Personal Data Usage and Management

Beatriz San Miguel, Jose M. del Alamo and Juan C. Yelmo
[1]*Center for Open Middleware (COM), Universidad Politécnica de Madrid (UPM),*
Campus de Montegancedo, E-28223 Pozuelo de Alarcón, Madrid, Spain

1 INTRODUCTION

Personal data about users (customers) is a key component for enterprises and large organizations. Its correct analysis and processing can produce relevant knowledge to achieve different business goals. For example, the monetisation of this data has become a valuable asset for many companies, such as Google, Facebook or Twitter, that obtain huge profits mainly from targeted advertising.

The increase of personal data is astonishing and the forecasts point to a sustained growth in the next years (Gantz and Reinsel, 2012). Nowadays, personal data is distributed across different Service Providers (SP) or diverse components of Enterprise Architectures (EAs). Moreover, due to users accepting the terms of service and the privacy policies to access services, they are transfering and giving their personal data to enterprises that can use and manage it as they please.

In this context, there is a great deal of controversy surrounding personal data. Different discussions, movements, forums and studies have been promoted, launched and created from many perspectives such as, legal, political, social, psychological, economic and tehcnical ones.

From the legal and political perspective, some goverments such as, United States and European Union (EU) are revising their policies and legislation to evolve data protection. As an example, the current proposal to reform the EU Data Protection Directive includes *Privacy by Design* principles. It means that personal data protection is taken into account in the development of business procceses for services or products, setting high level of privacy by default. Moreover, regarding personal data processing, there is an advice paper (Kohnstamm, 2013) that suggests the inclusion of additional elements in order to provide for a balance approach on personal data processing and mitigate the risks for users. It involves:

- More transparency to users, as they will be able to understand what and why their data is being processing.
- Increasing user control over his personal data and the obtained knowledge about them.
- More responsibility and accountability from SPs or data controllers.
- Allowing a balanced approach, evaluating at the same time the SP interests and the user rights and freedoms.

There are more legislations and organizations, such as, (Schwab et al., 2011) or (Digital Enlightenment Forum, 2013), that defend and promote these principles and fundamental elements. Moreover, users are becoming more concerned about their privacy and they require tools and mechanisms to manage and control their data, while allowing them to use services.

These new demands will entail a greater workload for enterprises. They must implement new features and functionalities that will probably be moved away from their core business. So, the personal data management can result in complicated time-consuming tasks, that maybe not all of the enterprise can assume.

However, it can bring advantages and new opportunities for enterprises. First, the quality of data will be enhanced because users would provide and modify their data and also give feedback. Furthermore, enterprises could access numerous data sources inside or outside their EAs, if the user considers it convenient. In this way, personal data would be reused and new knowledge would be obtained in order to be exploited. Finally, new entities from the private or the public sector could appear to allow enterprises to manage personal data and here, a wide variety of new and novel business models and technical solutions would flourish.

The PhD project presented here addresses different issues that ease the change to the new paradigm. We envision the appearance of personal data frameworks where enterprises can share personal data and knowledge about it, respecting user preferences about his privacy.

San Miguel B., del Alamo J. and Yelmo J..
Personal Data Usage and Management.

2 STATE OF THE ART

Personal data is the digital data created by and about one person (Schwab et al., 2011), including three main types:

- Volunteered: explicitly created by the user, for example, as he fills in a registration form, he is willingly giving demographic data or expressing his preferences.
- Observed: captured by enterprises when users use their services, e.g., location data, browsing history, temporal viewing behaviour, ratings, or purchases.
- Inferred: created by processing volunteered and observed data. For example, interest predictions or purchase purposes.

The set of personal data is denoted as *identity*, *"user profile"* or *"user model",* depending on the application area. The three terms are closely interrelated. However, their origin and therefore, their techniques and technologies differ considerably.

An identity is the representation of a person in form of one or more information elements that allows its distinguishing within a context (International Telecommunication Union, 2009). It can contain volunteered, observed or inferred data but its main feature is the data (or set of it) that identifies a person. In this way, an identity is traditionally applied to functions and capabilities related to Identity Management such as authentication, access control, discovery, security or privacy issues.

Identity Management systems define a set of functions and capacities used for the assurance of the person identity while supporting business and security applications. We can find diverse models to carry out the previous one but generally, they consider three entities:

- Identity Provider that creates and secures the identity
- SP provides users with services that need the user identity to do it
- User who accesses services

On the other hand, the terms *"user profile"* and *"user model"* are applied in Human-Computer Interaction (HCI) area. They are usually regarded as synonyms but sometimes one of them is used meaning both. However, they can be distinguished. Taking the differentiation that many authors have done (Fröschl, 2005) as the starting point, we can define a user profile as a collection of volunteered and/or observed data about a person. It is raw personal data, without any processing or interpretation. Depending on the business goals and the amount of personal data that a user profile contains, it can be processed and interpreted to obtain a user model. This process is called user modelling.

A user model is the interpretation of a person in a specific context for an enterprise, a representation that includes what an enterprise thinks that a user is, prefers, wants or is going to do. It comprises mainly inferred data but can also include some volunteered or observed data. In an EA, a user model can be used to predict user behaviour, recommend new contents or services, personalize user interaction or adapt user experience, among others. In the end, it represents a core business element, as *"user model"* is becoming a substantial part of customer satisfaction: it can improve user experience and engage users; therefore, it can provide enterprises with better-quality services. These improvements will develop into competitive advantages and economic benefits.

User modelling process covers different stages that can take place cyclically (Barla, 2010). First, a data collection stage analyses what, how and where the personal data is obtained. Then, an inference process obtains relevant knowledge with the previous data and creates a user model. Finally, an application of the user model is done. This application is the user model purpose or use and can generate new personal data to give feedback to the first stage.

Figure 1 shows the relationship between the previous terms. Here, we can observe that a person, our user, has different identities that are used to access diverse services. Moreover, each SP can produce one or more user models that represent user with different goals.

3 RESEARCH OBJECTIVES

The main project objective is to establish a personal data framework that allows enterprises to register, discover, access, recover and use personal data and specifically, user models, while providing users with tools and mechanisms to manage and control their data. It involves numerous challenges from Identity Management to User Modelling areas, including Privacy field.

This PhD project is focused on three specific objectives:

- Definition of an interoperable user model that represents user in different contexts and allows the incorporation of new personal data, user profiles and user models from diverse data sources.

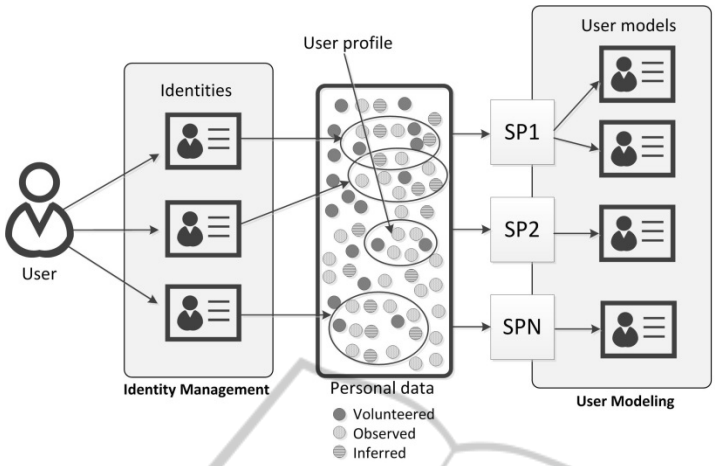

Figure 1: Personal data and related terms.

- Identification of mechanisms necessary to empower users to effectively manage their data.
- Research of the threats and opportunities in a personal data framework where the previous results apply.

Our main focus is to contribute with technological solutions related to user models that allow the creation and exchange of them and facilitate enterprises to manage them inside the new personal data framework. It will involve the actualization of existing business models and the appearance of new ones that we will also analyse.

4 METHODOLOGY

We are going to follow an iterative and incremental methodology, where stages of analysis, design, validation and refinement are performed cyclically. We will combine a top-down and bottom-up approaches. The top-down approach will allow us to analyse the global personal data ecosystem, identifying lacks and opportunities continually. On the other hand, we will mainly adopt a bottom-up approach to integrate and improve the different contributions of each objective into the global vision that we will be creating.

Specifically, we have defined the next tasks:

- Comparative research of the existing user models to identify common characteristics and lacks to cover.
- Formal definition of an interoperable user model that allows enterprises to represent and incorporate new personal data and user models from diverse data sources.
- Validation of the interoperable user model in a case study that includes personal data of social networks and other services.
- Comparative research of the existing solutions that allow the registration, access, management, control and usage of data. It will not be limited to personal data and other fields such as, open data, linked data or social good data will be considered.
- Specification and design of a framework that allows users to transfer and control their personal data and user models while they can be used and/or exploited.
- Validation of the global solution in the context of a project called POSDATA of the Center for Open Middleware (COM), where personal data from authenticated bank users will be analysed.

Research and definition of business models, value proposition, value chain, entities, roles, relationships, revenues, etc. that are applicable to the new framework will be continuously performing and evolving during the project.

5 STATE OF THE RESEARCH

We have previous expertise in user-centric approaches to identity management and relevant contributions in this area, for example these international patents (Monjas et al., 2012) and (Monjas et al., 2013), that can be applied. These patents deal with different issues related to the personal data flow:

- The former defines a set of methods for allowing users to carry out privacy management in an identity network. This kind of networks includes different SPs and a special one called identity provider, forming what is called a “circle of trust”

in the Identity Management area.

- The other patent relates to the selective distribution of information in a communication network such as the Internet. In particular, it centres around the methods that SPs must implement to perform the distribution of personal data, reducing the operational burden on users.

On the other hand, we are currently focusing our efforts on user modelling process. Specifically, we have an ongoing study of standards, vocabularies and specific personal data representations that are used by main SPs and Identity Management Systems. Here, we have observed that there is a wide variety of solutions. Moreover, the trend points towards silo approaches. The revised solutions focus on personal data but not on user models due to the almost non-existent information about them. In this sense, we are going to research tools and mechanisms that allow generic user modelling (Kobsa, 2007).

The so-called generic user modelling systems allow the reuse of user modelling components for others systems, isolating or separating the user modelling techniques. Many features differentiate these systems (Carmagnola et al., 2011) and we want to go in depth with the methods that use them to acquire and represent personal data and user models, and to create the user models.

6 EXPECTED OUTCOME

This PhD project sets out a change in the current context related to personal data and knowledge about it. This change in itself involves numerous challenges varying from legal to technical issues and our efforts will be focused on the last ones.

We expect to define a new personal data framework that integrates different contributions from Identity Management and User Modelling areas, including Privacy field, to empower users regarding their personal data, while enterprises are still able to exploit it. Nowadays, the most relevant challenge that we face is to find the appropriate methods and techniques to define an interoperable user model that represents and incorporates new personal data and user models from diverse data sources.

REFERENCES

Barla, M., 2010. *Towards Social-based User Modeling and Personalization.* Dissertation Thesis at Slovak University of Technology in Bratislava. Available at < http://acmbulletin.fiit.stuba.sk/theses/barla-thesis.pdf > (31 January 2014).

Carmagnola, F., Cena, F. Gena, C., 2011. User model interoperability: a survey. *Journal User Modeling and User-Adapted Interaction*, Vol. 21, issue 3, pp. 285-331. Springer Netherlands.

Digital Enlightenment Forum, 2013. Available at: http:// www.digitalenlightenment.org/. (31 January 2014).

Fröschl, C., 2005. *User Modeling and User Profiling in Adaptive E-learning Systems.* Master's Thesis at Graz University of Technology. Available at http:// www.iicm.tugraz.at/0x811bc82b_0x00029076. (31 January 2014).

Gantz, J., Reinsel, D., 2012. *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, IDC iView. Available from: <http://idcdocserv.com/1414>. (31 January 2014).

International Telecommunication Union, 2009. *Recommendation ITU-T X.1250: Baseline capabilities for enhanced global identity management and interoperability.* Telecommunication Standardization Sector of ITU.

Kobsa, A., 2007. *Generic User Modeling Systems.* The Adaptive Web Lecture Notes in Computer Science, Vol. 4321, pp. 136-154. Springer-Verlag, Berlin Heidelberg New York.

Kohnstamm, J., 2013. *Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation.* Article 29 Data Protection Working Party, Brussels. Available at: <http://ec.europa.eu/justice/data-protection/article-29/>. [31 January 2014].

Monjas, M. A., del Alamo, J. M., San Miguel, B., Trapero, R., Yelmo, J. C., 2013. *Method for selectively distributing information in a computer or communication network, and physical entities therefor.* Application US 13/809,503. PCT/EP2010/059264.

Monjas, M. A., del Alamo, J. M., San Miguel, B., Yelmo, J.C., 2012. *Method for Privacy Management in an Identity Network, Physical Entities and Computer Program Therefor.* Application US 13/263,669. PCT/EP2009/054223.

Schwab, K., Marcus, A., Rico, J., Hoffman, W., 2011. *Personal Data: The Emergence of a New Asset Class*, An Initiative of the World Economic Forum in Collaboration with Bain & Company, Inc. Available at:<http://www.weforum.org/reports/personal-data-emergence-new-asset-class>. (31 January 2014).