Building an Eco-System of Trusted Services via User Control and Transparency on Personal Data

Michele Vescovi, Corrado Moiso, Mattia Pasolli, Lorenzo Cordin, Fabrizio Antonelli

▶ **To cite this version:**

Michele Vescovi, Corrado Moiso, Mattia Pasolli, Lorenzo Cordin, Fabrizio Antonelli. Building an Eco-System of Trusted Services via User Control and Transparency on Personal Data. 9th IFIP International Conference on Trust Management (TM), May 2015, Hamburg, Germany. pp.240-250, 10.1007/978-3-319-18491-3_20 . hal-01416231

HAL Id: hal-01416231
https://inria.hal.science/hal-01416231v1

Submitted on 14 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Building an Eco-System of Trusted Services via user Control and Transparency on Personal Data

Michele Vescovi [1], Corrado Moiso [2], Mattia Pasolli [1], Lorenzo Cordin [1,2], Fabrizio Antonelli [1]

[1] Telecom Italia, Semantic & Knowledge Innovation Lab (SKIL),
via Sommarive 18, I-38123 Trento, Italy
[2] Telecom Italia, Future Center,
via Reiss Romoli 274, I-10148 Torino, Italy

{michele.vescovi, corrado.moiso, fabrizio.antonelli, mattia.pasolli, lorenzo.cordin}@telecomitalia.it

Abstract. The amount of personal information that is generated and collected on a daily basis is rapidly growing due to the increasing number of activities performed online and, in particular, in mobility. The availability of such a huge amount of data represents an invaluable opportunity for organizations and individuals, respectively to enable precise business intelligence and innovative services. Nevertheless, it represents the commodity of a flourishing "market of data", mostly fostered by the biggest ICT companies, from whereof benefits users are almost excluded, significantly increasing the public concern on data privacy. In this scenario we developed a framework, based on a personal data store, enabling the development of an eco-system of trusted application, which allow users to full transparency and control on the exploitation of their data.

1 Introduction

The increasing adoption of smartphones and their capability of collecting personal and contextual information have generated a tremendous increment in the production of Personal Data (PD). The amount of PD that is available and generated on a daily basis is rapidly growing also due to the increasing number of activities performed online and, in particular, in mobility. Nowadays, a constantly increasing number of users access the internet mostly (or exclusively) by means of their smartphones/tablets; they use an innumerable variety of online/Web services, often through specifically designed mobile applications (shortly apps). Very often these apps, which possibly connect to external devices/sensors, are their self-generated source of novel types of PD that are transmitted and collected server-side.

The availability of such a huge amount of data (ranging from locations or interactions record, to the content produced by users, e.g. describing choices, preferences, etc.) represents an invaluable opportunity for organizations and individuals to enable new application scenarios and to benefit from innovative

services. Nevertheless, they represent the commodity of a flourishing market mostly fostered by the biggest ICT companies. The collected data are exploited for internal business analytics or sold to drive third party business intelligence or advertisement.

However, it has also significantly increased the public concern on data privacy. In fact users have, in general, very scarce opportunities to control how their data are accessed and collected and to use them for their purposes. Some Operating System (OS), such as Android, informs users about the accessed resources at installation time but, mainly, such information is not user-friendly and users can either avoid installing the App or grant unrestricted access to all the required resources, without means to further control permissions or to audit the access to PD. Increasingly this phenomenon is widespread due to wide adoption of the so called *social logins* (i.e. the use of the credentials of a particular social network to log also into third party services) through which, third party service-providers can directly access the plethora of users' information collected through their use of social networks. This further sharpens the incongruity between the information consciously and transparently disclosed by users and the users' personal data concretely accessed by third parties. Nevertheless, the benefits of the usage of personal data are always more imbalanced between the users and, mainly, the world-wide biggest ICT companies.

This is far from the desired scenarios in which PD enable and contribute to the generation of widespread socioeconomic benefits to the collectivity. In order to reach these benefits, we believe that we need a fair PD management, where individuals, empowered with control and awareness over their PD, are enabled to actively and knowingly participate (in)to a PD eco-system. We present the design of *My Data Store*, currently validated in a living-lab: a privacy preserving service that enables users to collect, control and exploit PD generated in mobility. We integrated *My Data Store* into an innovative framework which enables the development of trusted and transparent (in terms of access and use of PD) services and apps: a user can control and audit their behavior. In this way, it is possible to create an eco-system of PD-based trusted apps/services. The framework, acting as a broker, would also potentially allow users to gain direct economic benefit from the disclosure/exchange of their data.

2 The Context

We are experiencing a rapid change of paradigm in technology and in business, where data are becoming an essential resource for the design of new and better services and products. The amount of data available, generated and processed on a daily basis is so huge and rapidly increasing. "Big Data" has become the keyword around which innovation, competition, and productivity in ICT are orbiting, so as to create a new data-driven society. One of the most interesting classes of data is Personal Data (PD, i.e., any information relating to an identified or identifiable person): they are data about people, their behavior, their preferences, etc. When handled and interpreted such data can describe an individual's actions, behaviors and habits [2].

While so far most of the PD had been static (e.g. socio-demographic profiles), the smartphones, jointly with many other connected personal devices (e.g., environmental sensors, wristbands, etc.), have enabled the collection of highly dynamic PD,

describing the behavior of people in the real life (e.g. locations, communication patterns, social interactions, apps usage, etc.). The exploitation of this data is a key element for enabling the design of novel personal services able to improve users' experience. These services, moreover, can generate novel types of PD, becoming also source, not only consumer, of PD that continuously enriches the user's "digital trace". Assuming it is possible to gather all these data from people, we have a perfect example of "Personal Big Data" with enormous potentials. The availability of such a huge amount of PD is an invaluable resource and opportunity for organizations and individuals to enable new applications and businesses. Organizations can leverage on these PD to have a deeper understanding of people's needs and behavior, either as single individuals or communities, and can provide tailored services. Accordingly, people (the actual data "owners") can benefit from the creation of novel personalized apps with an enhanced user-experience that help them measure/track and improve the quality of their life.

The current adopted models of managing PD often do not fully allow a controlled and effective exploitation of these opportunities; in addition, people are currently excluded from the life-cycle of their data, relegated to the role of PD producer with limited ability to control and to exploit them. PD are collected by several services in a fragmented (often redundant) way and then spread in the data centers of a multitude of organizations, which manage them according to the specific agreement signed by people. This results in several limitations: (i) it is not possible to have a holistic view of individuals, as their PD are collected and stored in several independent silos; (ii) there is a limited involvement of people, thus resulting in a scarce possibility for them to understand how their PD have been used; (iii) people cannot manage a copy of their PD, with great limitation on the possibility for them to fully exploit their PD.

In order to overcome the drawbacks of current "organization-centric" approach, a new user-centric model for PD management has been proposed [11], [7]. In general these initiatives promote the possibility for people to have a greater control over the lifecycle of their PD (e.g., collection, storage, processing, sharing) and they recognize the crucial and active role played by a person into a righteous and fruitful PD ecosystem. As mentioned by [1] a user-centric paradigm should complement and not replace the organization-centric one. A key aspect of the PD user-centric model is the right of the user to have copy of all their PD [7]. While this is a first step this right does not necessarily create value for people, if not combined with tools for their PD management and for easily and dynamically control how PD must be accessed and exploited by the services. A Personal Data Store (PDS) platform [10], [4] delivers a set of services enabling the owners of PD to collect, manage, track and control their data according to their wills and needs.

3 The Mobile Territorial Lab Experience

We designed and experimented *My Data Store*, a PDS platform able to manage heterogeneous PD, from those collected by apps and sensors on smartphones or on connected devices (e.g. environmental sensors, etc.), to those gathered from online services (such as social networks) or organizations in relation with their

customers/users to whom they offer services (e.g. network operators, service/utility providers, retailers, etc.). We then devised an innovative framework for a trusted PD management built around the role of *My Data Store*, as far as any other PDS. Its architecture supports developers in building trusted and transparent apps compliant with a user-centric PD management model. Specifically, it provides users with a mobile UI allowing them to dynamically control and audit the accesses, collection and usages of PD by means of the compliant apps. In this way we aim at creating an eco-system of trusted and transparent apps, feeding and exploiting the data stored in the PDS, pushing forward paradigms and common practices in PD management.

In cooperation with other partners, we are currently experimenting [9] the user-centric PD paradigm within the "Mobile Territorial Lab" (MTL, www.mobileterritoriallab.eu), a long-term living lab where a real community (involving about 150 families) experiments this new paradigm in a real living environment: in fact the participants to MTL collect, manage and use their PD while they act in their real life, e.g., by interacting and performing digital activities through their smartphones, and by using ad-hoc designed apps exploiting their PD. The project and its objectives have been included in the World Economic Forum reports of the Rethinking Personal Data initiative (pg. 28 of [11]). Participants to MTL are provided with a smartphone empowered with a sensing SW continuously and passively collecting users' data in independent users' silos. The data automatically sensed consist of: i) call and SMS logs, ii) proximity data scanning for near-by devices, iii) locations from GPS and WiFi. Additional PD such as iv) mood and v) expenses done by the participants are collected through experience sampling methods by means of ad-hoc apps. MTL participants collect also vi) the air-quality in their surroundings (e.g. CO and other gasses levels) by geo-referencing the values measured by an environmental sensor connected to the smartphone. Each MTL participant is then provided with a private *My Data Store* account, through which they can access, visualize and transparently manage all the mentioned PD collected about them.

4 My Data Store

My Data Store offers a set of tools to manage, control and exploit PD by enhancing an individual's awareness on the value of their PD. The *My Data Store* development has been driven by principles emerging from existing studies on PD management, as in [6], [8]. In particular, the principles followed are:

- *Participant Primacy*: users should be provided with appropriate functionalities that enable them to have complete control over the management of their PD. This includes a nuanced process of permission-granting ensuring that users can easily move through the collection and sharing settings and take clear actions over them (decide which data gather in their space, share/delete sub-sets of them, etc.);
- *Data Legibility*: users should be supported in understanding the meaning and the potential of different kinds of data, as well as the risks and the consequences associated to PD usage (e.g. the data which can be inferred, also from aggregated or anonymous collection of data, when PD are shared with 3rd parties);

- *Long-term Engagement*: it is necessary to provide users with technologies for controlling their data (collection, sharing, visualization, etc.) but it is equally important that the system and the process of collecting and managing PD is perceived to be relevant. Services and tools provided to users over their data and exploitation opportunities can help and enhance such engagement.

While the above general principles inspire the design of any PDS service, the methodology followed during the design of *My Data Store* included a specific focus-group study whose goal was to identify elements and guidelines relevant for users in a real setting, to be used as drivers. The results highlighted the following:

- *PD Awareness*: participants did not really realize the extent of the PD impact on their life and initially struggled to understand the value of PD and thus the need of a technology to manage them. This raised the need of providing support for guiding the users in the discovery of the PD risks and opportunities to make them aware of the meaning, the value and potentials of their data (as in *Data Legibility*). One of the solutions adopted is the usage of visual elements that lower the barrier for non-skilled;
- *Personally Meaningful Data as Triggers*: one goal in the focus groups was to identify the scenarios perceived as valuable by the users and that ensure the *Long Term Engagement* general principle. The scenarios emerged as more relevant to the users are those related to time or cost savings. The design of the system should hence consider personal values to enforce users' engagement with PD and thus must provide relevant and intuitive data management features. The increasing exploitation of PD contextually to the growth of an eco-system of services/apps built on top of PD further foster the effects of this driver;
- *Social Comparison* of users with other users (single, similar or particular groups/communities of users): it can both work as a tool to improve awareness and to stimulate the user engagement.

4.1 My Data Store Services

My Data Store is a Web portal with a controlled access that makes available to the granted users a set of tools for managing their PD, collected from several sources. The design of *My Data Store* was driven by the principles described above. In particular, we focused on three drivers: empowering people with full control over the life-cycle of their PD, improving their awareness on the data and enabling the exploitation and use of PD in accordance to their needs and willingness. The design aimed also at simplifying the user experience by providing people a limited, but clear and powerful set of capabilities.

4.2 Data Regions

In order to increase users' awareness and their ability to control, the data collected in *My Data Store* are organized in Data Regions (DR). Under the principle of Data Legibility DRs are created by grouping in the same region different data sources w.r.t. their (i) perceptiveness, i.e. considering the information that can be inferred from the

data, abstracting from technical details (e.g., both GPS and nearby WiFi AP, bring along the localization concept, etc.) (ii) sensibility to privacy, i.e. data with comparable levels of privacy-risks (e.g., data concerning interactions between individuals, such as calls, SMS, Bluetooth contacts, etc.). Every DR is associated with a brief description of the information brought by the PD and its list of data sources (Fig. 1). At this stage all the *My Data Store*'s features operate in an uniform way w.r.t. the DR, but in future it will allow expert users to customize their settings w.r.t. the single sources or service/application generating data.

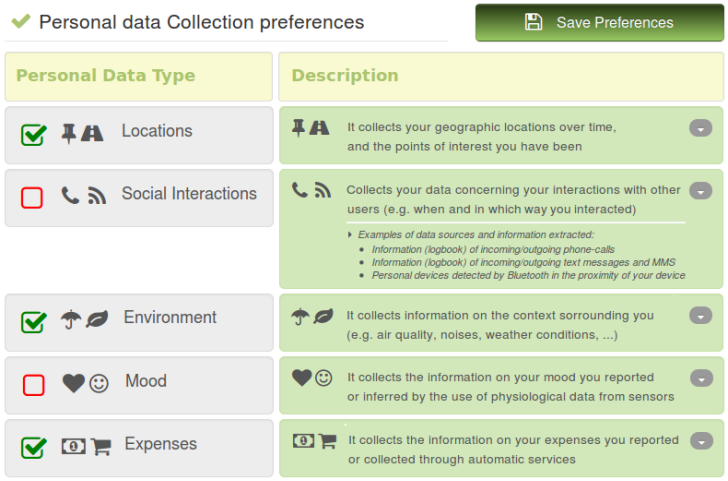

Fig. 1. Example of *Collection Area* (organized by Data Regions).

4.3 Main functions

My Data Store includes PD Management features to fulfill the *Participant Primacy* principle over the entire PD life-cycle (i.e., from the collection to the deletion of a data record). Its main functions are:

- *Collection Area*: In the Collection Area users can choose how DRs are collected and stored (Fig. 1). Users then have a complete set of controls for tuning the settings the best fit with their privacy concerns, exploitation or usage wills (indeed, the PDS is associate to a collector application, running on the users' devices, which is the responsible of collecting and sending the information desired by users);
- *Sharing Area*: Users can set the disclosure level of the collected data by granting those who can access them and the level of detail. So far the choice concern only the disclosure with the participants of our experimentation community, but further options will be considered in future providing finer granularity;
- *Deletion Area*: users can delete single records or all PD collected in a specific DR and time interval.

Fig. 2. Examples of *Individual Views*: (a) Expenses (auditing), (b) Locations (aggregated), (c) Mood (aggregated), (d) Expenses (social).

My Data Store makes use of visualization elements such as graphs, chart, etc. to show the data in the above areas at different levels of details/aggregation. First, this choice aims at increasing users' awareness on the value of their PD, in terms of perception on the data informative power and on the level of risks arising from PD disclosure and exploitation (Data Legibility principle). Second, being able in providing intuitive, interesting and representative visualizations play a crucial role for a *Long-term Engagement* of users and in stimulating users toward the exploitation and use of their PD (Data as Triggers), e.g. by sharing them for apps/services or social comparison. The types of visualizations provided by *My Data Store* are (Fig. 2):

1. *Detailed "Auditing" Views* in tables or maps, where every available piece of raw data for every data source is represented in detail;
2. *Aggregated Individual Vie*ws with aggregations, at different levels, of the PD owned by a single person (e.g. charts, pies, clusters of frequent locations, distance travelled, quantity of contacts, etc.) which aims at increasing his/her consciousness on daily behavior);
3. *Social Views* built from the PD voluntarily shared (with different levels of details) by other users enabling a social comparison of a person's behavior (e.g., "*How much am I social?*", "*How my spending pattern does compare with others' one?"*) with the ones of similar users (e.g. by gender, age).

My Data Store differs from similar systems, such as [3], for the wide types of dynamically collected PD and for the flexibility and the user-friendliness of its Personal Data Management (PDM) features.

5 Toward an Eco-System of Trusted Personal Applications and "the Bank of Personal Data"

On top of the PDM features provided by *My Data Store* we designed a framework to manage applications (e.g., offered either as online services or as apps on devices) accessing PD in a trusted way. The goal is dual:

1. to provide users with a way to discriminate PD *Trusted and Transparent Applications*, i.e., those compliant with the user-centric approach, and to empower them with a clear description of the potential impact of each of such apps on PD (e.g. the generated/accessed PD, declared granularity/quality) and with tools to monitor/audit PD access at real time;
2. to provide apps developers with a workflow process and architecture enabling them to be compliant with the requirements of a trusted and transparent user-centric approach, easily interacting with solutions for PD collection and control, such as a PDS.

To achieve the goal the applications that want to be compliant with the framework must define (and keep updated) a "Statement", which declares the list of PD types handled by the application and for each of them:

- the quality and granularity of the data acquired or generated (type, accuracy, tolerance, etc.);
- the terms by which these PD will be handled by the application (frequency, purposes, etc.);

and distinguishing also among different usages:

- *Dynamically accessed* (i.e. required instantly at run-time) by the application (e.g. accessing the current user location in order to provide location-aware hits when searching the Web);
- *Collected* (i.e. stored for future uses into a repository such as the PDS);
- *Accessed "historically"* (i.e. the requirements and usages of historic data accessed from the repository, e.g. for providing maps, charts, or inferences based on long-term analysis of user behaviors);
- *Shared* (i.e. used and transmitted –possibly aggregated and/or anonymized– into applications which are not only personal but also shown or used, for any reason, to/by other users or third parties).

Potentially the Statement could include further information "describing" the application capabilities that require each specific PD usage, i.e. those that could stop properly working on the chance that the user revokes the grant to perform that action on that particular data (so that the user can be properly informed).

Users are provided with a Mobile or Web UI (the *Apps Area*) allowing them to check the list of enabled (e.g., those installed on their devices) trusted applications (Fig. 3, left) and to control at any time (granting or denying usages) how they can access/use the different PD (Fig. 3, right) accordingly to their Statement.

The framework provides a set of APIs for the development of compliant trusted applications, thus for the run-time access, collection and historical usage of the user PD. These APIs could be provided either server-side by the PDS, or device-side as a layer on top of the device OS or directly ensured by it. These APIs guarantee that any access to the PD is performed (and thus contextually monitored and logged) in

compliance with the application Statement and the user settings. Applications could be certified on the basis of their exclusive usage of "trusted and transparent" APIs.

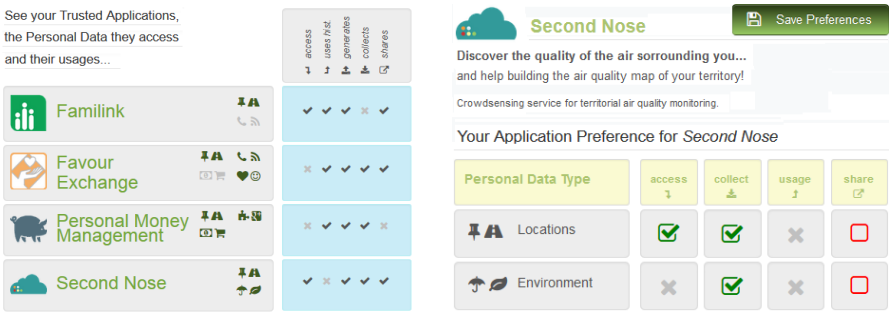

Fig. 3. Example *Trusted Applications*: app list (left); user settings for air-quality app (right).

The framework middleware, thus, automatically carries out the retrieval and collection in/from the PDS of the PD for which the user requested/granted the collection (so that the user is empowered with full transparency on the accesses PD, exploiting the PDM features of the PDS, and further put into value/exploit the collected data) and the collection of the auditing information. Similarly (mobile-) browsers and Web applications could provide the same set of APIs enabling the collection and monitoring of other PD (e.g. browsing history, bookmarks, published content, etc.). On the basis of the auditing logs the framework can provide to the user reports on PD collection and usage, on services/applications behaviors and reports if a service/application behaves in line with its statement.

Interestingly our technology enables a very diverse variety of applications. One specific application scenario could concern the direct monetization of users' personal data, gaining personally an economic remuneration from the disclosure of their personal data. In this last scenario the PDS could play the role of the *"Bank of the users Personal Data"* [5] where, not only the data of the user can flow and be stockpiled from different data sources, but also various exploitation opportunities can be proposed to the users as a data brokerage platform, on the basis of their choices/policies, with all the appropriate protections (e.g. anonymization and/or aggregations of data from large, selected user bases). The framework, in fact, allows for the complete auditing of data accesses and usages enabling, contextually, the automatic accounting and the dynamic user control on data exploitations.

6 Conclusions and Challenges

Introducing *My Data Store* we tried to push further user-centric Personal Data Management, extending the concept of Personal Data Store (PDS) so as to provide users with full control and awareness along the whole life-cycle of their PD from data collection, to data exploitation into added-value services and even data monetization.

In particular we focused our effort in letting the PDS be the core element of a novel framework enabling a new generation of personal apps and services. Our framework provides apps developers a way to access PD in a fully transparent way, consistently with the user choices, and aims at supporting users and developer in the process of PD collection (improving user awareness and widening the set of controlled exploitation opportunities) and in monitoring the apps/services real behavior. This framework could be included, e.g., as middleware of (mobile) platform or OS, while the PDS features could be provided as cloud services.

Even if it cannot prevent from misuses (e.g. illegal copies) of the PD to which the access has been granted and cannot prevent from the coexistence of "non-trusted and non-transparent" applications, our framework shows the way toward the provisioning of transparent, privacy-preserving applications. This solution will fully satisfy the user "right of copy", enabling them to benefit from data reuse, data fusion or monetization scenarios. The main challenges consist of creating the user demand (and thus developer and OS provider availability) towards this kind of applications and to let the PDS-like services become "the Bank of users' PD" [5].

We believe that a change of paradigm in PD management and the construction of a fruitful eco-system of data producers and consumers pass only through such an enhanced transparency and the empowerment of individuals; in particular, individuals should have the possibility of controlling and exploiting their PD, consciously and actively participating in the eco-system, In this way, the value of PD is unlocked and transformed into business and societal value,

References

1. Doc Searls: The intention economy: when customers take charge. www.searls.com/time2grow.html (2012)
2. Dong, W., Lepri, B., Pentland, A.: Modeling the Coevolution of Behaviors and Social Relationships using Mobile Phone Data. Proc. of ACM MUM (2011)
3. Hong, J., Landay, J.: An Architecture for Privacy Sensitive Ubiquitous Computing. Proc. ACM MobySys (2004)
4. Moiso, C., Antonelli, F., Vescovi, M.: How Do I Manage My Personal Data? - A Telco Perspective. Proc. of the Int. Conf. on Data Technologies and Applications (2012)
5. Moiso, C., Minerva, R.: Towards a User-Centric Personal Data Ecosystem – The Role of the Bank of Individuals' Data. Proc., ICIN 2012 (2012) 202–209
6. Mun, M., Hao, S., Mishra, S., et al.: Personal Data Vaults: A Locus of Control for Personal Data Streams. Proc. of ACM CoNext (2010)
7. Pentland, A.: Society's Nervous System: Building Effective Government, Energy, and Public Health Systems. IEEE Computer, 45(1) (2012) 31-38
8. Shilton, K., Burke, J., Estrin,* D., et al.: Designing the Personal Data Stream: Enabling Participatory Privacy in Mobile Personal Sensing. Proc. Communication, Information, and Internet Policy (2009)
9. Vescovi, M., Perentis, C., Leonardi, C., Lepri, B., Moiso, C.: My data store: toward user awareness and control on personal data. Proc. ACM UbiComp 2014 (2014) 179–182
10. Wang, J., Zhongji, W.: A Survey on Personal Data Cloud The Scientific World Journal 2014 (2014)
11. World Economic Forum: Unlocking the Value of Personal Data: From Collection to Usage. www.weforum.org/issues/rethinking-personal-data (2013)