My Data Store: Toward User Awareness and Control on Personal Data

Michele Vescovi
Semantics & Knowledge Innov. Lab. (SKIL) – Telecom Italia
via Sommarive, 18 - Povo
Trento, I-38123 ITALY
michele.vescovi@telecomitalia.it

Bruno Lepri
Fondazione Bruno Kessler (FBK)
via Sommarive, 18 - Povo
Trento, I-38123 ITALY
lepri@fbk.eu

Christos Perentis
Fondazione Bruno Kessler (FBK) & SKIL - Telecom Italia
via Sommarive, 18 - Povo
Trento, I-38123 ITALY
perentis@fbk.eu

Corrado Moiso
Future Center – Telecom Italia
via Reiss Romoli, 274
Torino, I-10148 ITALY
corrado.moiso@telecomitalia.it

Chiara Leonardi
Fondazione Bruno Kessler (FBK)
via Sommarive, 18 - Povo
Trento, I-38123 ITALY
cleonardi@fbk.eu

UbiComp'14 Adjunct, September 13-17, 2014, Seattle, WA, USA
ACM 978-1-4503-3047-3/14/09.
http://dx.doi.org/10.1145/2638728.2638745

Abstract
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. The availability of such a huge amount of data represents an invaluable opportunity for organizations and individuals to enable new application scenarios. However, it has also significantly increased the public concern on data privacy. In this paper, we present My Data Store, a tool enabling people to control and share their personal data. We tested My Data Store with 63 participants that used it in order to manage their own data collected from mobile phones and through experience sampling applications. Preliminary results show improvement over the users' awareness of their personal data and the perceived usefulness of the tool.

Author Keywords
Personal Data; Awareness; Living Lab; Human Factors

ACM Classification Keywords
H.1.2. User/Machine Systems: human factors; H.5.2. User Interfaces: user-centered design; K.4.1. Public Policy Issues: privacy.

Introduction
Personal Data (PD), i.e. data about individuals' behavior and preferences, are one of the richest classes of data. PD can be used to better understand the behavior of

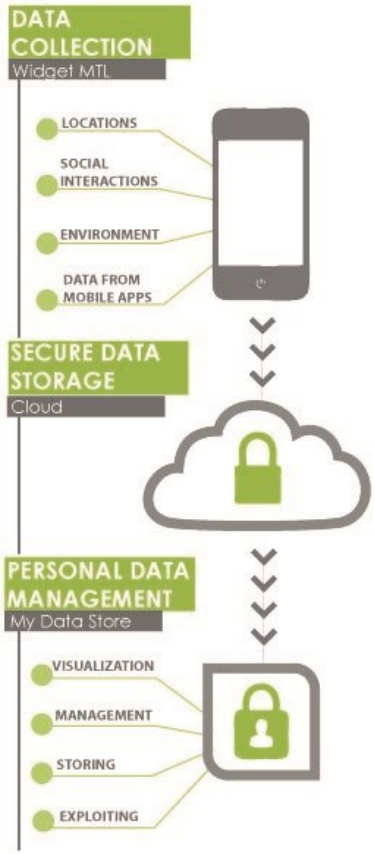

Figure 1. Personal data platform organization and data flow.

communities and individuals and to support the design of novel personalized applications to enhance user experience and improve quality of life (e.g. self-quantified applications) [1,2]. So far, most of the collected PD were static (e.g. socio-demographic profiles), but smartphones enable the collection and the processing of highly dynamic PD in a pervasive environment, describing real life user behavior (e.g. locations, communications).

Despite the potentials of PD, the current management models do not fully allow an effective exploitation of these opportunities. PD are usually collected by several applications in a fragmented and often redundant way. Furthermore, the limited involvement of users in their data life-cycle results in lack of transparency, concerning to whom and for what purpose their data has been used.

In order to overcome this situation, a new user-centric model for PD management[1] has been proposed enabling higher control over the lifecycle of PD [4]. Researchers and companies have started developing repositories (e.g. openpds.org) for various types of personal data collected online or ubiquitously [3, 5, 7].

However, up to now few studies explored how people approach this user-centric paradigm. In this paper, we first describe a Personal Data Store (PDS) platform, called My Data Store, delivering to people a set of web-based services enabling them to collect, manage, control and exploit their ubiquitously collected PD. Then, we present preliminary findings gathered through a 15-weeks longitudinal study involving a Living Lab of 63 participants that used the service to manage mobile-phone data (call and SMS logs, proximity interactions detected by Bluetooth hits, and location data) continuously collected by a mobile sensing platform and daily data on mood and spending records collected through experience sampling applications.

[1] www.weforum.org/issues/rethinking-personal-data

The Mobile Territorial Lab

The *Mobile Territorial Lab* (www.mobileterritoriallab.eu) project (*MTL*) is a long-term living lab focused on human dynamics understanding based on mobile phone data analysis. MTL aims at exploiting smartphone sensing capabilities to allow the unobtrusive and cost-effective access to PD sources related to daily social behavior (location, communication, proximity, etc.).

Participants are provided with an Android smartphone equipped with a sensing widget (including and largely extending Funf: www.funf.org) that gathers the sensed PD from mobile phone into user's PDS (My Data Store), where they flow into PD collected from other sources (e.g., experience sampling applications, environmental sensors, etc.). Collected data are stored and secured in the cloud in independent user's silos, from where they are accessed, i.e. My Data Store space, or exploited by personal/social services (Figure 1).

My Data Store

My Data Store is a secure digital space, owned and controlled by the user, acting as repository for personal information. Three main drivers were identified to guide the design of My Data Store: (i) empowering people with control over their PD life-cycle, (ii) improving people's awareness on their data, that is to assist the user to better understand the PD potentials by looking into her data and comparing her behavior with others (e.g., a subject might gain more awareness on her daily

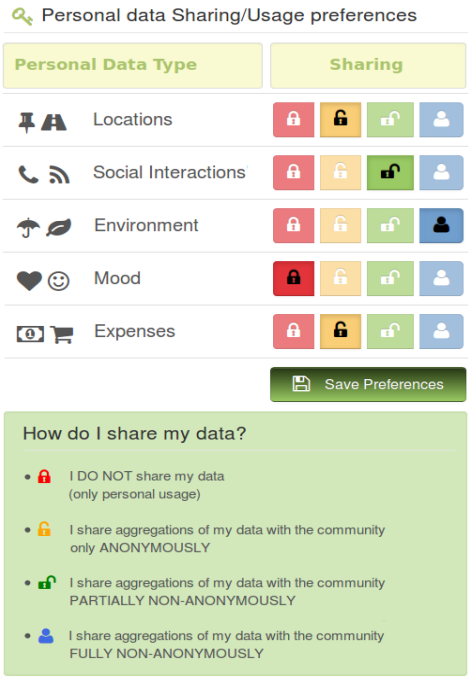

Figure 2. PD sharing preferences.

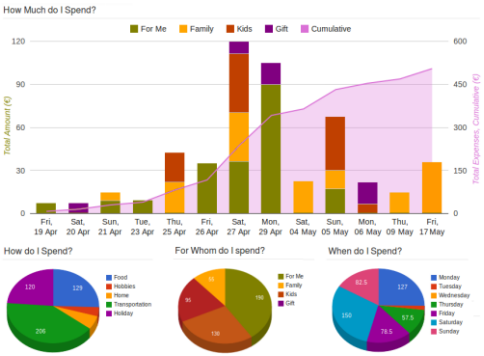

Figure 3. *Aggregated Individual Views*.

and weekly expenses, etc.), (iii) enabling the exploitation and use of PD (e.g., sharing effectively PD with 3rd parties, accessing and using new generation smart city apps, etc.) [4]. In this context, security is a supporting element, being investigated separately, and is out of the scope of this work.

Functionalities. My Data Store provides:

- *Collection*: users can determine which data to automatically collect and store in the PDS.
- *Sharing*: users can choose whether to disclose or not their data and in which detail (e.g. anonymously).
- *Deletion*: users can delete single records or all data records collected in a specific region and time interval.

Moreover, users are provided with visualizations having different levels of aggregation:

1) Detailed Views in tables or maps and *Aggregated Individual Views* (Figure 3) such as charts and timelines increasing user's consciousness on daily behaviors.

2) Social Views (Figure 4) using also data shared by and enabling social comparisons of individual behaviors.

Data Organization. The PD automatically sensed by smartphones consist of: (*i*) call and SMS logs, (*ii*) proximity data obtained by scanning near-by Bluetooth (BT) devices, (*iii*) locations from GPS or Wi-Fi, (*iv*) media activity events (e.g. pictures taken, recorded videos, etc.), and (*v*) devices status information. Additional external sources of PD are integrated in the platform. Participants are provided with an environmental sensor (www.senordrone.com), which periodically senses (vi) the *air quality* and simple

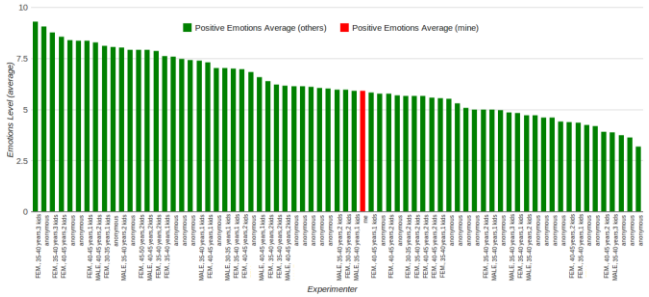

Figure 4. Social Views: users can decide how to share enabling the opportunity to compare with other participants.

meteorological information (humidity, temperature, etc.). Moreover, data on (*vii*) daily mood, stress, quality of sleeping and (*viii*) daily expenses are collected through experience sampling methods by means of ad-hoc mobile-Apps integrated with the collecting platform. Specifically, daily mood (affective) states were collected using the 10-items version of the well-known PANAS scale, addressing positive and negative affective states.

In order to simplify user experience and control, data have been organized in Data Regions (DR) by putting collected data, thus, were grouped in 5 DR (Figure 2): (a) *Locations*, including data from GPS and Wi-Fi, (b) *Social Interactions*, including data about calls, SMS and Bluetooth hits, (c) *Environment*, including data acquired from external sensors (air quality) synchronized with the smartphone, (d) *Mood* and (e) *Expenses.*

Evaluation and Results

My Data Store is provided to the 150 participants of the MTL living lab enabling control and awareness over the huge volume of data collected, consisting so far of, e.g., more than 30M locations, 1M call/SMS logs, 2M Bluetooth hits, 7M running apps records, 20K records of personal expenses, etc. From July to October of 2013,

After the experience with My Data Store	Agree + Strongly Agree
I feel more aware on PD relevance, criticality	**30%**
I feel more aware on PD meaning	**49%**
My perception about PD value has increased	**30%**
I consider transparency in PD manag. more important	**59%**
I consider PD collection/ control more useful	**30/40%**

Table 1. Impact of My Data Store.

	Ranked Critical	**Feel More Aware**
Locations	**1st**	**70%**
Inter. C/S	**4th**	**59%**
Inter. BT	**3rd**	**46%**
Expenses	**2nd**	**41%**
Mood	**5th**	**19%**

Table 2. Gained awareness on personal data types using My Data Store.

we ran a preliminary study with the 63 MTL's participants to measure the impact of My Data Store on people's awareness of PD value and meaning. To this end, we asked users to fill several surveys and we analyzed the usage-logs of My Data Store.

Surveys revealed that the 32% of participants were *very satisfied or satisfied* with their experience with My Data Store, while only the 14% of them were *unsatisfied*. Among the different functionalities, 68% and 51% of the participants found *very relevant* the *Individual Aggregated Views* and the *Collection, respectively*. *Cancellation* was found relevant only for 33%. Overall, visualizations were valued as very relevant and easy to understand from the majority of the participants. In particular, *Individual Views* of communication, daily expenses, daily mood states, and interactions were found relevant by more than the 75% of the involved users. The functionality of comparing own behaviors with the behaviors of other participants was found useful/very useful (37%) or adequately useful (43%) by the majority of people. The most useful Social Views were those comparing individual spending behaviors with other participants (90%). Instead, a significant amount of participants did not consider useful the Social Views related to mobility.

Regarding the effect of using My Data Store, almost 50% of users self-reported to have gained *more awareness* on PD and to consider transparency in PD systems more important (Table 1). In particular, their awareness has been improved especially for location, interactions via calls/SMS and Bluetooth data (Table 2). Moreover, more than 33% stated that their approach to privacy changed after using the tool.

Conclusion

We presented My Data Store: a user-centric tool for PD management empowering users with control and awareness of their PD in a ubiquitous context. Our preliminary results show that users' awareness of their personal data improved by interacting with the PD management tool. We are continuously enriching our platform with new data sources from smartphone or wearable sensors (e.g., wristbands, NFC, etc.) and providing a mobile version of My Data Store.

References

[1] Aharony, N., Pan, W., Ip, C., Khayal, I., and Pentland, A. Social fMRI: Investigating and Shaping Social Mechanisms in the Real World. *Pervasive and Mobile Computing*. 7(6), pp. 643-659, 2011.

[2] Dong, W., Lepri, B., and Pentland, A. Modeling the Co-evolution of Behaviors and Social Relationships using Mobile Phone Data*. In Proc. of ACM MUM*, 2011.

[3] Hong, J., and Landay, J. An Architecture for Privacy Sensitive Ubiquitous Computing*. In Proc. of ACM MobySys*, 2004.

[4] Moiso, C., Antonelli F., and Vescovi M.: How Do I Manage My Personal Data? - A Telco Perspective. *In Proc. of DATA*, 2012.

[5] Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., Hansen, M., and Govindan, R. Personal Data Vaults: A Locus of Control for Personal Data Streams*. In Proc. of ACM CoNext*, 2010.

[6] Pentland, A. Society's Nervous System: Building Effective Government, Energy, and Public Health Systems. *IEEE Computer*, 45(1), pp. 31-38, 2012.

[7] Van Kleek, M., Smith, D.A., Shadbolt, N.R., and Schraefel M.C. A Decentralized Architecture for Consolidating Personal Information Ecosystems: The Web Box. *In PIM*, 2012.