mQoL Smart Lab: Quality of Life Living Lab for Interdisciplinary Experiments

Alexandre De Masi, Matteo Ciman, Mattia Gustarini
Quality of Life lab,
University of Geneva, Switzerland,
{Alexandre.DeMasi, Matteo.Ciman,
Mattia.Gustarini}@unige.ch

Katarzyna Wac
Quality of Life lab
University of Geneva, Switzerland
HCC Group
University of Copenhagen, Denmark
katarzyna.wac@unige.ch
wac@di.ku.dk

Ubicomp/ISWC'16 Adjunct , September 12-16, 2016, Heidelberg, Germany
 ISBN 978-1-4503-4462-3/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2968219.2971593

Abstract
As a base for hypothesis formulation and testing, accurate, timely and reproducible data collection is a challenge for all researchers. Data collection is especially challenging in uncontrolled environments, outside of the lab and when it involves many collaborating disciplines, where the data must serve quality research in all of them. In this paper, we present own "mQoL *Smart Lab*" for interdisciplinary research efforts on individuals' "Quality of Life" improvement. We present an evolution of our current in-house living lab platform enabling continuous, pervasive data collection from individuals' smartphones. We discuss opportunities for mQoL stemming from developments in machine learning and big data for advanced data analytics in different disciplines, better meeting the requirements put on the platform.

Author Keywords
Data Science; Data Analysis; Data Collection; Platforms; People Centric Sensing; Smartphones

ACM Classification Keywords
H.4.m Information Systems Applications: Miscellaneous

Introduction

In the last years, smartphone penetration in individuals' life enabled acquiring data about their behaviors and context for extended periods of time and in their natural environments, i.e., "in-situ." Social and behavioral, as well as medical, scientists, are now leveraging it –performing experiments in which it is necessary to assess individual's state repeatedly "in-situ". They often require collecting data throughout the individual's day in a non-invasive manner, i.e., without influencing their lifestyle and habits. To design and develop new experiments and to be able to understand and model the variability of human behavior "in-situ," a *simple computer science approach* to user studies, in which we (computer scientists) collect data, derive algorithms, and then experiment with them, providing final answers to social and behavioral scientists, is not sufficient. It becomes necessary to introduce an interdisciplinary approach and to build an adequate technological infrastructure to let social and behavioral scientists acquire accurate, timely and reproducible data. We motivate this statement given our research experience, as follows.

In March 2012 we have launched the *mQoL Mobile Computing and Communications Living Lab* [1,2] in our Quality of Life group at the University of Geneva. It enables researchers to run studies and collect data from smartphone users in the Geneva region in Switzerland. The Living Lab has more than 40 individuals participating in different studies; some of them are participating in the lab since 2012 to date. We have published several results about mobile computing technologies, Human Computer Interaction and behavior assessment "in-situ" - from Quality of Service (QoS) [3], Quality of Experience (QoE) [4], to intimacy [5], stress [6] and context awareness [7]. Even if the mQoL platform performed well for our experiments and data collection, along our research efforts we understood that to fulfill the new requirements of the interdisciplinary approach in data collection and experimentation a full redesign of the current platform is necessary. The ultimate goal of the platform is to enable experimentation that facilitates the improvement of Quality of Life of its users. In this paper, we present the current *mQoL Living Lab* platform, its current shortcomings and the redesign and prototype of the new *mQoL Smart Lab*. The redesign is specifically aiming at improving the platform scalability and flexibility, which will facilitate interdisciplinary research efforts.

Some other platforms for "in-situ" experiments and data collection exist. For example, the University of Buffalo "Phone-Lab.org" platform [8], or the "AWARE" framework of the University of Oulu [9] are frameworks and infrastructures for data collection. They mostly focus on research aspects in computer science like scalability, accuracy and privacy-preservation for collecting vast amounts of raw data. OpenmHealth.org is a longitudinal framework aiming at collecting raw data related to the individual health and wellbeing, but it is intended only for computer scientists collaborating in diverse mhealth projects. The MITs Human Dynamic Lab [10] focuses on computational social sciences; however, it does not release its framework for data collection or the data itself. Finally, the Kalvi's HUMAN [11] project focuses on socio-demographics and urban factors influencing individuals' life, involving 2.500 families for 20 years. The biggest shortcoming of all these platforms is that they do not embed in their

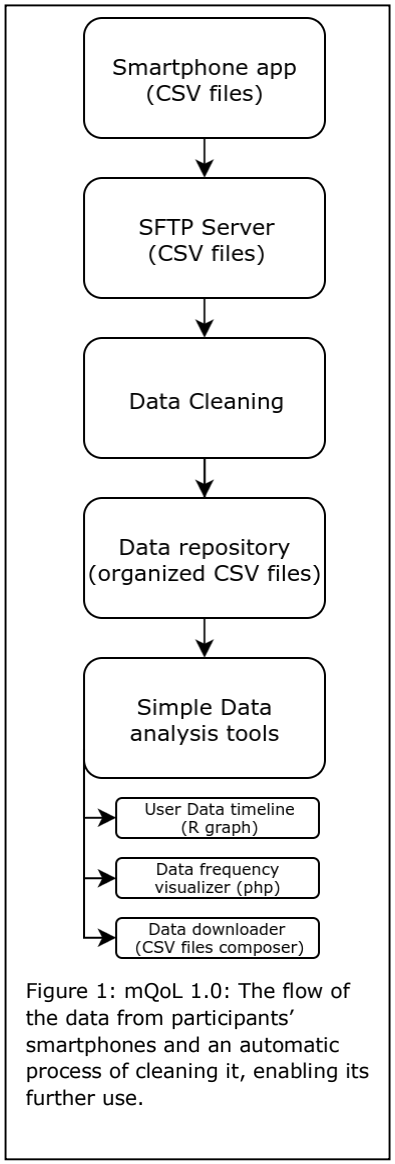

Figure 1: mQoL 1.0: The flow of the data from participants' smartphones and an automatic process of cleaning it, enabling its further use.

design an interdisciplinary approach, but focus only on data acquisition and the specific data analysis.

Current Platform: *mQoL Living Lab* 1.0

The platform comprises two main components: the *mQoL logger* and the *mQoL server*. The mQoL logger is an Android app installed in our living lab participants' phones. The app collects the data we are interested in, including, but not limited to screen touches, screen ON-OFF events, apps use and duration, wireless access network type and delays. The mQoL logger synchronizes with the server, which then processes the data and allows us to create simple visualizations and download the data for further analysis (the simple data flow is in Figure 1). We built this platform at the beginning 2012 given the requirements of research projects at that time.

mQoL 1.0 platform requirements (2012)

The main requirements for the current platform were: fast to build, own hosting, security and privacy, reliability, and being non-intrusive for study participants. It had to be fast to build because of an urgency of own research. We avoided any complex server logic and used SFTP (security requirement) to send the data stored as simple CSV text files from the mQoL logger to our server (Figure 1). To comply with privacy laws, we hosted the data in our university servers. To further preserve the privacy of our participants, we do not store any personal identifiers along with the data on the server. The data collection and transmission had to be reliable. Therefore, we decided to store data on users' devices first and transfer it only over WiFi. We do not remove the data from the users' devices until the uploading operation is successful. Otherwise, we retry the whole upload. The mQoL server has a nightly process that cleans the data from duplicates and broken records (Figure 1). The cleaning operation prepares the data for the simple visualization engine and easy download (e.g., we can select time periods, data kind, and users' identifier to download CSV ready for the analysis).

mQoL 1.0 platform limitations

Having this platform for the experiments was precious for our simple computer-science driven research along the past years; we also understood its limitations. Right now, we have the goal of supporting an interdisciplinary approach, including use of the platform for experiments run by psychologists, sociologists or economists, and at the current stage, we have two main challenges: scalability and flexibility. The current architecture does not allow us to scale to more than 80-100 users. We cannot distribute our *mQoL logger* out of the controlled environment of our living lab. Server resources are limited (e.g., memory and CPU), and the data collection through SFTP in CSV files does not allow us easily to scale our experimental and data analysis needs. The platform is not flexible enough: neither we (nor our interdisciplinary collaborators) cannot easily plug in new components such as event-based collection triggers, new experiments (with open to the world onboarding campaigns), machine learning tools (ML), more complex visualizations engines, and so on. We decided to evolve our current platform to a next level.

New mQoL Platform: *mQoL Smart Lab* 2.0

The interdisciplinary approach goal is to allow any type of researcher to experiment with datasets and run different algorithms, e.g., a sleep analysis from smartwatches data could be process in a few clicks. We identified new components able to fulfill the new

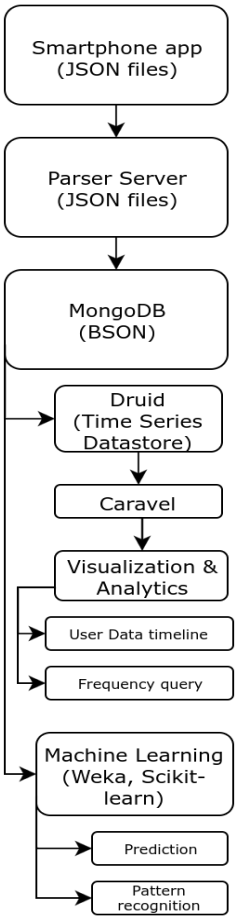

Figure 2: The flow of the data in the mQoL Smart Lab platform. Data comes from participants' smartphones and the Parse server manages the data cleaning, storage, visualization and so on.

requirements, as follows: server (Parse – see sidebar for links), database (MongoDB), data visualization tool (Caravel), ML tools (TensorFlow, Scikit-learn, and Weka), experiment builders (our own modular Android library, Gradle and Google Play API), and marketing tools (Facebook Ads and Google Adsense). All of these components enable scalability and flexibility, as follows. The updated data flow is presented in Figure 2.

Parse: Open source cloud solution (scalability)
Parse is an open source backend solution for infrastructures requiring scalability. It is based on Node.js and requires a MongoDB database. Parse supports user authentication, enabling creation of an anonymous login for study participants. It also features a key based client verification to verify the genuineness of the data coming from participants. It allows sending notifications to smartphones, to communicate with study participants (e.g., notify them about a new experiment). It also features live queries to subscribe to database updates and trigger events in the server or client apps. This feature allows us to automate future data analysis tasks, or to respond to specific events with actions. Parse offers clients API (e.g., Android, iOS, PhP, Javascript, and more) to easily implement the data transmission between our participants' smartphones (or any other device we may use) to our Parse instance.

MongoDB: Database (scalability and flexibility)
The new architecture will use MongoDB, a NoSQL database. It offers features that our previous data management scheme does not have: high scalability, load balancing, fault tolerance, replication, aggregation, and querying. Multiple programming languages support access to MongoDB database and allow for fast processing of the data. For example, our collaborators in e.g., psychology and sociology would get access to data via their regular interfaces (e.g., R) enabling them to run data analytics and conduct statistical significance tests directly on the data. Since most of the collected data are time series, we are also exploring the area of Time Series Data Store (TSDS), a special class of NoSQL database highly optimized for handling time series data. Overall, having different programming languages supporting the access to the MongoDB database, it is easier for researchers with diverse backgrounds and data analytics skills to have direct access to the data source and the database.

Visualization tools (flexibility)
The interpretation of data is often the next step after the collection procedure. We may need to visualize complex data to detect a particular trend, or to allow the researcher to validate or refute their hypothesis. There exist new powerful open-source technologies to deal with big data visualization coming from millions of users. For example, AirBnB Inc. designed Caravel as a data exploration platform to be visual, intuitive and interactive. It allows the researcher to write multiple, easy, and fast query to a database and output the data in a powerful visualization form, e.g., distribution of mobile applications usage of participants aged between 18 and 23.

Machine Learning (scalability and flexibility)
Machine Learning (ML) enables the design of complex algorithm simplifying manual data analytics task (which we have followed so far). We plan to integrate ML into our data analytics platform, enabling the platform to provide predictive services for the researchers running the experiments with participants. For example, one of

Platforms in this paper

OpenmHealth Project
http://www.openmhealth.org

Parse
https://parse.com/

MongoDB
https://www.mongodb.com/

Caravel
https://github.com/airbnb/caravel

TensorFlow
https://www.tensorflow.org/

Scikit-learn
http://scikit-learn.org/

Weka
http://www.cs.waikato.ac.nz/ml/weka/

Gradle
http://gradle.org/

Google Play API
https://developers.google.com/android-publisher

Facebook Ads
https://www.facebook.com/business/products/ads

the future goals of the platform will be to predict the application use experience of nomad smartphone user via ML, or to assess intimacy or stress levels of individuals "in-situ".

Experiments Builders (flexibility)
We envision a special module that allows researchers of any discipline without coding skills to create their apps to collect data and perform "in-situ" studies. The tool, e.g., a web platform, will use our modular Android library to package the sensors in the app, Gradle to automatically create an Android apk file and Google Play API to deploy the app for distribution. For example, an economist could create an app to perform self-reports based on the Experience Sampling Method (ESM) [12], about situational marketing, launch it on the Google Play store and distribute it to participants.

Targeting Potential Users (scalability and flexibility)
One of the difficulties to run studies is to attract a representative group of participants. In the past, we used the mQoL website to promote the living laboratory and the benefits of participating in our studies. Our new approach consists of using the powerful online ad platforms, e.g., Facebook Ads, Google Adsense, to display ads for specifics studies to a representative part of the target population. For example, a psychology researcher could need twenty new participants between 20 to 45 years old who live in the Geneva canton and are likely buying luxury goods. With Facebook Ads, we can send the ad about the new study directly on Facebook feeds of people fulfilling these criteria.

Discussion and Conclusive Remarks

Based on the previous experience and the new requirements to support an interdisciplinary approach, especially including social and behavioral research, it became apparent that we require a new platform. Therefore main goal of the mQoL Smart Lab approach is to redesign the existing smartphone-based research platform (MQoL Living Lab) to able to support the collaboration between scientists from different research fields, e.g., computer science and social and behavioral sciences. In particular, the platform will take into account the fact that different research areas, in the future, could require the integration of various modules, visualization tools, and machine learning algorithms. It will support the ad-hoc add-on of new instruments for data collection analysis, to improve the collected data quality and timeliness, and reproducibility of the results acquired based on this analysis.

Thanks to the interdisciplinary approach and the collaboration with researchers from different research areas, the *mQoL Smart Lab* will boost behavioral and computer science research. We will develop interactive, smartphone-based behavior assessment technologies, tools, and algorithms that are unobtrusive, that we can deploy "in-situ" and longitudinally, and allow to combine self-reports, experimental tasks, and data from the smartphone sensors. Using this platform, we will perform large-scale innovative research in different areas, by bringing experimental paradigms into real-world situations. All these insights will provide an essential basis for research towards enhancing people's life quality following the mission of our lab.

The *mQoL Smart Lab* platform we research aspires to become a base upon which a whole interdisciplinary ecosystem can be build.

References

1. K. Wac, M. Gustarini, J. Marchanoff, M. Fanourakis, C. Tsiourti, M. Ciman, J. Hausmann, G. P. Lorente, "mQoL: Experiences of the 'Mobile Communications and Computing for Quality of Life' Living Lab", in Proceedings of the *International Workshop on the Living Lab Approach for Successful Design of Services and Systems in eHealth (LivingLab'15), co-located with the IEEE HealthCom 2015*, Boston, USA, October 2015.
2. M. Gustarini, S. Ickin, K. Wac, "Evaluation of Challenges in Human Subject Studies "In-the-Wild" Using Subjects' Personal Smartphones", in Proceedings of the *International Workshop on Ubiquitous Mobile Instrumentation (UbiMI)*, co-located with the *UBICOMP* conference, Zurich, Switzerland, September 2013.
3. K. Wac, G. P. Lorente, M. Gustarini, J. Marchanoff, "More Mobile & Not so Well-connected yet: Users' Mobility Inference Model and 6 Month Field Study", in Proceedings of the *7th International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT 2015)*, Brno, Czech Republic, October 2015.
4. S. Ickin, K. Wac, M. Fiedler, "Quality of Experience-Based Energy Reduction by Controlling the 3G Cellular Data Traffic on the Smartphone", In Proceedings of the *22nd International Teletraffic Congress Specialist Seminar on Energy Efficient and Green Networking (22 ITC SSEEGN)*, co-located with the *ATNAC*, Christchurch, New Zealand, November 2013.
5. M. Gustarini, K. Wac, A. Dey, "Anonymous Smartphone Data Collection: Factors Influencing the Users' Acceptance in Mobile Crowd Sensing", Personal and Ubiquitous Computing, 20(1), 65-82, Springer, 2015/2016.
6. M. Ciman, K. Wac, O. Gaggi, "*iSenseStress*: Assessing Stress Through Human-Smartphone Interaction Analysis", In Proceedings of the *9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, Istanbul, Turkey, May 2015.
7. M. Gustarini, K. Wac, "Ubiquitous Inference of Mobility State of Human Custodian in People-Centric Context Sensing", in Proceedings of the *International Workshop on Ubiquitous Mobile Instrumentation (UbiMI), co-located with the UBICOMP conference*, Pittsburgh, PA, USA, September 2012.
8. A. Nandugudi, A. Maiti, T. Ki, F. B. M. Demirbas, T. Kosar, C. Qiao, S. Y. Ko and G. Challen, "PhoneLab: A Large Programmable Smartphone Testbed", in Proceedings of *SENSEMINE'13* (2013).
9. D. Ferreira, V. Kostakos and A. K. Dey, "AWARE: mobile context instrumentation framework", in *Frontiers in ICT* 2015.
10. MIT Media Lab: Human Dynamics Group, [Online] Available at: http://hd.media.mit.edu/, [Accessed 7 Jun. 2016].
11. Kavli HUMAN Project, [Online] Available at: http://kavlihumanproject.org/, [Accessed 7 Jun. 2016].
12. J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi, "Experience sampling method: Measuring the quality of everyday life", Sage, 2007.