D4.16 Personal Data Stores for Sensor Data

Project	SWELL
Project leader	Wessel Kraaij (TNO)
Work package	WP4
Deliverable number	D4.16
Authors	Bob Hulsebosch, Arnout van Velzen, Henny de Vos & Laura Claas (InnoValor)
Reviewers	Maarten Wegdam (InnoValor), Maya Sappelli (TNO)
Date	30 May 2016
Version	1.1
Access Rights	External
Status	Final version

SWELL Partners:

Noldus, InnoValor, Philips, TNO, Radboud Universiteit Nijmegen, Roessingh Research and Development, Sense OS, Almende & Universiteit Twente.

Summary

In SWELL different solutions for privacy in sensor-based applications have been reviewed. This deliverable addresses whether Personal Data Stores are a suitable solution to address privacy issues with sensor data for well-being and well-working applications.

Sensor based applications exploit (often mobile) device technology that register its surroundings, often aimed at behaviour of people, as opposed to conventional computer systems that generally register input devices and internal processes. Aside from general privacy concerns, such sensor data may encroach on privacy in a number of ways: ubiquitous sensors create significantly larger bodies of people-centric data, new types of personal data are uncovered, and in combination with other data may synergistically reveal unforeseen information about human behaviour. Moreover, subjects are often continually tracked over time and may be unaware of the data gathering as sensor operate concealed.

A Personal Data Store is an application that allows subjects of personal data to view and allow or deny access to their personal data. A Personal Data Store as such may serve different functions, including viewing sensor data to different levels of detail, adjusting various aspects of the data, and allowing full or partial access to personal data. Through providing an overview of personal data and delegated access management, a Personal Data Store helps to safeguard privacy, establish trust and meet legislative demands. Another main benefit is centralization of access management for multiple data sources and relying parties in a single application.

A Personal Data Store allows data subjects to curate data gathering and access, hence they may mitigate privacy concerns as to the properties of personal sensor data usage, such as duration, volume, accuracy, and type of data collected and accessed. PDS's may also alleviate to a lesser extend privacy concerns dealing with uncertainty about sensor data gathering and usage, such as user awareness, unexpected or wrong inferences, purpose of data gathering, (re-)identification, and meaningful consent.

In conclusion, PDS's are suitable as a solution for some of the privacy issues that relate to sensor data. Mainly because of its core functionality to provide control mechanisms and transparency. However, there are many factors that determine whether a PDS is a successful tool to preserve privacy in sensor-based applications, such as a reliable trust framework, friendly user interaction design and the business case.

Contents

1 Introduction

Recently the Dutch Data Protection Authority[1] (DPA) struck down on digital sensor applications because of privacy concerns. In November of last year the popular NIKE+ running app was audited by the Dutch DPA, that explained this move as setting an example[2]. Also, in March of 2016 the DPA ruled in two cases that employers processing health data of employees is against privacy regulations, after both employers provided bracelets to their employees that measure their physical activity[3]. As supported by the Dutch DPA, protecting privacy in sensor based applications is gaining traction, in the margin of a larger trend toward stricter privacy regulation[4].

This deliverable is a product of SWELL, a project of the Dutch National Research program COMMIT. In the SWELL-project[5] the focus is on preserving privacy in well-being (meaning supporting personal health) and well-working (referring to wellness at work) applications that rely on sensor data. So far in the project, different transparency and control mechanisms to safeguard privacy have been introduced. In this white paper, we introduce another solution to safeguard privacy of personal data: Personal Data Stores. This solution intermediates between personal sensor data and applications that exploit this data by allowing the subject to manage access rights to their personal sensor data in an absolute fashion.

1.1 Objectives and Approach

The objective of this deliverable is to assess the use of Personal Data Stores (PDS's) for user controlled privacy preservation of sensor data in the context of well-being and well-working applications. The main research question hence is whether PDS's are a suitable solution to resolve the privacy issues associated with sensor-based data. In order to do so, we have formulated the following guiding questions:

- What are the privacy issues of sensor based data?
- What are Personal Data Stores and how do they impact privacy?
- How would PDS's help to resolve the identified privacy issues relating to sensor data?
- What are possible architectures for PDS's for sensor data?

To this end, we have executed a review of relevant literature on Personal Data Stores, sensor-based data and privacy. Furthermore, we have conducted interviews with technology partners to debate these findings.

[1] Autoriteit Persoonsgegevens, fmr. College Bescherming Persoonsgegevens.
[2] College Bescherming Persoonsgegevens (November 2015). Onderzoek naar de verwerking van persoonsgegevens in het kader van de NIKE+ Running App door Nike Inc.
https://cbpweb.nl/sites/default/files/atoms/files/onderzoek_nike_running_app_november_2015_1.pdf
[3] https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-verwerking-gezondheidsgegevens-wearables-door-werkgevers-mag-niet
[4] Notably, the adoption of the European General Data Privacy Regulation (GDPR) in April 2016
https://en.wikipedia.org/wiki/General_Data_Protection_Regulation
[5] http://www.swell-project.net/

1.2 Reading Guide

Firstly, in the following chapter the privacy issues regarding sensor data and sensor-based applications are explained. In the third chapter Personal Data Stores are introduced, as well as the privacy issues they may renege. These two chapters are brought together in the fourth chapter that outlines how Personal Data Stores may help solving the privacy concerns associated with sensor data. The fifth chapter presents two cases illustrating the conceptual and architectural implications of Personal Data Stores, based on interviews. The paper closes with a critical discussion of the findings.

2 Sensor Data for Well-being and Well-working

2.1 COMMIT/ SWELL

Computer devices, including your smartphone and wearables such as a smartwatch, are increasingly context-aware through sensors. These sensors could be a camera or microphone, but many mobile devices also contain sensors such as GPS sensors, accelerometers, gyroscopes, or magnetometers. Context-aware applications that exploit device sensors collect a host of information about individuals and their environment. Moreover, as more and more devices are connected through communication protocols such as Wi-Fi, they form the Internet of Things. Part of this revolution of the data landscape are context-aware problem solving applications that aid self-management, sometimes referred to as e-coaches. Some are employed to increase users' health: well-being apps, whereas others focus on professional performance: well-working apps.

How users may assure privacy in such context aware well-working and well-being sensor based applications is the topic of the SWELL-project. Previous deliverables presented the requirements such privacy solutions should meet, such as user friendliness or granularity of control, based on privacy principles from the OECD6, FTC7, AICPA8 and APEC9, among others (D4.1). Recommended privacy approaches for context-aware applications are asking consent, layered privacy controls, adjustable data granularity, applying group permissions, withdrawing permissions, providing an overview of data collected, a privacy mirror (seeing what data others see), and giving privacy notifications (D4.2). To implement user-controlled privacy settings, the access control protocols XACML and UMA were assessed (D4.3). A guideline for asking meaningful consent was proposed, containing the following criteria: minimal intrusiveness, clear agreement, requiring reasonable competence and understanding, be given voluntarily, and providing all the necessary information to make an informed decision (D4.4). Design patterns for providing transparency about privacy, such as a tutorial, settings dashboard or reminder, were researched (D4.9-D4.12). Finally, two Privacy Impact Assessments were conducted and a guideline for PIAs for sensor based applications was developed (D4.13-D4.15).

In this next sections the privacy risks of sensor data in the context of well-being and well-working applications are described.

2.2 Privacy in Context-Aware Applications

Context-aware applications formalize much and detailed personal information about people and their behaviour, often from many different sources, hence this data is inherently privacy-sensitive. Privacy-sensitive information includes personally identifiable information, usage data and unique device identities. For example, multimedia camera phones harvest acute privacy concerns since they reveal much about subjects' personal and social environments.

The concept of privacy may be intuitive, but it is opportune to touch on a more precise explanation. One definition describes privacy as the ability of individuals to keep their lives and personal affairs

[6] http://oecdprivacy.org/
[7] https://en.wikipedia.org/wiki/FTC_Fair_Information_Practice
[8] http://www.aicpa.org/InterestAreas/InformationTechnology/Resources/Privacy/GenerallyAcceptedPrivacyPrinciples/Pages/default.aspx
[9] http://www.apec.org/Groups/Committee-on-Trade-and-Investment/~/media/Files/Groups/ECSG/05_ecsg_privacyframewk.ashx

out of public view, or to control the flow of data about themselves[10]. A fundamental right ratified in the United Nations Declaration of Human Rights, privacy may take various forms such as the right to be left alone, control of information about ourselves, or avoiding harm from its violation. As such, there are different classes of privacy: first, information privacy, dealing with the generation and processing of personally identifiable information (PII) and intellectual property; second, bodily privacy, concerning integrity of the human body; third, territorial privacy, being free from observation within spatial bounds; and fourth, communications privacy, referring to private correspondence.

Outside of ethical responsibilities, more instrumental reasons to protect users' privacy are to comply with law and regulation, as well as consumer demands. To the subject, loss of privacy may have all sorts of emotional, real-life and virtual consequences. To service providers in a broad sense, consequences of privacy breaches may entail non-compliance and legal procedures, loss of reputation, erosion of user trust, and undesirable exploitation of the information lost; hence privacy has strategic value.

Sensor data is often stored somewhere in the cloud. Users of cloud-based services express high levels of concern for unaware usage of their personal data, and organizations and governments have become more amenable to the need for privacy. For example, a 2012 U.S. survey[11] found that 57% of app users rejected or uninstalled an app over personal information sharing concerns. Similarly, a survey in the U.K.[12] showed 68% choose to not even download apps they do not trust (instead of just limiting usage), and an Australian survey[13] found 69% refuse apps or websites that use too much personal information. The upswing in privacy concern is reflected in legislation, e.g. the European Union Data Protection Directive 95/46/EC.

Following the susceptibility of context-aware applications to breaches and the growing necessity for privacy, such applications best ensure privacy protection. This straddles a balance between privacy, usability and functionality. In the back-end, i.e. the make-out and functions of an application, privacy may be safeguarded through Privacy Impact Assessments (PIAs) that score an application on privacy principles and assert recommendations for improvement. Another means for privacy assurance in the back-end are Privacy Enhancing Technologies (PETs), such as privacy management tools, online access mechanisms, and pseudonymization. Also, it is prudent to assess privacy throughout software design phases and implement privacy-enhancing features. This is known as Privacy by Design, and entails for example minimizing the data used.

In the front-end, i.e. the interface with the user, primary components of privacy assurance are transparency and user control. User control refers to the ability of the user to effectuate preferences as to the functionality of the application. One such tool is the ability to adjust the granularity of

[10] Hafiz, M. (2006, October). A collection of privacy design patterns. In Proceedings of the 2006 conference on Pattern languages of programs (p. 7). ACM.
[11] Boyles, J.L., Smith, A. & Madden, M. (September, 2012). Privacy and Data Management on Mobile Devices. Pew Research Center. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_MobilePrivacyManagement.pdf
[12] TrustE (July 2012). Consumer Privacy Attitudes and Business Implications. Retrieved from http://www.truste.com/window.php?url=http://download.truste.com/TVarsTf=7EDO6P8Z-187
[13] Arnott, C. & Andrejevic, M. (February 2012). Internet privacy research: report. University of Queensland. Retrieved from http://cccs.uq.edu.au/personal-information-project

personal data collection, e.g. to define the accuracy of location information being processed by an application. Giving control to users is complicated as it often requires fine-grained and complex privacy policy management that goes beyond the skills and ambitions of the average user.

Trust in an application is enhanced when procedures are clear, transparent and reversible, so users feel in control. Before users may take control of their personal information, they have to be informed of what information is handled, by whom, for what reason, and with what effects. A previous user study in SWELL found that privacy information has a positive effect on trust in context-aware systems[14]. Privacy transparency in this respect refers to the insight provided about all facets of the data generated and processed by an application.

2.3 Privacy Issues of Personal Sensor Data

This sections deals with the several potential privacy issues of sensor data, divided into two categories: specific privacy concerns for sensor data and more general concerns.

Sensor Data Specific Privacy Concerns

1. Volume of personal data. In the Internet of Things, computers are ubiquitous and sensors everywhere, hence large quantities of personal data are mined. As more privacy sensitive data is collected, whether from a mobile device, wearables, or internet of things-devices, more privacy encroaching inferences could be deduced.

2. Different types of data. Also, such pervasive people-centric computer environments gather different types of personal data that reveal individuals' interactions with their environment in new ways. Think of sensors such as heart rate monitors that collect personal data that was conventionally restricted to medical contexts. As new types of data are collected, new inferences can be made.

3. Unexpected personal information. Combinations of even unrelated contextual information may reveal previously undiscovered insights into our behaviour. Similarly, a translation from raw sensor data to information per some semantic reference always has to be performed. This may render seemingly harmless sensor data privacy invasive; in a manner unobtrusive to the user. Collecting large sets of sensor data can yield unforeseen insights about an individual's personal life. For example, heart rate and GPS location may appear harmless to privacy in isolation, but combined may unexpectedly show an individual's heart rate to increase around certain people or in certain location, which may cause embarrassment.

4. Continuous tracking. Moreover, sensors have the propensity to track subjects continuously in their context. Interestingly, sensor data collected and stored over a longer period of time may become classified as medical data, e.g. heart beat data after 30 days. The reasoning behind this classification is that it may reveal medical conditions if measured over a certain period of time.

5. Unawareness data gathering. Lastly, subjects are often unaware of the data gathering since sensors tend to operate concealed.

[14] Koldijk, S., Koot, G., Neerincx, M., & Kraaij, W. (2014). Privacy and User Trust in Context-Aware Systems. In: *Proceedings of the 22nd Conference on User Modeling, Adaptation and Personalization (UMAP 2014)* (Aalborg, Denmark, 7-11 July 2014).)

General privacy concerns relevant to sensor data

6. People centric data. Sensor data collected by well-being or well-working apps is people-centric: it is often data about one or more individuals. Sensor data collected by well-being apps stems from one individual in the first place. This data can be aggregated with other individual data-sets to a dataset about a group of people. Aggregated data about groups of people are generally considered less privacy invasive. However, the initial individual and very personal nature of the sensor data make it privacy sensitive. Inferred sensor data is statistical in nature, i.e. for large groups and may single out the individual user and they may fall 'victim' to discrimination, ridicule or repercussions.

7. Data accuracy. Sensor data can be highly specific and accurate, which makes it more privacy sensitive because it leads to more accurate inferences. GPS data for instance reveals an individual's position with an accuracy of a few meters. As sensor data is more precise, the certainty of inferences also increases.

8. De-anonymization. Datasets of individual data subjects are often aggregated into more encompassing datasets (e.g. combining heartrate data with location data) in order to discover new information. Likewise, multiple individual datasets can be aggregated to create a dataset about a group of individuals. This can be helpful to discover general patterns in the data. When aggregating data of individuals, this data should be anonymized so it cannot be traced back to the individual data subjects, to preserve privacy. However, it may be possible to re-identify individual data subjects by combining or cross-referencing different data sources. For example, when collecting large amounts of anonymous (sensor) data from various sources, profiles can be created of unique data subjects. By further recombination with other available information (for instance about place of residence, or education) the identity of the data subject of the profile can be established.

9. Wrong conclusions. Sensor data can easily be interpreted wrongly. This can either be caused by incorrect data (missing data, low data quality) or by defects in the way conclusions are drawn from the data. Data can be incorrect because it is incomplete; for instance, technical errors can cause missing data. Incorrect conclusions from sensor data can harm the privacy of the individual. Furthermore, in wellbeing apps, sensor data from individuals will often be compared against norms or averages derived from aggregated datasets from large groups of people. Such comparisons don't do justice to unique individual circumstances. This may lead to wrong interpretations or conclusions.

10. Purpose unclear. It occurs that data is being gathered for unclear purposes. Users may understand that data is being gathered, without understanding what the goal of the data collection is. Such indefiniteness may stem from improper communication from the app's side, which can be solved by improving the privacy statement. Or it can be caused by a lack of understanding from the user.

11. Fine-grained data control and consent. Data can be gathered without the user's consent. Or users feel forced to provide their consent, although they do not agree with all data collection, but because they definitely want to use an application.

12. Security. Lastly, improper security forms a general privacy risk when processing sensor data. When sensor data is not processed securely by the app, there is a high risk that the data can be accessed by people with malicious intent, or that the data is accidentally made public. In either case,

the sensor data is involuntarily available to people whom the data subject did not authorise to have access to this data. This constitutes a breach of privacy.

3 Personal Data Stores

3.1 What are Personal Data Stores?

A Personal Data Store (PDS) is a service with which a user can manage the sharing of his own personal data from one centralised place. Personal Data Stores are also known as Personal Data Lockers, Personal Data Services, and Life Management Platforms. As of yet, there are over a hundred different initiatives around PDS's globally, but the adoption differs, with most being experimental. No work has been done focussing specifically on Personal Data Stores for sensor data. In this white paper we assess the use of Personal Data Stores for user controlled privacy preservation of sensor data.

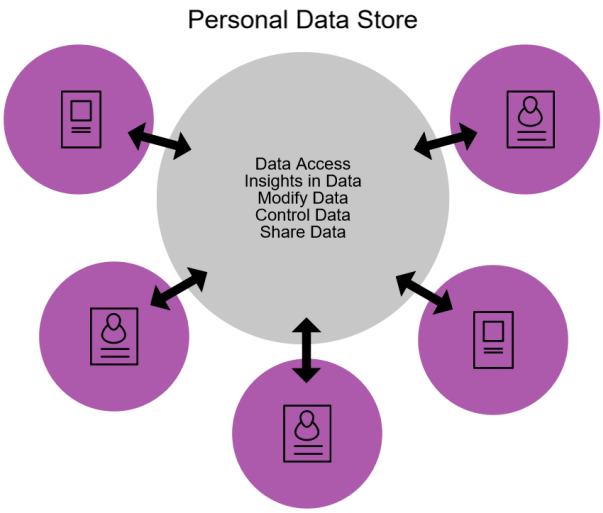

Figure 1: Personal Data Store overview

PDS functionalities

Personal Data Stores (PDS) are services that empower individuals to control and maintain personal information in order to be able to share it with others. A PDS enables a user to view which personal data is stored in what location and by whom it is used. PDS's offer the user control over the sharing of data by enabling providing or withholding permission to other parties to access this data. PDS's function as a central place where all personal data (virtually) comes together, thus providing the user an overview of his personal data.

PDS's concern the use of personal data, rather than but not excluding storage of data. As such, the data isn't necessarily stored within the PDS; the PDS can be a central access point to dispersed data sources. Thus two models for data storage are possible for a PDS. First is a central model, where all (relevant) data is copied into the Personal Data Store, that functions as a database. The second option is a decentral model, where the data remains stored at the original data source. The Personal Data Store functions as a central point of control over the data, without actually storing any data.

Data types. Ideally, PDS's encompass all (types of) data about an individual, such as health data, financial data, or educational data. The user can control the access to these various data sources. For instance, when applying for a job, the user could provide his prospective employer temporary access to his diploma information.

However, PDS's can be smaller in scope and specific to one domain, while still disclosing various data sources. A PDS for the medical domain for instance could disclose data about medicine use, dentist appointments, test results from diagnostic institutions, or information provided by specialists in a hospital.

Typically PDS's may disclose several types of data (Kearns, 2009), that are either stored in the PDS or accessible through the PDS and stored otherwise:

- My Data: within an individual's domain
- Your data: within an organisation's domain

- Their data: gathered by third parties
- Everybody's data: public data
- Our data: one or more of the above, accessible by means of a transaction or relation

Ecosystem. A PDS is the linking pin in an ecosystem of several parties. On the one hand there are data providers, i.e. governments, commercial parties, devices or other parties gathering, processing or managing personal data. On the other hand there are parties that depend on access to this personal data. Such parties rely on personal data to provide services to users or for their own purposes. The PDS is an intermediary through which users manage access rights of service providers to their personal data from the data provider. PDS's may store (personal) information, but personal data may also reside elsewhere, as long as regulating access to the data happens via the PDS. The PDS needs to interface with data providers and data users to facilitate data transactions. This does not necessarily entail that data is transferred via the PDS, but access requests and permissions will flow through the PDS. Such interfacing needs standards and data management rules (provided by the users) for data handling. The data-ecosystem is visualized in Figure 2.

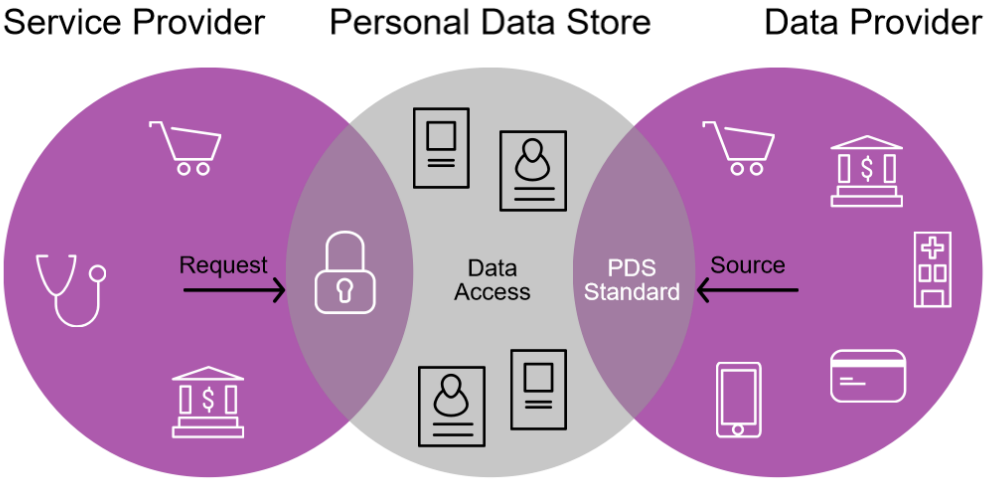

Figure 2: Personal Data Store Ecosystem

Additional Functionalities

PDS's can allow for data management, enabling the user to modify his data or possibly delete it. However, it depends on the implementation of the PDS in what way data can be modified by the user. Ideally, the original data in the data source is modified; the user however must be warranted to do so. Alternatively, users may modify data that is being stored in the PDS itself; but this entails merely modifying a copy of the original source data.

Personal Data Stores may function as a discovery platform for application developers to identify (i.e. 'discover') and access sensor data. PDS's are linked to multiple data sources and multiple service providers; i.e. applications or services that want to use one's personal data. Thus PDS's fulfil the function of a hub connecting numerous applications to each other. As such PDS can function as a central "marketplace" providing an overview of various types of user-generated data from various sources, where service providers can easily find what data is available.

Advantages of PDS's

PDS's shift data management from service providers to users. This can be advantageous for both users and service providers.

The most important added value of a PDS for users is that they are able to gain control over their personal data and that data use by third parties is transparent. Research from InnoValor (2015) shows that PDS-services fulfil a market need in this respect: 57% of Dutch people above 18 are interested in using a PDS.

The (assumed) added value for service providers is that PDS's increase users' trust and subsequently increase their willingness to share data. The quality of data may improve by allowing users to correct wrong data, although with an authoritative source, such as sensor data, this may not be possible or desirable. As such, PDS's can decrease errors in subsequent data analysis. PDS's can increase the efficiency of data handling for service providers, on the one hand because data management is delegated to users, on the other hand because data quality improves and less time needs to be spent on correcting errors. Lastly, PDS's may support service providers in complying to privacy laws and regulations.

Another advantage of PDS's for service providers it that PDS's may serve as a hub or intermediary between data and applications, implying that by linking to a PDS, application providers may have easily access to data and that data can be offered via the PDS to applications.

Trust Frameworks & PDS

Aside from technical standards, PDS's need a set of agreements that determine responsibilities; a trust framework. Such agreements may have additional implications for the technology behind a PDS application.

Since PDS's are by their very nature part of an ecosystem with many parties, mutual agreements that govern the relations between these parties are virtually inescapable. To avoid an unorganised system where parties come to a cacophony of agreements amongst themselves, a trust framework can be employed. A trust framework is a set of policies and technical standards and protocols that enables parties to trust each other. The trust framework specifies the roles that different parties can assume, and the rights and responsibilities that correspond with these roles. It ensures that parties assuming the same roles are treated equally. Parties can only join the ecosystem if they agree to operate under the regulations specified in the trust framework.

Examples of such frameworks are Qiy[15] and the Respect Network[16]. Qiy is a trust framework under development that enables the delegation of access management of personal data to the data subject. Hence application developers that use personal data in their applications can use the provision of the Qiy standard in their application and arrange that data management is delegated to the user. Qiy's ambition is to become an open, publicly available standard.

The Respect Network is a network of private clouds; web-based storage spaces for personal data that only the user can access. From these private clouds, users can securely communicate with any other users of the Respect Network, allowing them to share data with end-to-end encryption. Crucial

[15] https://www.qiyfoundation.org/
[16] https://www.respectnetwork.com/

to the Respect Network is that all data created within the network, for instance by apps, is the property of the data-subject.

3.2 Business Case of the PDS

A PDS has several stakeholders, each of which can potentially finance the operation of the PDS:

- End-users
- Operator of the PDS, commercial or non-commercial
- Data providers
- Service providers
- Third parties: commercial, governmental or NGO

Of course, one party could take on multiple roles, for example an service provider could also be a data provider. Below we shortly describe seven possible business cases for the PDS. In these it is assumed that the PDS is fully functional and implemented; thus we disregard the development process and potential sale of the PDS as a software product.

1. User Subscription. Since a Personal Data Store provides a service to end-users, a business case is possible where the end-users finance the operation of the PDS. The party that operates the PDS can aim to make a profit by charging the users a fee for using the PDS service, or can aim to merely cover their costs and break even. The fee could for instance be in the form of a monthly subscription fee.

In order for this business case to be viable, it must be very clear to (potential) users what the added value of the PDS is. Not only must this value be high enough for them to use the PDS, end-users must be convinced that the PDS service is worth paying for.

2. Commercial PDS operator. The party that operates the PDS may do this for commercial reasons; as such, they are the party that finances the operation of the PDS. Rather than selling the PDS as a service to end-users, the PDS provider may choose to finance the operation of the PDS by selling the data to service providers. This could take the form of a buy-and-resell model, where the PDS operator buys data from data providers and resells it to the service providers. Or it can follow an affiliation model, where service providers buy the data from data providers, and the PDS provider charges a small fee for facilitating this transaction.

In an economy where data is referred to as "the new oil" [17] it is probable that a business case where data is sold is successful. The PDS needs to be connected with enough data providers, service providers and end-users to create a blooming locus of data trade.

3. Non-commercial PDS operator. It is possible that the operating party of the PDS is not looking to make a profit, but is still the party that finances its operation. The operator could for instance be a non-profit foundation or NGO that operates the PDS for idealistic reasons. Such parties can finance their entire operation, including the PDS, by other means.

17 See for instance The New Oil: Using Innovative Business Models to Turn Data into Profit by Arent van t Spijker. The introduction describes Neelie Kroes' use of "data as the new oil" to refer to data as the driving force of the economy.

Due to the potential of a PDS as a privacy-preserving mechanism, non-profit privacy organisations will likely be interested in the concept of the PDS. This however does not necessarily mean that it is in their power or desire to develop or operate a PDS ecosystem.

4. Commercial third party invests. It is not uncommon for startup companies developing and/or selling a successful innovative product to be bought, in their entirety, by large corporations[18]. Google for instance acquired YouTube and Android, that are now solid elements in their business offering. The rationale behind such acquisitions is both to eliminate competition, as well as investing in potential future successes. It is possible that a large third party either acquires or develops a PDS service for the same reasons; to be the first (large) company to operate a PDS, so as to invest in future potential and beat competitors.

This business case is not unlikely, but it heavily depends on the perceived value and commercial potential of Personal Data Stores.

5. Non-commercial third party finances. In the UK, the government initiated and financed the development of a PDS ecosystem called Mydex. In other countries this example may be followed. This constitutes a business case where a large non-commercial third party finances the operation of the PDS. Such a party can be a governmental organization or a non-profit domain organisation.

Of all the business cases sketched here, we consider this one the most likely. Governmental organisations have the power to set up a PDS ecosystem, including participation of data providers and service providers. If need be, they can even require such participation by law. Unlike commercial parties operating a PDS, they do not need to rely on persuading other parties and users with either the allure of financial gain or other benefits.

6. Data Provider Finances. It is possible that the data providers operate and finance a PDS, but not likely. Apps that gather sensor data already have a business case. Unless they see great opportunities in selling data, a PDS is not in direct interest of data providers.

7. Service Provider Finances. The same holds for the operation and financing of a PDS by service providers. Although creating a marketplace for data trade is in their interest, a PDS is not the most likely way for a service provider to do so. Service providers are more likely to seek direct trade with data providers, without the end-user standing in between.

3.3 PDS and Privacy Legislation

Needless to say, a Personal Data Store must comply with applicable legislation. But a PDS can also be a means for other parties to comply with privacy legislation. The General Data Protection Regulation was ratified on April 14th by the European Commission. In this Regulation are some data privacy requirements that a PDS can contribute to. Below are a few examples.

1. Consent. Data processors are required to obtain informed, specific and freely given consent for the use of data. A PDS is a means for users to provide such consent.

[18] See for an impression the List of Aquisitions by Google and the List of Acquisitions by Facebook on Wikipedia

2. Subject Access Request. Data-subjects have the right to request insight in what data about them organizations possess and how it is used. A PDS can be a means to provide such insight.

3. Subjects' "right to be forgotten". Data subjects can request that organisation delete the personal data about them. A PDS can provide insight in what data an organisation has, and possibly offer means of communicating such requests.

4. Subjects' right to data portability. Data subject have the right to request a copy of their personal data (in a common machine-readable format) and to transfer this data to another organisation, so it may be used by them. A PDS can facilitate transferring data between parties.

5. Subjects' right to restrict processing. Data subjects have, under specified circumstances, the right to restrict the processing of their personal data. This means that although the data remains stored at the organisation, the data subject can strictly limit the purposes for which the data can be used. A PDS is a suitable means for data subjects to exercise such control over data use by companies.

3.4 Examples of Personal Data Stores

An example of a PDS from the United Kingdom is Mydex[19], which evolved from the Midata project of the British government. In Mydex, personal data is created by the user or by authorised parties within the PDS. Next to personal data storage, Mydex enables storing of official documents such as passports. If desired, the user can share this data with other organisations. Both commercial and governmental organisations are connected to Mydex. Furthermore, users can analyse their own data with Mydex. Lastly, Mydex functions as a platform, which facilitates integration with new applications via API's.

Another example of a PDS is Personal.com[20], which is an online PDS service based in the United States. In this PDS users can store all their personal information, and share this data with various commercial parties. However, users must actively insert their data into Personal.com themselves, rather than for instance retrieving sensor data from a data source such as a smartphone app. Personal.com offers services for a personal cloud, a data vault, automatically filling online forms, and an online notepad. Since 2013 Personal.com has partnered with FileThis, an online secure mailbox and data vault.

3.5 PDS as a Privacy Solution

Personal Data Stores claim to provide user control over personal data. However there are other alternative service offerings that claim similar value. In this section we compare PDS's to other privacy and personal data solutions, such as privacy dashboards and personal clouds. To further investigate how PDS's relate to privacy, we have analysed how PDS's relate to the OECD privacy principles.

PDS versus Other Privacy Solutions

In this paper we address the concept of a PDS as a privacy solution, but there are of course many other privacy solutions. Privacy Enhancing Technologies can, depending on the circumstances, be used as an alternative to a PDS, but may also form part of a PDS solution. For instance, privacy

[19] https://mydex.org/
[20] https://www.personal.com/

control mechanisms such as privacy control layers, fine grained control and removing policies can be employed within a PDS system. We refer to the SWELL paper D4.2 Approaches for adaptive and intuitive privacy control and privacypatterns.org for more detailed information on the variety of PET's and privacy design patterns.

Next to that, there are other personal data managements systems that function in a similar fashion to a PDS. Below we discuss three examples of these and their differences with the PDS: Privacy Dashboards, Personal Clouds and Data Vaults.

Privacy Dashboards. A privacy dashboard provides to the user a dashboard or control panel that presents privacy settings related to an application. Such dashboards have become a standard practice, applied in numerous applications. Social media websites for instance have privacy dashboards that allow for fine-grained control over which social media items (e.g. posts, personal information, profile picture, photos one is tagged in) are visible to which groups of people. Most web browsers have a privacy dashboard that enables the user to control settings regarding cookies, browser history, search history and tracking protection.

Often privacy dashboards are multi-layered, meaning that settings are displayed in multiple levels of detail. Commonly, three layers are used, where each deeper layer provides more detailed control options. The browser Firefox for example offers three options regarding browser history: 1) remember history, 2) never remember history, and 3) use custom settings for history. Only when a user selects the third option, more detailed settings are displayed.

A privacy dashboard can be seen as an application-dependent Personal Data Store. A specific feature of a PDS however is that it discloses personal data from and to multiple applications. The added value of a PDS lies in the ability to provide data to an application that the application in itself would not be able to obtain. Therefore, connection to multiple data sources and service providers is a differentiating characteristic for a Personal Data Store.

Personal Clouds. A personal cloud is a personal online storage space for storing text documents, photos, spreadsheets and other kinds of documents and files. Most personal clouds allow for easy sharing of these files with other individuals and between devices. Well-known examples of personal clouds are Dropbox and Google Drive.

While personal clouds and PDS's have the functionality of sharing in common, what is shared and with whom differs very much. Personal clouds concern sharing of entire documents, to individual people whom by viewing the content of the documents, make sense of the information contained in it. PDS's concern sharing of (raw) data with applications; these applications process this data. Eventually, the application can present the outcomes of this data processing in a manner that yields information understandable for humans, or alternatively produce machine readable output.

Data Vaults. Data vaults[21] allow users to store data in a secure place in the cloud. They function very similar to Personal Data Stores, with one important exception: data vaults do not allow the user to (easily) share his data. Whereas PDS's let the user control access to his data, Data Vaults do not

[21] Note that the term "Data Vault" is widely used to denote different things; firstly in the meaning discussed here, namely a secure personal cloud for storing personal data; secondly, Data Vault is a data modelling method used for long-term, historical storage of data; and thirdly, Data Vault is used as a brand name for various companies offering backup services.

allow any other party than the user itself to access the data in the vault. Off course the user can download his data and manually share it with another party, as a workaround. In contrast, the PDS is explicitly aimed at facilitating data sharing.

Confronting PDS's with OECD's Privacy Principles

Freemantle and Scott (2015) endorse the concept of user managed access as an important approach to safeguarding privacy. The Privacy Principles[22] from OECD are commonly used as guidelines on how applications or services should deal with privacy-sensitive data, in order to safeguard the data-subjects' privacy. In this section an overview is provided of the privacy principles, and how PDS's are assessed with respect to these principles.

PDS's provide transparency and user control on (external) data sources and applications/services that use data. The primary data collection of PDS's themselves is related to user's data management rules and to providing the user feedback on data usage. Data being controlled via the PDS can remain at its external source location (decentral model) or can be imported into the PDS (central model). Furthermore, PDS's can offer the possibility to store personal information manually entered by the user (thus not from a data source). This makes PDS's complex and/or uncertain in terms of which parties (or applications) function as data storage, data processors, data owners, etc. The analysis below is aimed to be both concise and complete.

Security Principle. *Administrative, technical and physical information security. Security should be in line with the sensitivity of the data.*

The nature of a PDS entails that the data is securely stored, transferred, communicated and/or accessed and only the user is allowed to control / manage data and data access. Proper administrative, technical and physical security measures should be in place.

Data Collection Limitation Principle. *Ensure that the data collected is limited to the data that is actually required to realise the application's requested results.*

PDS's can enhance data collection limitation of the data providers. Users are enabled to reject data collection of data providers if they regard it as unnecessary. In order to fulfil their function as intermediary between data providers and service providers, PDS's must collect some data from the users, regarding their requirements for data sharing. Depending on whether data is stored at the data providers (decentral model) or is duplicated into the PDS (central model), the PDS may or may not collect user data. It must be ensured that this data is limited to the data necessary for the PDS to fulfil its function.

Use Limitation Principle. *Ensure that data that has already been collected is only used for the purposes that were previously specified and consented to.*

PDS's can effectuate Use Limitation of data sources, as the user may control data usage by service providers by consenting to specific uses or not. Since PDS's can provide insight into data, users are able to monitor data use by providers. Processes and procedures have to be arranged to prevent data use bypassing the PDS. Furthermore, a PDS cannot technically enforce how the data is used

[22] http://oecdprivacy.org/

within the systems of the service providers or data providers. So even though a PDS is a good means of communicating about the use of data, the PDS itself cannot guarantee restriction of data use.

Purpose Specification Principle. *Explain the reasons for the collection of data and intended uses.*

The purpose of data usage can be approved by the user with the help of a PDS by providing transparency on the purpose of data use by service providers, and the means to consent to specific purposes or not.

Quality and Availability Principle. *Ensure that information is accurate and that it is not inappropriately modified during transmission, storage, or processing.*

Depending on the technical implementation of the PDS, it can be possible for a user to correct errors in their own data. If the user can modify the source data of the data provider, this can contribute to higher quality data. Regardless of this option, the PDS and surrounding data ecosystem should ensure data quality and availability are maintained. The ecosystem should avoid for instance data loss through too much compression.

Individual Participation Principle. *Ensure users can exercise their rights: to know whether data controllers have data relating to them; to have complete insight into this data; to challenge a denial of insight into data; and to challenge the data and subsequently have data erased or altered.*

A PDS is focussed on user participation and provides transparency and control to the user. By employing high quality information visualisation and good user experience design, the rich amount of information available in the PDS can be made insightful and easy to use.

Transparency Principle. *Be open about what kind of data is collected, how this data is maintained, and how it is used.*

PDS providers inform users about personal data collection, usage, and purpose of data providers and service providers connected to the PDS. As such, parties that function as data sources or service providers provide transparency through the PDS.

Accountability Principle. *Be accountable for complying with the principles listed above.*

Accountability has to be established within technology, business operations, processes and procedures. When PDS technology is combined with a trust framework, the trust framework can ensure compliance with agreements and accountability regarding privacy. Examples of such trust frameworks are Qiy and the Respect Network, discussed in section 3.1.

4 Personal Data Stores for Sensor Based Data

In Chapter 2 we discussed the privacy issues around sensor data, and in Chapter 3 we discussed the Personal Data Store. In this chapter, we discuss PDS's specifically in the context of sensor data. First, we describe how Personal Data Stores can form a privacy solution for sensor-based data. In the second section we discuss the architecture possibilities for PDS that specifically deal with sensor data.

4.1 PDS as a Privacy Solution

In this section we will address how Personal Data Stores may be able to mitigate the privacy issues stemming from sensor data as outlined in the before chapter.

Privacy issue 1: volume of personal data
Privacy issue 4: continuous tracking
Privacy issue 6: people centric data
Privacy issue 7: data accuracy

Through the use of Personal Data Stores, the data subject has full transparency and control over what sensor data is collected and processed to what extent. Therefore the properties of data gathering, such as volume, accuracy and duration, may also be adjusted to what the data subject is comfortable with.

Privacy issue 3: unexpected personal information

While unexpected revelations from sensor data may not be avoided by Personal Data Stores, despite curation by the data subject, upon discovery the subject has the ability to immediately cancel further data gathering. PDS's also provide the opportunity to block re-use of inferred data by other applications.

Privacy issue 5: unawares data gathering
Privacy issue 10: unclear purpose

A PDS acts a single point of access for personal data, and therefore as a nucleus of control and discovery for the data subject. This implies that, at least when PDS-associated sensors are concerned, all data gathering can be reviewed by the subject. Applications that exploit sensor data have to ask permission to access sensor data through the Personal Data Store, ideally also informing about the purpose of data usage, eliminating the possibility of unknown data usage. If the purpose of data gathering is unclear, it is at the subject's discretion to allow or deny the application access.

Privacy issue 9: wrong conclusions

Due to faulty or missing sensor data, wrongful interpretation or comparison to aggregates, users may incur a breach on personal integrity, and thereby privacy, from sensor based applications. While not negating the possibility of such occurrences, Personal Data Stores help to mitigate this issue on two levels. Firstly, by heightening awareness of sensor data being processed. Secondly, by allowing insight, and possibly options for adjustment, into possibly erroneous sensor data, e.g. a broken sensor may be signalled through the Personal Data Store.

Privacy issue 11: fine-grained data control and consent

Traditionally, sensor data applications offer few options to control whether or not certain data is gathered and/or used. Through Personal Data Stores data subjects have some form of absolute control over sensor data usage, at the least, albeit only for data usage of sensors that are connected to the PDS.

Privacy issue 12: security

Personal data stores add another layer of security by expanding access management. However, the PDS environment itself does become an attractive target for agents illegitimately pursuing access to personal information. Hence, it cannot be sustained that a Personal Data Store has a positive impact on security per se; depending on the actual implementation however PDS's can contribute to security.

4.2 Architecture

A PDS for sensor data would centralize access management of applications to sensor data; in contrast to decentralized settings per application or sensor. Architecturally, the PDS application acts as a hub between different sensors and applications looking to use this sensor data. The PDS application may be located on a sensor device (e.g. a smartphone) or computer, but more likely resides on a server (in the Cloud).

In order for PDS's to carry out their function of hubs in a data ecosystem, they must be able to communicate with data sources and service providers. To ensure that all interconnected systems "speak the same language" it is prudent to employ technical standards. Two notable technology standards for privacy policies are User Managed Access (UMA) and eXtensible Access Control Markup Language (XACML)[23]. UMA and XACML are authorisation standards that can be utilized in a variety of systems, including but not limited to PDS's. Lastly, an architectural standard specifically for PDS's is OpenPDS.

Access to personal data in a PDS is for example realized with a technical protocol such as User Managed Access (UMA)[24]. This is a standard for delegated authorization based on the Oauth[25] standard for access. UMA was developed by the Kantara initiative, a non-profit organization for innovation in the domain of digital identity management. See Figure 3. UMA has been implemented by several organizations, for example ForgeRock's identity platform[26].

The eXtensible Access Control Markup Language (XACML) is an XML-based language, or schema, designed specifically for creating policies and automating their use to control access to disparate devices and applications on a network. The power of XACML lies in the fact that access control is no longer something that resides inside the application, but is externally managed using a standardized policy language.

[23] See SWELL 4.3 chapter 3 from http://www.swell-project.net/results/deliverables
[24] For more on UMA see this video: https://www.youtube.com/watch?v=4nZ5DvCFp8U or Kantara Initiative https://kantarainitiative.org/confluence/display/uma/Home
[25] http://oauth.net/
[26] https://kantarainitiative.org/confluence/display/uma/UMA+Implementations

A standard for a PDS application as a whole is provided by OpenPDS, an open-source architecture developed by MIT. OpenPDS only communicates anonymous, processed data, and thus protects privacy by not sending raw data. With this mechanism privacy-sensitive processing of data is restricted to the OpenPDS environment. OpenPDS itself uses the Funf open-sensing framework to collect sensor data from Android smart phones and send it to the OpenPDS environment. An application that runs on the OpenPDS data is OpenAnswers, see this video. MIT and Denmark Technical University are currently doing pilot studies based on OpenPDS.

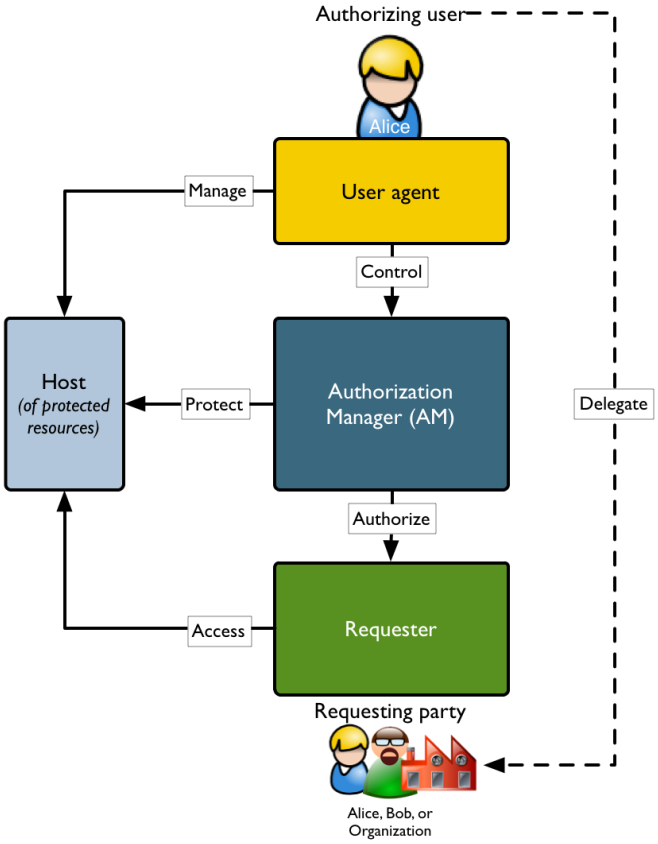

Figure 3: User Managed Access Architecture

Architecture Scenarios. Below we will sketch two scenarios for a PDS for sensor data; in the first scenario, a Cloud-based PDS directly controls access to sensor data, in the second scenario all sensor data is ported from the sensor devices to a server that is controlled by a Cloud based PDS.

Scenario 1. The PDS would control proprietary sensor hard- and software, for example the sensors on a smartphone, smartwatch or fitness tracker. However, most proprietary operating systems, such as Android by Google, Windows Mobile by Microsoft or iOS by Apple, do not allow applications to manipulate the access permissions for other applications. Hence, unless the developers of these proprietary operating systems either allow a PDS to manipulate access permissions to sensor resources, or develop a PDS application themselves, this scenario to realize a PDS for sensor data is unlikely. Android version 6 and iOS already have a tab where users can view and control access permissions for mobile sensors. Alternatively, a PDS application may be developed for dedicated hardware (sensors). This entails setting up a completely new data ecosystem, rather than integrating a PDS in an existing data ecosystem.

Scenario 2. A PDS client application ports all sensor data from the origin device to a server. A PDS server application then acts as a platform that other applications may draw sensor data from. Also, the PDS server application would offer users an interface to view and alter access permissions for such applications.

An example of scenario 2 is provided by Synergetics. Synergetics[27] is a privacy-protecting personal data ecosystem for health care systems that revolves around a platform that serves as a Personal Data Store. Users are able to decide what data and insights they share using the Synergetics platform. Health related data from sources such as medical records, health tracking devices or social

[27] www.synergetics.be

media are stored in the Synergetics database and can be disclosed to such service providers as medical professionals, health related software services or insurance firms.

5 Casus Sense-OS

Sense[28] provides a platform for sensor data storage. It stores raw data from mobile phone sensors and data from external sensor sources (e.g. fitbit, smart watch). Applications can use this data and provide users with insights and coaching. In an interview with Sense we have discussed the opportunities and threats of PDS's and how sensor data within the Sense platform can be approached by means of a PDS.

Within the Sense platform, the 'raw' data from e.g. mobile phone sensors or other (external) sensors is interpreted into meaningful measurements, e.g. sleep, stress, physical fitness, top locations. The data can be used in applications, currently mainly focussing on well-working and well-being domain. These apps are helping people to change their lifestyle by providing insight into their habits and health, and tips and coaching to stimulate new behaviour. Sense is a service provider for such applications.

Two examples of Sense-applications are Brightr and Goalie. The Brightr-app is a well-working app that builds on Sense platform data. The application tracks a user's sleep and activity with accelerometer, microphone, and location sensors in the phone. The app provides individual employees with insights into their own health; mental health as well as physical health (e.g. sleep, fitness). Through company-wide challenges and digital coaching, employees are stimulated to improve their health. A dashboard is available for the employer or HR staff, which provides insights in the collective health of employees through aggregated data that cannot be traced to individual employees.

Goalie is an app that helps the treatment and recovery of patients in psychological healthcare. The app supports patients achieving all kinds of goals throughout their daily life, which the therapist and patient have set together. The app monitors, amongst others, the physical activity, sleep rhythm and emotional states of the patient. Based on these measurements, the Goalie app provides personalized coaching to the patient. Through a dashboard, the patients' therapist can monitor their progress.

Current privacy situation. The sense platform and Brightr and Goalie apps deal with privacy in a consent-based manner. The user is asked to provide an all-covering consent once, so as not to keep prompting the user. The terms of service are explicitly written to be easy to understand. Within the Brightr application, a set of privacy transparency patterns was implemented to communicate the privacy policies to the user:

- A privacy policy text was written in a format explicitly aimed to be easy to read and was provided to each user before registration.
- Right after registration users went on a short tour through the app, which included a page explaining which sensors in the phone were used.
- Users received a welcome e-mail, mentioning privacy related aspects of the app.
- The application contained a section with frequently asked questions (FAQ), which amongst others targeted privacy.

[28] http://www.sense-labs.com/

- The app contains a general text about the goals of the app, and privacy related aspects (under 'About this app').
- Aside from these touch points with the user, the standard iOS and Android privacy controls are applicable, e.g. approve location tracking and use of microphone data.

The approach of using multiple privacy transparency patterns lets users become aware of privacy issues and be able to control sharing of personal information. A user evaluation in SWELL [29]showed that one or more of the patterns were noticed by user. However most users indicate that they are not very interested in privacy issues and might have stuck with noticing privacy explanations instead of reading their contents.

Opportunities for PDS. We discussed the vision of Sense-OS on the concept of Personal Data Stores. The way the sense platform currently operates has some overlap with a PDS data ecosystem. The Sense platform stores sensor data from hardware sensors and shares these with apps that use the data; the data and the application are linked via proprietary software. The main difference however is that the user has no fine-grained control over this data sharing.

Because of the similar manner of functioning, the Sense platform could be converted into a PDS by integrating fine-grained data control for the user. Or, the Goalie and Brightr app could be linked to an external PDS to be able to receive more or different data.

Sense sees three main opportunities for PDS's.

Firstly, a PDS that functions as a true integrator of many data streams and many functionalities. The PDS would be the hub linking data and applications: on the one side there are many data sources that all connect to the PDS, which then connects to many applications on the other side. In this way new data providers can easily find and match application providers that might be interested to use their data and, vice versa, new application providers can easily find and use data.

Such a PDS has a few prerequisites. To function optimally, it must have many users, must connect with many applications, and thus have many data streams. It follows from this, that in order to create the desired magnitude, only a few PDS's can exist. This means it is likely that these few PDS's are initiated by companies that already have a very large user base, such as Google or Facebook.

The second opportunity for Personal Data Stores lies in domain-specific PDS's. Rather than bringing together all personal data of an individual, a domain-specific PDS discloses personal data of one domain, such as healthcare data, educational data, or financial data. Currently, initiatives of domain-specific PDS's are on the rise, likely because the need and added value for such domain-specific PDS's are higher. For example in the medical domain, the need for data sharing is high; e.g. a dentist needs to know what medication his patient uses. A PDS can satisfy this need for data sharing. Furthermore, since medical data is privacy-sensitive, a data sharing solution that promotes privacy has added value over a data-sharing solution that does not take privacy into account. Finally, in some domains there are parties that can stimulate or even enforce the use of PDS's (e.g.

[29] D4.12 Evaluation of privacy transparency and control functionality in SWELL. Bob Hulsebosch (InnoValor), Henny de Vos (Innovalor), Joris Jansen (Sense OS). March 2015

governments or other regulating parties), making their implementation in these domains more likely.

Currently, data ecosystems of such domains are based on many-to-many data streams, where many applications connect to many other applications via API's. In such an ecosystem, applications that connect to the most other applications are the most successful. Assuming that currently successful apps keep getting more and more popular, a PDS-hub following an hourglass model can evolve from the current many-to-many model. The successful app with numerous connections will form the central point of the hourglass and direct data streams from and to other applications. It must be noted however that this does not necessarily entail that the central app gives users control over their own data.

The last opportunity is for a PDS to gain momentum because it is connected to a "killer app"; an app so unique and useful that it attracts immense popularity. A PDS on its own is likely not enough to persuade people to start using it; it must be evident what the added value is for them. In principle, users aren't interested in data only; graphs and statistics don't remain interesting for long to the average user. When it comes to sensor data in the health domain, even becoming healthier may not be a strong enough incentive in itself. Sense found that people in general are bad at maintaining a health programme for a longer period of time. Their apps counteract this by nurturing people's intrinsic motivation, enhancing their ability to follow the programme, and by providing the right triggers for behaviour, at the right moment. From their example it becomes clear that in order for a PDS to be successful, it must be embedded in a broader system of useful functionalities going beyond "just" data and clear added value for the user.

6 Discussion

In this chapter we briefly recapitulate the findings outlined earlier, before discussing the implications of a PDS for sensor data with respect to the privacy issues.

Summary. Sensor based applications exploit (often mobile) device technology that register its surroundings, often aimed at behaviour of people, as opposed to conventional computer systems that generally register input devices and internal processes. Aside from general privacy concerns, such sensor data may encroach on privacy in a number of ways: ubiquitous sensors create significantly larger bodies of people-centric data, new types of personal data are uncovered, and in combination with other data may synergistically reveal unforeseen information about human behaviour. Moreover, subjects are often continually tracked over time and may be unaware of the data gathering as sensor operate concealed.

A Personal Data Store is an application that allows subjects of personal data to view, alter, and allow or deny access to their personal data. A Personal Data Store as such may serve different functions, including viewing sensor data to different levels of detail, adjusting various aspects of the data, and allowing full or partial access to sensor data. Through providing an overview of personal data and delegated access management, a Personal Data Store helps to safeguard privacy, establish trust and meet legislative demands. Another main benefit is centralization of access management for multiple data sources and relying parties in a single application.

A Personal Data Store allows data subjects to curate data gathering and access, hence they may mitigate privacy concerns as to the properties of personal sensor data, such as duration, volume, accuracy, and type of data collected and accessed. PDS's may also alleviate to a lesser extend privacy concerns dealing with uncertainty about sensor data gathering and usage, such as user awareness, unexpected or wrong inferences, purpose of data gathering, (re-)identification, and meaningful consent.

Discussion. In this white paper so far it has been furthered that Personal Data Stores are able to resolve most of the privacy issues stemming from or involved in sensor data, and overall have a positive impact on privacy. In this section we will assess whether PDS's are indeed a suitable solution to resolve privacy issues of sensor based data.

A PDS has attractive benefits outside of privacy preservation. A Personal Data Store may offer additional functionality, aside from privacy transparency and control features, such as users improving data quality by supplementing new data or fixing errors, or coupling services with the Personal Data Store. A PDS could entice users to disclose more and better data, leading to better services and less errors. Compared to privacy enhancing technologies per application or data source, a centralized environment such as a PDS that offers the same functionality for multiple applications and data sources is more efficient to the user. Also, as a centralized institution that governs multiple data sources and applications in a single environment, a Personal Data Store may be used by application developers or service providers for discovery of data and applications included within the Personal Data Store.

Before any implementation of a PDS for sensor data may be a suitable and sustainable privacy enhancing solution, some demands, however, have to be met. A Personal Data Store is a relatively complicated solution, befitting sensitive data sources. The privacy sensitivity of sensor data is

deemed to warrant such an approach. It also demands a certain scale of data usage, or would otherwise be too much of an investment. Moreover, whether users will be able to complete the relatively demanding tasks Personal Data Stores ask of them poses a challenge to UX design and communication. One such consideration is whether to offer fine grained controls or predefined settings, for example in the form of profiles.

The PDS becomes a key component in a service system network, therefore it requires a high level of trust in the party developing and hosting the PDS. There are complicated legal issues involved, such as questions of data ownership and responsibility. Hence, a trust framework has to be installed that captures agreements and procedures for all parties involved to reliably engage with the Personal Data Store. Moreover, PDS-applications require rigorous security as it is a gateway to sensitive personal data and has a central function in a service system network.

There are several alternative solutions to privacy transparency and control that may arguably result in a similar level of user autonomy and protection, and may be simpler to implement, such as a settings dashboard or privacy policy. There may also be a wide range of non-privacy related challenges, such as effective marketing or a positive business case. Finally, the Personal Data Store has to be technically integrated in existing IT-infrastructure. Such technical integration requires standardization between different service providers to be interoperable. It may not be feasible to deliver on all the above demands at once, hence the personal data functionality and other demands may likely be implemented in stages.

On a nuancing note, this deliverable addresses Personal Data Stores conceptually, as there are many implementations imaginable that all have different implications for privacy. To illustrate, a PDS could include one or multiple sensors and/or applications, could be a third party or embedded application, could arrange absolute or partial access to data sources, could provide different granularity of control, ranging from absolute settings to risk appetite profiles, et cetera, depending on the context and choices made during development. Also, a PDS is likely to encompass more than just sensor data, but a combination with other data sources for the same or different processes.

Industries that see the most initiatives for personal data management are the health care and public sector. A Personal Data Store for sensor data may be expected for well-being and mobility applications, since these are heavy on sensors and privacy is an important issue.

In conclusion: there is a case to be made for Personal Data Stores as it has merit to protect privacy from sensor data, but it is far from a one-size-fits-all solution, given the many requirements that have to be redeemed.

Sources

We have consulted the following sources:

CTRL-Shift (2011). *The new personal data landscape*. Retrieved from: https://www.ctrl-shift.co.uk/wp-content/uploads/2011/11/The-new-personal-data-landscape-FINAL.pdf

CTRL-Shift (2014). *Personal information management services – an analysis of an emerging market.* Retrieved from: https://www.nesta.org.uk/sites/default/files/personal_information_management_services.pdf

De Montjoye, Y. A., Wang, S. S., Pentland, A., Anh, D. T. T., & Datta, A. (2012). *On the Trusted Use of Large-Scale Personal Data*. IEEE Data Engineering Bulletin, 35(4), 5-8.

Fremantle, P., & Scott, P. (2015). *A security survey of middleware for the Internet of Things*. Retrieved from: https://peerj.com/preprints/1241.pdf

Hardjono, T., Greenwood, D., & Pentland, A. (2013). *Towards a trustworthy digital infrastructure for core identities and personal data stores*. In Global Forum on Identity. Retrieved from: http://www.findthomas.net/blog/wp-content/uploads/2013/05/hardjono-greenwood-coreid04C-ID360.pdf

Kearns, D. (2009, July 7). *Five datasets of the personal data store*. Networkworld.com. Retrieved from: http://www.scientificamerican.com/article/3-projects-prove-privacy-is-not-dead/

Narayanan, A., Toubiana, V., Barocas, S., Nissenbaum, H., & Boneh, D. (2012). *A critical look at decentralized personal data architectures*. Retrieved from: http://arxiv.org/pdf/1202.4503v1.pdf

SWELL D4.15 PIA guidance for sensor-based and big data applications.

Van Kleek, Max and O'Hara, Kieron (2014). The future of social is personal: the potential of the personal data store. In, Daniele Miorandi, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt & James Stewart (eds.), *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*. Berlin, DE, Heidelberg, DE, Springer-Verlag, 125-158. (doi:10.1007/978-3-319-08681-1_7).

Van Velzen, A., Jansen, J., De Vos, H., Siljee, J., Janic, M. & Hulsebosch, B. (October 2015). *Design patterns for privacy transparency in context-aware applications*. Proceedings of the Amsterdam Privacy Conference 2015 [SWELL deliverable D4.10].

World Economic Forum (February 2013). *Unlocking the value of personal data: from collection to usage.*

World Economic Forum (May 2014). *Rethinking personal data*.