A data architecture for research in mobile health

Anna van der Zalm
Robbert Jan Beun
Sandor Spruit

Technical Report UU-CS-2013-006
May 2013

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

A DATA ARCHITECTURE FOR RESEARCH IN MOBILE HEALTH

Anna van der Zalm
Utrecht University
a.vanderzalm@uu.nl

Robbert Jan Beun
Utrecht University
r.j.beun@uu.nl

Sandor Spruit
Utrecht University
a.g.l.spruit@uu.nl

Abstract.

Using smartphones in healthcare research is an upcoming topic, but it raises issues in terms of security and privacy. We have a case in which we want to research a mobile application, using an online participant registration system and online questionnaires. In this article we describe the problems that we faced building a research data architecture and the solution we have created.

Keywords: *M-health, data architecture, security, research data*

1 INTRODUCTION

Mobile health or m-health, as an enhancement or even replacement of e-health and regular healthcare, is becoming more popular; the amount of applications for smartphones ('apps') that have a medical purpose is growing rapidly(Dolan, 2012).

The percentage of people owning a smartphone has been steadily growing, in the second quarter of 2012 58% of the population owned a smartphone (Telecompaper 2012). The availability and ubiquity of smartphones makes applications a suitable medium for therapies and interventions in which face to face contact is not required. M-health faces the same problem as e-health: there has been a proliferation of applications, regardless of real user needs and hardly any treatment or technique of these products is evidence-based (Choe et al. 2011). Therefore, the effectiveness, both in health outcomes and in costs, needs to be researched.

To effectively and securely research the user needs and effectiveness of a smartphone health application, an extensive research data architecture is needed. This architecture should support a secure and reliable data flow from different sources and should also ensure privacy and accessibility of various stakeholders.

In this paper we will present a data architecture for research that was developed in the context of the Sleepcare project. In this project, we aim to implement cognitive behaviour therapy for insomnia (CBT-I) into a smartphone application. We want to construct a scientifically-proven framework that integrates persuasion strategies for sustainable behaviour change in the domain of sleep and the technology to support these strategies. Previous research shows that self-help CBT-I can be effective (Vincent & Lewycky 2009) and as described in Free et al. (2010) and Beun (2012), there could be many benefits in using smartphones for CBT and CBT-I in particular. These benefits include the assessment of relevant momentary information, the generation of personalized feedback, the obtainment of sleep data by sensory measurement and information exchange between various stakeholders, such as peers and caregivers.

We are in the early stages of the project, our starting point is the communication between the user, the application and the researcher but in future implementations we want to involve other stakeholders. The stakeholders and communication lines are displayed in Figure 1.

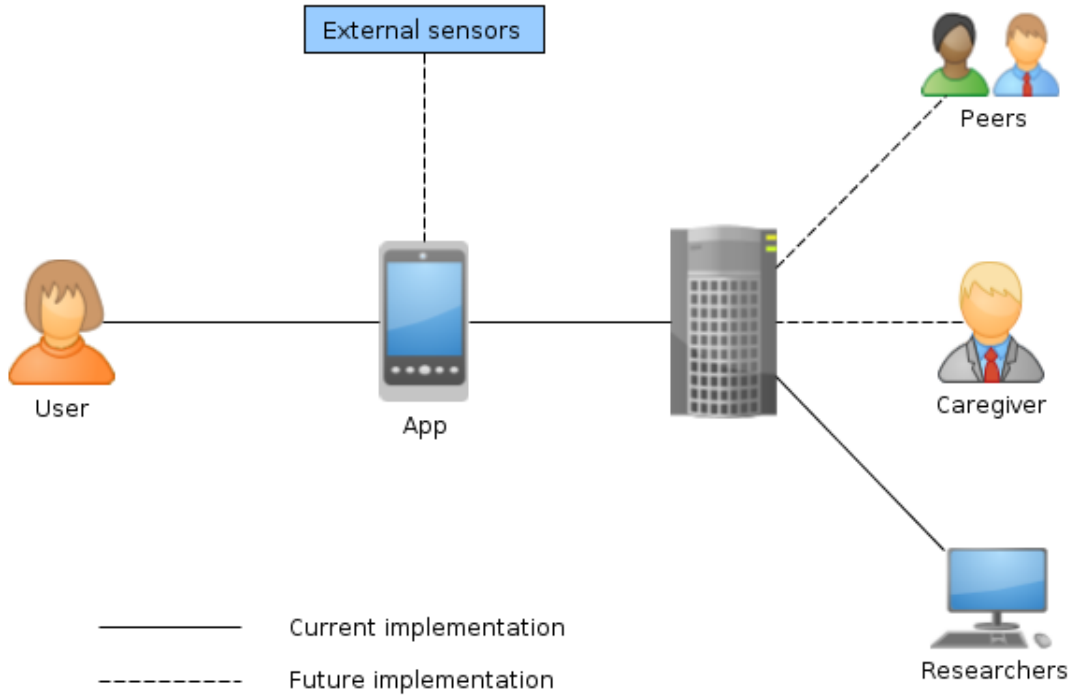

Figure 1. Basic information flow between different stakeholders

We follow an incremental approach in the design process with short evaluation cycles. Each year the design will go through several iterations where the requirements baseline will continuously be refined as new insights are acquired through (formative) prototype evaluations and reviews with different stakeholders. In the final year of the project, a summative user study will be performed in a Randomized Controlled Trial (RCT). Each evaluation circle includes one or more user tests.

Testing an application requires that multiple participants, meeting the criteria for a specific test, install the application on their phone, use it for a certain period of time and answer relevant questionnaires. To do these type of tests, we need the aforementioned data architecture. In what follows, we will describe the requirements of this architecture, discuss the design used to meet these and finally describe its implementation.

2 REQUIREMENTS

To do aforementioned user studies we have to integrate information from different sources, namely:

- Mobile application data transfer (user input and sensor/usage data)
- Online questionnaires
- Online registration

In this section, we will explain why we convert these sources to specific elements in our architecture and what their requirements are. The communication of data from these elements towards the researchers should be secure; the privacy of participants is important, not only from an ethical point of view, but also to increase participants' willingness to partake in the user studies. Security and protection of privacy are the main requirements of the architecture.

2.1 Mobile application data

Our user studies require information from the mobile application; more specifically the user input, data on the actual usage of the application and data gathered by sensors, both those embedded in the phone and in a later stadium, external sensors, such as actigraphy monitors. These data have to be registered within the application and sent to the researchers over the Internet. During a study, the application should send data to the researchers, if consent is given by the participant. Also, researchers should be able to request the data if needed.

It is important to pay attention to how the data is saved in the application. For example, in case of self monitoring, a user registers information everyday; he or she might make a mistake in entering this information and correct it. The application should be designed in such a manner that edits (and the time on which an edit is made) are saved, but the latest data is visible for the user. From the data on edits we can deduct whether there are problems with the user interface but they can indicate insecurity or socially desirable answers of the user.

2.2 Online questionnaires

A second type of information that is needed to perform an effective user study is the users' opinion on their experience with the application. To gather this information, questionnaires need to be used. Questionnaires can also be used to assess user wishes and needs.

A number of online questionnaire systems already exist, which we can use (e.g., Wufoo, NetQ, Surveymonkey). It would be preferable if the data from such a system could be automatically transferred to the researchers, but it is acceptable if a system requires a manual action (such as logging in and/or downloading an export file), as long as this manual action does not compromise the security of the system. Accepting systems in which a manual action is required makes the architecture more flexible, because most systems do not offer this functionality. It also allows for multiple systems to be used. There are considerable differences in functionality between existing systems, so the ability to use different questionnaire systems might be beneficial.

2.3 Online registration

As described there will be multiple user studies, for which we will need different participants. Sometimes one participant can partake in several studies and some studies will have specific inclusion criteria (such as diagnosed insomnia). To be able to easily call upon the participants who meet the criteria for a certain study, we want to create a database to register participants and their relevant in- or

exclusion criteria. We plan to recruit participants online (via websites, fora and social media), so they should be able to register online.

To enable online registration, we need a website which allows people to register and where they should also be able create an account, so that participants can easily contact us and have a place where we can provide information and results to them.

3 DESIGN

The data from the elements discussed in the previous section needs to be securely transferred to one place and that place must only be accessible to the right stakeholders. In this section, we will describe how we designed the database to store all data from the different elements and what architecture we used to establish a secure but fluent data flow.

3.1 Database

To use data from different elements in research, researchers need to link the data from the different sources. However, to protect the privacy of participants, it should not be possible to trace back the data from the different elements to a person or to data from a another element.

To ensure this separation between data from the different elements, our database design consists of two main tables, a research-table and a so called key-table. The research-table consists of all research data from the elements, saved as it was received. The key-table consists of the identifiers for each of the three elements. As soon as someone is registered, a research number is created to identify the person.

Identifiers for the questionnaires and the mobile application are randomized and saved in the key-table.

Each entry in the key table also has a field that contains permissions. By default, these are set to the research settings of which people are made aware in a informed consent. When we, in a later stage, implement access for other stakeholders than researchers, we can create permission settings for certain roles (such as: care givers should not get permission to access questionnaire data, but should get access to application data. Researchers should get access to all data, but anonymized). If it proves to be useful, a permissions-role can even gain access to a specific part of the application or questionnaire data. This enables researchers of different projects to work in the same system, but it can also be used to create the system of peer support, in which peers have access to a summary of the application data.

The key-table can be extended with columns which include who the user has permitted to fulfill certain roles. Again, the user is the manager of these permissions and can request to withdraw consent at any given time.

The whole database (both tables) is backed up regularly.

3.2 Architecture

As stated earlier, the data architecture needs to be designed to ensure the security of the data and the privacy of users. Separating the data from different elements is an important step, but to ensure the data is not easily accessible, it is also required to use encrypted data where possible, making the system as a whole more secure.

The research and key table (from now on: research database) needs to be completely secure; this is where all the data is stored and linked together. The best way to ensure this is to make it inaccessible

from the outside world. Most companies and universities have a secure network, Utrecht University is no exception. The research database was placed on a dedicated server within this network. This entails it is only accessible from the physical location of the university (or VPN) and even then, it requires access to a secure system.

But even though the database is not supposed to be accessed from the outside, we still need to enter data from the online questionnaires, the mobile application and the online registration. So, we do need a way to safely enter data.

We realize this using a server in the so called demilitarized zone (DMZ), a sub-network contained within the university network, yet completely separated from it using a firewall, so that it functions as a buffer. The DMZ is reachable from the internet, so we can use the server in the DMZ to host the registration website and use it to send and receive mobile data and distribute the mobile application itself. A schematic overview of this is given in Figure 2.

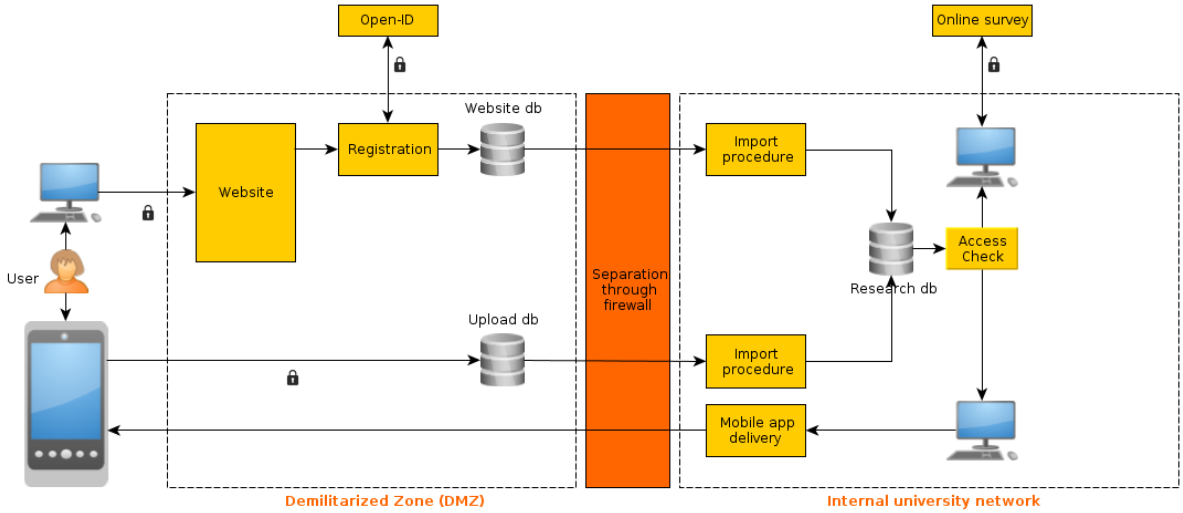

Figure 2. Schematic overview of the architecture

As shown in Figure 2, all communication of data towards the database is encrypted (encryption is depicted as a lock). In this figure the separation between the DMZ and the secure internal university network is obvious. The internal network contains a server with not only the research database on it, but also import procedures. Since the server in the DMZ cannot reach the server in the internal network, but the server in the internal network can reach the one inside DMZ, the data has to be requested by the server, instead of being sent from the databases. This is done by the import procedures.

The requests for data from the import procedures are scheduled multiple times a day, so that available data is quickly stored and privacy related data can be deleted from the databases in the DMZ.

4 IMPLEMENTATION

In this section we will describe what implementations we use to meet the requirements of the different architectural elements, as well as how we integrated them into the architecture.

4.1 Online registration

As described, the participants should be able to create accounts on the online registration website. To implement this, we have decided to use the Open-ID authentication mechanism, which allows participants to reuse a variety of existing on-line accounts to identify themselves. This is convenient for the user, but also reduces the amount of sensitive information on our server by delegating the authentication procedure to a trusted third party, over a secure network connection.

The procedure results in a unique identifier for each user that gets stored in our website's database, supplemented with an e-mail address for communication purposes.

For people who refuse to use Open-ID, we also have a regular account system, which includes as little personal information as possible.

Registration for the research includes filling in a short list of questions (related to inclusion criteria), which is shortly stored on the server in the DMZ, but deleted as soon as it is transferred to the research database. All the network traffic towards and on the registration website is encrypted through the use of a SSL-certificate.

4.2 Mobile application data

To meet the requirement of registering edits, a table in the application database was created to store them. Each edit entry includes a time stamp, the field that was edited and the value before and after the edit. This table of edits enables the application to only sent the aggregated data, which means that less data has to be sent and consequentially less of a risk of privacy violation.

To transfer the data, there is a server in the DMZ set up to temporarily store the data, before it will get transferred to the research database. The mobile application can generate its own triggers to send the data to the DMZ. At the moment, we use time, the availability of wifi and an above 20\% battery level to create a trigger. The battery level is self explanatory, and need for available wifi connection stems from the small data-limit some users have on their smartphones. But we also have a built-in safe mode so that the data will be sent at least once a day (even there is no trigger generated). We plan to fine-tune this further via user studies.

To be able to use the data sent by the application in the research data base, the data will have to include the mobile identifier of the user. The identifier will be sent to the participants (separately from the application), and they will have to enter it into the designated field before they can start using the application. The identifier has a built-in parity check, which enables the detection of input errors and invalid mobile identifiers. This reduces errors in the data sent to our database and, consequential, in the research data.

4.3 Online questionnaires

As described in the requirements, different systems might be used for the online questionnaires, but it could also be that a specific system is used multiple times. Therefore, it was decided to encrypt the identifier for the online questionnaires differently for each new questionnaire. This way, questionnaires cannot be linked to one another, except in the research database.

5 CONCLUSION

There has been rapid increase of health related applications, which are neither evidence based or developed with user needs in mind. This is why research into these applications is important, and to do this research in an effective way, a research data architecture is needed.
This architecture has to protect privacy, provide security measures and enhance accessibility for all stakeholders, including researchers.

In this paper we have presented an architecture that meets these requirements. We have used a combination of existing techniques to create an environment in which communication towards the researchers is secured by SSL and in which research data is only accessible to researchers in a secure environment. By the separation and anonymisation of the data until it reaches the secure database, the risk to privacy violation is minimized.

In the future we will expand the system to include different stakeholders (peers and caregivers). This has already been taken into account while creating the architecture.

On top of what's essentially a communication architecture, we will need to build a system with reasoning capabilities on both the phone and the server to meet privacy and functionality requirements. This allows the phone to be self reliant and respond directly to sensors, while we improve privacy protection by sending preprocessed data to the server that handles peer-to-peer communication. Therefore, we have to formalize the cognitive behavior therapy for insomnia - as mentioned in our introduction - in such a way that it works on a distributed system, which will be interesting but challenging.

References

Beun, R.J., 2012. Persuasive strategies in mobile insomnia therapy: alignment, adaptation, and motivational support. *Personal and Ubiquitous Computing*, pp.1–9.

Choe, E.K. et al., 2011. Opportunities for computing technologies to support healthy sleep behaviors. In ACM Press, p. 3053. Available at: http://portal.acm.org/citation.cfm?doid=1978942.1979395 [Accessed January 16, 2012].

Dolan, B., Report: 13K iPhone consumer health apps in 2012 | mobihealthnews. *mobihealthnews.com*. Available at: http://mobihealthnews.com/13368/report-13k-iphone-consumer-health-apps-in-2012/ [Accessed November 12, 2012].

Free, C. et al., 2010. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. *BMC Research Notes*, 3(1), p.250.

Telecompaper, 2012. Telecompaper Press Release: Smartphone penetration increased to 58%. *telecompaper.com*. Available at: http://www.telecompaper.com/pressrelease/smartphone-penetration-increased-to-58--899740 [Accessed November 12, 2012].

Vincent, N. & Lewycky, S., 2009. Logging on for better sleep: RCT of the effectiveness of online treatment for insomnia. *Sleep*, 32(6), pp.807–815.