The wearable Living Lab: how wearables could support Living Lab projects

Tanguy Coenen
iLab.o, iMinds-SMIT
Vrije Universiteit Brussel

Lynn Coorevits
iMinds-MICT
Universiteit Gent

Bram Lievens
iLab.o, iMinds-SMIT
Vrije Universiteit Brussel

Keywords: wearables, living lab, evaluation

Abstract
Part of the added value of living lab or other in-situ research resides in its capability to capture contextual and personal data of end-users in real life environments. In this paper, we present an architecture for the wearable living lab (WELLS), allowing the collection of user experience feedback in a more permanent and unobtrusive way. It enables mainly the user researcher to capture contextual and experience data during a Proxy technology assessment, a formative or a summative software evaluation.

Introduction
Wearables provide intrinsically different opportunities for the research process of several application domains ranging from health to consumer research. We perceive wearables as any computing device that can be worn on the body (from smart-textiles to smart-bracelets, -jewelry and –watches). They enable the monitoring, tracking and controlling of several human activity types. A distinguishing feature is that, unlike other devices such as smartphones, wearables are often worn almost permanently and therefore allow the continued and longitudinal capturing of data. The market for wearables is looking promising and wearable computing devices are increasingly hitting the market, like smart and sport watches (e.g. Pebble, Apple Watch,...), fitness trackers (e.g. Fitbit, Jawbone UP3,...) or smart glasses (e.g. Google Glass).

Consumer-grade wearables allow the measurement of different types of user data in a less obtrusive and often more objective way than current, mainstream living lab data collection methods, i.e. observation techniques or surveys. Also, software development is more and more taking place in the context of mobile devices, where the interactions that the user has with the system are mediated by the context in which the use takes place (at work, at home, in the car, in bed,...). The assessment of such systems requires an in-situ or “in the wild” evaluation.

Current application developers and researchers try to evaluate the software product by logging the actions of the users. However, logging data on itself has its limitations as it cannot track all contextual elements nor the individual’s personal experiences related to the usage of the application. Therefore, (additional) experience sampling is used to capture in the moment experiences of users. Both techniques allow to track users ‘in the wild’, but requires specific research methods that afford the continuous evaluation of the user’s response to the system.

Katz (2001) concluded that there is an invisible side of emotions that cannot be induced from observations of user behavior alone. One way to measure emotional fluctuations is by researching their physiological changes via skin conductance in order to understand arousal. Due to the specific set-up and technology required, such measurements were only possible in a dedicated physical location. Because of the ongoing commercialization of wearables allowing measurement of behavior as well as psychological changes, wearables provide research opportunities in living lab environments. To allow this, the challenge which was already addressed in e.g. the work of Kocielnik et al (2013) or Sano & Picard (2013), to make such measurement work outside of controlled lab conditions, must be further taken up for application in living lab conditions. In other words, we must ask how these measurements can be made in the in situ evaluation conditions that distinguish living lab projects from other design approaches.

In this position paper, we discuss how wearables can be used in living lab projects (and by extension other types research with an in-situ component) to perform novel types of systems evaluation. To do this, we first discuss the process structure of living lab projects, which informs us on what living lab stages can benefit from unobtrusive systems evaluation. Then, we proceed to discuss how this can be done with the means that are currently or in the near future on the consumer market. We focus on consumer-grade devices, as it is important that the user of the device become as accustomed to it as possible, in order not to
bias the data (Henandez et al 2014). Consumer-grade devices are designed to accomplish exactly such accustomedness. Another reason for focusing on consumer-grade wearables is that evaluations in living labs are often large-scale. This can only be achieved if the wearables to be used are already owned by the user, or because their cost of acquisition is low. In sum, this paper discusses how the need for in-situ evaluations that exist in living labs can be met by the opportunity of new wearables becoming available as consumer devices.

**Prototype evaluation in Living Lab projects**

Living lab-research is a state-of-the-art methodology aiming at the involvement of end-users in the innovation process. Although there are several ways of integrating consumers in the innovation process, living labs are interesting because of the context for understanding customer co-creation. Living labs are experimental platforms where end-users can be studied in their everyday context (Eriksson et al. 2005). They confront (potential) users with (prototypes or demonstrators of) products and/or services in the innovation process (Schuurman & De Marez 2012). As a distinguishing factor of living labs is that the systems they produce are tested in non-laboratory conditions, but “in the wild” (Westerlund & Leminen 2011, Ballon 2015). By introducing this in a real life environment, experience based learning and discovery (facilitating serendipity) is made possible. Through real-time iterations the design and development actions are constantly being validated. This will help developers to make more informed decisions, thus increasing the likelihood of success (Trimí & Berbegal-Mirabent, 2012).

Various types of process models for living labs have been proposed in literature (e.g. Pierson & Lievens 2005, Schaffers et al 2008, Tang & Hämäläinen 2012, Bergvall-Kåreborn et. al 2009). Recently, Coenen & Robijt (2015) proposed the Framework for Agile Living Labs (FALL) process model as a way to guide both researchers and practitioners in how to perform a living lab project. This process model was shown to be compatible with most of the existing process model literature.

In the FALL approach to a living lab project, there are 2 stages in which user feedback can be of use, i.e. the problem formulation and the so-called BIEL phase. In the problem formulation phase, users can be brought in contact with existing technologies that are configured to mimic the behaviour of the prototype that the project team has in mind. This is done through a Proxy Technology Assessment (PTA) (Pierson et al 2006). In the iterative build, intervene, evaluate and learn (BIEL) phase, two different types of evaluations can be performed to get user feedback on a prototype: formative or summative evaluation. In formative evaluation, the prototype is of a very provisional nature and the aim is to get feedback that allows the living lab team to create better prototypes after it. In summative evaluation, the objective is to evaluate a more stable prototype and find out how effective and efficient it is to be able to report this to some project stakeholder (client who commissioned the project, customers, societal actors,…). These three types of prototype evaluation are the domains where we see potential in living lab projects for the application of research methods that are supported by wearables.

**WELLS or how wearables can be used to support prototype evaluation in living lab projects**

Based on our experiences in conducting real life evaluations of prototypes in living labs, as well as the literature on wearables, usability research and affective computing, we propose the architecture for a Wearable Living Lab Software Evaluation System (WELLS) depicted in Figure 2. Various roles are described as part of FALL, of which the user researcher is the role that is involved in carrying out the evaluation of prototypes. It is therefore the user researcher that will be the main consumer of the data generated by WELLS. Taking into account the opportunities of the wearable devices and the research needs for in-situ, contextual user-centered living lab research we focus on two main types of data-sets collected through wearables: bio-data and context data.
Bio-data
In our choice of sensors for measuring bio-data to be included at the core of the WELLS system, we concentrate on sensors that can provide a measure of the arousal experienced by the user. Arousal is a general term describing an emotional response by the user. The valence of the response can be either positive or negative representing emotions such as excitement and frustration. We interpret arousal according to the two factor theory of Schachter Singer (1962), assuming that arousal is the product of physiological arousal and the cognitive processes that respond to a particular situation (=context and stimuli) provoking the emotion. The cognition of the user will determine whether the physiological arousal will be perceived as anger or happiness. As sensors can detect arousal but not the valence of the emotion, the WELLS model also adds contextual data in order to enrich and by so interpret the captured bio data better. Measuring arousal will detect positive or negative experiences and therefore identifying moments and actions where a prototype can be improved on.

Various physiological measures can be used to indicate arousal, among which blood pressure, heart rate, heart rate variability (HRV), electrodermal activity (EDA), pupil diameter (Sano & Picard 2013) as well as facial expression (Albert & Tullis 2013). However, not all of these measures can be collected using unobtrusive devices. For example the measurement of pupil dilation requires the constant training of a camera on the subject’s pupil. By investigating current hardware offerings, we found that what can currently be measured in the least obtrusive way are electrodermal activity and heart rate.

Electrodermal activity
“Electrodermal activity is a measure of sweat excreted by the eccrine glands, which are innervated by the sympathetic nervous system”. (Hedman et al 2012). Electrodermal activity has been used widely to measure the response of the sympathetic nervous system to events in our environment. In the past, devices that were used to do this were often cumbersome to the wearer, resulting in feelings of stress resulting from wearing the device itself. Recently, devices like Philips’ Discrete Tension Indications (DTI-2) or the Empatica E4 Wristband have become available that allow EDA measurement by wearing a relatively unobtrusive, watch-like device. The latest in wearable fitness trackers, like the Jawbone UP3 wristband allows to measure electrodermal activity through the use of a device type that is becoming more and more mainstream on the consumer market. These mainstream devices will reflect a natural setting for participants and will allow Living Lab researchers to maintain a real life research setting.

Heart rate
Heart rate and heart-rate variability, or the fluctuations in beat-to-beat intervals, can be used to detect stress and arousal (Albert & Tullis 2013, Choi et al 2012). There are more devices that can measure heart rate than EDA, making it a pragmatically useful addition to the WELLS architecture. Indeed, heart rate can be measured by mainstream fitness wearables by wearing a sensor band on the torso. However, wearing such a band is obtrusive and few are the people who would want to wear it permanently, beyond their fitness workout. But as for EDA, new devices are becoming available that can be worn on the wrist and that also measure heart rate.

Context data
Video and audio
Video can be used to capture the interactions of the user in a rich way and has been proven to be highly valuable to detect possible improvements of the user experience. Within usability research, video has been used in various way to capture and detect user interactions and experience. However, this has either been in a controlled lab environment or with very obtrusive equipment when performing this in-situ. For a pervasive game called “Playground”, we collected user data with smart glasses, as they are able to capture the users context when using the system in a rather easy and less obtrusive way. Players were asked to play the game on smartphone, while wearing these smart glasses and performing the think aloud protocol. The latter is a standard method in which participants thinking aloud as they are performing a specified amount of tasks. Because the data is recorded, developers can look back at the experiences and reactions of the users and interpret the data. As pervasive games are not only about the use of the game on a mobile device, but also about the experience of the player with the physical environment in which the game is set. By using smart glasses, both dimensions (use of device and interaction with physical context) could be captured in rich detail.

Our experience within the Playground project indicated that the use of smart glasses is promising but still a number of issues need to be tackled. First, they are less unobtrusive than expected. For example, Google Glass constantly shows a display in one’s field of vision. It is therefore hard for the user to “forget” the presence of the recording device. In addition,
battery life can be a hurdle in smart glasses. The battery life time of Google Glass is limited (one hour of video recording), making long experience sessions difficult. In addition, there were issues with the video framing. As the camera in the Google glass is located in the top-right corner of the glass frame, the recorded video did not always catch the interface on the smartphone. Finally, often the video was not focused on the interface, but on other objects in the field of view. This resulted in blurry recordings of the interface, making it hard to identify certain interface elements.

An alternative to using smart glasses is devices that are used for “LifeLogging” or the long-term recording of video in daily life. One example is the Narrative clip 2, which allows up to 30 hours of video recording through a device that can be clipped on clothing and is around 10 by 10 cm large. However, many of these devices do not record audio. Our experience within Playground learned that audio is often essential as extra information layer to allow users to express their frustration and provide additional context on what it is that they are experiencing.

**Experience sampling**

Experience sampling is a research methodology in which users are asked to self-report on various elements (such as experience, context,...) on specific moments (this could be time-based, event-based,...). Huang et al (2014) and Sano & Picard (2013) e.g. used experience sampling, via short questionnaires distributed through a smartphone app, to query the user on their emotional state of mind. In their research they demonstrated that gathering survey data can be of great use in identifying emotions such as stress related to a specific context. The captured data also linked with a whole batch of collected metrics like EDA, heart rate and others. However, in systems evaluations in living labs, taking place in real-life, naturalistic environments, the identification of stress needs to take place within a brief time interval from the event that caused the affective response. This needs to be done to be able to identify instants that caused an affective response, making aggregate experience sampling per day ineffective as retrospective self-reporting will not reveal these specific moments. Therefore, more fine-grained approaches need to be included to measure affect. Two options are possible: (1) based upon real-time data of the wearables in which the reporting is triggered by certain data-points. But this would require a permanent online connection and could also result in too many interactions and request. (2) based upon predefined intervals in which the user is asked to evaluate her aggregate affect level in a previous time interval. Such an approach would not produce labels on the exact moments when stressful events took place, but being able to search down the search space would be useful in analysing the large amounts of data that will be produced by a large-scale evaluation using the WELLS infrastructure.

**Location, event logs and time**

Keeping track of location is important to be able to gather context data. Such location data can come from geolocation based on GPS, wifi or bluetooth data as is often done in mobile, geolocative apps. A well-known drawback of such apps is the rate at which they drain a smartphone’s battery. Situations where location is less accurate but more energy efficient can be built based on iBeacon solutions. In such cases, the captured data only shows in what general area the user resides. An infrastructure cost is associated with such a solution, as the area in which the user interaction takes place needs to be fitted with iBeacons. Next to location, movement is also an important contextual element, which can easily be determined due to the accelerometers within smartphones and wearables. This data stream is also important, as it can determine if a user is moving or not. Keeping event log data is another important aspect of evaluating data during in the wild trials. Such log data can indicate bugs and can provide useful context information to be able to further make sense of the user’s behavior. Finally, keeping accurate time is essential for synchronizing data in the research interface when data is coming from multiple sensors (Banaee et al 2013). Therefore, having timestamps that are in sync for all measurements is critical.

**Smartphone**

The smartphone plays an important central role. Firstly, most wearables need a smartphone to make connection to various (cloud-based) services. Therefore it can be used to bundle the incoming data and send it over to the data repository in an efficient way (stability, robustness...). In the future, wearables may become directly connected to internet, bypassing the need for smartphone. Secondly, the smartphone can help in keeping all sensor data sets synchronized.

**Data repository**

The data repository should store the data coming from as many sources as required. It should be able to receive real-time or batch data from the smartphone. In addition, this data should be stored in such a way that privacy requirements are met. It is important to keep in mind that other metrics can be of use when measuring user affect. As more wearable sensors will become available, it will be important to make the architecture as extensible as possible, to be able to quickly add new sensors to the system. The
data repository, and the way in which it is accessed, provides the main point in the architecture where such future compatibility needs to be realised. This will require flexible API’s that can accommodate data coming from a wide array of sensors as well as an efficient identification mechanism.

**Research interface**

The purpose of the research interface is to allow user researchers to make sense of the data that is produced by the sensors and the smartphone. Especially relevant are moments in the user’s experience that have generated arousal while using the application being evaluated. In order to support sense making by user researchers, the WELLS data should be visualised, mined and made queryable.

![Figure 3: Wearable Living Lab research interface mockup, adapted from Kocielnic et al (2013). Galvanic Skin Respons (GSR) is synonym to eletrodermal activity (EDA). Foto is a still from a Google Glass recording of a user session in Playground Kortrijk](image)

**Data mining**

The amount of data produced by sensors over a longer period of time can be very large. In order to make sense of such sensor data, which comes in the form of a time-series, analytical methods are necessary. Banaee et al (2013) provide a survey of the methods used in data mining and machine learning in the area of medical applications, that is relevant to the aims of this paper. In their literature review, they found support vector machines, decision trees, neural networks, hidden markov models, gaussian mixture models and rule-based models to be in use.

It is important to distinguish EDA and HRV fluctuations resulting from movement from electrodermal activity resulting from emotional response. This can be achieved by applying rule-based filtering, one of the main data mining approaches discussed by Banaee et al (2013). The data from the 3-axis accelerometer is crucial in distinguishing both types of EDA and HRV response. Rules can be built that disregard EDA data from periods in which the user was moving along one of the 3 axis in a value that exceeds a certain threshold.

The main objective in the analysis of data coming from user experience evaluation in living labs should be on finding instants that coincide with increased arousal from users. These instants can then be labeled as possible candidates for further analysis. By juxtaposing data that has been labeled in such a way from EDA or HRV with video, audio and location data, a rich picture will emerge of what features and contextual circumstances can be stressful or enjoyable to the user.

**Visualisation**

In order to be useful, the system needs to visualize the data in a way that facilitates sense making. The video and audio data will be the main way in which the user researcher will come to understand what exactly was going on at a particular point in time. However, the video and audio data of a certain evaluation can be extensive and if the trials are large-scale, user researchers can not be expected to go over all the produced video and audio for each participant. Therefore, combining the video/audio feed with data mining techniques on data coming from other sources can be powerful to discover moments in the data where the user got frustrated with using the software.

Kocielnic et al (2013, 2014) experimented with collecting work stress-related data using EDA over a prolonged period of time. A visualisation was created that combined sensor data from EDA and accelerometer with labels that were extracted by interfacing with online agendas. Not only did this approach deliver an instrument for the analysis of the data, it was also able to present the findings back to the user and in this way obtain better self-reporting. By combining a visualisation of EDA data with calendar data and questionnaire data, users were able to make sense of their experiences in terms of the relative stressfulness of different events in their day-to-day professional life.

**Privacy issues**

Continuous tracking poses many privacy challenges that neither users nor stakeholders are ready to deal with. Wearables constantly collect, transmit and store data that is considered personal, private, sensitive and confidential by users. This brings many benefits to the researchers, as it can provide them with access to end-users’ latent experiences and fill the gap between the saying versus doing paradigms. Yet for users this constant surveillance and sousveillance can lead to privacy threats and risks. Sousveillance is the recording of an activity by a participant in the activity typically via small wearables or other small portable
technologies (Mann, Nolan & Welmann, 2003). Users are currently less aware of these privacy concerns because wearables are a relatively new thing. In addition, wearables are often shaped as accessories that never posed any privacy threats to users such as glasses, watches, etc. Motti & Caine (2014) identified several privacy concerns from users about wearables related to the device and the application itself, the data being captured and the sensors. Especially their last two concerns are of importance to living lab research. Data specific privacy concerns are related to the issues of users that certain data, when combined, could have critical implications. Next to having the explicit consent of the users, it will also be necessary to inform the users on which type of data will be collected during the scope of the research, what will be done with that data as well as to foresee the ability to delete all data afterwards (in line with the EU policy on the right to forget).

**Discussion**

We have described the general architecture of a wearable living lab and touched upon certain opportunities and challenges. However we are still in the phase of putting the integrated WELLS architecture into practice. The benefit of WELLS is that there are no aspects of the architecture that can not be implemented using today’s consumer-grade wearables technologies, web-based programming languages, database systems, machine learning and visualisation techniques. What remains to be investigated, however, is the accuracy of the different devices that are and will become available and the quality of the data. In addition, some other concerns exist in terms of battery power, data heterogeneity, form-factor, ruggedness and ease of use. **Battery power** has been a constant issue in many recent technology developments. Certain applications, like geolocalisation, can drain a smartphone battery at a much increased rate compared to a normal smartphone application. The WELLS system will only remain operational for as long as all of its components remain active. How long we can expect this to be still remains unclear, but is an important unknown to figure out in order to allow long-term measurements of user experience. **Data heterogeneity** will result from different devices creating different data in a variety of formats. Handling this so that for example all data feeds can be combined and accurately synchronized is a hurdle which we still need to tackle. **Form-factor** is an important aspect, as it influences the obtrusiveness that is experienced by the user. For example the form factor of the Google Glass proved to be sub-optimal, as users would not feel comfortable with wearing the device. For devices in the wearable living lab to function well, they should be as invisible as possible, both for the user as its environment. Video-capturing devices are available that allow more discreet video-recording, but many of them also have their limitations, like lack of audio recording capability. Still, previous research on wearables shows that even for more common devices such as wristbands it takes a while for users to get accustomed to the device and start behaving naturally (Hedman, 2011). Therefore we will need to investigate how the use of a WELLS architecture impacts the natural use of the systems under investigation. **Rugedness** is important, as the wearables living lab needs to remain operational in various environmental condition. Especially continued active when it is raining seems essential. Finally, the WELLS architecture should be coordinated through an app that runs on a smartphone and that is able to handle all the interactions that are needed with the user. As with all applications, it will be necessary to make sure that the user experience design of this app, and the way in which it integrates with the wearables, is created in such a way that it is as **easy to use** as possible.

The WELLS architecture can be of use in each of the three living lab phases discussed above: a PTA, a formative prototype evaluation or a summative prototype evaluation. WELLS will be most useful in situation in which mobile technologies are evaluated. Indeed, these are the technologies that are hard to evaluate in usability lab conditions, as many of the user’s reactions will result from the combination of the software and the context in which the user resides. Ballon (2015) found that, although there is heterogeneity in the existing living lab approaches, they share four characteristics: (1) the discovery of unexpected usage and new service opportunities, (2) the evaluation or validation of new digital technology solutions by users, (3) a familiar usage context, and (4) a medium or long term research angle. We believe it is clear that WELLS can add novel aspects to each of these characteristics but can not go into detail on how this is the case due to space restrictions.

We plan to use this architecture in many of our ongoing research and development projects. One example is a pervasive city game in which players can use their smartphone to play games in the urban environment. Another is a project in which the aim is to provide resistive schizophrenic patients with cognitive behavioral therapy over wearable devices. Both projects are examples that will need users to experience the system in a prologued way, underlining the need for a research approach that can collect, analyze and visualize in the wild and continuous data, originating from an extensible set of sensors.

**Conclusion**
We have presented an architecture for the wearable living lab, a system to collect user experience feedback in living lab projects. The user researcher is the main consumer of the WELLS data and can use the system in a living lab project during a PTA, a formative or a summative software evaluation. Its relevance for Living Lab or other in-situ research resides in its capability to capture contextual and personal data of end-users in real life environments. The next steps are to build, gradually, a concrete instance of WELLS and to explore and evaluate the different elements addressed in this paper. This will entail creating a system that can support a living lab research approach (and in-situ research in general) that can collect, analyze and visualize in the wild and continuous data, originating from an extensible set of sensors.

References
Tang, T., & Hämäläinen, M. (2012). Living lab methods and tools for fostering everyday life innovation. In Engineering, Technology and Innovation (ICE), 2012 18th International ICE Conference on (pp. 1–8). IEEE.