learned that letting users use their own mobile phones for the tests is preferable since they do not have to learn a new phone, they do not have to carry two phones with them everywhere they go, and the handling of the tests becomes easier.

In addition, when evaluating innovations such as mobile services during a short test period it is difficult to get results on the actual impact that the innovation has in its context. However, if the test period is extended the risk of ending up with evaluation results without an impact on the innovation becomes even more increased.

Based on or experiences, user needs can be met at different levels. This means that a product, or a service, can meet the need of a user concerning one aspect, but still the user might not be aware of their need of the product, or service, as such. Therefore, a need of an innovation might exist, but the users do not use it anyway, since the users fulfill their needs by a different means. An innovation can do the right thing and in the right way, but still remain unused due to people's general unwillingness to change their behaviour. Consequently, if an innovation does what the users need it to do, a change in user behaviour needs to be encouraged to help the users change their actions [11].

Finally, the experiences of applying a Living Lab approach in these projects have provided us with a diversity of valuable insights which might not been possible by other approaches. For example, in terms of gaining insights about users thoughts, their everyday life and their needs, these insights have supported the development process as the rich data could be used to redesign the mobile innovations. In addition, applying a Living Lab approach also means to be open and to have the endeavor to really understand the users situation even though the situation, context and usage cannot be observed [11]. Spontaneity is another important factor that is supported by the Living Lab approach and which really is an important aspect when the aim is to involve users in mobile Living Lab activities [13]. Here, we identified the importance of support for spontaneous input when the users are set of to use a mobile service in their own home. In sum, applying a Living Lab approach in a Living Lab milieu has provided us with many valuable experiences and insights, not only in the projects referred to here, but in others as well [13, 14].

3. REFERENCES

1. Ives, B. and M. Olson, *User Involvement and Mis Success: A Review of Research.* Management Science, 1984. 30(5): p. 586-603.
2. Barki, H. and J. Hartwick, *Rethinking the Concept of User Involvement.* MIS Quarterly, 1989. March: p. 52-63.
3. Olsson, E., *What Active Users and Designer Contribute in the Design Process.* Interacting with Computers, 2004. 16: p. 377-401.
4. Ståhlbröst, A., *Forming Future IT - The Living Lab Way of User Involvement*, in *Department of Business Administration and Social Sciences*. 2008, Luleå University of Technology: Luleå.
5. Bekker, M. and J. Long. *User Involvement in the Design of Human-Computer Interactions: Some Similarities and Differences between Design Approaches*. in *HCI2000: People and Computers XIV*. 2000.
6. Eason, K., *Information Technology and Organisational Change*. 1987, London: Taylor and Francis.
7. Kaulio, M.A., *Customer, consumer and user involvement in product development: A framework and a review of selected methods.* Total Quality Management, 1998. 9(1): p. 141-149.
8. Larsson, U., *Designarbete i dialog - karaktärisering av interaktionen mellan användare och utvecklare i en systemutvecklingsprocess*, in *Institutionen för datavetenskap*. 2004, Lindköpings Universitet: Lindköping.
9. Eriksson, M., V.P. Niitamo, and S. Kulkki, *State-of-the-Art in Utilizing Living Labs Approach to User-centric ICT innovation - a European approach*. 2005, Centre of Distance Spanning Technology at Luleå University of Technology, Sweden, Nokia Oy, Centre for Knowledge and Innovation Research at Helsinki School of Economics, Finland.
10. Mirijamdotter, A., et al. *The European Network of Living Labs for CWE - User-Centric Co-Creation and Innovation*. in *Integrating Visions of Technology - Proceedings of the 12th Annual Working Conference of CPTS*. 2007. Maarssen, The Netherlands.
11. Ståhlbröst, A., *Human-Centric Evaluation of Innovation*, in *Department of Business Administration and Social Sciences*. 2006, Luleå University of Technology: Luleå.
12. Karat, J. and C.M. Karat, *The evolution of user-centred focus on the human-computer interaction field. IBM Systems Journal*, 2003. 42(4).
13. Bergvall-Kåreborn, B. and A. Ståhlbröst, *Living Lab - an Open and Citizen-Centric Approach for Innovation.* International Journal of Innovation and Regional Development, 2009. 1(4): p. 356-370.
14. Bergvall-Kåreborn, B., M. Holst, and A. Ståhlbröst. *Concept Design with a Living Lab Approach*. in *HICSS-42*. 2009. Big Island, Hawaii.

LivingLabs as Real-World Co-creation Platforms in Development of ICT in Rural India: A Reflection.

Naveen. L. Bagalkot
IT-University of Copenhagen, Rued Langgaards Vej 7, DK-2300 Copenhagen S
nlba@itu.dk

ABSTRACT
By understanding the emerging notion of LivingLabs as an arena for co-creation and not just as a test bed for ICT applications, I have tried to argue for the potentials of employing this notion in developing ICT applications for rural India. Based on my own research experiences of the visual phonebook for low-literate users [6], I have outlined some of the shortcomings of the research methodology that we employed and discussed how LivingLabs can overcome these by active user involvement during all phases of design as co-creators. Further I have discussed on some of the challenges that setting up of such a LivingLab in rural India might face.

Categories and Subject Descriptors
WS [**Mobile Living Labs 09**]:

General Terms
Design, Human Factors

Keywords
LivingLabs, ICT4D, Co-creation, Literacy, Rural India, Mobile Human Computer Interface

1. INTRODUCTION
Developing countries like India have a very high number of people who are not literate and who have vast differences in their socio-economic and cultural leanings. Developing and designing ICT applications for this population can be very challenging and involves a lot of complex issues, especially for the fields of Human Computer Interface and Interaction Design. This could be particularly so for the mobile applications, as mobile phones have broken the barriers of other forms of computing in these areas. The design of Mobile and ICT applications for these demographics was tackled by 'User- Centered' design influenced by ethnographic methods of understanding the 'needs' of people, to develop applications that suit their 'contexts of use' better. Jan Blom et al [2] have specifically talked about the 'Contextual and Cultural' approach for effective user research in mobility. Divya Ramachandran et al [8] have identified, through different case studies of ICT in developing nations, the role of local stakeholders in being part of the design process by contributing the needs and practices while interacting with technologies. They define 'early stage co-design' as the phases through which designers understand the needs and practices of potential users, by using HCI formative design and evaluation methods like contextual inquiry, task analysis, etc.[8, p1087]. Hence, as identified by the above studies, the involvement of the user and the settings of use in the design process is very important in designing effective and appropriate ICT for the people of developing nations. The above examples use the more traditional approach of 'selective ethnographic' methods to involve users in the early framing and understanding phase and in the later evaluation phase. In this paper, I offer an alternate position that the use of LivingLab model is an effective method for active user-involvement, as it involves the user and other stakeholders in the entire design and development process including the ideation and co-creation phases. Also, apart from Co-creation, the 'real-world' situations of use get reflected in the notion of LivingLabs leading to more appropriate, situated designs. I state this position by reflecting on my experiences in being a researcher and co-author for the paper by Joshi et al [6]. To give a better perspective of my position, I come from a background of Industrial design, moving into the field on Interaction Design and ICT for everyday life, and the research for the visual phonebook started this movement.

In the next section I summarize the intent, process and findings of our research [6], where we used the above mentioned methods in understanding the needs and issues of using a mobile phonebook by low literate users. We also used the more formative methods of evaluation in the homes of users to evaluate and validate the designed application in the 'contexts of use'. In section 3, I detail out the idea of LivingLab, which I subscribe to, of the many current notions of LivingLabs and present an alternative model of user-involvement informed by this notion of LivingLabs. In section 4, I specify the potentials that this model offers by reflecting on what I term as 'missed opportunities' of the Visual phonebook research [6]. I conclude by opening up for a discussion on some of the challenges that I perceive lie in taking up the LivingLab model as an effective method of user-involvement in Co-creation of ICT for rural India.

2. SUMMARY OF THE VISUAL PHONEBOOK RESEARCH
Our paper [6] was one of the outcomes of the Nokia University Grant funded research to understand the Language and Literacy issues that people of rural India face when using mobile phones. Our paper specifically concentrated on the issues in using a phonebook to store contacts, leading to the design and evaluation of a Visual mobile phonebook (see figure 1). During our initial user studies, we came across unique and interesting instances of how people with low-literacy save their contact details and how they circumvent the restrictions posed by the current application on their mobile phones. One instance, where the user had saved his friend's vehicle registration number instead of his name [6, p 219]., highlighted the means of appropriation rooted in the everyday situations and circumstances. We further looked at various ways of categorization of contacts according to color, shape, taste, relation, location, etc. leading to a design of the

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

mobile phonebook based on colors and icons. We then demonstrated successfully the advantages of a color and icon based organization in a mobile phone book over the current alphabetically ordered applications. The results of the formative evaluation showed that the low-literate users took very less time to access the contacts on the Visual Phonebook and also created fewer errors in finding the right contact [6, p 222].

Figure 1: Screenshots of the Visual phonebook prototype

We concluded by saying that we "cannot have one interface for all illiterates" [6, p 223] and that we need to design for better appropriation by using local metaphors and organizing principles.

The research methodology was inspired by Hugh Bayer's notion of 'Contextual Inquiry and Design' [1]. Based on this underlying theory of the need to understand the contexts of use for more usable design, we employed various adaptations of the different user research methods, like contextual inquiries, card-sorting exercises, etc. These methods were conducted in the homes of people in rural India, giving us substantial information through observations and insights on the various strategies that the users employ to save contact details. The research phases involved User studies, Categorization study, Design and building of Prototype, Pilot evaluation, Second prototype and Final evaluation. Out of these phases, User studies, Categorization studies and the evaluation were conducted in the homes of the people, while the design, analysis and the prototyping activities took place in the Interaction Design Lab at the university. Figure 2 models the involvement of users in the various phases of the research.

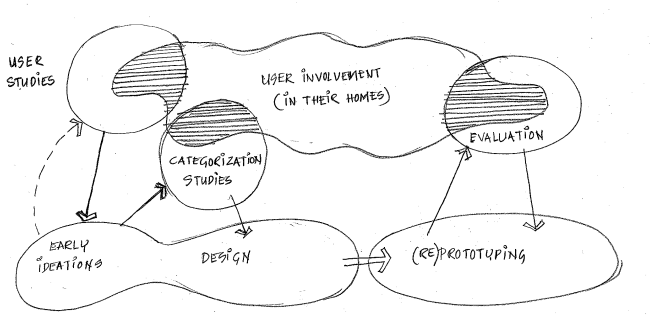


Figure 2: Model of user involvement in the phases of design

It was because of the involvement of the users in the initial understanding phases that the designed application was successful in the evaluation phases. Also the early studies and the testing were conducted in the homes of the people, allowing for a much better understanding of the user strategies and methods of appropriation in as 'natural' settings as possible.

3. THE NOTION OF LIVING LABS

There are various notions and definitions of Living Labs based on the underlying theories and traditions on which they are built and managed. Asbjørn Følstad in his Literature review of LivingLabs [4] has identified three basic types of LivingLabs – LivingLabs to experience ubicomp, LivingLabs as open innovation platforms and LivingLabs as exposing testbed application to users. He further identifies the emerging trend of LivingLabs as a platform for innovation by co-creation, which is grounded in the actual 'contexts of use'. Eriksson et al [3] of the Open LivingLabs also identify with this trend and define their initiative as "an R&D methodology where innovations are created and validated in collaborative multi contextual, empirical real-world environments." [3, p5]. They then go on to differentiate their notion with other user-centered methods by arguing that the LivingLab approach attempts to break the traditionally held idea of user or consumer as an object for research by enabling a co-design process where users and developers actively work towards new solutions [3, p5]. Winthereik et al [5] add to this notion of LivingLabs the Scandinavian participatory design tradition of involving users in the early stages of framing the purpose of design.

They stress the need to clarify the conditions for active user participation by highlighting the specific activities that users are involved during the LivingLab life cycle and also how the users are represented. Their study is a reflection on challenges involved in the conceptual framework of setting up a LivingLab for active co-creation and evaluation throughout the cycle. I will attempt to reflect on some these challenges in section 5 below.

It is this notion of Living Labs as a real world environment enabling co-creation through active user participation in all phases of its life cycle that I adhere to in this paper. In figure 3, I have drawn a model of the above-mentioned notion of Living Lab. It provides an alternative model of user involvement and gives a clear indication of the involvement of users in all activities of the development process in varying degrees. It also places the LivingLabs as situated in the everyday life of people intended to use the systems.

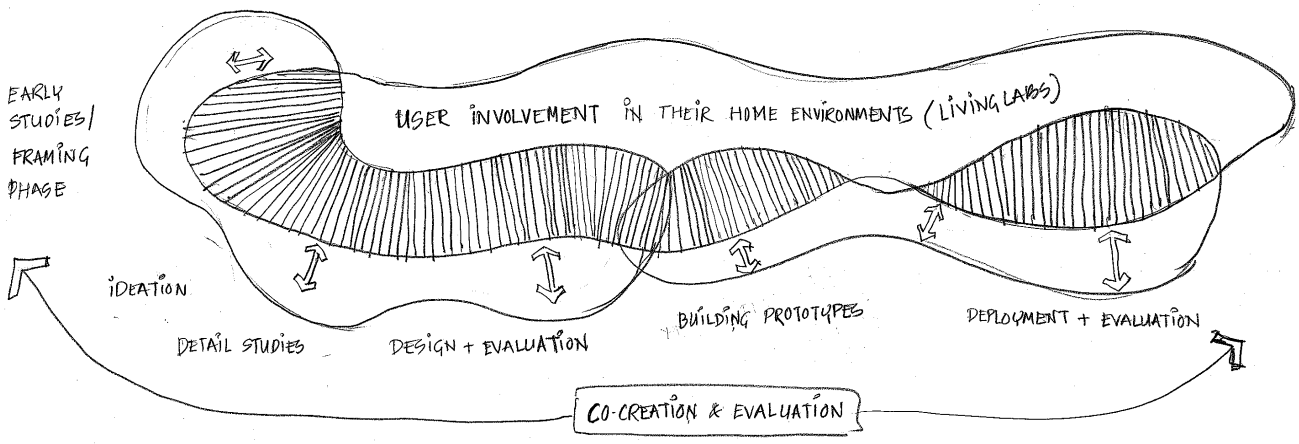


Figure 3: Model of the Living Lab notion as a Real-World setting for Co-creation.

4. REFLECTING ON

In this section I attempt to look back on my experiences during the visual phonebook research and reflect on what potentials the above-mentioned notion of Living Lab would have opened up.

Leaving apart the ideation, design and prototyping phases, we, otherwise, spent a lot of time in the villages trying to first understand how people save their contact numbers and then to identify different methods that they can employ to categorize their contacts, (see figure 2). It was during these two phases that we came across unique instances of different strategies that the users did to save contact details. But this whole set of knowledge gained by user involvement in the actual settings of use was not carried into the final prototype(s) as much as it could have been.

Let us go back to the instance of the user saving his friend's vehicle registration number in the phone instead of his name (see section 2). The users action is a strategy to circumvent the challenges posed by the application and device. As it was easy for him to enter the digits than the alphabets, he found this situation of only one of his friends using a motorbike to use the unique identification number as the identity of his friend. This action of the user embodies the notions of the situations that lead to his action and the meaning that he could make out of it – that the 4-digit number relates to his friend's name. Similar such instances were observed which stressed the need to carry forward the inherent notions of user actions rooted in everyday situations.

Reflecting on the research process through the lens of the Living Lab model as described above, I argue that we missed out on these opportunities mostly because of the below main shortcomings. One, though some phases of the process was set in the 'context of use', we did not have a methodology to understand and take forward this understanding of the complex nature of the situations and circumstances that the everyday world provides the user to base her actions and generate meanings. This resulted in only one idea ending up as a prototype, which allowed some degree of flexibility for the users to use the situations of everyday life in saving contact details. Taking the observation of the user saving the vehicle reg. number, we found it very interesting as one of the many other user strategies, but failed to take it forward into the design phase, ending up with the colors and icons as our only prototype.

Secondly, after the initial studies and card sorting, the users and the settings of use were not involved in the actual ideation and design phases, which filtered some of the insights gained on user strategies during the earlier phases.

Also, the evaluation was specifically goal oriented, comparing two applications with respect to time and errors made in using them. This did not allow a proper understanding of the possibility this one prototype had for allowing users to appropriate it in their actual situated settings of use.

The final prototype that we reached resolves some of the issues like quicker recollection of saved contact details for a low-literate user in using a mobile phonebook. But we lost out on many 'opportunities' to develop a larger set of applications situated in the everyday life of users. I argue that this set of designs will inform the development of ICT for rural India and move the process forward.

Hence, the LivingLab notion as a Co-Creation and evaluation platform situated in the everyday life of users (figure 3) becomes important in dealing with the complexities that the socio-cultural use settings of rural India offer in the development of ICT.

4.1 Potentials of the LivingLab Model

The above-mentioned drawbacks of the research methodology of the visual phonebook can be taken care by the notion of Living Labs as described in figure 3. This notion of Living Lab will provide a better infrastructure in understanding the situations and circumstances of use of the mobile phone, as it is located in the actual everyday life of the users. As it actively involves users even in the creation phases, the existing user strategies of saving contact numbers will get carried into the later designs and these will evoke discussions on what could be other strategies leading to novel designs of phonebook applications. The ideas that don't hold value for the users will be dropped and hence each idea for an application of phonebook generated will be discussed, enacted and evaluated simultaneously in the LivingLab environments by all the stakeholders. Thereby the evaluation will be a means not just to inform and push the design process forward but also provide knowledge for the other developments of ICT in rural India. Also, the other aspect of the LivingLabs as open platforms for innovation, allows for a more flexible approach of defining the intent of each phase and hence the knowledge generated is not filtered. To summarize, given below are some of the potentials that I foresee in using the LivingLab as an open platform for co-creation and innovation.

- Living Labs open up possibilities for development of a larger set of ideas rooted in the socio-cultural practices and actual everyday life of the users by involving users actively in all phases of design and development of ICT solutions in the same everyday life.

- The evaluation is inherent in the ideation and design phases and this informs and pushes forward the design development in a non-reductionist way. The design contribution will be a larger set of ideas and artifacts, which embody the knowledge of appropriation by users rather than looking at one application and its performance in isolations.
- As the whole design process happens in the everyday life of the intended users, the LivingLabs will generate a varied set of design artifacts and prototypes, which embody the intents, the cultural and situational inspirations and the use patterns, thus contributing to the larger set of knowledge of developing ICT applications for developing countries.

5. DISCUSSION ON CHALLENGES

In general, the Living Labs face the challenges of setting and clarifying the conditions of user participation as mentioned by Winthereik et al [5]. They also stress the importance of issues involved in proper representation of the users in the LivingLabs. Specifically for the LivingLabs as innovation platforms in developing nations, Mulder et al [7] have highlighted the challenges involved in transporting the European notion of Living Lab to the cultural complexities of rural South Africa and exchanging of methods and practices between the LivingLabs. They have also mentioned the challenges that lie in the actual setting up of a LivingLab in terms of resources and infrastructure.

With my understanding of conducting research in the rural India along with the understanding of Living Labs as described in this paper, I fore see many interesting challenges for setting up a successful LivingLab platform for ICT design in India. Firstly, the conditions available for active user participation within the LivingLabs have to be clear in the intent of setting up the lab. To extract the best of a LivingLab, people have to be encouraged to be active participants in the processes and this encouragement can be tricky due to the varied cultural leanings. For instance, during the user research for the visual phonebook, we decided to pay a nominal amount for the time people spent with us, but hardly anyone accepted it. Their encouragement was the 'fun' that the card-sorting exercises brought.

The next challenge is representation of users as mentioned by Winthereik et al [5]. Even in rural India, there are multiple demographical changes based on parameters like caste, religion, language, education, and profession, within the users at a same location, age and gender. Hence it is very important to involve the right kind of users in the LivingLabs for effective development.

The infrastructure, both physical and technical, in setting up a LivingLab in rural India can bring in a marked difference in the output. For instance, mobile phones, satellite TV and DVD players are the most spread out infrastructure while Internet and desktop computers are quite thinly spread. Also most of the times, families share one mobile phone within them as observed in the visual phonebook research [6].

More generally, I would like to discuss what is 'co-creation' in a LivingLab construction. While Divya Ramachandran et al [8] see 'co-creation' as the involvement of users to understand their needs, I seek a much more active role for users in a LivingLab. I understand 'co-creation' as an active involvement of users in the design activity of creating artifacts, beyond just contributing to the 'need requirements' and 'evaluation' phases. Hence, as suggested by shown in figure 3 above, by making the 'creative design' phase an integral part of the LivingLab along with the initial 'understanding' and the later 'evaluation' phases, we can attain a much more 'fruitful' partnership with the stakeholders in creating digital technologies to support their daily life.

Finally, I ask a broader and basic question, but a very important one in the LivingLab settings. We need to reflect on the challenge about defining role of designer in the setting of LivingLabs. How much of a control should a designer have over the creation process and the ideas generated, whether a designer is a facilitator or a stakeholder? And more specifically, how to take advantage of the real-world everyday environments provided by the LivingLabs to embody the designs in these environments of use?

6. ACKNOWLEDGEMENTS

I thank my advisor Prof. Tomas Sokoler for clearing of my concepts and helping with compositional problems of the paper. I also thank the research team of the Visual-phonebook, including Prof. Aniruddha Joshi, for giving an opportunity to open my mind towards the challenging field of Interaction Design. I also thank the two reviewers of this paper for their valuable and encouraging comments. Finally, I am grateful to all the people and users who shared their unique insights and practices.

7. REFERENCES

[1] Beyer, H. 1997 Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann Publishers Inc.

[2] Blom, J., Chipchase, J., and Lehikoinen, J. 2005. Contextual and cultural challenges for user mobility research. Commun. ACM 48, 7 (Jul. 2005), 37-41. DOI= http://doi.acm.org/10.1145/1070838.1070863

[3] Eriksson, Mats; Niitamo, Veli-Pekka; Kulkki Seija. 2005. State of the art in Utilizing Living Labs approach to user-centric ICT innovation – a European approach. Unpublished Working Paper, http://www.openlivinglabs.eu/

[4] Følstad, A., 2008. Living Labs for Innovation and Development of Information and Communication Technology: A Literature Review, eJOV, Vol. 10, 99-131. http://www.ejov.org/

[5] J. C. T. Winthereik, L. Malmborg, and T. B. Andersen (forthcoming). Living Labs as a Methodological Approach to Universal Access in Senior Design. Accepted for publication at HCI International 2009, San Diego, CA, 19-24 July 2009.

[6] Joshi, A., Welankar, N., BL, N., Kanitkar, K., and Sheikh, R. 2008. Rangoli: a visual phonebook for low-literate users. In Proceedings of the 10th international Conference on Human Computer interaction with Mobile Devices and Services (Amsterdam, The Netherlands, September 02 - 05, 2008). MobileHCI '08. ACM, New York, NY, 217-223. DOI= http://doi.acm.org/10.1145/1409240.1409264

[7] Mulder, I., Bohle, W., Boshomane, S., Morris, C., Tempelman, H., Velthausz, D., 2008. Real World innovation in Rural South Africa, eJOV, Vol. 10, 7-20. http://www.ejov.org/

[8] Ramachandran, D., Kam, M., Chiu, J., Canny, J., and Frankel, J. F. 2007. Social dynamics of early stage co-design in developing regions. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 1087-1096. DOI= http://doi.acm.org/10.1145/1240624.1240790

The Mobile Oracle – an on Demand Wizard of Oz Tool

Charlotte Magnusson
Department of Design Sciences
Lund University
PO Box 118, 221 00 Sweden
+46 46 222 4097
charlotte@certec.lth.se

Margarita Anastassova
CEA, LIST
18, route du Panorama, BP6
Fontenay-aux-Roses, F-92265 France
+ 33 1 46 54 92 99
margarita.anastassova@cea.fr

Konrad Tollmar
Department of Design Sciences
Lund University
PO Box 118, 221 00 Sweden
+46 46 222 8130
konrad.tollmar@design.lth.se

Martin Pielot
OFFIS Institute for Information Technology, Escherweg 2
D-26121 Oldenburg, Germany
+49 441 9722 188
martin.pielot@offis.de

Kirsten Rassmus-Gröhn
Department of Design Sciences
Lund University
PO Box 118, 221 00 Sweden
+46 46 222 0350
kirre@certec.lth.se

Samuel Roselier
CEA, LIST
18, route du Panorama, BP6
Fontenay-aux-Roses, F-92265 France
+ 33 1 46 54 92 99
samuel.roselier@cea.fr

ABSTRACT
This paper describes a novel tool for eliciting user requirements early in the design process of mobile applications. The "Mobile Oracle", as we have called it, is intended to help developers and designers obtain a better understanding of what the user wants at different points in space and time. It is an extension of a lo-fi version of the well-established Wizard of Oz technique, but it also adds an "on demand" component to force users to explicitly request the information they need. The technique has been tested in an investigation involving 15 users (sighted, visually impaired, and elderly). Results show it to generate valuable information concerning the ways people ask about directions and distances, as well as the services they would like to have in future mobile applications.

Categories and Subject Descriptors
D.2.1 [Requirements/Specifications]

General Terms
Design, Human Factors

Keywords
Design, on demand, wizard of oz, user requirements

1. INTRODUCTION
The study presented in this paper is performed within the framework of the HaptiMap project [17]. In this project, we target the design of systems making use of geospatial data for helping different future user groups in orientation and wayfinding. Thus, one of the questions we need to address is "what kind of information does the user need during the process of wayfinding and exploration, at different points in time and space". The project concentrates on a large number of future users (e.g. visually impaired, elderly and users without any particular sensory impairment) and also on multiple usage scenarios which include city navigation, cycling, hiking etc. Thus, we need methodologies for user requirements elicitation that are flexible enough to adapt to different user profiles and usage environments. Since we have a particular focus on the non-visual interaction design within HaptiMap, we need to be able to handle also these types of interaction. Furthermore we want to obtain a rich picture of the user's activities in order to derive a large number of implications for our future designs.

A literature review showed that previous research used mainly interviews, questionnaires, and experimental evaluations of cognitive wayfinding strategies or of prototypes of pedestrian navigation systems [e.g. 12, 18]. Though these methods provide valuable knowledge and input for design, they are often applied out of the user's context of daily activity. For this reason, relevant elements of this context may be overlooked. In order to get insight into these contextual elements, we decided to use ethnographically-informed observations of users' wayfinding strategies applied in naturalistic settings. In the literature, there are only few studies based on this methodology. Still, it could provide very rich, relatively objective and contextualized information, potentially useful for the design of navigational systems [e.g. 4, 5, 13]. The study presented in this paper adopts the approach of combining a lo-fi version of the Wizard-of-Oz (WOZ) methodology with an "information on demand" technique [20]. These two techniques and the resulting new tool are presented below. We also report an on-going user study within this framework and discuss its methodology and results.

2. RELATED WORK
The WOZ methodology was first introduced by Kelley [15] as a research technique for the design and creation of natural language interfaces. It is a well-established way of testing systems before the actual design is done. The basic idea of the technique is that a human called a *wizard* would act out the responses expected from the not yet existing system and would give feedback to the user whenever needed (just as was the case in the old movie "The Wizard of Oz"). This technique is well-suited to involve users in the early design stages of mobile applications, as the ones designed in the HaptiMap project, since a human can fairly easily impersonate functionality requiring features such as context

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

sensitivity, knowledge of position and related geospatial information.

In the WOZ technique, the *wizard* can be designed with different levels of detail implying different roles for the human simulating the system to be designed [11]. A *wizard* acting as a "controller" fully simulates an unbuilt system function (e.g. system intelligence), whereas a "supervisor" simply oversees a working system and, if necessary, overrides decisions made by the user or the system [11]. An intermediate role is the one of the *wizard* "moderator" who supplements the functioning of an already working system, which cannot be fully trusted. The variety of *wizard's* roles allows the designer to obtain a similar kind of sketchiness or "fudgeability" [9] as the one that has been found to be useful in e.g. lo-fi prototyping [19]. Sketchy or "fudgeable" systems allow designers to explore a larger range of design issues than detailed but rigid systems. This feature is important since the level of detail will affect the way the user responds to the interaction: detailed designs tend to result in comments on usability issues and details, while more sketchy designs tend to provide comments on the concept and the overall design [21]. Specifically in early design stages feedback of the latter type is more important, and we decided to use a lo-fi WOZ design where the responses allowed to the *wizard* are not specified in detail, but rather given as information about areas of competence.

In order to add even more "fudegeability" (i.e. openness and flexibility) to the WOZ method, we decided to combine it with the technique of "information on demand" [20]. The idea of this technique is that, instead of looking for information himself, the user would ask questions to the experimenter (the *wizard*, in our case) when needed. The underlying assumption is that seeking advice is not only matter of question-answer dialogue, but also a matter of asking questions, as people frequently make specific claims about the answer in their own query [7]. Consequently, the "information on demand" technique allows investigating the pieces on information selected to solve a problem, the order in which they are used by the person, as well as the variety of individual problem solving strategies [11]. In addition, asking questions encourages people to make their information needs explicit. The resulting tool, called the "Mobile Oracle" which combines the advantages of a lo-fi version of the WOZ technique and the "on demand" technique, is introduced below.

3. THE MOBILE ORACLE IN THE FIELD

We employed the proposed "Mobile Oracle" tool in a test design where the user is instructed to request information from a mobile navigation service when he/she feels this is necessary. Our assumption was that such a combination could result in a potentially useful tool for providing additional insights into the kind of information users are interested in during navigation and exploration. To further strengthen the idea of the "on demand" part of the design we decided to call the individual impersonating the imagined device the "Oracle", since an oracle is held to respond wisely when consulted and asked questions.

The Mobile Oracle tool could be useful for design for three major reasons: 1) the richness and dynamics of dialogues as cooperative activities, in which protagonists' objectives are co-constructed, transformed and refined in order to gradually concord [14]; 2) the openness of a traditional WOZ task, in which there is no single correct answer to a question asked by the user; and 3) the additional freedom and initiative given to the user thanks to the "on-demand" part of the technique.

4. TEST CASE

To evaluate the "Mobile Oracle" we had to find an appropriate scenario. We wanted to have a relatively complex task, because, obviously, any simple task which only requires few interactions between the two protagonists will not give much data to analyze [10]. We also had a particular interest in more exploratory navigation. Thus we decided to make use of a shopping scenario where the user was asked to navigate in a shopping mall, to investigate three different types of items and select one of these. A shopping mall is an interesting place because shoppers often have problems in finding their way in a mall, while at the same time there are few studies on this topic [8, 25], especially as far as the needs of visually impaired people are concerned. At the same time we expected to obtain general information about wayfinding in urban environments, because a large part of the cognitive wayfinding strategies are similar in both settings. Since we target several user groups we included 3 sighted, 4 visually impaired and 8 elderly in the test. We conducted 2 pilot tests before the actual testing to make sure that the test design worked as intended.

Each test followed a fixed scenario:

> "You are to attend a birthday party. You have not yet bought the present and have a limited amount of money to spend. The person you are buying the present for is Oscar, 13 years old and a fantasy fan. He has a wish list that contains the following items:
>
> - A t-shirt/sweater with a fantasy motif (dragons/skulls or similar)
> - A necklace with a dragon/skull pendant ("cool" male type necklace)
> - A fantasy book with dragons in it
>
> You want to check all three types of things and buy the cheapest. But you can only spend 30 minutes on this task in order to make it for the party.
>
> An ORACLE - a person, who simulates a mobile navigational service, follows you. Please ask the ORACLE as soon as you feel you need some service or information. If possible we appreciate if you also try to "think aloud" about your navigational decisions/considerations."

We gathered test data by using a mobile phone to record audio and a small digital camera to take pictures (due to security regulations, video recording was not allowed, but we were granted permission to take pictures). During the pilot testing we had one observer that took notes – but this generated unwanted attention. Thus, during the actual test only a mobile phone (held by the test person) was used for the recording. By holding a mobile device we also wanted the test persons to be aware of the mobile technology.

One of the experimenters took the role of the "Mobile Oracle". It was designed so that it knew the layout of the mall and could say things about the general type of content in different shops. The oracle could also remember places or things for the user and can point out where things are in general (like different sections in a shop). Moreover, once at a particular stand the oracle could pretend to be a shop assistant/sales person (we did not want to

disturb the personnel in the shops since we did not intend to buy anything). To avoid influencing the user, we decided to leave much of the workings of the oracle open and only encouraged the user to ask as soon as some information was needed.

Figure 1. Pictures from the test environment

After the actual test we interviewed the participants about the navigational experience and their thoughts on potential navigational services. They also filled in a questionnaire containing questions about how familiar they were with the mall in question, their visual ability and their sense of direction.

5. EARLY RESULTS

The method used was seen to work well for all the users involved in the study. The resulting questions were recorded and could be grouped in the categories "Content overview", "Spatial layout", "Direction/route", "Distance", "Notification/prompts", "Confirmation", "Content", "Recommendation", "Memory", "Time" and "Capability of the device". Added to this grouping, the individual questions themselves were seen to provide rich and detailed input for the design process.

As one example, the preliminary analysis shows that all three user groups were interested in a service that prompts you as you pass by interesting things (although it has to be well-designed not to be annoying). Specific requirements for more severely visually impaired users were information about the precise location of the entrances and obstacles as well as the possibility of optimizing the way to a shop. Landmarks, distances, directions and orientations were other things of interest, as were the kind of shop indicating service that helps you noticing things as you move from A to B (eg. I'm going from here to the bookshop but I would like to add suitable shops for clothes on the way).

Overall the scenario caused exploratory behaviors and all three groups of users asked the Mobile Oracle relevant questions:

- The Mobile Oracle was able to engage the subjects in such as way that they could provide us with interesting and relevant information about "what kind of information does the user need during the process of wayfinding, at different points in time and space".
- The kind of questions asked (and the discussions generated) provided information both for the type of content that will be needed as well as for the type of services suitable.

Finally, the way the data was recorded did not disturb the environment. Two or three people walking around talking, where one of them is holding a mobile phone, fitted well into the environment and did not attract much attention.

6. DISCUSSION

On the general level, the results we obtained agree well with the available knowledge on which environmental features are important for wayfinding [1, 6, 18, 23, 24]. Added to this we get detailed, situated [16] comments and questions. We observed that the Mobile Oracle technique works well in studies involving non-visual interaction and a wide range of user groups (including visually impaired users). One of our initial concerns was that the sighted participants might neglect the Oracle and just solve the task on their own. This turned out not to be the case – even users stating that they did not like to ask for help and preferred to walk around on their own asked for help from the Oracle (in total 5 sighted persons did the test, since both participants of the pilot studies were sighted).

Although the main source of information is the user questions, interesting information can also be obtained from the interactions with the Mobile Oracle. It was interesting to see how the participants reacted to the directions given by the Mobile Oracle and how the Oracle (being human) adapted these descriptions to the user in question. In the presented study the Oracle was quite free to define and construct himself/herself through the interaction with the user using both verbal descriptions and physical pointing. Further work on this technique could examine in detail more specific guidelines for the way the Oracle responds.

In the presented field study one of the experimenters filled out the role of the oracle. This ensured that the Mobile Oracle had sufficient knowledge about the environment and the task. While the participant representing the user gives us insights into what information is needed, the insight we get about how to provide that information is limited since in every session the same set of persons (the experimenters) acts as the Oracle. Therefore another possible extension to the described tool could be to have participants also acting as the Oracle. This would allow investigation of how the needed information is conveyed by humans serving as a guideline for the information presentation design.

Nevertheless, in order to get valid findings, the participants acting as Mobile Oracle need to have deep enough knowledge about the environment to give helpful and timely suggestions. In the pilot tests performed before the actual study we tried having participants fulfilling both roles, and came to the conclusion that this did not work that well due to lack of knowledge of the environment, even though we provided the Mobile Oracle with an annotated map. One solution for this is to have a "Mobile Oracle training session" before the actual test, but this also puts higher demands on the amount of time the test participants have to spend. Another alternative is to find participants who are already very familiar with the environment in question.

7. CONCLUSION

Our results show the usefulness of a technique consisting of a combination of a "sketchy" or less detailed version of the well known Wizard of Oz technique with an "on demand" type of interaction design. The tool, which we call the "Mobile Oracle" has been tested in a recent user test in an explorative shopping scenario, and was seen to work well for sighted, visually impaired

and elderly users. The individual questions were seen to provide rich and detailed input for the design process while the overall categories resulting from a grouping of these questions agree well with what is found in earlier studies [1, 6, 18, 23, 24] confirming the validity of the "Mobile Oracle" approach. Thus this type of approach can indeed provide useful information in the early stages of a mobile application design process, by providing an open framework in which users can present, negotiate and construct their requirements for the future technology.

8. ACKNOWLEDGMENTS

The authors are grateful to the European Commission which co-funds the IP HaptiMap (FP7-ICT-224675). We also want to thank VINNOVA for additional support.

9. REFERENCES

[1] Allen, G.L. 2000. Principles and practices for communicating route knowledge. Appl. Cog. Psy. 14, 4 (July/Aug 2000), 333-359.

[2] Bisseret, A., Sébillotte, S., and Falzon, P. 1999. Techniques Pratiques pour l'Etude des Activités Expertes. Octarès.

[3] Bradley, A. and Dunlop, D. 2005. An experimental investigation into wayfinding directions for visually impaired people. Pers. Ubi. Comp. 9, 6 (Nov. 2005), 395-403.

[4] Brown, B. and Laurier, E. 2004. Designing electronic maps; an ethnographic approach. In Map-Based Mobile Services – Theories, Methods and Implementations, L. Meng, A. Zipf, T. Reichenbacher, Eds., Springer Verlag. Berlin Heidelberg, 241-257.

[5] Brown, B. and Laurier, E. 2005. Maps and journeys: an ethnomethodological investigation. Cartographica, 40, 3 (Fall 2005), 17-33.

[6] Caduff, D. and Timpf, S. 2008. On the assessment of landmark salience for human navigation. Cog. Proc. 9, 4 (Nov. 2008), 249-267.

[7] Carroll, J. and Aaronson, A.P. 1988. Learning by doing with simulated intelligent help. Com. ACM, 31, 9 (Sept. 1988), 1064 – 1079.

[8] Chebat, J.Ch., Gélinas-Chebat, C., and Therrien, K. 2005. Lost in a mall, the effects of gender, familiarity with the shopping mall and the shopping values on shoppers' wayfinding processes. J Bus. Res., 58, 11 (Nov. 2005), 1590-1598.

[9] Cooper, A. and Reimann, R. 2003. About Face 2.0 – The Essentials of Interaction Design. Wiley Publishing.

[10] Dahlbäck, N. and Jönsson, A. 1989. Empirical studies of discourse representations for natural language interfaces, In Proceedings of the 4th conference on European chapter of the Association for Computational Linguistics (Manchester, England, April 10-12, 1989).

[11] Dow, S., MacIntyre, B., lee, J., Oezbek, Ch., Bolter, J.D., and Gendy, M. 2005. Wizard of Oz support throughout an iterative design process. Perv. Com., 4, 4 (Oct./Dec. 2005), 18-26.

[12] Giudice, N.A., Bakdash, J. Z., and Legge, G. E. 2007. Wayfinding with words: spatial learning and navigation using dynamically updated verbal descriptions. Psy. Res., 71, 3 (May 2007), 347-358.

[13] Jönsson, B., Malmborg, L., Svensk, A., Anderberg, P., Brattberg, G., Breidegard, B., Eftring, H., Enquist, H., Flodin, E., Gustafsson, J., Magnusson, C., Mandre, E., Nordgren, K., & Rassmus-Gröhn, K. 2004. Situated Research and Design for Everyday Life. CERTEC Report, LTH Number 2:2004, University of Lund, http://www.certec.lth.se/doc/situatedresearch/situated.pdf.

[14] Karsenty, L. and Falzon, P. 1992. Spontaneous explanation in cooperative validation dialogues. In Proceedings of the Workshop on Improving the use of knowledge-based systems with explanations ECAI'92, P. Brezillon, Ed., Université Paris VI, Paris.

[15] Kelley, J.F. 1984. An iterative design methodology for user-friendly natural language office information applications. ACM Trans. Info. Syst. 2, 1 (Jan. 1984), 26-41.

[16] Lueg, C. and Pfeifer, R. 1997. Cognition, situatedness, and situated design. In Proceedings of the 2nd International Conference on Cognitive Technology CT '97 (Aizu-Wakamatsu City, Japan, August 25 - 28, 1997). IEEE Press, Washington, DC, 124-135.

[17] Magnusson C., Brewster S., Sarjakoski T., Roselier S., L. Sarjakoski T., Tollmar K 2009. Exploring Future Challenges for Haptic, Audio and Visual Interfaces for Mobile Maps and Location Based Services, LocWeb 2009, Acm press

[18] May, A.J., Ross, T., Bayer, S.H., and Tarkiainen, M.J. 2003. Pedestrian navigation aids: information requirements and design implications. Pers. Ubi. Comp., 7, 6 (Dec. 2003), 331-338.

[19] Nilsson, J. and Siponen, J. 2005. Challenging the HCI concept of fidelity by positioning ozlab prototypes. In Proceedings of the 14th International Conference on Informations Systems Development ISD'05 (Karlstadt, Sweden, 15-17 August, 2005), Springer, Berlin Heidelberg.

[20] Rimoldi, H.J.A. 1963. Processus de décision et fonctions mentales complexes. Rev. Psy. Appl., 13, 2, 65-81.

[21] Rudd, J., Stern, K., and Isensee, S. 1996. Low vs. High-Fidelity Prototyping Debate. Interactions, 3, 1 (Jan. 1996), 76-85.

[22] Sá, M., Carriço, L., Lessons from Early Stages Design of Mobile Applications, MobileHCI 2008, Amsterdam, The Netherlands

[23] Spiers, H.J. and Maguire, E.A. 2008. The dynamic nature of cognition during wayfinding. J. Env. Psy., 28, 3 (Sept. 2008), 232-249

[24] Strothotte, T., Petrie, H., Johnson, V., and Reichert, L. 1995. MoBIC: user needs and preliminary design for a mobility aid blind and elderly travellers. In Proceedings of the 2nd TIDE Congress (La Villette, Paris, France, 26-28 April, 1995).

[25] Tlauka, M., Brolese, A., Pomeroy, D., and Hobbs, W. 2005. Gender differences in spatial knowledge acquired through simulated exploration of virtual shopping centre. J. Env. Psy., 25, 1 (March 2005), 111-118.

Understanding tourists on a bicycle trip "in the wild"

Martin Pielot, Benjamin Poppinga
OFFIS Institute for Information Technology
Oldenburg, Germany
{pielot|poppinga}@offis.de

Susanne Boll
University of Oldenburg
Oldenburg, Germany
boll@informatik.uni-oldenburg.de

ABSTRACT
In this paper we describe and discuss a requirements analysis that aimed at informing the design of a pervasive application. We approached the requirements analysis by conducting a survey and a covered field observation to understand tourists on a bicycle trip. Both studies yielded significantly different results. We therefore join arguing that studies of different types should be mixed in order to get a complete understanding of the target situations. We showcase that the field observation, although supposedly being less valid than the survey, yielded important results that we would not have found by concluded the survey only. Our field observation also highlights that there is a need for guidelines about ethics when evaluating pervasive applications in the field.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Evaluation/methodology

General Terms
Human Factors, Experimentation

Keywords
Requirements Analysis, Pervasive Applications

1. INTRODUCTION
In order to design successful and useful pervasive applications it is important to understand the user's needs beforehand. Currently it seems to be agreed that a single method for understanding users "in the wild" is not enough. They rather have to be triangulated for countering the disadvantages of each method alone [4].

In our previous work we designed a pervasive exploration and orientation aid for tourists on a bicycle trip. At first, we aimed at understanding how tourists on bicycle trips can be supported by pervasive applications. We conducted two supplementary studies: a survey where details about a trip were asked and a covered field observation where we observed tourists on their actual bicycle trips. The two studies drew considerable different pictures about the tourists' situation.

Mobile Living Labs 09 September 15, 2009, Bonn, Germany

In this paper we present the conducted studies and highlight the different results of the survey and the field study. In addition we discuss the topics that would be relevant for any requirements study for pervasive applications, comprising the different results of both studies, and validity and ethical issues with the field study. This discussion aims at contributing to the discussion about methodologies for requirements studies for pervasive applications.

2. RELATED WORK
"If I'd asked what they wanted, they would have said a faster horse" is a well-known sentence by Henry Ford. It stressed out that designers not only need to understand the superficial needs of users' but also have to reveal the users' underlying motivations to produce successful designs. This section provides an overview of methods that can be used to elicit requirements for pervasive applications.

Surveys, such as e.g. questionnaires, interviews, or focus groups, are established means for understanding the user's needs. While there are usable for gather the user's superficial needs they suffer from relying on the subjective views of the informants [7]. Consequently they might not reveal the users' underlying motivation.

According to Kjeldskov et al. [7] the most common practice in mobile HCI is engineering prototypes and evaluating them in the lab. Field evaluations are rare, but can yield useful and context-related insights (e.g. [1]). Still, engineering prototypes is useful for getting specific feedback, but it's rather ineffective in understanding the user's general needs, since participants tend to focus on the presented prototype. In early design stages, where prototypes are not yet existing, methods such as the Wizard-of-Oz Kelley [6] can be used to inform the first iterations of the design. The prototypes logic is then simulated by an experimenter, allowing to investigate requirements and potential design issues before the actual system is built. While the Wizard-of-Oz method does not depend on an existing prototype, it still requires an idea about the system to simulate and thus has the same problem as evaluating engineered prototypes.

To avoid bias by an existing prototype, one can investigate current strategies and practices. For example, May et al. [8] studied how people describe routes in order to understand the structure and properties of good and usable route descriptions. While this approach produces valid, reliable, and generalisable requirements, it can mostly be only applied to

parts of the design. Adopting methods from ethnographic research (e.g. [2, 3]) is another option for requirements analysis which has the advantage that it yield natural and rich results. As this kind of studies is very costly and time consuming, low cost variants of ethnographic research have been employed, such as cultural probes [5] or diary studies [10].

3. UNDERSTANDING TOURISTS ON A BICYCLE TRIP - THE STUDIES

In order to understand what kind of system would support tourists on a bicycle trip the most, we conducted a requirements analysis "in the wild" combining two different methodologies. The goal was to inform a design of a system that should support tourists on their bicycle trips in unfamiliar environments. The methodology of the conducted studies, the results, and the conclusions we drew are discussed in the following.

3.1 Methodologies

We approached the requirements analysis by conducting two complementing studies: first, we assessed how tourists plan and conduct their bicycle trips by a survey. Second, we conducted a covered field obervation where we observed tourists at parts of their bicycle trips. Both studies took place at the North Sea Island Borkum, Germany. Due to its size of about $30km^2$ and its infrastructure, it is perfectly suited for exploring it by bicycle. The local tourist agency has optimised routes and signs for tourists bicyclists. Typical tourists are families with smaller children as well as people with chronic cardinal or respiratory problems. The large number of bike rental outlets confirms that it's common practice for tourists to rent a bicycle for a trip.

Survey

The survey was administered by semi-structured questionnaires. We cooperated with a local bicycle renter at Borkum who handed out these questionnaires to people that rented bicycles. Thus, our sample contained a random set of tourists and other visitors that did bicycle trips on the island. The questionnaire was split into three sections. The first section asked the informants about their plans with respect to the trip. The second section focused on the experiences during the planned trip. The third section gathered relevant participants' details, e.g. how familiar they are with the island by the time they were doing the trip.

Field Observation

The field observation also took part at Borkum. Tourists making a bicycle tour were observed for a part of their trip. An observer rode around busy public areas and followed random cyclists, mostly in the range of audibility. The observer did not attract attention, since around the time the study took place the island was full of cyclists. Tourists were observed for a part of their route and only as long as they stayed in public space. Once they decided to stop e.g. at a cafe, the observer discontinued the observation. The observations were written down from the observer's memory. No personal information was recorded.

3.2 Results

In the following we present a brief summary of the results of both studies.

Survey

Ten questionnaires were returned. Eight informants reported that they used navigation aids on their trip. These were mostly paper maps and seldomly public "you are here" maps. None of the participants used electronic navigation aids. Although eight of ten informants had given a destination for their trip, only five of them actually reached that destination. In general, the destinations described rather large areas, such as a beach area that spans across half of the island. Seven informants stated that they had lost orientation at least for a short period of the trip. Nevertheless, they still expressed that the navigation aids were found helpful.

Field Observation

Notes were taken about six groups and four single persons during the field study. The use of paper maps was observed twice. We did not observe the use of any other navigation aid. Route choices often seemed loose and spontaneous. Sometimes they were heavily discussed within the group. Efficient navigation did not seem to be much of a deal. One family, for example, turned around three times during the observation. They seemed to make those decisions, because they did not like the environment. Overhearing a few conversations we discovered that often there were no definite destinations for the trips or they were re-planned during the trip.

3.3 Discussions

After having analysed both studies, we compared the results in order to get a more complete picture of what a good support system should offer the tourists. The survey indicated that destinations were often imprecise and not always reached, and despite using map, disorientation events occurred. The field observation showed that the nature of navigation was mostly undirected and spontaneous, while mostly no navigation aids were used. We discovered four central themes:

Limited map usage

In the survey, seven of ten informants reported to have used a map, we could only observe two map usages in ten observations. We suggest that this discrepancy might be due to the fact that the observed people had a map with them but used it at times where we did not observe them anymore. We also suggest that there might be differences between the observed people, and those who participated in the survey. As participating in the survey was voluntarily, a different kind of people might have participated, who are also more likely to use maps. The survey also showed that maps are perceived helpful while the informants still saw room for improvement.

Frequent loss of orientation

Seven of ten survey participants and seven of ten observed tourists lost orientation at least once and for a short time. Thus, both studies consistently indicate that tourists often frequently experience disorientation events.

"Easy-going" navigation

The observation showed that people were uncertain at decision points but nevertheless spontaneously decided how to

proceed. This indicates it was not that important for the tourists to take the most efficient route. This might be attributed to the holyday experience in general. One survey participant commented a case where he lost orientation with the words: 'One time, I chose the wrong way, but it did not matter. I am on holiday!'

Spontaneous change of destination

While eight of ten survey participants specified a destination for their trip only five of those actually reached that place during their trip. Additionally, most destinations were not specific places but larger areas. From the observed tourists only few seemed to have a specific destination. Therefore, reaching specific places does not seem to be a high priority for tourists. It also shows that those tourists were willing to alter their plans and accept unexpected changes.

4. DESIGN OF THE TACTICYCLE

These results of the requirements analysis were applied to the design of a pervasive application that should support tourists on their bicycle trips. We derived four design implications for such a system: They high number of reported and observed disorientation events and the limited use of maps suggested that such a trip companion should **provide orientation help**. The spontaneous nature of navigaton indicates that preplanning of trips is not desired by the user. Instead, **planning trips on-the-fly** should be supported. The field study also showed that people are open for spontaneous deviations from their current goals. In order to improve the experience, a travel companion should therefore **highlight interesting places** nearby. Since destinations were rather denoted by large areas and reaching those goals seemed mostly optional, a **drift towards the destination** should be supported rather than providing detailed route instructions.

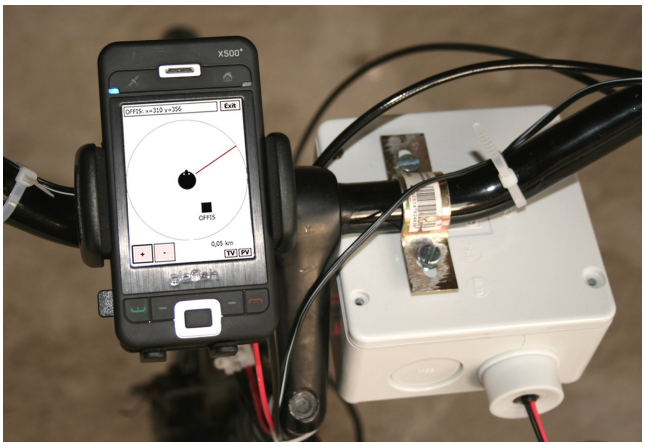


Figure 1: The hardware of the resulting Tacticycle system attached to the steering rod of a bicycle.

These implications drove the design of the Tacticycle. The Tacticycle is a pervasive application designed for improving the experience of a bicycle trip through implementing the above identified design implications. Surrounding landmarks are displayed by a PDA for providing orientation help. The integrated GPS receiver is used to obtain the cyclists position. A self-developed hardware platform connected by Bluetooth allows obtaining the cycles orientation by an electronic compass as well as driving two vibration motors that are fixed to the steering rod handles (see Figure 1). These are used to convey the direction of a selected destination (see Figure 2) and announce nearby interesting spots via tactile feedback, and thus supporting the drift towards a destination as well as supporting planning trips on-the-fly The PDA application allows changing destinations so the cycle trip can be planned on-the-fly. A more detailed description of the system can be found in [9]

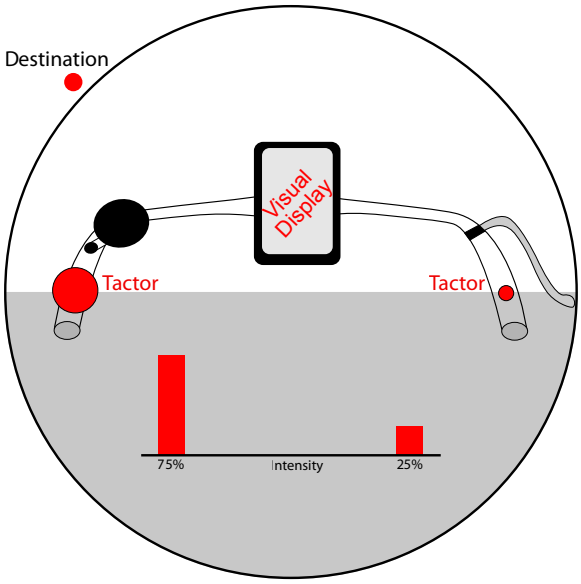


Figure 2: The destination (at 315°) gets divided into two intensities applied to the actuators.

5. CHALLENGES

The survey and the field study yielded significantly different results. In addition, the field study had issues with the validity of the results as well as privacy and ethical concerns. Our reflections on these challenges are discussed in the following.

5.1 Survey vs. Field Study

While the survey results might suggest that tourists mostly have a pre-defined route and destination, the observations made in the field study contradict that conclusion. There are several explanations for these results: 1) The survey participants did not want to admit that they felt lost for some reason, 2) the survey participants did not feel like being badly oriented, or 3) the survey participants represent a different population which actually does not get lost so easily compared to the observed people. The aspects 1) and 2) are always problematic in surveys, as they rely on the participants being honest and capable of self-reflection. Aspect 3) is a problem of voluntary participation, as people that are more likely to volunteer for participating might also be more likely to plan their trips in more detail. Thus, our samples of participants might reflect different populations.

The fact that the majority of the survey participants reported to be content with their navigation aids despite having lost their orientation supports the second aspect, that

people did not feel badly oriented. Only in the combination of field study and survey we learned that people might be willing to accept unexpected deviations from their plans. This supports the conclusions of previous workshops on evaluating pervasive applications [4] that triangulating methods is necessary to compensate for the weakness of each.

5.2 Validity of Observations

The field study lacked a direct communication between the observer and the informants, such as thinking aloud. Instead, any observation is in fact an interpretation of the observing person. The observer acts as a second filter which can distort results further. Thus, the quality of the results strongly depends on the observer's capability of classifying the observed behaviour correctly. Given the fact that the survey and the field study yielded significantly different results (e.g. map usage) there is a good chance that the observations from the field study where partially misinterpreting.

Nevertheless, in retrospect, the field study yielded important results, such as the spontaneous und undirected nature of navigation observed in many cases, which would not have been uncovered by the survey alone. Most likely, only few other methods would have yielded the same results. Thus, even if the validity of findings is questionable, such studies can greatly contribute in understanding the design space when building pervasive applications.

5.3 Privacy and Ethical Considerations

If possible, experiments should gain informed consent from the participants of their study. The consent is required if personal information are collected during a study. Failing to due so may be considered unethical and even counter laws. On the other hand, informing participants that they are being observed will most likely alter their behaviour. In the case of our study, the participants might have focused much more on navigation, leading to the impression that there is actually no need for an orientation aid.

To avoid unethical behaviour, two aspects had to be considered. First, informed consent is required when collecting personal information. Personal information can be defined as information, where the person's identity can "reasonably be ascertained" from the information. Examples are names, video and audio material, but also the context of observation. In our study we approached this issue by forgoing any recording tools and only taking written notes from the observer's memory. Second, if people are observed their privacy may be violated. However, in public places, there is no general expectation of privacy. Thus, we restricted our observations to public spaces, namely to the highly frequented roads and bicycle trails. The observed people were aware that their actions could potentially be observed by anybody.

This issue shows that there is a need for clarifying the ethical and privacy aspects of evaluation "in the wild". Future methodology for understanding requirements of pervasive applications should provide guidelines for researchers in order to keep their work ethically sound.

6. CONCLUSIONS AND FUTURE WORK

In this position paper we presented a requirements study aimed at informing the design of a pervasive system and the challenges we faced during the study. It comprised a survey and a field study which yielded significantly different results. While especially the validity of the field study is questionable, it still provided us with helpful insights. We support previous conclusions that combining different types of studies is necessary to get a complete picture about a pervasive application. We also highlight that conducting such field studies raises ethical issues which should be addressed when building a set of methodologies for evaluation pervasive applications. In the end, reaching a "safe spot" with any number of studies might be impossible. We therefore advocate that prototyping designs and evaluating them "in the wild" is not an option, but always needed. Thus, in our future work we will to evaluate the Tacticycle against the presented requirements. Besides validating the requirements we also intend to evaluate the validity of our requirements methods.

7. ACKNOWLEDGMENTS

The authors are grateful to the European Commission which co-funds the IP HaptiMap (FP7-ICT-224675). We thank the consortium partners for sharing their thoughts with us. We would also like to thank Georg Poppinga, who is the owner of the bicycle rental station where we issued our requirements questionnaire.

8. REFERENCES

[1] I. Aslan, M. Schwalm, J. Baus, A. Krüger, and T. Schwartz. Acquisition of spatial knowledge in location aware mobile pedestrian navigation systems. In *MobileHCI '06*, 2006.

[2] B. Brown and E. Laurier. *Map-based Mobile Services*, chapter Designing Electronic Maps: an Ethnographic Approach. Springer Berlin Heidelberg, 2004.

[3] B. Brown and E. Laurier. Maps and journeys: an ethnomethodological investigation. *Cartographica*, 4(3):17–33, 2005.

[4] K. Connelly, K. Siek, I. Mulder, S. Neely, G. Stevenson, and C. Kray. Evaluating pervasive and ubiquitous systems. *Pervasive Computing, IEEE*, 7(3):85–88, 2008.

[5] B. Gaver, T. Dunne, and E. Pacenti. Design: Cultural probes. *Interactions*, 6(1):21–29, 1999.

[6] J. F. Kelley. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.*, 2(1):26–41, 1984.

[7] J. Kjeldskov and C. Graham. A review of mobile hci research methods. In *MobileHCI '03*, 2003.

[8] A. J. May, T. Ross, S. H. Bayer, and M. J. Tarkiainen. Pedestrian navigation aids: information requirements and design implications. *Personal Ubiquitous Comput.*, 7(6):331–338, 2003.

[9] B. Poppinga, M. Pielot, and S. Boll. Tacticycle - a tactile display for supporting tourists on a bicycle trip. In *MobileHCI '09*, 2009.

[10] T. Sohn, K. A. Li, W. G. Griswold, and J. D. Hollan. A diary study of mobile information needs. In *CHI '08*, 2008.

Touring in a Living Lab: some methodological considerations

Abdallah El Ali
HCS - University of Amsterdam
Science Park 107
1098 XM Amsterdam
+31 20 525 6121
abdo[dot]elali[at]gmail[dot]com

Frank Nack
HCS - University of Amsterdam
Science Park 107
1098 XM Amsterdam
+31 20 525 6377
lastname[at]uva[dot]nl

ABSTRACT
This paper presents a number of thoughts about the use of the Living Lab methodology, which are based on experiences gained in the ongoing MOCATOUR project. The central topic of the MOCATOUR project is to establish novel computational methods to facilitate tourists with personalised and contextualised access to and annotation of cultural and historic information while they freely explore a city. We present a brief description of the scenario in which the Living Lab methodology is applied. We then outline the positive as well as problematic aspects of this research methodology for mobile environments with a focus on affective computing.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: *Theory and methods, User-centered design, Evaluation/methodology, Interaction styles.* H.5.1 [**Multimedia Information Systems**]: *Evaluation/methodology, Artificial, augmented, and virtual realities.*

General Terms
Experimentation, Human Factors, Theory.

Keywords
Experience capture and representation, ubiquitous, tourism, mobility, Living Lab, context.

1.INTRODUCTION
In the middle of June, Alex and Amy are tourists in Barcelona, Spain, walking towards Parc Güell. The sun is particularly cruel on this summer day, and despite Alex's reservation to wander about the city in the outdoors under constant threat of the scorching heat, he had promised Amy, a Modernist architecture enthusiast, that he would go with her to Parc Güell. Arriving at the entrance, Alex has now heard Amy muttering about the greatness of Antoni Gaudi, the architect who designed this park, at least half a dozen times. Annoyed by Amy's overzealous behavior, Alex's mobile device (having sensed the location and picked up Gaudi auditory speech cues) gently asks him whether he would like some information on Gaudi. By now only slightly interested in knowing more, Alex reluctantly accepts his mobile device's request to give him the general information on Gaudi it promised. After quickly skimming through the information, he learns that Gaudi belonged to the Art Nouveau movement, he gets acquainted with his Gothic phantasmal architectural works, what led to his artistic vision, and his embeddedness in this Catalonian city. Hardly thrilled by the visual, auditory and textual depictions his mobile device provided, Alex realizes Amy had wandered off. Upon a quick predefined gesture, his mobile device notifies him of Amy's location. Alex heads towards Amy, finally arriving at Gaudi's serpentine bench, providing a welcome opportunity for him to sit and steam his exhaustion. Finding the design of the bench particularly distasteful, Alex leaves on the bench a virtual experience trace using his mobile device of a textual annotation "cruel reality" overlaid on a quickly made sketch of a blue, poorly hand-drawn, surreal-looking sun (Figure 1). Not out of intrinsic dislike for Gaudi's work, but as a composite expression of frustration from the day's events. After breaking another sweat, Alex's frustration on this hot day marks the end of his company with Amy, the park, and the works of Gaudi that have now been deeply associated with negative affect – in search of a cooler, more indoors place.

Figure 1. Experience graffiti left by Alex at Gaudi's bench.

2.MOCATOUR: Graffiquity design and development challenges
The foregoing scenario is representative of work in progress under the MOCATOUR (Mobile Cultural Access for Tourists) project. In this project, the aim is to supply tourists with more personalized cultural and historic information access upon their interaction with cultural institutions in a city, such as an outdoor art exhibition or museum. This necessarily involves being able to adequately capture human experiences so that deeper insight is gained into what kind of system representations are necessary to enrich the tourist experience of being in a city. The project context is mobility and hence great emphasis is placed on interaction

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

using a mobile device both in an indoor as well as an outdoor setting. In order to sufficiently inform the design of a system and/or application that can augment the tourist experience, extensive testing and evaluation is required. In addition, existing as well as novel methods of human-mobile interaction need to be well understood. To increase understanding of interaction behavior, it is very important to understand 'natural' interaction outside the walls of the laboratory. It this aspect of research that poses an ongoing research challenge: how can we extract salient elements from human experience in a noisy, natural environment, especially when we still do not fully understand what experiences are, how they are formed, and why they occur as such [1]?

In the mentioned scenario, this roughly translates to being able to adequately explain 'why' the experience trace left was left as it is - i.e., what were the socio-cognitive-emotional aspects in that particular context that gave rise to such an expression. Currently, attempts are being made to design and develop the Graffiquity (Graffiti for Ubiquitous Tourist Experiences) application that allows for the capture and representation of an individual experience using the mobile device as a medium that allows the user to leave a virtual trace in the physical world but also to allow users to experience annotations by others. The application relies extensively on the graffiti metaphor, under the notion that graffiti is a form of self-expression that allows the capture of one's experience at a particular time and location.

The aim with Graffiquity is to allow tourists to leave an experiential trace out there in the world, making use of location as a canvas or wall by which the graffiti can be placed. In place of the spray can, the interactional medium in Graffiquity is the mobile device, which relies on gesture movement data and mobile button press (as can spray-head) for augmented reality annotation expression (free drawing, textual annotation, photo placement, etc.) at some location in a city. Lastly, just as in real-world graffiti you would have to go to the particular location to view the work of the graffiti artist, in Graffiquity the augmented reality graffiti can only be viewed through the eyes of the mobile device at the same location the graffiti was made. Ultimately, the aim is to establish a model that facilitates the capture of an experience (as graffiti), and once captured, how to extract the *relevant* contextual information from that experience at a concurrent or later point in time for presenting or sharing this particular experience.

3.MOCATOUR – The Living Lab embrace

Going back to the introductory scenario, it remains ambiguous to an observer of the graffiti what was meant by the experience annotation Alex had left: was the graffiti trace left directed at the art style, the weather conditions at the moment, Alex's internal cognitive, motor, and emotional life in that particular context, his social interaction with Gaudi-fanatic Amy, the crowd at that particular interval of time, or a mixture of each. It is quite likely that each of those reasons were behind Alex's experience trace; the problem is understanding the actual dosage and the respective interaction of each, the salient elements that are really 'about' the Gaudi bench, representationally sculpted at the place he left the experience graffiti. Given the difficulty in teasing out the causal factors of an experience trace, how can a mobile application reminiscent of the above scenario be evaluated so as to gain a deeper appreciation of the causal intricacies in human-mobile interaction? How can we, as designers and developers be sufficiently informed about the cognitive and especially emotional lives of people during interaction with a mobile application so that the application in question can seamlessly and ubiquitously be merged with their lives?

As a starting point, what is required are more human-centric evaluative methods, that can aid in identifying and the subsequent application of interaction methods in context-sensitive situations. However it is unclear how such methods fare under different contexts. Are these methods accurate and reliable within and between users, under different contexts, such as indoor and outdoor settings? The questions surrounding mobility research in general, and the MOCATOUR experience capture and representation aspects in particular, are at the heart of what the Living Labs paradigm [3,4] promises to deliver: the real-life testing and evaluation of humans interacting with technology in their natural behavioral, cognitive, and emotional context so as to adequately inform the development of ecologically valid prototypes, the capture and subsequent refinement of natural human-system interaction, and experimentally valid interaction ideation in the wild. This methodology is representative of a shift in mobile system and application development that attempts at transcending classical testing and evaluation under controlled laboratory conditions in order to better inform design decisions sprouted from what real-life users want, so that technology development becomes an intimate three-way dance between designers, developers, and users.

Such a working definition should clarify that the Living Lab approach[1] is not simply a reinvention of anthropological techniques of field testing, where phenomena of interest are *observed* in the wild. What the Living Lab approach seeks is not only to observe and measure *existing* behavior, but also to understand hitherto unknown forms of behavior made possible due to technology use in everyday living. In the present context, of particular interest is how mobile devices and application advances are accomplished given what the user (an active participant in the system design process) desires. This is of course a two-way iterative feedback process, with human-machine adaptation on the one hand, and the design of human-centric technologies on the other, whereby the ultimate aim is to reach a harmonious and invisible interactional equilibrium between human and machine, akin to Weiser's [2] vision of ubiquitous computing.

More specifically, the Living Lab as a research methodology provides a window by which we can observe the natural manifestation of different kinds of human experiential behavior, under the implicit assumption that the contextual cues and user intentions in interaction are truly representative of uncensored, seamless 'wilderness' behavior. In the context of the MOCATOUR project, this would allow research to more deeply tap into the cognitive, emotional, and behavioral resources called upon in leaving a virtual experience trace (through the Graffiquity application) at a particular location in physical space (context

[1] The careful reader would have noticed by now that the Living Lab paradigm has been rendered under different terminological constructs that appear to be synonymous yet pack subtle meaningful distinctions, some of which are: Living Lab as methodology, Living Lab as framework, Living Lab as approach, and Living Lab as paradigm. Here, we are using the term Living Lab in the context of a research methodology, and mean it to reflect a way of approaching user testing and evaluation.

sensing of frustration from the heat, mockery of Gaudi's surrealism, etc.). This is especially important since the perspective adopted in Graffiquity takes experiences as essentially dynamic in character (i.e., changing over time) that are contingent on interaction time line. Here then, a Living Lab approach provides a great advantage in that experiences can be tracked over long periods of time (longitudinal analyses), allowing for further understanding of the dynamic aspect of human experience. Yet this seemingly novel approach to testing and measurement of human-computer interaction is not without intrinsic limitations. These will be discussed primarily in the context of novel experience capture and representation when mobile.

4.MOCATOUR – The Living Lab doubts

A general limitation of the Living Lab approach concerns the stage of system or application development at which such 'wild' testing can inform design. During the early development stages, there should be greater reliance on informed and explicit evaluation methods such as user-interviews, questionnaires, partially constrained free-recall feedback, etc. This is primarily due to the near-limitless magnitude of the application design space – only through extensive user-based interaction knowledge acquisition can such a design decision space be manageable. In MOCATOUR this resulted in the decision to making use of field-study methods (administering questionnaires and observation), which inform about the user wishes. This approach is hypothesized to reveal that users may have problems readily grasping new interaction methods and as a result exemplar interactions may be necessary, through explicit application demonstration or usage guidance. This however poses a trade-off between relying extensively on exemplars to guide interaction with new mobile technology on the one hand, and the risk of losing interesting insights that would not have been cut off had it not been for explicit interference in the interaction process that was initially setup to ensure desired human-mobile interaction behavior.

In short, the design space is too big in early stages of development, and the most informed way of carving that space into manageable and feasible decisions is through explicit and controlled testing of human-technology interaction. Within the MOCATOUR context, this would mean the subtle integration of human feedback in the application at hand; of particular concern is how to ‘calmly’ embed this feedback request within the Graffiquity application. Here, a ‘silent’ data collection mechanism should be implemented that collects for example the choice of graffiti color or spray stroke for long periods of time; this kind of information can prove to be highly valuable especially in revealing undiscovered correlations between emotio-cognitive and behavioral patterns under different modalities (e.g., gesture-speed in graffiti drawing and the respective correlations with fineness or coarseness of spray diameter).

At later stages however, the kind of evaluation required to truly assess the application functionality and user-experience should be invisible from the perspective of the user, or minimally simple yet effective feedback request in the form of for example quick 'yes or no' questions presented after the completion of some task. This kind of *informed* data acquisition has two requirements: first, it should be long-term and continuous so as to truly arrive at subtle human interactional elements that develop over (system and/or application usage) time. Second, this data acquisition should be 'invisible' to the user, so that the stream of human behavior is not suddenly a manufactured product of (conscious) bias that is not representative of raw human-machine interaction.

At a more pragmatic level, the long-term acquisition of data from actual users under a living lab approach faces a number of limiting factors in the kinds of evaluative methods that can be employed. Consider for example system or application adaptation - ideally, a user can be tested over a long period of time, which reveals how well this system integrates into a particular user or type of user profile and lifestyle. The problem then is that if different interfaces are to be evaluated, a clear bias is manifest in evaluation within one person after this person has used a particular system for quite some time. A competing alternative will by necessity be less intuitive, user-friendly, useful, etc. by virtue of adaptation to a similar system at an earlier point in time (resulting in adaptation conflict). One way to circumvent such an inherent limitation is to measure the difference in interface usage by two different people for the same period of time; yet this approach suffers from reliability breakdown[2] and raises the question of how do we know that we are in fact measuring the same thing in these two people, given that they are let loose in their natural environment and are allowed to exhibit the entire range of human behavior?

This shortcoming also strangles classical laboratory experimentation, albeit in a different form: under general laboratory conditions, a user would be provided with two competing interfaces (counterbalanced among recruits) and extensively asked for feedback after brief usage of each. While users do end up evaluating two competing interfaces, they do so for a short period of time; this is clearly less revealing than longitudinal analyses that can inform us more accurately about the human-system interaction lifetime of the interface in question. Yet, the test in the laboratory offers the advantageous option to test desired ‘micro’ features for a particular user group that would otherwise be cluttered with environmental noise. This is why in MOCATOUR, for certain application interface elements (e.g., ratings represented on a 10-point scale or 5-point scale, using stars versus smileys for data rating visualization, etc.) the testing should be confined to the laboratory. Nevertheless, for general experiential dimensions that strongly require long-term observation and analysis, the testing should be taken out in the world.

Lastly, testing under a Living Lab methodology raises concern when considering the demarcation of events that the user, as a primary actor in some action or actions, is part of. The fact that the user is allowed to freely use the application in question risks 'over measurement' – what is required is the encapsulation of events into a few small measurable units that are immune from the mostly unpredictable character of unsupervised human interaction. The notion of supervision here is paramount in permitting the measurement of the interaction phenomena in question, without of which extraction of meaning from continuous flux of data proceeds in an ad-hoc manner, risking efforts at extracting meaning from the collected data to be lost in interpretation. This is another way of saying that evaluation of

[2] The primary objection raised here does not concern randomized controlled trial testing per se, but rather the practice of it under a Living Lab where the object of testing is a mobile application and its experiential usage which *may* differ across *uncontrolled* users.

human-machine interaction is orders of magnitude more difficult to accurately and reliably measure than *user*-machine interaction, in which the latter is a constrained subset more favorable to controlled laboratory conditions. Without at least minimal control exerted on the testing conditions during evaluation at least in the early stages of application development, there should be no reason why the user model that is being built up from such evaluations is trustworthy enough to inform our cognitive and emotional apparatus and knowledge about such apparatus to inform user-tailored system and application design (cf., attentional breakdown in field testing [5,6] and application neglect) – such should be the message from a more human-centric approach to design and development. This is also evident in (partially) automated experience capture applications (e.g., SocioXensor [7] and GREATDANE [8]), where there is to some extent interjectional user prodding. For MOCATOUR, this requires systematic reflection and analysis over which interface aspects and their corresponding interactional manifestations that arise should be left inside the walls of the experimental laboratory, and which of them let loose in the city. To be more specific, not all possible data streams should be stored that come from using Graffiquity to lay an experience trace, precisely because we do not yet fully understand what aspects of human experience are truly relevant to increase our understanding of affect, cognition, and interactional behavior on the one hand, and towards more informed experience-tailored application design on the other.

5.CONCLUSIONS

It has been highlighted here that the Living Lab paradigm represents a research methodology that allows us to observe the range of human experiential behavior in interaction with a system in a natural, non-artificial and non-obtrusive manner. This was grounded in current and potential human-mobile interaction in particular. It was argued that in the context of the MOCATOUR project, which aims at refining existing and establishing novel methods of mobile interaction behavior that allows for more personalized cultural/historic information access behavior for tourists in a city, can strongly benefit from such an approach to user testing and evaluation. In particular, the dynamic quality of experiences highlights the need for long-term 'out in the wild' observation and measurement. These methodological features were further grounded in the Graffiquity mobile application currently being designed and developed. This application allows tourists in a city to leave experience traces both outdoors and indoors using a gesture-based interface where the mobile device is a metaphorical surrogate for a spray can. These experience traces can then be shared, given the right representation, with others who revisit the same location the graffiti was made.

Despite some of the limitations that the MOCATOUR project faces in adoption of a Living Lab methodology – namely, early design space requires scaling down, interface adaptation bias and conflict, over measurement and ad hoc interpretation – the Living Lab method can nevertheless strongly aid in gaining deeper insights about experience capture and representation, especially with regard to the observation and measurement of affective, cognitive, and interactional behavior taking place in a natural setting over time. Simply put, such an approach provides us with greater access to raw and uncut human experiences. Thus, it is especially fruitful in cases where the objects of measurement are both identifiable and amenable to measurement without observer and interference bias.

From the opening scenario, it becomes clear that human experience factors such as Alex's cognitive, affective, and behavioral makeup at that particular day due to a set of primary causal factors (scorching sun, tiredness from wandering about a city, sweating and dehydration, frustration with Amy's constant rambling, etc.) can influence the perceptual and affective judgment of architectural works belonging to great figures like Gaudi. Here, it can be said that there is more to perceiving and interpreting context than meets the (artificial) senses. For the MOCATOUR project, especially the affective computing component provides an ongoing challenge for arriving at a scientifically well-validated user model that can adequately predict human-mobile behavior in a range of contexts (indoors, city outdoors). In turn, this also presents the Living Lab approach with some methodological considerations that require further thought, especially in accommodating the capture and representation of intangible and dynamic phenomena such as human experiences.

6.ACKNOWLEDGMENTS

The MOCATOUR project is funded by the Dutch Ministry of Economic Affairs and Amsterdam Topstad. The authors also wish to express gratitude to Amsterdam Innovation Motor for their support. Also special thanks to Patrick Mast and Ricus Smid for their implementation efforts.

7.REFERENCES

[1] Frank Nack. Capturing experience - a matter of contextualising events. ACM Communications, 2003.

[2] G. D. Abowd, E. D. Mynatt, and T. Rodden. The human experience. IEEE Pervasive Computing, 1(1):48–57, 2002.

[3] M. P. de Leon, S. Balasubramaniam, and W. Donnelly. Creating a distributed mobile networking testbed environment-through the living labs approach. In Proceeding on Testbeds and Research Infrastructures for the Development of Networks and Communities, pages 134–139. 2nd International IEEE/Create-Net Conference, 2006.

[4] M. Eriksson, V. Niitamo, and S. Kulkki. State-of-the-art in utilizing living labs approach to user-centric ict innovation – a european approach.Technical report, Luleå University of Technology, Sweden, 2005.

[5] Antti Oulasvirta. The fragmentation of attention in mobile interaction, and what to do with it. Interactions, 12(6):16–18, 2005.

[6] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 919–928. Conference on Human Factors in Computing Systems, ACM New York, USA, 2005.

[7] G. H. Ter Hofte. What's that hot thing in my pocket? socioxensor, a smartphone data collector. In Proceedings of the e-Social Science 2007, pages 7–9. Third International Conference on e-Social Science, Ann Arbor, MI, USA, 2007.

[8] K. Jensen. Evaluating mobile and ubiquitous applications in the field by automated capture and analysis of reality traces. In Proceedings of 8th Danish HCI Research Symposium, 2008.

Evaluating a Mobile Multimedia Application in Field Trials: the cost-benefit of self-report methods

Michael Leitner[1], Peter Wolkerstorfer[1], Arjan Geven[1], Norman Höller[1], Manfred Tscheligi[1,2]

(1) CURE, Centre for Usability Research & Engineering
Hauffgasse 3-5,
1110 Vienna
++34.1.743.54.51
{michael.leitner;Wolkerstorfer;geven;hoeller;tscheligi}@cure.at

(2) ICT&S Center
University of Salzburg
Sigmund-Haffner Gasse 18 5020 Salzburg
+43.662.8044.4811
manfred.tscheligi@sbg.ac.at

ABSTRACT

In this paper we discuss future perspectives and upcoming methodical challenges in the realm of Living-Labs, with emphasis on methodological issues in user experience research and self-report methods. In detail, we report some of the findings of a three month field trial with 100 users evaluating a mobile multimedia application. This application provides users with video content in form of clips consumable with the mobile device's internet browsers. An "online diary" and an adapted "experience sampling method (ESM)" were used to get overall feedback on the users' experience. Comparing the two methods, the less complex "online diaries" were able to reveal nearly the same findings as the more complex ESM. Analysis still needs to go in further detail to better prove this argument, but these preliminary results let us think about the cost-benefit of chosen methods in living-labs.

Categories and Subject Descriptors

K.4 COMPUTERS AND SOCIETY

General Terms

Human Factors

Keywords

Living Labs, Experience Sampling Method (ESM), Online Diaries

1.INTRODUCTION

In this position paper we want to lay emphasis on methods used in Living Labs setups and to discuss the possible impacts of probing and testing in the field for HCI studies. In detail we want to raise a discussion on cost-benefit factors of user experience research methods in Living-Lab setups. As an example we discuss a field trial recently done in the area of Mobile Multimedia applications. In this short report we discuss two self-report methods used (online diaries vs. experience sampling method) and highlight some of the resulting methodical questions raised in the setup and the analysis phase.

1.1 Methodology in Living-Labs: Challenges and Research Streams

The value and challenges of Living Labs for a HCI driven approach and from a methodological point of view lie in the following three points:

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

1) Support User-Driven Innovation in the early stages of idea generation. One of the main challenges is to reveal the innovative potential of the user. Even if we get them involved in the process, this doesn't guarantee that our participants go beyond being "criticizers" and become "creators". The challenge is to choose the right strategy and the right methods to involve them into the idea generation phase and make them actively participate to the process throughout the project, in the focus group and during the trial phases (including the use of feedback tools like ESMs or diaries, cultural probes, or others). A cost-benefit analysis may help to choose which methods to use, in order to get feedback what researchers might expect from the methods and what they have to invest to get the data available for analysis.

2) Iterative prototype evaluations in Living Labs: Nowadays the evaluation and assessment of software applications and their purposes become multifaceted and therefore more challenging. We see a main challenge in choosing the right methods for evaluation that bring valuable results instead of collecting huge amount of data without any particular meaning. Researches have to select the most appropriate ones from different methods and evaluation tools in order to stick to the questions and hypothesis defined. In our point of view there is still a need for further knowledge to solve methodical issues, especially in the selection of qualitative and quantitative methods at different prototype and project stages. Likewise, the evaluation and use of technology in particular contexts still needs further research.

Due to the nature of Living-Labs the infrastructure for rapid deployment of prototypes into the real user world is a challenge. Different setups provide different pros and cons (e.g. web-based deployment can function rapidly; TV-set-top-boxes requires infrastructure support; mobile deployment can be performed over-the-air; physical equipment has to be replaced physically, which not possible to do this in large quantities). To define best ways to face those structural problems a close cooperation between existing Living-Labs have to be set-up.

3) Getting the user involved and motivated: Living-Lab coordinators and researches have to find the right people, participants and testimonials for the planed Living-Lab trials. The best tools and best solutions will not be used within trials if the participant is not motivated. Hence, we see a strong challenge in finding the right way of incentive-giving to get the users involved. Different kinds of (persuasive) strategies are needed to be followed and tested that motivate, involve and re-involve the users to participate in the evaluation of technology.

1.2 Comparison of methods in field trial use: The cost-benefit of methods

In the rest of the paper we report findings on a field trial carried out to evaluate "MAMTAM", a Mobile Multimedia platform providing video content to users via a mobile streaming channel. The aim of the trial was to explore the overall users' experience interacting with the multimedia platform and to find out significant interaction and consumption patterns related to different contexts and contents. In this paper we want to concentrate on the methods that we used to gather the users' experiences; to be precise, User Diaries and ESM [8]. The first results we got by analyzing field trial data, indicates that it is quite equal using desktop questionnaires or mobile questionnaires even as the former could be considered as a media break. Interpreting these results we ask ourselves about the cost-benefit of these methods, as the online diary was much easier to setup and further caused less ongoing effort during the trails.

2.RELATED WORK

To gather user experience many methods has been used in HCI. An overview on emerging methods can be found in Hagen et al. [3]. These methods are collected in three groups:

1.) Mediated Data Collection: Self-reporting, diaries, probes, Use/data logs, Video Observation.
2.) Simulations and Enactments: Lab tests, scenarios, heuristics, prototypes, NASA TLX, emulators, simulators, Scenarios, role-playing, work shopping, prototyping, storyboarding.
3.) Combinations: existing methods, and/or mediated data collection and/or simulations and enactments are combined to allow access to complementary data." [3].

Experience Sampling Method is a method that can be used for "frequency and patterning of daily activity, social interaction, and changes in location" among others [8] One of the biggest problems of this method is the time investment of participants [10]. An example to cover this problem can be found in Hsieh et al.[9].

In 2007 Froehlich et al. [4] introduced "my Experience" "a system for capturing both objective and subjective in-situ data on mobile computing activities". The tool uses mixed methods and presented a combined technique of "passive logging of device usage, user context, and environmental sensor readings", and "active context-triggered user experience sampling to collect in situ, subjective user feedback." To cover some drawbacks of existing in situ self-report methods it increases user attention and provides extensible framework for combining automatic logging and user sampling on a participant's personal device. [4]. The Context-Aware Experience Sampling (CAES) [7] tool incorporates sensor data to trigger self-report surveys at specific moments of interest. But it lacks the ability to avoid prompting during inopportune moments and it does not offer the flexibility of specifying dynamic trigger conditions and generic actions [4].

Diary studies have a long history among these methods due to its natural relation with the user. New generation of diary studies using mobile phones has several examples, e.g. voice-mail diary studies [1] is a method where users observe each other by taking photos and explain these with notes. Isomursu et al. [2] evaluated a technique called „experience clip" where two users (e.g.: friends) were given a PDA and a camera phone. While using the PDA, observer make clips of him/her and later they were encouraged to storytelling over the process. In the control version observer was not a friend but a researcher. This study provided rich data on emotions, feelings and experiences.

Mobile diaries [5] are consisting of "a diary pack with prompts, instructions and digital tools for recording reports". The combination of video, camera, text and audio enables the collection of rich data and means that people were able to adapt the reporting method to modes that most suited their personal expression and technology preferences. A similar idea without real time exchange of recorded media can also be found in Carter and Mankoff [6]. They reported the following pipeline for these kind of studies according to their results to optimize the method: "1) A participant takes a photo 2) The participant annotates the photo with an audio recording 3) The participant uses a tool to log the photo and audio and add more annotations 4) The researcher provides feedback about the captured data 5) The researcher holds an elicitation interview with the participant using the captured media as prompts. " [6]. Carter and Mankoff revealed a need for situated annotation of captured event in elicitation studies. They also found that the best approach to feedback studies may be to pair media capture with structured, question-and-answer based annotations. Their studies also revealed the usefulness of different media in different situations [6].

According to the state of the art in diary studies and similar data gathering methods like ESM in HCI, optimization possibilities of these methods are clear. These are getting real time feedback from users without interrupting them much, categorizing data gathered by these methods to have a better insight about usage patterns, enlarging the sample size by using semi-automatic mixed methods next to user feedback.

3.TRIAL & METHODS

3.1 The Study Object: MAMTAM

MAMTAM is a mobile multimedia platform foreseen to be used on a mobile device. The application uses the mobile internet browser installed on the device. The platform provides Video content (news, entertainment and programs), which is content provided by a TV broadcasting station. Via MAMTAM the users were able to browse the broadcasting stations program date back to the beginning of the field trials. This means that during the three-month trials period the participants could access all programs broadcasted in this time. By this setup a real usage scenario was build, especially for news and up-to date information. Apart form content, MAMTAM provides interactive components, as users are able to post comments and ratings of content they viewed. As special feature the application is able to apply textual search within the video content.

3.2Trial Setup

In the field trials 100 participants took part, each of them was equipped with a mobile tool (Nokia 6210) with all necessary software already installed to view the video content. The users got some introductive information on the devices and the MAMTAM interface. The field trial was followed by a focus group with some selected participants to gather more qualitative experiences directly reported.

Graph 1: The MAMTAM interface on two NOKIA phones.

As mentioned, the main goals of the study were covering aspects on technical issues (performance, scalability), usage numbers and user habits. Beside these issues the following methods were used to gather related user experience in the study. These methods are "online diaries" and a slightly adapted "user experience sampling (ESM)" method.

3.3 Online Diaries vs. Experience Sampling

The study setup assessed 50 participants using the "online diary" and the other 50 participants the "experience sampling method (ESM)" [8]. Both methods included the same questions, which were:

- How did you like the last clip? (likert scale [1-5] good – bad)
- How did you experience the lighting condition watching the clip? (likert scale [1-5] light – dark)
- How was the noise environment? (likert scale [1-5] noisy – silent)
- Did you watch in company, if yes with how many people? (1- 5, 6-11, 12-20, more than 20, alone)
- Did you feel concentrated or distracted? (two options [concentrated/distracted])
- Were you on the move? (yes/no)

Online Diary: The online diary was completely web based and designed for a desktop browser. "Online diary" participants received an e-mail once a week serving as a reminder for the online questionnaire. The participants were supposed to fill in the questionnaire using their desktop browsers and – by the nature of the desktop online diary – they were mainly at home or in their offices filling in the diary. In fact this is a reportable media break as participants were watching the videos on a mobile and were filling the questionnaires on a desktop. Further, it is supposed that due to this setup there is a certain time span between filling in the questionnaire and the consumption of the videos.

Adapted Experience Sampling Method (ESM): In order to implement a suitable technical solution for the ESM we scanned different tools like the ESP, the experience Sampling Program (www.experience-sampling.org), the "My Experience Tool (myexperience.sourceforge.net) and "Momento Start" (http://momento.sourceforge.net/documentation.html). These tools are powerful software applications, but often are dependent on a particular operating system or need to be installed on the phone itself. As for some technical an organizational reasons we could not use any of these tools and we finally had to implement an ESM tool by ourselves. We then build a questionnaire that was filled in by using the mobile browser of the field trial device. However, there was no direct linking to a video (= answering a questionnaire after watching a video). Because of that we call the method "adapted" as classic ESM foresees this direct linking between video and experience. In contrast our trials participants were triggered two times in a day by SMS containing a link to the questionnaire sent to the participants. The questionnaire needed to be filled in by the users using the mobile browser. We suppose that by the number of triggers the participants are more likely to report on their experiences closer to the watching session.

Summarizing, the main differences of the methods are displayed in the following table:

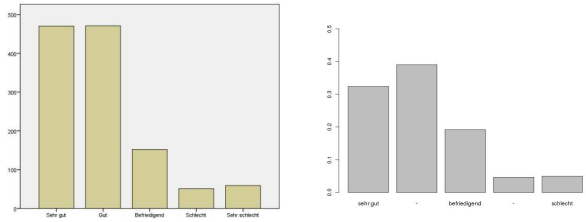
Table 1: Differences between methods Online Diary and ESM:

Online Diary	Adapted ESM
Triggered once a week	Triggered twice a day
Triggered via e-mail	Triggered via SMS Link
Fill in using desktop browsers	Fill in using mobile browser

An obvious trade-off of for both methods is that there is no 1:1 relation between experience and a particular video. In contrast we might assume that participants will rather evaluate the "whole experience" that they have with the content provided by the MAMTAM application. Analyzing the interdependence between triggers and answering rates/times shall also be a future challenge.

3.4 Comparing Results

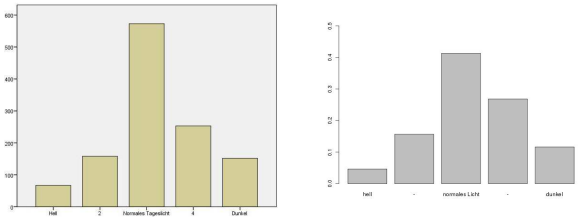
As the data material of our study is yet not been analyzed in detail only some pre-example results shall be discussed in this section (Graph 2, 3 and 4). However, the results show that there is little difference to report comparing the two users groups: Those using the online diary and those using the adapted ESM.



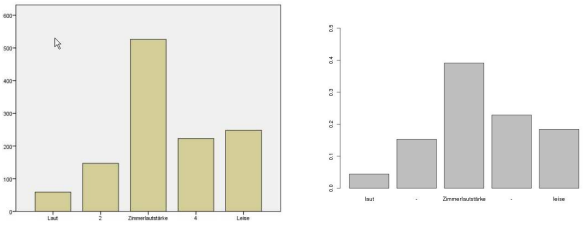
Graph 2: How Did you find the last clip you saw (likert scale [1-5] good – bad / left online Diary in absolute numbers; right adapted ESM results normalized)

Graph 2 shows that most of the people like the clips they selected. This finding is not very surprising as we might assume that people select those clips that they like best according to topic and description. The comparison of online diary and adapted ESM nearly show the same picture. There is a little difference observable. The very left column indicating that in the online diary participants were more likely to rate the highest value for the seen clips.

Graph 3 shows that people experience the lighting condition quite normal (normal lighting conditions like daylight) with a tendency to too dark surroundings.



Graph 3: How did you experience the lighting condition watching the clip (likert scale [1-5] light – dark / left online Diary in absolute numbers; right adapted ESM results normalized)



Graph 4: How was the noise environment watching the clip (likert scale [1-5] noisy – silent / left online Diary in absolute numbers; right adapted ESM results normalized)

Again, nearly the same picture is shown at the question of noise environment. Participants by trend rate the environment in which they watched videos as "silent". Some minor differences, again, are observable at the right side where "diary users" rated the surrounding more silent as adapted ESM users did.

4. CONCLUSION

Overall these findings show that the analysis of cost-benefit factors could help researches choose the right methods. In detail, our discussion on self-report user experience research methods shows that for the particular case simple methods might be better applicable causing less effort. At the same time they are less obtrusive and less interrupting. Our example comparing "online diaries" vs. and adapted "experience sampling method (ESM)" showed that triggering users once a week (online questionnaire) leads to nearly the same findings as triggering twice a day. Even the media break (watching on mobile devices and filling in questionnaires using desktop browsers) had little or no effect. So, the cost-benefit question arises asking if researches should spend efforts on complex Experience Sampling Methods that are harder to implement and to adapt to the mobile devices – at least this is the impression that we got from our studies and attempts to set-up the ESM methods in reference to set-up effort. The findings reported show that as well an "online diary" can reveal the same results (as long no in-situ information like photos, sound capturing, etc. is necessary) and all major trends are traceable in both methods. In any case, the data gathered requires further analysis in order to prove this argument into more detail, investigating possibly influencing variables.

However, the question is which methods should be used in order to gain optimal cost-benefit. The results presented indicate that a smoother integration with less triggering might provide enough information to draw conclusions. In general, this is no critique or downplay of the ESM method, as value of this method for particular cases is of immense importance. It is clear that these implications are not valid for 1:1 experience sampling method (meaning that the users rates the experience of a particular spot or video – in this case "online diaries" are not applicable).

5. ACKNOWLEDGMENTS

The research herein is partially conducted within the competence network Softnet Austria (www.soft-net.at) and funded by the Austrian Federal Ministry of Economics (bm:wa), the province of Styria, the Steirische Wirtschaftsförderungsgesellschaft mbH. (SFG), and the city of Vienna in terms of the center for innovation and technology (ZIT). The reported activities and results are supported through the Living-Lab Austria initiative.

6. REFERENCES

[1] Palen, L. and Salzman, M. Voice-Mail Diary Studies for Naturalistic Data Capture under Mobile Conditions. In: In Proceedings of ACM CSCW'02, (2002), 87-95.

[2] Isomursu, M.; Kuutti, K. & Väinämö, S.Experience clip: method for user participation and evaluation of mobile concepts *PDC 04: Proceedings of the eighth conference on Participatory design, ACM*, **2004**, 83-92

[3] Hagen, P.; Robertson, T.; Kan, M. & Sadler, K. Emerging research methods for understanding mobile technology use OZCHI '05: Proceedings of the 17th Australia conference on Computer-Human Interaction, Computer-Human Interaction Special Interest Group (CHISIG) of Australia, 2005, 1-10

[4] Froehlich, J.; Chen, M. Y.; Consolvo, S.; Harrison, B. & Landay, J. A. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones MobiSys '07: Proceedings of the 5th international conference on Mobile systems, applications and services, ACM, 2007, 57-70

[5] Hagen, P.; Robertson, T. & Gravina, D. Engaging stakeholders: mobile diaries for social designDUX '07: Proceedings of the 2007 conference on Designing for User eXperiences, ACM, 2007, 2-14

[6] Carter, S. and Mankoff, J., When participants do the capturing: the role of media in diary studies. Proceedings of CHI 2005, (Portland, Oregon, USA, 2005), ACM Press, 899 - 908.

[7] Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., and Bao, L. A context-aware experience sampling tool. In CHI '03 Extended Abstracts (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003). CHI '03. ACM Press, New York, NY, 972-973.

[8] Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. New Directions for Methodology of Social and Behavioral Science, 15, 41-56.

[9] Hsieh, G.; Li, I.; Dey, A.; Forlizzi, J. & Hudson, S. E. Using visualizations to increase compliance in experience sampling UbiComp '08: Proceedings of the 10th international conference on Ubiquitous computing, ACM, 2008, 164-167

[10] Scollon, C., Kim-Prieto, C., and Diener, E. Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. Journal of Happiness Studies 4, 2003, pp. 5-34.

Features for the future Experience Sampling Tool

Vassilis-Javed Khan
Eindhoven University of Technology
Den Dolech 2
5612AZ Eindhoven
+31 40 247 5256
v.j.khan@tue.nl

Panos Markopoulos
Eindhoven University of Technology
Den Dolech 2
5612AZ Eindhoven
p.markopoulos@tue.nl

Berry Eggen
Eindhoven University of Technology
Den Dolech 2
5612AZ Eindhoven
j.h.eggen@tue.nl

ABSTRACT
Based on our experiences with Reconexp, a distributed application which partly runs on a mobile device and partly on a website, and a review of tools developed to help researchers survey user attitudes, experiences and requirements in field studies we present a list of requirements for future experience sampling tools.

Categories and Subject Descriptors
H.4.m [Information Systems]: Information Systems – *Miscellaneous.*

General Terms
Human Factors

Keywords
Experience Sampling, Day Reconstruction, Diary Studies

1. INTRODUCTION
Existing research and evaluation methods have been shaped to address the evaluation of task-oriented interaction, usually contained within a short time span. Extending characterizations and evaluation methods to address user experiences as they occur in context reflecting social interactions between several people and various environmental and technical contingencies, requires scaling up the sampling of data in frequency, duration and in the richness of the records made.

The objective of the evaluation has also changed significantly. Transcending usability, evaluations of ambient applications need to consider higher level aspects of user experiences and user needs relating to persuasion, fun, engagement, trust, etc. Evaluation practices must be able to account for, capture and investigate the variability of contexts described above allowing experimenters to manipulate and control those environments or, when working in the field, to capture sufficient contextual information them.

Contextualized methods of data collection should allow reports of attitudes, opinions, or appraisals of subjective experiences to be captured close to the moment that a particular experience occurs, in the context that events and activities unfold. Also, such sampling of user attitudes can occur repeatedly over time, allowing the study of behaviors and experiences over medium or long periods of time, to uncover temporal patterns or to examine patterns of use over time.

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

One well established method that addresses these requirements to a large extent is the diary method whereby informants are asked to keep a journal or a log, where they record events, activities and experiences regularly over a specified period of time. In traditional diary studies informants record data, usually in writing, but often combining or even replacing written records with other recording media, see for example [2].

In diary studies, the initiative for capturing information is left completely up to the informants who have to remember and take the initiative to report in their diaries. This may be detrimental to the quality of the data collected for several reasons. Informants may forget to enter information in diaries, or entries may be made at moments that they have the time and appetite to do so, rather than the ones of interest to the researcher. This can lead to loss of data and systematic response biases.

For these reasons, the Experience Sampling Method (ESM) [7] is gaining ground in human-computer interaction studies for understanding human behavior to design better products and services and for studying use in the field. The ESM is a quasi-naturalistic method that involves signaling questions at informants repeatedly throughout the sampling period. For example, informants may be given a pager or even another notification device through which they may be reminded to fill in a set of questions in a diary. To allow for flexible sampling and surveying approaches, researchers have turned to the use of handheld computing devices (Smartphones or PDA's), that participants are required to carry through the study period and through which the question-asking protocol is applied.

The ESM method is gaining in popularity in the field of human-computer interaction. Consolvo and Walker [4] have used the ESM for evaluating an Intel Research system called Personal Server. Hudson et al. [8] have used the ESM to explore attitudes about availability of managers at IBM Research. Froehlich et al. [6] used ESM to investigate the relationship between explicit place ratings and implicit aspects of travel such as visit frequency.

2. Challenges and Pitfalls of ESM
Although very useful in prompting the reporting of subjective experiences over time and in context, ESM also has shortcomings such as interrupting the subject at inappropriate moments, the onus of repeatedly answering the same or similar questions, the difficulty of entering self-report data in some social and physical contexts, the need to sample more frequently when some activities take place and less frequently otherwise, etc. Moreover, ESM is

expensive; it puts high burden on participants, and provides little information about uncommon or brief events, which are rarely sampled [11]. These problems lead to loss of data, inaccurate reporting and nuisance to participants. Current research in this field is concerned with developing methodological innovations and corresponding tools to remedy these shortcomings.

Loss of data seems to be a major problem. Froehlich et al. [6] report completion rate of 80.5% similar to Consolvo and Walker [4] who report an 80% completion rate (on average 56 out of 70) with as low as 28.5% (20 out of 70). Even worse, these numbers are silent regarding the significance of the data lost. It is reasonable to assume that the data loss occurs when people are busy or engaged in social or professional activities. Depending on the goals of the investigator, these might be precisely the situations that researchers are interested in studying.

On the other hand, the unique advantage of ESM is its ability to capture daily life as it is directly perceived from one moment to the next [5], providing a rich set of data to researchers.

3. Methodological variations to address ESM shortcomings

An alternative to ESM, proposed by Kahneman et al. [11] is the Day Reconstruction Method (DRM), which was designed to assess how people experience their various activities and settings of their lives. Subjects in this case are asked to record a detailed diary of activities and events during one day. These do not relate directly to the focus of inquiry of the researcher, which is not disclosed to them at this point, but are meant as a memory aid, a kind of scaffolding, to allow informants to recall and reconstruct the experiences and feelings of the last day during a follow up interview the day after. This is an in-depth semi-structured interview, during which the researcher probes regarding experiences and feelings that the investigation aims to explore. Kahneman produced strong evidence regarding the efficacy of this method; however DRM suffers from low accuracy regarding factual aspects of the experience (e.g., time when events occur, factual details and environmental aspects which are easy to capture with ESM), similar to those of diary studies reported above.

By its nature, DRM is appropriate for short studies. Its efficacy for providing rich and contextualized accounts of user emotions in the last 24 hours is achieved by means of an elaborate interview which is not meant to be carried out repeatedly in a study and is practically difficult to repeat over longer sampling periods. Field studies in the domain of ambient intelligence typically exceed two weeks in duration, reaching some times even half a year. For such cases, DRM can help understand only a small fraction of the activities and experiences of informants, missing out a lot of information regarding the context in which it takes place.

A combination of ESM and DRM has the potential to compensate for their complementary weaknesses. Such a combination is the Experience Sampling and Reconstruction Method (ESRM) introduced below. Following this hybrid method participants follow procedure as with ESM through which a partially complete (given the data loss issues discussed above). Furthermore, at regular intervals (e.g., daily) participants are required to complete, elaborate and even reflect on the reported experiences using the partially complete ESM log as a scaffolding. Reconstruction is done partly by completing gaps in the data collection of the day and partly by elaborating and reflecting on this recent data. Crucially, this stage is still lightweight enough that it can be repeated daily for some. The queries which are missed during the sampling day can then be recovered through an interface with the log.

4. Reconexp

The "Reconexp" ("reconstructing experience") tool [12] was developed to support the ESRM method. It is a distributed application partly running on a mobile phone (from now on mentioned as "device") and partly on a website.

The procedure is described below from the perspective of the participant.

The exact procedure for an informant is as follows (Figure 1):

- Personalization of experience sampling protocol
- Combined ESM and DRM data collection
- Debriefing interview

These steps are discussed in detail below.

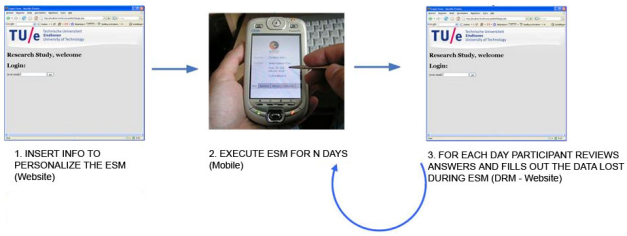


Figure 1: The steps of the ESRM method. First, participants insert information to personalize the experience sampling step. Then the experience sampling is executed using a hand-held device and for each experience sampling day participants are asked to access the web application to review their answer and fill out the data which were lost during the sampling day

4.1 Personalization step

The motivation for having the personalization step is to subsequently minimize the time and effort needed for participants to respond to the mobile device when prompted to do so. Reducing the effort required is expected to help prevent data loss but also encourage accurate reporting.

Personalization can mean a few things: adjusting the timing of the sampling procedure, personalizing defaults and choice items offered to the user and thus reducing the effort for entering data. Personalization has several advantages. From a research point of view it enables the researcher to compare the answers given for the personalization step with the experience sampling itself. This comparison can be interesting to analyze since it relates participants' expectations with what actually happens in situ. It also allows more intensive text entry to be done on the website rather than in situ with a pen (an inherently slower and more cumbersome way of entering data). The information thus obtained is used to populate list-boxes offering choices to participants at sampling time. Personalization helps also set some parameters for the sampling protocol thus allowing better timing for the sampling events.

4.2 Sampling and reconstruction step

During the day the device prompts participants to enter information as in a standard experience sampling approach. As with all computer supported experience sampling, prompting can be programmatically controlled to occur in regular intervals, at random moments or when some conditions regarding the context and the informant activity have been specified.

The informant can respond by selecting between choices of items describing his/her activity, context or emotions, or even by free text entry to answer more open questions.

The information entered on the handheld device is stored on the online database and is available for retrieval and review directly.

The reconstruction step should happen as close as possible to the collection of data through experience sampling, e.g., within 24 hours. It requires the visualization of the experience sampling logs, the ability to edit them and provide extra information. The interaction requirements for the tool support are different than those applicable for experience sampling: whereas mobility and speed of entry of some brief information is the priority during experience sampling, it is now required to have a good visualization, and efficient ways of editing and inputting text, e.g., using a desktop computer. Of course, one could also allow revision and editing of answers using a small handheld device also for the reconstruction, but this could be at the expense of obtaining richer and more extensive descriptions from informants.

Appropriate visualization of earlier answers can help informants reconstruct their experiences and provide richer descriptions/information about them. Also important, such visualization can help researchers track the progress of the study, opening up the possibility to adapt the sampling protocol while the study unfolds. Researchers can, for example, provide additional incentives or further instructions if they notice that a particular participant is not responding to the daily queries. It also enables researchers to prepare questions for debriefing interviews while the sampling is still unfolding.

4.3 Debriefing interview step

During the debriefing interview participants are asked to reflect upon their opinions to the queries posed during the sampling period. The interview can be utilized to understand in depth the reasons participants responded to the research questions. This step becomes even more useful if the logs of answers are reviewed before approaching each participant. For example, researchers might spot in the log a pattern in the way a participant had answered to a particular question. Based on such an observation the researcher has a unique opportunity in discussing the pattern in detail with the participant. Moreover, the researcher conducting the interview can go through the logs together with each participant and let the participant give further explanations of the underlining reasons behind the participant's answers.

5. Features of the future ES tool

Reconexp was used in an investigation of intra-family communication needs and the way pervasive computing would be able to support family members have awareness of each other through the day [12]. Based on the Reconexp study findings and the review of tools used in research studies a list of requirements for can be drawn. The event-triggering of queries, for example when a participant enters a location, is supported by some of the tools reviewed however it is either a built-in function, in the case of Reconexp for example, or in the best case (MyExperience [5]) it is programmable by using XML. Although XML is in many ways easier than programming in C or VB it still requires a certain expertise in markup up languages. Therefore, a requirement for an even more accessible tool would be the even-triggering of queries to be end-user programmable. End-user programming would be much easier to learn and apply when having a plethora of events that could potentially be available to the researcher.

In the reviewed studies tools presented the queries either in the screen of a mobile device or a desktop. However, a participant might have left the device in her bag while she is working in front of her desktop computer. A future tool can certainly include the possibility of smart presentation of queries. More specifically, queries can be prompted in multiple devices, in either desktop PCs or mobile devices according to which device is more accessible and available to participants.

Support for multimodal participant input has been already included in several tools. Text, audio, photos and video can provide richer data to the researcher [3]. On the other hand, participants can choose the most efficient and convenient modality for addressing the query.

In the case of CAES [10], MyExperience [5] and Reconexp [12] among others, user context factors are automatically captured. The location and possibly the activity of a participant can serve as examples. Automatic capture of participants' context would provide different perspectives for researchers to look at the gathered data and obviously provide more in-depth results.

All of the reviewed tools require extensive installation procedures and in most cases management of database or web servers. In tools that combine mobile phones and desktop PCs installation procedures have to be followed in both devices. This fact brings another obstacle to researchers who lack technical skills. Extremely quick and easy installations on mobile devices and even no installation procedures for desktops would be another important requirement for such tools.

Another requirement is support for optional, user initiated input. In the case of Reconexp, participants could not initiate the queries. That would be useful in cases where participants would recognize the importance, in terms of research, of the context they currently are and initiate the research queries. In that way salient information will be saved.

An important shortcoming of Reconexp was the difficulty participants had in synchronizing the data. Automatic synchronization of captured on the device data to a remote server would both secure the data as well as provide the grounds for feeding the data back to participants as the case with Reconexp was. MyExperience already supports such a feature and according to our experience of Reconexp it is a must have feature.

Automatic and configurable information visualization tools of the collected data would be a crucial feature for helping researchers disambiguate the data and quickly provide useful results. Alternative visualizations in the form of graphs can enable researchers to view the data in new, fresh ways and provide therefore opportunities in identifying new results. This analysis

tool should be able to support visualization of events that occur both frequently and infrequently [1].

Participants might become less motivated during the course of the research study. Programmable by researchers email or SMS notifications to the participants can help to keep participants highly motivated. In addition, support for notifications for researchers when certain events occur would also be of added value [1].

In most cases, ESM tools which use a mobile device force participants to carry another mobile device along with their personal one. It would be even more convenient if such tools would run on participants' phones. In this way such studies could be widespread and the reliability of results enhanced. However, a researcher would thoroughly need to have tested the tool so that it would not interfere with the participant's device. Moreover, agreements with the mobile service provider must be made in advance so that participants are not buried with the cost of the service. Therefore, agreements with mobile phone providers must be in place to easily refund costs participant occurred during the study.

Another important feature beyond the ability of participants to review the collected data would be the ability to participants in annotating the data and also to fill out the gaps. Moreover, giving feedback to participants during an experience sampling study has already been proven to be beneficial to participants' motivation [9].

In case researchers need to visually explain something to remote participants the support of video streaming between participant and researchers (video phone calls) would be handy.

Furthermore, mobile devices in comparison to desktops have limited processing and memory. The data collection tool on the mobile device should not noticeably impact the performance of the participant's mobile phone [3]. If that happens it might affect the results of the study since participants will experience a lag in the presented queries.

In a previous study [3] it was found that the tool on the mobile device should provide mechanisms to avoid interruptions at inopportune moments. Inappropriate moments will create frustration to participants and negatively bias them in answering the queries.

In case where a mobile device is lost, the tool on the mobile device should offer mechanisms to protect the security and privacy of the data [3].

Finally, in a more abstract level the tool on the mobile device should be easy to use. Thus, participants should be able to increase the color contrast, the font size [3].

6. ACKNOWLEDGMENTS

The research reported in this paper has been funded by the Dutch Ministry of Economic Affairs, under its Innovation Programme on Man-Machine Interaction, IOP MMI. We would like to express our gratitude to Senter IOP-MMI. Moreover, we are thankful to Wijnand IJsselsteijn and Boris de Ruyter for the discussions in the earlier phases of the Reconexp project.

7. REFERENCES

[1] Barrett, L.F. and Barrett, D.J. 2001. An Introduction to Computerized Experience Sampling in Psychology. Social Science Computer Review., V. 19, No. 2, S01, pp. 175-185

[2] Carter, S. and Mankoff, J. 2005. When participants do the capturing: the role of media in diary studies. CHI '05, pp. 899-908

[3] Carter, S., Mankoff, J., and Heer, J. 2007. Momento: support for situated ubicomp experimentation. CHI '07. ACM, New York, NY, 125-134

[4] Consolvo, S. and Walker, M. 2003. Using the Experience Sampling Method to Evaluate Ubicomp Applications. IEEE Pervasive Computing 2, 2 (Apr. 2003), 24-31

[5] Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., and Landay, J. A. 2007. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. MobiSys '07. ACM, New York, NY, 57-70

[6] Froehlich, J., Chen, M., Smith, I., and Potter, F. Voting With Your Feet: An Investigative Study of the Relationship Between Place Visit Behavior and Preference. Ubicomp 2006

[7] Hektner, J. M., Schmidt, J.A. and Czikszentmihalyi, M., "Experience Sampling Method: Measuring the quality of everyday life.", Sage, 2007, ISBN: 1412925576

[8] Hudson, J. M., Christensen, J., Kellogg, W. A., and Erickson, T. 2002. "I'd be overwhelmed, but it's just one more thing to do": availability and interruption in research management. CHI '02

[9] Hsieh, G., Li, I., Dey, A., Forlizzi, J., and Hudson, S. E. 2008. Using visualizations to increase compliance in experience sampling. UbiComp '08, vol. 344. ACM, New York, NY, 164-167

[10] Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., and Bao, L. 2003. A context-aware experience sampling tool. CHI '03. ACM, New York, NY, 972-973

[11] Kahneman D., Krueger A.B., Schkade D.A., Schwarz N., Stone A.A., A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method, Science 306, 1776 (2004)

[12] Khan, V.-J., Markopoulos, P., Eggen, B., IJsselsteijn, W., and de Ruyter, B. 2008. Reconexp: a way to reduce the data loss of the experiencing sampling method. MobileHCI '08 pp.471-476. DOI= http://doi.acm.org/10.1145/1409240.1409316

Experience-Sampling Tools: a Critical Review

Joel E. Fischer
The Mixed Reality Laboratory
University of Nottingham
Nottingham, NG8 1BB, UK
jef@cs.nott.ac.uk

ABSTRACT
In this position paper I argue for the merits of the experience-sampling method for HCI research and a wider scientific community and for the support of the method through publicly available, freely configurable tools. I will take a critical look at some of the relevant tools available. On that basis some recommendations for the design of ESM tools are given and a design space for ESM tools is sketched.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Evaluation/methodology*

General Terms
Design, Human Factors, Measurement

Keywords
Experience-sampling method, ESM tools, evaluation, studies

1. INTRODUCTION
The Experience-Sampling Method (ESM) lends itself well to the study of experiences in the wild – in fact, it was originally designed for that purpose [4].

As the settings of human-computer interaction (HCI) become increasingly mobile and temporally fragmented [11]; and its constituent parts become increasingly distributed across devices and locations [3], new challenges for the study and evaluation of these settings arise.

ESM has been used to study interruptibility in office settings [9], and mobile settings [8]; or to study users' information needs as part of a requirements analysis for a ubiquitous computing application [2]. We studied the change of player engagement over time in a long-term SMS-based experience by means of an ESM-based experiment [5].

Traditional ethnomethodologically-informed ethnographies focus on the *observable, overt action* by using techniques such as the observation of interaction *in situ* [3]. ESM in turn, is concerned with the *experience* that is *covert* to the eye of the observer, as it is *subjectively perceived* [7]. It is also a method that allows for *longitudinal* studies, as the participants are repeatedly prompted to assess their experience over a desired timeframe.

ESM has been applied to study such diverse fields as the quality of experience in everyday life, the experience of work, the examination of cross-cultural differences, and to educational and clinical research questions [7]. With a strong tradition in psychology and diverse application fields, it has just more recently been adopted by HCI researchers to study mobile and ubiquitous applications in the wild.

Hence, the method is being used by researchers from diverse backgrounds, not all of which have sufficient technical skills to implement an ESM-based study. The implementation of studies that use ESM is often costly and would benefit from supporting software. What tools are out there to support ESM studies? What are their benefits, what are their shortcomings? More generally speaking, what are the processes that a researcher conducting an ESM based study is involved in and how could they be supported?

2. REVIEW OF EXISTING ESM TOOLS

2.1 ESP and iESP
The experience-sampling program version 2.0 (*ESP*) has been reported as a early as 1999 [1] and is still available to the public[1], at the time of writing in version 4.0. It is a software package that contains a native application to trigger and run the ESM questionnaires on the PDA Palm Pilot, and a desktop application for Windows or Linux to create the logic for the timing of the prompts and the content and structure of the questionnaires in a browser-based application and to facilitate deploying the studies to the PDAs.

Intel Research continued the development of ESP into iESP [2] but a lack of information suggests that its development is discontinued – and a note on the ESP website says that its latest version includes and improves all features of iESP.

Whereas the feature that the content and the logic of the ESM study are created through familiar browser windows may contribute to the user's ease-of-use, it is the choice of the device eventually carried by the user that is dubious. The Palm Pilot is outdated since 1998. By choice of ESP, the researcher is locked into using an outdated device without connectivity or any communication functionality, not to mention sensors of location or acceleration. In fact, it does not even have a colour screen.

2.2 CAES
Context-aware experience sampling is the term a project at the MIT gave their contribution to research in ESM-based studies, with the twist that the questionnaire triggering may be influenced by contextual clues derived by additional computation of data captured with the device[2]. They also developed a tool [10] for a

[1] http://www.experience-sampling.org

[2] http://web.mit.edu/caesproject

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

PDA running Microsoft's PocketPC, Windows Mobile's predecessor.

Unless you are forced to use a device running PocketPC, there is no reason to use this tool. This tool is probably the least developed of the ones reviewed here; in fact, the authors say that the tool is no longer supported and that they have now joined forces with the *MyExperience* initiative.

2.3 MyExperience

MyExperience is open-source software that runs on devices with Windows Mobile 5.0 [6]. In addition to explicit data from questionnaires, it can be configured to collect sensor data collected with the device (e.g. GSM cells, GPS positions) alongside with user activity on the device (address book access, photo capture, phone calls, SMS usage etc.).

The tool does provide the researcher with interesting features to study experiences in the wild. Questionnaires cannot only be triggered by time (random or scheduled), but also by additional sensor data gathered from the environment (e.g. GSM cells), the devices position (acceleration) or the phone activity. So to speak, it combines automatic logging software such as ContextPhone [12] with the possibility to trigger questionnaires. However, the creation of the studies requires the researcher to be technically savvy, as it is done by editing XML files.

2.4 Conclusions

Let us consider and evaluate the tools presented from the perspective of the users, both the researcher that designs the study and the participants of the study. For reasons outlined above, we will only consider the ESP and MyExperience tool.

In all cases, the researcher has to familiarize himself with the tool-specific ways of creating a study. In *ESP*, she can use a browser-based series of forms to configure the logic and content of the questionnaires and to facilitate deployment to the devices, e.g. by defining the number of devices. In MyExperience the researcher has to edit XML files to configure the logic and the content of the questionnaires and duplicate them as many times as devices she wishes to deploy the questionnaire to. Additional configuration is necessary to employ other than the time trigger for the questionnaires.

The fact that both tools are native applications entails costs for the researcher. To recruit participants for the study, the researcher has two options: Recruit only people that own the required device or obtain the devices and provide the participants with them. With usually sparse resources in research, how large a sample can you reach with that approach?

Even though MyExperience includes a facility to post data to a server, this functionality does not come out 'of the box'; at least, a server would need to be set up for this. The (default) design decision to store the gathered questionnaire responses on the device instead of making standard use of connectivity over the air gives rise to another problem for the researcher: Not only has she to collect the data from the single device and aggregate it for analysis, she also has no means to monitor the distributed experiment's progress while the experiment is running. Do the participants continue to engage in the experiment? Do some people need motivation or assistance? It will be a black box for the researcher.

3. RECOMMENDATIONS FOR THE DESIGN OF ESM TOOLS

3.1 Think client-server

Why make the deployment of a native application a requirement for the study? Today's mobile devices have data connectivity that should be used. Instead of putting everything on the device, give the researcher his own server and a gateway so that she has to configure everything just once. The devices can access resources on the server and transmit data to it for further computational processing or for later statistical analysis.

3.2 Design for authoring

A tool that supports ESM studies should be designed for easy-to-use creation of the studies. An authoring interface should be designed so that the widest possible range of users is able to use it effectively and efficiently. A browser-based interface to a series of forms could let a researcher configure the timing of the prompts and the content and structure of the study.

3.3 Make use of people's own devices

People have mobile devices. Let them use their own devices for the study; it is less expensive and chances are higher they will actually carry the device on them when you prompt them to answer your questionnaire. You can recruit more people and don't even have to meet them physically to enroll them in the study.

3.4 Design for different levels of study complexity

Include different configuration options for your experience-sampling study. Chances are that your participants will either be able to receive SMS or email. Use that channel to prompt the participant to answer your question. Your questionnaire is easily accessed via the device's browser, your server will figure out if it should show the questionnaire as WML or as (X)HTML. Provide native applications as an additional option if the researcher wishes to collect phone usage or sensor data, or if sophisticated triggering is required.

3.5 Separate logging and questionnaires

Where phone usage data or data from sensors in or connected to the device is desired, create client software to be deployed on the phone. Sophisticated triggering could be deployed directly to the client as well, supporting scheduling either by asynchronously downloading trigger schedules from the server, or by waking up the trigger application remotely, based on logic derived by the server. Still, the questionnaires can be accessed online from the device's browser.

3.6 Make wise client choices

Designing and implementing client-side applications is costly. Be aware of the consequences your design choices may have. If your aim is to develop for interoperability and develop, e.g. a J2ME or Python application, be aware that it may not be easy to access all the phone and sensor data you require. If your aim is to develop for a native experience and develop a native application, e.g. for the iPhone, be aware that your application cannot currently run in the background and thus be 'woken up' to trigger a question, or that the deployment of the software may incur extra costs or is subject to a political selection process.

3.7 Support orchestration

A server-based solution would also allow a researcher to orchestrate the study, i.e. monitor the progress of the study, motivate participants to engage more, help out with technical problems, alter the content or structure of the questionnaire, or even expel participants.

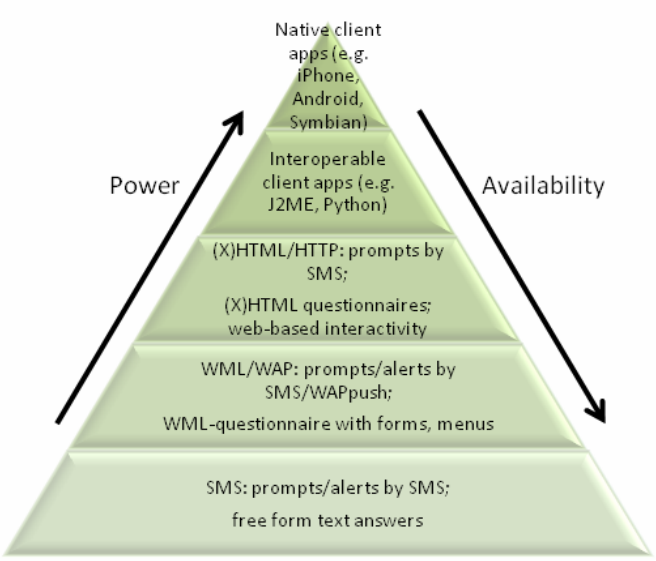


Figure 1: The design space for a proposed experience-sampling tool. With increasing power of functionality, the availability of the required technologies decreases.

4. DESIGN SPACE FOR A FLEXIBLE ESM TOOL

Following from the recommendations in the previous section, a successful ESM tool would be flexibly configurable by a researcher so that it would operate successfully in the design space depicted in figure 1. In a nutshell, the 3 lower levels would not require client-side software, while the two upper levels enable the researcher to gather data from the device such as phone usage and sensor data. While the power of the experience-sampling study increases in terms of complexity and richness of data gathered towards the pyramids narrow end; the availability of required technologies in the real world is broader at the pyramids broad base.

On its lowest level, a flexible ESM tool would support a study solely relying on communication by SMS. SMS from the server configured by the researcher would serve as prompts and as questions, SMS send back by the participants would be the answers. Of course, this is error prone, as it would require the users to type.

The next level would allow the researcher to ask questions in a more user-friendly and constrained way as simple WIMP elements such as lists, buttons and hyperlinks become available in the device's WML browser. Prompts to answer questionnaires should still be send by SMS, even though some devices support WAPpush messages.

On the third level, full web capabilities become available to the researcher constrained only by the devices browser implementation (e.g. currently Flash is not supported by many mobile browsers), enabling data transfer over HTTP. The embedding of pictures, or even audio and video becomes possible. Prompts however, are still most reliably sent by SMS.

If and only if the researcher requires data than can only be accessed natively on the mobile device should additional software be installed on the participant's mobile device. Client software can use the devices HTTP connection autonomously to communicate with the server and be instructed to 'wake up' in order to trigger a question for the participant; and it can access the data on the device and transmit it to the server for further computational processing or for later statistical analysis. Whereas clients developed to run with the support of middleware on several platforms may not easily access all of the required resources on the device, the development of native applications is more costly and entails platform specific risks and problems, as outlined above.

5. REFERENCES

[1] Barrett, L. F., Barrett, D. J. An Introduction to Computerized Experience Sampling in Psychology. *Social Science Computer Review* 19, 2 (2001), 175-185.

[2] Consolvo, S., Walker, M. Using the experience sampling method to evaluate ubicomp applications. *Pervasive Computing*, 2, 2 (2003), 24-31.

[3] Crabtree, A., Benford, S. et al. Supporting ethnographic studies of ubiquitous computing in the wild. *Proc. of INTERACT 2006,* ACM Press, (2006).

[4] Csikszentmihalyi, M., Larson, R. et al. The ecology of adolescent activity and experience. *Journal of Youth and Adolescence* 6, 3 (1977), 281-294.

[5] Fischer, J. E., Benford, S. Inferring player engagement in a pervasive experience. *Proc. of CHI 2009,* ACM Press (2009).

[6] Froehlich, J., Chen, M. Y. et al. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. *Proc. of MobiSys 2007,* ACM Press (2007).

[7] Hektner, J. M., J. A. Schmidt, et al. (2007). Experience Sampling Method: Measuring the Quality of Everyday Life. Thousand Oaks, CA, Sage.

[8] Ho, J., Intille, S. S. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. *Proc. of CHI 2005,* ACM Press (2005).

[9] Hudson, J. M., Christensen, J. et al. "I'd be overwhelmed, but it's just one more thing to do": availability and interruption in research management. *Proc. of CHI 2002,* ACM Press (2002).

[10] Intille, S., Kukla, C. et al. Eliciting user preferences using image-based experience sampling and reflection. *Ext. abstracts CHI 2002,* ACM Press (2002).

[11] Oulasvirta, A., Sakar, T. et al. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. *Proc. of CHI 2005,* ACM Press (2005).

[12] Raento, M., Oulasvirta, A. et al. ContextPhone: a prototyping platform for context-aware mobile applications. *Pervasive Computing*, 4, 2 (2005), 51-59.

SizzleLab: Building an Experimentation Platform for Mobile Social Interaction

Martti Mäntylä, Matti Hämäläinen, Kimmo Karhu, Airi Lampinen,
Kasimir Lehväslaiho, Esko Nuutila, Antti Oulasvirta, Olli Pitkänen,
Risto Sarvas, Emmi Suhonen, Jani Turunen, Seppo Törmä, and Antti Virolainen
Helsinki University of Technology
Otakaari 1, 02015 TKK, Finland
+358 50 351 2160
Martti.Mantyla@tkk.fi

ABSTRACT
We give an overview of the SizzleLab campus experimentation platform for mobile social interaction services. The paper focuses on the design, architecture, and implementation of the platform, and also discusses its future outlook.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human factors; H.5.2 [**User interfaces**]: Evaluation/methodology, User centered design; H.5.3 [**Group and Organization Interfaces**]: Evaluation/methodology

General Terms
Design, Experimentation, Human Factors.

Keywords
Social interaction, mobile communications, living lab.

1. INTRODUCTION
During the recent years, social interaction services such as MySpace, Facebook, and Twitter have become popular on the wired Internet. A similar phenomenon is starting to appear also on smartphones. Indeed, we expect that the success and significance of mobile social interaction services will outshine the wired ones.

We have studied mobile social interaction since 2002. Over the years, we have built and experimented with many research prototypes [Raento and Oulasvirta 2008] [Jacucci et al. 2007] [Sarvas et al. 2005]. Typically, our studies involve a small number of subjects (6-20) who use the test apparatus (mobile phone + a test system) for 1-4 weeks. The study procedure combines automatic logging of system events with qualitative methods such as direct observation and interviews.

In our experience, this research strategy exposes interesting and relevant phenomena on the scale of a single user to a small group of users. However, small-scale field experiments are susceptible to noise and random events that threaten the validity of the conclusions drawn. In addition, post-trial data does not shed light on the actual use, and provides no hard baseline for assessing whether the observed changes were for the worse or for the better.

More alarmingly, the few cases where we have been able to study data from large user communities suggest that there are also larger-scale phenomena especially related to social interaction services that do not expose themselves in a small experiment.

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

The SizzleLab mobile social interaction experiment platform is intended to overcome these limitations. With it, we aim to conduct experiments cost-effectively with communities up to thousands of users, while having a wide control on the depth of data logging and a total freedom of changing the services studied. With this, we hope to achieve nothing less than a Copernican revolution in the science of mobile social interaction by creating a research instrument that can see further and deeper.

This paper describes our progress so far. First, we outline the research interests that motivate our work. This is followed by a discussion on our key design choices and trade-offs. Next, we outline the technical environment of the SizzleLab platform. The experiments linking the platform to campus life and its stakeholders are discussed in the next section. Finally, we discuss some insights of our work.

2. OBJECTIVES
Several lines of research have motivated us to embark on the development of the SizzleLab platform: the added value of mobility, privacy and publicity, service adoption, diffusion, and appropriation, and user innovations.

Added value of mobility: As a platform for social interaction, the mobile phone has a number of unique properties:

1. Mobile phones are inherently "personal" and "social". They are on 24/7, and are carried by their users everywhere. They are used for communication with family, friends, and workmates, and store information on users' social networks and other data reflecting the personal and social lives of their users, such as calendar items, photographs, and music.
2. All mobile phones know their location up to the granularity of a GSM cell. Many phones also include WLAN, Bluetooth, or GPS functionalities that can be used for more accurate indoor and outdoor positioning.
3. Through embedded light, sound, and motion sensors mobile phones can perceive their immediate environment to gain data that can be used to infer cues on the prevailing context and social situation of the phone user.

At the present, these characteristics are still underutilized in actual services. The gap between the potential and the actual offers a first-rate opportunity for explorative research aiming to

understand precisely how these features should be best utilized to create added value to the end user.

Privacy and publicity: Mobile phone characteristics expose severe issues related to privacy and location disclosure [Vihavainen et al. 2009]. As users at the same time desire to be visible to some users and hidden from others, they evolve and fine-tune strategies to manage actively and dynamically their own desired level of privacy and publicity, and instinctively also the privacy of their social network friends. How these strategies should be supported by various service characteristics while observing human factors, social norms, and legal regulation is largely an unexplored topic.

Service adoption, diffusion, and appropriation: Experience from wired services shows that social networks can be a powerful engine for service adoption and diffusion: people are much more likely to sign on a new service if a real-world friend invites them to it, or demonstrates directly its usefulness. Network effects are expected to increase the power of the social networks with deepening penetration. After that, continued service use is determined by how well the service can be adapted to the everyday life needs of the new user. Moreover, the appropriation of a new service in user's service portfolio is likely to change how existing services are used: some may become disused, while others work well in association with the new service, and win increased use. In effect, users will innovate opportunistically and adaptively new composite services through emergent practices where two or more component services are used jointly.

User innovations: If a mobile service is successful at all, it is because users contribute to the service with their own content and ideas, making it fit for their everyday life needs. In a social interaction service, their contributions become visible and add value also to other users; again, network effects are in play. This creates a powerful incentive for durable user relation; therefore, research on user innovations is needed, such as incentive and reputation systems that make the benefits of user investments tangible and easily understandable.

3. SIZZLELAB DESIGN

The design challenges of the SizzleLab platform revolve around three interacting sets of issues: the *scalability* of the platform for a user community of thousands of users, its *transparency* for research with sufficient and relevant instrumentation and data gathering tools, and the *attractiveness* of its services so that a durable and stable community of users can emerge and the intended network effects can start working.

The need of *scalability* breaks down in several dimensions. Along the dimension of *service provisioning*, we decided that our services should to the extent possible run on user's own mobile devices, without extensive configuration from our part. We also provide management tools so that users can themselves sign on the services, maintain their accounts, and link their identities to a identity management service for single sign-on. While they will still need helpdesk services, these choices should simplify service provisioning considerably.

Unfortunately, this choice nevertheless meant that we had to face another dimension of scalability: the *diversity of user's terminals* and the consequent limits to our services' depth of penetration in the user population. To address these issues, we decided to build our services to run in the WWW browser of the mobile device. As a result, our services work not only on Nokia terminals, but also on iPhones and other mobile phones with sufficiently modern browsers. Moreover, users who cannot use the services on a mobile device can use a Web browser on their PCs. A downside of the this approach is that it complicates services' access to terminal resources such as the camera, network status, or calendar information. Moreover, even after this choice, not all terminal browsers provide the full functionality we need.

To address the second set of issues, the *platform transparency* for research, adequate instrumentation and data gathering tools must be included in the platform. By virtue of our restriction to WWW-browser based services, these requirements could be satisfied to a considerable extent by instrumenting the server components of the services and logging the data passed between the browser and the server. To complement this data, we also collect horizontal service data from users' terminals through a terminal-resident software component that logs all user actions at a desired level of accuracy, and reports the logs periodically to a server.

The *attractiveness* of the services is the most complex of the main challenges, not least because it is aligned with our intended research theme of service adoption, diffusion, and appropriation discussed above. Thus we faced the dilemma that to study service features critical for attractiveness, we somehow had to jump-start a set of services sufficiently attractive by design alone.

In this respect, the decided to design our services from the baseline of earlier work by ourselves and others, incorporating our informed guesses of how the various aspects of the added value of mobility might be realized. As we went along, we also became increasingly aware of the need to base our design effort to an understanding of campus life and its challenges from the viewpoint of our intended first-line users. We also expect to learn by doing: by observing closely the actual success of the services and the use patterns, and reacting opportunistically to them by further design, we should be able to find the winning formulae.

4. SIZZLELAB PLATFORM

As implemented presently, the SizzleLab platform consists of user-level services implemented on top of a shared run-time infrastructure providing user profiles, social networks, groups, locations, points of interests, *etc*. To enable rapid service development, the infrastructure also provides building blocks for commonly needed functionality, such as messaging channels, location sensing, authentication, and authorization. In addition, we expect interesting interactions between the services to emerge from this data sharing. Figure 1 (overleaf) gives an outline of the platform (with extensions currently in progress).

4.1 End-user services

At the present, we have implemented two online social interaction services Ossi (http://ossi.sizl.org) and Kassi (http://kassi.sizl.org).

Ossi (Figure 2) is a mobile online social interaction service for high-end mobile phones (e.g., Nokia N95, N97, iPhone), aimed to expose "what is sizzling" in users' social and physical environment. Currently Ossi provides facilities for creating social networks and exchanging messages between users in public and private channels. Functions for sharing location information, and for creating and managing user groups are currently being developed for release in September 2009. We have also created an "iFramed" version of Ossi that can be embedded in any WWW page environment (widgets, FaceBook app).

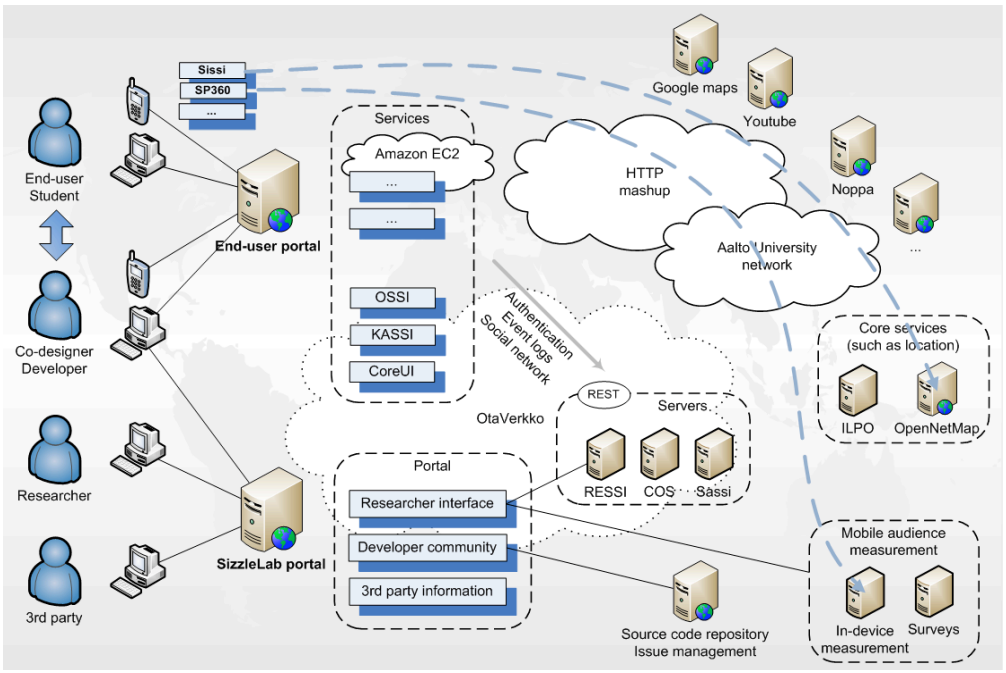


Figure 1: Overview of SizzleLab platform

Kassi is an online crowdsourcing service for exchanging resources in a local community. The users are considered to have three types of resources in their disposal: goods, skills and time. With Kassi, users can publish in their tradable resources, such as items possessed and skills, for other users to see. If another user cannot find a needed resource by search, she can post an announcement of what is needed. The announcements are shown either to all other users, just to author's friends, or for the members of a group.

Ossi and Kassi share the social network; hence users who are "friends" in one service are that also in the other. This is intended to encourage natural service diffusion and service composition.

In addition, portals www.sizl.org and www.sizzlelab.org aimed at end users, co-developers, experimenters and 3rd parties are being implemented. The end user portal will also include the CoreUI service intended for setting advanced privacy controls.

4.2 SizzleLab Infrastructure

The end-user services run on a shared infrastructure, intended to provide useful resources to the services and to facilitate collecting empirical and experimental data for research.

The joint service execution environment *Common Services* (COS) provides tools for creating user identities and groups, building social networks, and launching communication channels associated to identities and groups. Thus COS allows a large amount of data on the users and the usage of the services to be gathered and studied. COS is closely associated with the researcher data gathering component *Ressi* that collects a full searchable log of all interactions between end-user services and COS, and the identity management component *Sassi* that provides a single sign-on service for the infrastructure.

Certain core services are provided by 3rd party components such as www.OpenNetMap.org, a WiFi access point and GSM cell radio signal collection and location calculation service. With its terminal component Sissi, it provides the facilities for locating terminals and building location-sensitivity in SizzleLab services.

The infrastructure also includes tools for collecting horizontal data on service use directly from end users' terminals by means of a terminal-resident monitoring component that observes phone use on operating system level and uploads its logs daily to a server [Verkasalo and Hämmäinen, 2007]. At present, we use the SP360 monitoring system by Nokia.

5. SIZZLELAB IN CAMPUS LIFE

The SizzleLab platform is being implemented on the campuses of the Helsinki University of Technology, the Helsinki School of Economics, and the University of Arts and Design Helsinki – three universities to be merged as the Aalto University from 1.1.2010. Our longer-term aim is to integrate the platform in the campus life, and provide added value to its participants.

Our present focus is on new students and the predicament they face as they are supposed to change their lifestyle to match the expectations of being a student at a university. A rudimentary version of the Ossi service was launched for a pilot group of 40 first-year computer science students in September 2008. In December 2008, an improved version of the service with open and closed message channels was released. The Kassi service was launched as closed beta in February 2009 for a pilot group of 30 users. In the end of April 2009, it was opened for general public.

In parallel with the Ossi and Kassi experiments, we conducted a horizontal measurement experiment on a group of 80 users with the SP360 system. We also conducted ethnographic field studies aimed at uncovering first-year students' actual and self-reported use of social interaction tools and their role in freshmen's integration to the campus society. Our results show the central role of the Internet Relay Chat (IRC) service in both actual use and as a symbol of becoming a member of the community.

Figure 2: Ossi mobile social interaction service

The next experiment will be launched in September 2009 with the start of the new academic year. It will include 400 students from all three universities forming the Aalto University, including both first-year B.Sc. students and first-year M.Sc. students. For more diversity, 50 international students will be included.

The subjects will be exposed to new versions of the services with improved functionality and with content tailored for new students. Two main use cases will be considered: (1) using the services in association with key courses offered to the new students and (2) using the services to link new students to student associations. To encourage active use, we plan to loan Nokia N97 phones to committed users. To reach other users, we will insert embedded instances of the Ossi service in as many as possible Web contexts: course home pages, home pages of student associations and special interest groups, and university department home pages.

To facilitate the experiment, we have built a collaboration network consisting of faculty responsible of teaching freshmen or 1st year graduate courses, student organizations offering mentoring services to their new members, and units in the university offering study counseling and other student services. We have also established links to key WWW services aimed at students such as the popular Noppa portal (noppa.tkk.fi) to information resources related to studies at TKK.

While SizzleLab is mainly intended as a platform for semi-controlled experiments, it will also be used as a platform for application development. So far, we have executed one "code camp" type experiment on end user development. To our pleasant surprise, five of six student groups participating the one-week code camp were able to produce a working prototype.

6. DISCUSSION

Is SizzleLab a living lab? We have avoided the term here so far, conscious of the thick cloud of hype around the term.

In Mitchell's original formulation, *"Living Labs represent a user-centric research methodology for sensing, prototyping, validating and refining complex solutions in multiple and evolving real life contexts"*. Our experimentation platform may not quite satisfy the spirit of this definition, in that we are limiting our focus on social interaction services and thus not following user-centricity to letter.

Other formulations [Schumacher and Niitamo 2008] [DG InfSoc 2009] emphasize the multi-stakeholder nature of a Living Lab, blending concepts related to open innovation and public-private partnerships to the definition. SizzleLab falls short of these criteria as well. While we are interested in end-user innovations, we so far have put little emphasis on multi-stakeholder innovation ecologies. SizzleLab is a university initiative, run and managed by us. While we plan to open it for third party experiments at a later stage, we prefer to remain the driving seat.

7. ACKNOWLEDGMENTS

This work is funded by the MIDE program at TKK. Apart from the authors, the following persons have contributed to it: Tero Hasu, Kai Huotari, Jaakko Kangasharju, Vilma Lehtinen, Vesa Luusua, Juho Makkonen, Petri Martikainen, Petteri Noponen, Lassi Seppälä, Hannu Verkasalo, and Veli-Matti Vikman. We also acknowledge the sponsoring from Nokia and Elisa to our work.

8. REFERENCES

[1] Jacucci, G., Oulasvirta, A., Ilmonen, T., Evans, J., and Salovaara, A. 2007. CoMedia: Mobile group media for active spectatorship. *Proceedings of CHI'07,* ACM, New York.

[2] Raento, M. and Oulasvirta, A. 2008. Designing for privacy and self-presentation in social awareness. *Personal and Ubiquitous Computing,* 12 (7), 527-542.

[3] Sarvas, R., Oulasvirta, A., and Jacucci, G. 2005. Building Social Discourse Around Mobile Photos – A Systemic Perspective, *MobileHCI 2005,* Salzburg, Austria.

[4] Verkasalo, H. and Hämmäinen, H. 2007. A handset-based platform for measuring mobile service usage. *Info*, 9 (17).

[5] Schumacher, J. and Niitamo, V.-P. (eds.). *European Living Labs. A new approach for human centric regional innovation.* Fachhochschule Vorarlberg, 2008.

[6] DG Information Society and Media. *Living Labs for user-driven open innovation.* European Commission, 2009.

[7] Vihavainen, S., Oulasvirta, A., and Sarvas, R. 2009. *"I Can't Lie Anymore!": The Implications of Location Automation for Mobile Social Applications,* submitted for publication.

A Reference Architecture for Living Lab Measurement Systems

Tom Broens, Henri ter Hofte, Peter Ebben
Novay
P.O. Box 589, 7500 AN Enschede, the Netherlands
{Tom.Broens, Henri.terHofte, Peter.Ebben}@novay.nl

ABSTRACT
This paper describes initial results of the Amsterdam Living Lab project. It discusses a reference architecture for measurement systems to determine user behavior in mobile living lab settings. Such a reference architecture has a goal to act as (i) a common vocabulary among people that communicate about living labs, (ii) a blueprint for future implementations and (iii) a guideline for comparison of living lab measurement systems. The architecture is a work in progress that will be evaluated and refined in the coming period.

Keywords
Reference architecture, mobile living labs, measurement systems.

1. INTRODUCTION
Living Labs have a growing research and business attention [4]. Amsterdam Living Lab (ALL) [3] is a project that researches ways for developers and users to co-design products using a living lab approach. Multiple developers and users in the Amsterdam region cooperate with knowledge partners to create methods and tools that enable the proper design and evaluation of products in a living lab setting.

In a living lab setting, you do not bring people into a lab, but instead bring the lab to the people. This is especially useful when you want to measure context-dependent aspects or when usage conditions are hard to simulate in a lab, e.g. when you want to measure the usability of a software application that changes its behavior based on the location of the user (a so-called location-aware/context-aware application [2]). Location can be hard to simulate in a lab in a realistic manner and hence getting representative measurements can be hard. Two main discriminating factors of a living lab are its high situatedness (i.e., situated in real life) and its strive for unobtrusiveness (i.e. minimal impact on the behavior of the user).

A key element in living lab studies is the use of some form of living lab measurement system (LLMS). Such systems are responsible for measuring the behavior of participants in living lab studies in their daily environment. In the next section we zoom in on such systems in more detail. In this paper we discuss a reference architecture for LLMS. Such a reference architecture has as goal to act as (i) a common vocabulary among people that communicate about living labs, (ii) a blueprint for future implementations and (iii) a guideline for classification of living labs.

Mobile Living Labs 09, September 15, 2009, Bonn, Germany.

In the remainder of this paper we first give an overview of the position of an LLMS in a living lab and discuss the kind of data these systems can collect. We continue by zooming in on a living lab measurement system by providing a reference architecture from two perspectives. Finally, we give some conclusions and an outline of future work.

2. Living Labs and LLMS
In our view, an LLMS is the central part of an operational living lab. Figuratively speaking, an LLMS collects measurement data from deployed measurement instruments ('thermometers') at the different elements in a living lab. Figure 1 describes the stakeholders and elements in an operational living lab.

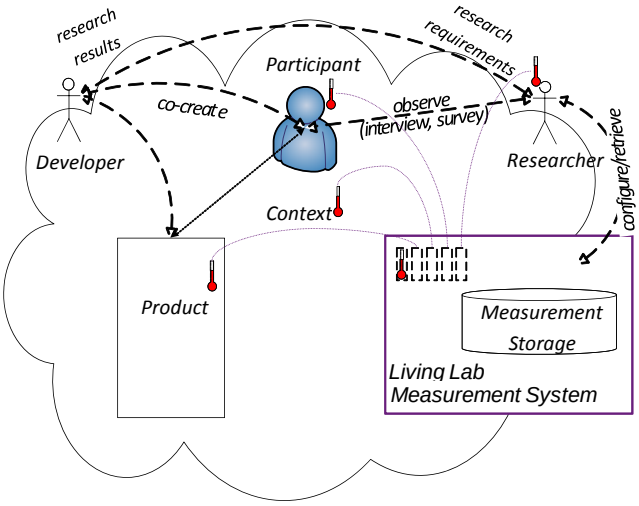


Figure 1: Living Lab Measurement System

A developer wants to evaluate the usability or user experience of a developed product. This product is used by a user (denoted as participant of the living lab study). A product can be a tangible product, a software application, but also a process, method, operational protocol etc. In this paper, we focus on software applications. A researcher determines, together with the developer (can also be the same person), the relevant research questions that should be answered using the living lab approach. The LLMS measures the behavior of the user, such that the research questions can be answered.

An LLMS can measure in two different ways:

- ***Manual logging:*** determining the behavior of the participant by asking questions to the participant (experience sampling, surveys, interviews) and/or observing the behavior of the participant (observation) by the researcher.

- ***Automatic logging*** through the use of sensors.
  - ***Application logging:*** determining the behavior of the participant through measuring the use of the application. For example logging interaction patterns (e.g. using custom made loggers or logging frameworks like Log4J), performance footprints (e.g. using system profilers), network usage (e.g. using protocol sniffers).
  - ***Context logging:*** determining the behavior of the participant through measuring environmental parameters that specify the situation that the participant is in. These parameters are called context parameters [2]. Some examples are location, movement, temperature, and number of people in the direct vicinity of the participant.

3. LLMS Reference Architecture

The goal of a reference architecture is to provide a high-level architecture of a class of systems. As such, our LLMS Reference Architecture does not provide technical details as a functional architecture would do. Additionally, it is not created with a specific implementation in mind, although it could be used as the basis of a functional architecture of a concrete implementation. Furthermore, several architectural components that we describe may be optional depending on the type of study being conducted. We envision this reference architecture to be useful for:

- Researchers who want to gain a better knowledge of Living Lab Measurement Systems and their use.
- Developers of applications that will be tested in a Living Lab environment.

The reference architecture is created based on the authors' experience of doing projects using several measurement systems (e.g. CMF [6], Xensor [5]). Additionally, architectures of current systems such as MyExperience have been the inspiration for this architecture.

We describe the reference architecture from two perspectives: (i) ***external perspective:*** provides an overview of the externally visible components of a Living Lab Measurement System and their interactions, and (ii) ***internal perspective***: provides a more detailed overview of the inner working of the components. This section is targeted towards developers and researchers.

3.1 External perspective

Figure 2 describes the external perspective on an LLMS. White boxes denote deployment units. Colored boxes / circles / cylinders denote components. Text along the lines indicates data flow: Data flows in the direction of the arrow. Control flow may be opposite.

We distinguish three deployment units:

- ***Measurement node***: required infrastructure at the measurement site (e.g., co-located with the participant on a mobile device) to perform the measurements. A study can contain multiple measurement nodes (e.g., multiple participants). Such a node collects data via the measurement engine from deployed sensors (e.g. application logging sensors, context sensors, experience sampling 'sensors'). The participant can use the dashboard to monitor and control (e.g. start/stop) the measurement.
- ***Central storage node:*** provides central, secure and persistent storage of the measured data coming from measurement nodes.
- ***Researcher node:*** allows the researcher to configure and control the study and to retrieve measured data for analysis.

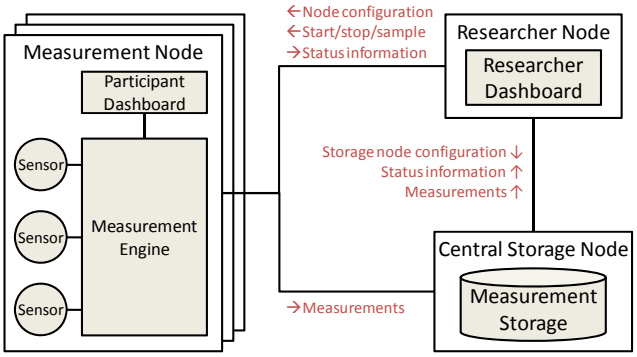


Figure 2: External perspective of an LLMS

3.2 Internal perspective

In this section we consecutively zoom into the different nodes defined in the external perspective. The notation is similar to the one used in the external perspective.

3.2.1 Measurement node

Figure 3 shows the internal perspective of the measurement node in which we detail the measurement engine, as defined in the external architecture.

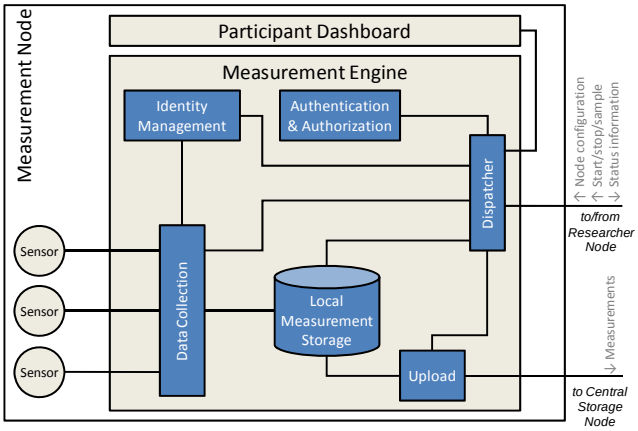


Figure 3: Measurement node

The measurement engine coordinates the life cycle of the different sensors and receives measurements from the sensors. It may perform simple processing (e.g. translating device IDs to participant IDs via the identity management component). Additionally, it delegates the (processed) measurements to the local measurement storage component for persistent storage (so that data is not lost in case connectivity is lost or no connectivity is available at all). It is responsible for secure upload of measurements to the central storage node. The engine receives configuration data from the researcher node via the dispatcher and configures the sensors, storage and upload accordingly. Both upload of measurement data and configuration of the node are subject to authentication and authorization.

Associated with the measurement engine is also a participant dashboard, which provides a minimal interface that allows the participant to monitor and control the status of the study, e.g.:

- Is the system still collecting data for the study properly?
- Suspend/resume data collection, erase recent data
- Initiate a self-report (for diary studies)

3.2.2 Central storage node

Figure 4 shows the internal perspective of the central storage node.

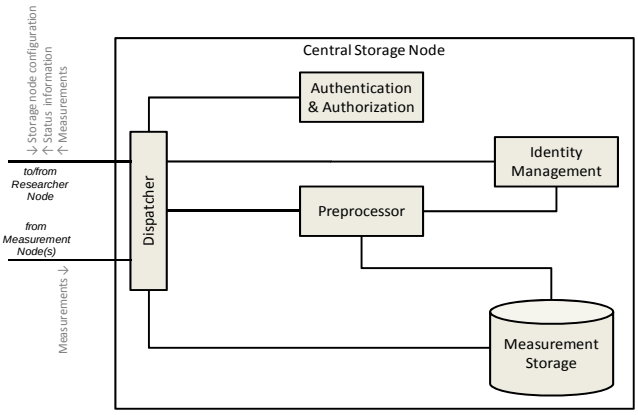


Figure 4: Central storage node

In the central storage node the dispatcher, authentication & authorization and identity management components have the same responsibilities as their counterparts in the measurement engine described earlier. The preprocessor can use the information provided by the identity management component to enrich the measurements before storage (e.g., annotate with metadata). The measurement storage component provides secure, persistent storage for all measurements such that they can be retrieved easily by the researcher.

3.2.3 Researcher dashboard

Figure 5 shows the internal perspective of the researcher dashboard.

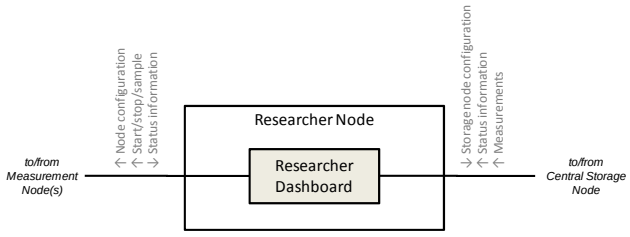


Figure 5: Researcher dashboard

The researcher dashboard gives the researcher a detailed overview of the status of the measurement system (as a whole and on a node-by-node basis if required):

- Is the system up and running?
- Is the system currently collecting information?
- Is there enough storage space left?

Additionally, it gives the researcher control over the measurement process: start/stop/initiate sample (on one or more measurement nodes). Finally it gives the researcher the opportunity to configure the measurement system (before the study starts) and to alter the configuration if required.

The researcher node could also contain data analysis tools. However these are out of the scope of this architecture.

4. Conclusions and future work

This paper gives a glimpse of the work performed in the Amsterdam Living Lab project. It proposes a reference architecture for living lab measurement systems. This paper is not exhaustive and we refer the interested reader to [1] for more details on the proposed reference architecture.

At the time of writing this position paper, we are working on descriptions of various existing living lab measurement systems in terms of components and functions identified in the reference architecture. Some are more full-blown systems, whereas others have more limited scope (e.g., a wearable sensor box) and could be used as a component of an LLMS.

Also, we are working on an overview of the role of context information for living labs in general and an overview of the context sensors employed by existing living lab measurement systems in particular.

By the time of the workshop, we expect to have initial versions of these overviews ready and we're more than happy to discuss the reference architecture and our overviews in the workshop.

5. ACKNOWLEDGMENTS

This work is part of the Amsterdam Living Lab project (http://www.amsterdamlivinglab.nl/).

6. REFERENCES

[1] Ebben, P., Broens, T., Hofte, H. ter, Living Lab Measurement System Reference Architecture v0.5, ALL D2.1, https://doc.novay.nl/dsweb/Get/Document-98640

[2] Broens, T., Context Bindings: Infrastructural Support for Context-Aware Applications, Telematica Instituut Fundamental Research Series, vol. 022. Enschede, the Netherlands: Telematica Instituut, 2008, https://doc.novay.nl/dsweb/Get/Document-92134

[3] Amsterdam Living Lab, http://www.amsterdamlivinglab.nl

[4] The European Network of Living Labs, http://www.openlivinglabs.eu/

[5] Ter Hofte, G.H., What's that Hot Thing in my Pocket? SocioXensor, a smartphone data collector. In: Proceedings of the e-Social Science 2007, the Third International Conference on e-Social Science, October 7-9, 2007, Ann Arbor, MI, USA. https://doc.freeband.nl/dsweb/Get/Document-75276

[6] Hesselman, C., Benz, H. et al., Bridging context management systems for different types of pervasive computing environments, In: Proceedings of the 1st international conference on MOBILe Wireless MiddleWARE, Operating Systems, and Applications (MOBILWARE '08), February 12-15, 2008, Innsbruck, Austria. ACM International. Conference Proceeding Series; Vol. 278. pp 1-8. ISBN 978-1-59593-984-5.