Touring in a Living Lab: some methodological considerations

El Ali, A.; Nack, F.

Publication date
2009
Document Version
Final published version
Published in
Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild: September 15, 2009, Bonn, Germany

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.