My Data, Your Data, Our Data: Managing Privacy Preferences in Multiple Subjects Personal Data*

Stefania Gnesi[1], Ilaria Matteucci[2], Corrado Moiso[3], Paolo Mori[2], Marinella Petrocchi[2], and Michele Vescovi[4]

[1] ISTI– CNR – Pisa – Italy
stefania.gnesi@isti.cnr.it
[2] IIT– CNR – Pisa – Italy
{ilaria.matteucci,paolo.mori,marinella.petrocchi}@iit.cnr.it
[3] Future Centre– Telecom Italia–Torino – Italy
corrado.moiso@telecomitalia.it
[4] Semantic&Knowledge Innovation Lab– Telecom Italia–Trento –Italy
michele.vescovi@telecomitalia.it

Abstract. The evolution of mobile devices, the success of social networks, and the digitalization of business/personal services have resulted in a huge and continuous production of Personal Data (PD). The creation of a balanced ecosystem of PD, where data act as the fuel for novel application scenarios, may drive the shift toward a user-centric paradigm, in which constraints should be imposed on the data usage, to protect the individuals' privacy. The possibility for people to directly collect, manage and exploit PD introduces both technical and regulatory new issues in PD management. Uncertainty especially arises in the case of PD related to multiple subjects, *e.g.,* containing identifiers referring to more than one person, each of which holds rights to control how these PD are treated. In this paper, we refer to this kind of valuable data as Multiple Subjects Personal Data (MSPD). The protection of MSPD in a user-centric paradigm is an undeniable requirement to ensure privacy to all MSPD right-holders. We discuss the relevance of MSPD, providing a technical approach to regulate their trusted management in a user-centric model context.

Keywords: Multiple Subjects Personal Data, Personal Data Management, Privacy policies management, User-centric Privacy-aware architecture.

1 Introduction

Nowadays data are becoming the fuel of the innovation and an essential resource for the design and development of new, or better, services and products for Society and Business and they are at the basis of all the modern applications, ranging from personal applications and social networks to the future "smart cities" and "smart spaces"

* The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant no 610853 (CoCo-Cloud) and the Registro.it funded project MobiCare.

B. Preneel and D. Ikonomou (Eds.): APF 2014, LNCS 8450, pp. 154–171, 2014.

solutions. They are the ingredient that is driving the evolution of the technology and the spark inspiring novel business. Moreover, data are also a source of information in order to better understand the behaviour of communities and of individuals, by means of data mining and social mining techniques.

The Directive 95/46/EC of the European Parliament and of the Council defines **Personal Data (PD)** as "any information relating to an identified or identifiable natural person (the data subject)" [13, 2]. Even if PD can (directly or indirectly) relate to an individual in several ways, in this paper we focus on PD intended to be pieces of digital information containing a Personal Identifier. In this paper, **Personal Identifiers (PIs)** are considered sequences of digital chars that uniquely identify a (natural) person within a domain[1]. Some examples are: names, phone numbers, e-mail addresses, passport numbers, driver's license number, credit card numbers, etc.

The amount of PD that nowadays is available and generated on a daily basis is rapidly growing due to:

- the increasing number of activities performed online or with a digital representation, due to a wide-spreader adoption of new types of personal devices (*e.g.,* smartphones, tablet), which enable people to access online services in an ubiquitous way and to interact with the real-world service (*e.g.,* payment, ticketing, check-in) in an innovative way (*e.g.,* by means of NFC solutions);
- the pervasiveness of sensors, either in the surrounding environment or integrated in the mobile devices, which enable the collection of contextual information in a transparent way with respect to people.

Gathering and processing PD enable organizations to a deeper understanding of people' needs and behaviour, while individuals can benefit from the creation of novel personalized applications with enhanced user's experience and improve their quality of life. Unfortunately, the current models of managing PD do not fully allow a rights-respecting, controlled, and effective exploitation of such benefits. In fact:

- PD are often spread and fragmented in the data centres of a multitude of organizations that an individual interacts with, either in the real or in the digital world. In this scenario, it is not possible to have a holistic view of individuals, as PD are collected and stored in several independent silos, each of which includes only the data concerning a specific domain.
- Individuals are almost excluded from the lifecycle of the PD which include the identifiers referring them (for the sake of brevity, hereafter we will refer to PD including the personal identifier of a data subject as "her" PD). Usually, they have a (very) limited possibility to manage their PD and exploit them according to their needs and wills, being mostly relegated to the role of producers of PD. This generated a lot of concerns in the users, leading to a loss of trust with respect to the collection and granting of their PD.
- As PD are mainly collected/stored by organizations, the focus of authorities has been more on PD protection, to reduce risks of uncontrolled use, than on the promotion of their full usage when paired with a higher control from data subjects.

[1] This concept, adopted in this paper, is not necessarily related to a legal terminology.

These factors imply a deadlock between the opportunities for exploiting PD in order to enable novel application scenarios and the constraints imposed on their usage to protect the individuals' privacy.

In order to overcome this situation, the shift towards a user-centric model for PD management has been promoted by several initiatives [42, 43], in particular by the WEF, in order to increase trust, to enable a higher control of individuals over the life-cycle of their PD [43], and encourage the creation of a balanced ecosystem of PD.

In the context of user-centric models, one of the undeniable requirement is to guarantee the right control and protection of PD also related to multiple subjects, *i.e.,* in which a subject can be the responsible of the storage of PD which refer to her and to other subjects Hereafter, we refer to this kind of data as **Multiple Subjects Personal Data (MSPD)**. Examples of MSPD are records of phone calls, co-location logs or reports of medical examinations. Hereafter, if one of the PIs of a user is in a MSPD, we refer to "her" MSPD, meaning that this user holds some rights on this MSPD.

Switching back to PDS services, if on one side a PDS should be able to collect and manage MSPD, on the other side such a service should adopt solutions for preventing and avoiding abuses performed by the "PDS-owner" *i.e.,* the PDS service subscriber, possibly damaging other individuals referred in the stored data. In this way, the PDS service is compliant to the Directive 2009/136/EC of the European Parliament and of the Council [14], and with the recommendation of an ENISA study [37].

Contribution. This paper aims at providing a technical solution to protect the privacy of the PDS-owners with respect to the other PDS-owners. The solution is based on privacy policies and it adopts a technical approach to regulate the storage, disclosure, and use of MSPD within a PDS model. In particular, we face the problem of personal data referring to more subjects. In general a *subject* could be a "single natural person", or simply a *person* that we refer in this work also with the terms *user* or *individual*, but it could be also a "legal person", *i.e.,* an organization (private or public) or a person acting in the name of the organization. In the following, we consider the case in which the subjects are natural persons. Furthermore, the protection of the privacy of the PDS-owners with respect to the PDS manager (and, in general, with respect to attackers) is not covered by this paper, and it is left as future work, although is a fundamental issue to be solved in order to design a real PDS system. The paper describes a PDS-based architecture that implements the proposed approach, detailing the interactions with the user and among the components of the architecture.

Structure of the paper. In Section 2, we discuss relevant aspects concerning PD and MSPD in the context of their storage and usage within PDSs. In Section 3, we describe a MSPD privacy-preserving architecture based on privacy policies. Section 4 analyses how the proposed solution could also apply to handle PD released back by organizations. Section 5 recalls related work in the area. Section 6 discusses about some pros and cons of the proposed approach. In particular, it provides an analysis of different privacy issues that may occur when we deal with not only natural persons but also legal persons, and when the PDS service provider is not a trusted party. In Section 7, we conclude with final remarks.

2 PD and MSPD in PDS Context

2.1 Personal Data in a Personal Data Store

A PDS is defined as a secure digital space, owned and controlled by an individual, acting as repository for PD, providing to her a set of services for the collection, management and the exploitation of her PD. PD can be collected from several sources and through different procedures. Some examples are:

- PD voluntarily introduced by the user, *e.g.*, uploading files in a personal cloud storage service (such as Dropbox), changing/filling the attributes of a user profile (*e.g.*, Facebook), content/information uploaded on particular Apps or services.
- PD automatically collected in mobility (from Apps or sensing platforms on personal devices of the user) or during online activities (*e.g.*, search/browsing history).
- PD uploaded (possibly in an automatic way) from organizations' data centers and returned in a digital, reusable format (*e.g.,* connectors to the social networks' APIs), as according to the "right of copy"/ "right of access".

PD are organized in records, grouping all the information related to the same object, action or event. We assume that distinct kinds of records stored in the PDS are predefined and, consequently, the format of each of these records is predefined too. In other words, for each kind of PD, a specific record type is defined. A record type declares its fields, each of which is characterized by type of its values (*e.g.,* a location, time, a sensor measure, etc.). For each record type, some fields store personal identifiers (PI). For instance, the records representing phone calls will include (at least) four fields: the phone number of the caller, the phone number of the callee, the call starting time and its duration. The values of the fields of the caller and the callee numbers are personal identifiers. The fields of the record which are not PIs, instead, become critical when they are stored in the record, because they can be refereed to (or have particular relevance/value for) the other subject(s) whose PIs are stored in the record. Roughly speaking, in the case of a phone call record including both the caller's and the callee's PIs, all the other information included in the record, such as the timestamp and the duration, become critical. They, in fact, may reveal personal information (such as actions, behaviours, etc.) concerning both the speakers.

A PDS provider manages an ecosystem of PDSs. A PDS subscriber (owning a PDS) can decide which PD have to be collected and stored in her PDS, can be passed as input to personal applications, or can be disclosed to other individuals or organizations. A PDS provider operates on behalf of its subscribers and should not perform any action on the stored PD according to autonomous decisions (unless these decisions have been authorised by subscribers). It is worth noticing that an individual can decide to store her PD in multiple PDS: this has the advantage to avoid a single point of failure on her privacy, but has the disadvantage of not having an integrated view of her digital footprint and of both increasing the complexity of data management and the data fragmentation and/or replication. For the scope of this paper, we concentrate hereafter on PDS owners storing their PD in a single PDS.

2.2 Multiple Subjects Personal Data

If a PD record includes PI fields referring to different subjects, this PD record is a **Multiple Subjects Personal Data (MSPD)** record: in this case, more than one subject could have control rights on (some fields of) such a record. It is worth noticing that MSPD are critical by itself whenever associated to other information; for instance, the exact time of an interaction coupled with one individual GPS location could reveal also the other individual location.

A simplified (possibly non-exhaustive) categorization of MSPD involving "natural persons" subjects includes *Interactions* and *Co-location*. In the following we provide definitions and examples of these two categories of MSPD.

Interactions. MSPD that contain parameters that identify two or more mutual individuals interacting, but also implicitly describe their relations, their social network, behaviour, and habits are classified as Interactions MSPD.

Examples of such MSPD are SMS, e-mails or messages exchange on social networks (which may involve simultaneously many actors and include sensitive content such as messages/mails' text). One of the most common is the Call Data Record (CDR), *i.e.,* the log of phone call. A CDR includes data such as: the speakers' (caller and callee)'s phone numbers, the time when the call was made, its duration, its type (received, unanswered, ..), etc. Therefore, a CDR includes PIs, *i.e.,* their phone numbers of (at least) two individuals, the caller (Speaker A) and the callee (Speaker B).

Co-location. We classified as Co-location MSPD those data that not only describe a relation (or, at least, a physical proximity) among two or more individuals, but also "benefit" of the property of being stackable with other personal information increasing the risks correlated to PD abuses. For instance if an individual A is co-located with B and this information is disclosed and combined with the location of B, the location of A is also inferred.

Examples of such category of MSPD are, *e.g.,* the logs of device-to-device interactions via Bluetooth (including the device name or univocal device id) as far as mutual tagging (*e.g.*, "I'm here with…") on social networks such as Facebook or Foursquare.

Let us consider more into detail the case in which an individual A (the PDS owner) wants to store in her PDS all the log records of the device-to-device interactions occurred via Bluetooth between her device and other Bluetooth devices in her physical proximity. In particular, using Bluetooth, A can continuously "scan" the area surrounding her device and monitor the presence of other visible Bluetooth devices (including personal devices like mobile phones or tablets). For every device-to-devices logged interaction, a record containing the two devices' MAC addresses, assigned names, and classes of devices can be stored together with the date and time of the interaction. In this case, the MAC addresses of the two involved devices are PIs, because the MAC address uniquely identifies the device and, thus, it might identify the device owner. Critical information contained in this kind of MSPD are, moreover, the name assigned to the device (which further can tell –but not uniquely identify– the identity of its owner) and the class of the device.

Thanks to the collected records, A can, *e.g.,* ask to some application to build the graph of her "face-to-face" interactions, to reckon her more frequent interactions

(in proximity) or, even more, to keep track of "where I met whom", by combining these information with her precise geographical location (e.g., from GPS sensor or Wi-Fi connections).

2.3 Rights and Permissions on MSPD

As shown in previous examples, in several cases, the PD stored in the PDS of a PDS owner contains PIs related to other subjects and thus they are MSPD. We think that the concept of "ownership" of MSPD should be considered as for the case of PD. As discussed in [42], "the debate over who owns PD has proven to be complex and a key source of tension. It is an emotionally charged debate in which stakeholders have radically different and valid points of view." In line with the Data Protection Directive 95/46, we will refer to "control rights", instead of "data ownership", also for MSPD.

Thus, each of the subjects, whom PIs are into the MSPD, has some rights on defining preferences on how those MSPD are managed, such as: how they are stored, processed, and disclosed.

Uncontrolled usage of MSPD could result in a violation of the privacy of some of the (right-)holders, for example caused by the disclosure of the MSPD with 3rd parties or by allowing applications to process them. A recent remarkable case is the one involving WhatsApp: the Office of the Privacy Commissioner of Canada and the Dutch Data Protection Authority, in a joint report, said the app violated privacy laws because users have to provide access to all phone numbers in their address book, including both users and non-users of the app [36].

To give an example of the kind of preferences that could be expressed for managing MSPD, we consider here the case of the Interactions MSPD Call Data Record (CDR), introduced in Section 2.2. If one speaker would like to store the CDRs related to her calls in her PDS, she needs the permission of the other involved speakers. Suppose, for example, that the subject owner of the PDS is the caller, *i.e.,* Speaker A (the other case is absolutely symmetric, thus equivalent). In this scenario, different cases may occur, such as:

- the callee (*i.e.,* Speaker B) allows the (specific or every) PDS owner to store the CDRs, including her PI, (*i.e.,* Phone Number);
- Speaker B grants the permission to the Speaker A to store such records, but with her phone number encrypted;
- Speaker B does not grant any permission to Speaker A to store the record including her PI.

In the previous example, we are focusing on the privacy between subjects of the PI. Whenever B required storing her PI not in clear, PI could be pseudo-anonymized. We assume that pseudo-anonymization is achieved through an irreversible hash function taking as input the PI referring to B and a key associated to A (*e.g.,* her internal ID). In this way, A can correlate records referring to B, but a 3rd party is not able to correlate records referring to B disclosed by different PDS owners. Even if this does not prevent the possibility that an entity can de-anonymize B by means of an inference attack, it reduces the risk. Moreover, B could also increase the level of protection by denying to A the possibility to disclose MSPD about B, without having previously removed all the B's pseudo-anonymized PI or other fields critical for B.

The PDS owner could then define further rules defining how the CDR should be stored in her PDS (*e.g.,* she could require that only the calls to people in a "white list" should be either stored or excluded).

Moreover, the callee (*i.e.,* Speaker B in the previous example) must be able to control how her MSPD are used. Indeed, the CDRs can be used by the PDS owner for several purposes, such as input to applications which, *e.g.,* determine her social graph, check the interaction level with a given person (possibly in combination with other interaction-related PD, such as SMS or e-mails exchanges), or determine the phone user specific usage profile.

Also in this case the callee can determine, *e.g.,* the level of detail according to which her identifier or other information are disclosed; these rules could be different according to the usage scenario (*e.g.,* processing performed by a "personal application" run by the caller, or the exchange). The rules of the callee can contribute to determine (jointly with the rules defined by the caller) the format of the disclosed CDR, for instance:

- <PhoneNumA, PhoneNumB, *null*, duration,…>: if the Callee does not want to disclose when a call is performed
- <PhoneNumA, PhoneNumB, day, duration,…>: if the Callee wants to reduce the level of resolution for data on time
- <PhoneNumA, *null*, time, duration,…>: if the Callee does not give the permission to disclose PhoneNumB(in this case, the record includes only information which are under the control of the caller)

Other options could be to define rules for disclosing CDR information in aggregated form (*e.g.,* number of calls between caller and callee in a given time interval).

Coming back to Section 2.2 and the example of the Co-location MSPD, we remind that an individual A can scan the area surrounding her device, monitor the presence of other visible Bluetooth devices, and store every device-to-devices logged interaction. However, in terms of rights and permissions, another individual B could choose to deny the storage in a PDS of the proximity interactions of his Bluetooth device, in order to avoid of being unconsciously "scanned" (and thus being co-located to other devices) when he activates and sets to visible his device. This should be also the default policy defined by the PDS provider. In a different case, the individual B could grant to the other users the right of storing and using (for personal applications) these data or, even more, to disclose them to third parties.

The PDS should implement mechanisms to enforce controls on the managed MSPD, according to the preferences defined by those subscribers of the PDS service which have rights on those MSPD, but also in protection of the privacy of all the other data subjects which have rights on MSPD too, and not necessarily are subscribing the service. Relying on such mechanisms, the subscribers of a PDS service (*i.e.,* the individuals that own a PDS) will be able to control which of their PIs, or, in general any of the data on which they have some right, can be stored in other PDS, used by other PDS owners, and disclosed to 3rd parties. Such kinds of control imply the capability for the subscribers to define preferences on how their PIs can be stored in MSPD, and how the MSPD on which they have some right can be stored, given as input to applications, or disclosed to 3rd parties. We propose here to express these preferences with privacy policies, and each time a PDS owner requests to perform an action concerning

a record and his data space, *e.g.,* store, give as input, or disclose, the set of people who have some rights on this record must be determined to enforce the proper privacy policies to decide whether the action on the record can be executed or not.

The PIs stored in the record (*e.g.,* the phone numbers and MAC addresses in the previous examples) are exploited to determine the ID of the referred person through the list of the PIs managed by the PDS manager and, consequently, to determine the privacy policy to be enforced. We assume that a default privacy policy is paired to individuals that are not subscribers of the PDS service, such that their PIs cannot be disclosed to the PDS service subscribers.

3 Architecture

We propose a framework for the privacy-preserving management of MSPD in a PDS context. The main goal is to define and enforce the privacy policies that regulate the storage, usage, and disclosure of MSPD within a PDS-based infrastructure.

Before introducing the privacy-preserving policy-based architecture, we present the actors involved in our framework. It is worth noticing that, hereafter, we concentrate on the storage of MSPD. A simple extension of the framework applies to manage the control on the usage and the disclosure of the MSPD already stored in a PDS.

The main actors of our framework are:

- *The PDS owner* that subscribed the PDS service and wants to store some PD records in her PDS. These records could be MSPD.
- *The MSPD right-holders*, *i.e.,* the individuals that have rights on the record the PDS owner wants to store in her PDS. The MSPD's right-holders are all the entities that are referred by some of the identifiers included in the record.
- *The PDS manager*, that is the entity that provides to individuals a PDS service; it is also contributes for the definition and the enforcement of the policies enabling people to control the collection, usage and disclosure of their MSPD. Moreover, it manages the registry for associating PIs to PDS subscribers.

When a new user subscribes the PDS, she can set her own preferences that regulate the storage of her data in the PDSs of other users (and the subsequent usage and disclosure to 3rd parties). However, since the subscriber does not know, at registration time, the exact set of subjects to whom she wants to disclose her data, she can set (or modify) her general and/or specific (*i.e.,* referred to one –group of– subject) privacy preferences at any time. The new preferences should be enforced both on the new MSPD that will be stored in her PDS or in the PDS of other users from that moment on, and on the existing MSPD, *i.e.,* the ones that have been already stored in PDSs. In the reference scenario, we protect the privacy of the PDS-owner w.r.t. the other PDS-owners only. Since all the accesses to the PDS are mediated by the authorization system (including the accesses of the PDS-owner to read the already stored MSPD) the right to execute an action on a MSPD is determined according to the current privacy preferences of all the subjects having some rights on this MSPD. Hence, in the case where the privacy preferences of one of the involved data subjects changed, the updated preferences are always used to determine the access right.

In order to identify its subscribers in the context of a PDS service, the PDS provider must assign to each of them a unique ID. An internal Id, or an hash of the user's phone number, or of another unique identification code (*e.g.,* the SSN in US, or the Fiscal Code in Italy and the National Insurance Number issued in the United Kingdom) could be chosen as unique ID.

Moreover, in order to easily determine the subjects referred by each MSPD, the PDS provider exploits a User Registry that manages a list of PIs of its subscribers. The PDS User Registry pairs each user ID with all the PIs of the user (and vice versa).

We also assume that "unknown" MSPD right-holders (*i.e.,* identified by PIs not included in the PDS manager registry) have associated the most restrictive policy, *i.e.,* that completely denies the disclosure of their PIs in the subscribers' PDS.

We focus on MSPD phone call data records (CDR). The caller A wants to store the CDR in her PDS. Both the caller and the callee B have rights on part of the CDR.

Let us suppose that B is registered to the PDS: she has not set any privacy preference at registration time and A calls her. A wants to save the CDR concerning this call in her PDS, but this record includes a PI of B (*i.e.,* her phone number). The PDS manager retrieves from the registry the unique ID of B from her phone number, evaluates B's privacy policy and finds out that it does not allow A to store B's data since no policy has been explicitly set to authorize that storage. Hence, A receives a notification that the CDR was stored with partial information (the CDR is stored without B's phone number). In the case where the record includes other data fields on which B holds some rights, these will not be stored as well. In the case where B is not registered to the PDS, instead, a default policy states that her PIs cannot be saved in the PDS of A. It is worth noticing that A is able to use the PDS. Indeed, A can store the fields in the MSDP that are only under her control. Hence, as soon as a bunch of people interacting one another starts to use the PDS, the amount of data stored in each PD starts to grow up. Finally, it is worth noticing that notifications are not stored in PDS.

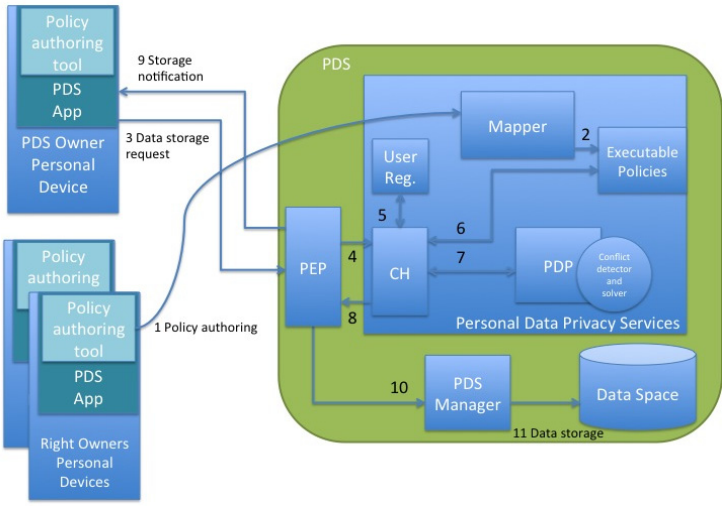

Fig. 1. Privacy-preserving policy-based architecture

3.1 Description of the Architectural Components

Figure 1 shows the policy-based architecture that we propose. The components of the architecture are the following:

1. The PDS App runs on user's device or machine and provides the user an access to the functions of her PDS. The PDS App allows users to control which types of PD are collected and stored in their PDS, to search, retrieve, and visualize the collected data, to delete some of them, and to control which data disclose to applications and the level of disclosure with other users or 3rd party organizations.
2. The Policy Authoring Tool (PAT) runs on user's device as well, allowing the user to edit her privacy preferences, on if and how MSPD referring to her can be accessed in the PDS of the other referred data subjects. The tool we consider is tailored for users not familiar with technical policy write up. To the best of our knowledge, there is a few works on privacy policies authoring tool tailored for non expert users. Some of them, *e.g.,* [19,22,7,35,8], study different aspects on the capability of common users to use such tools. Here, we consider the authoring tool we proposed in [11]. It has been designed and implemented in a customized way, in order to provide different levels of granularity when specifying the preferences. The graphical interface provides i) an easy and quick way for a common user to set privacy preferences on her MSPD in a few click; ii) the capability to set privacy preferences using a device of common use, such as a smartphone or a tablet; iii) an advanced mechanism to compose fine-grained privacy preferences for users that want to set up which MSPD category could be disclosed into which PDS, or could be used by which PDS owner applications, or could be disclosed to which 3rd parties. For example, users would simply like to set their preferences in few clicks, just giving a broad consent to sharing MSPD about them in other PDS, whatever the nature of such MSPD is. Instead, other kind of users would like to set, *e.g.,* the category of MSPD whose disclosure is allowed, or the period of time over which their preferences should be considered applicable.
3. The PDS Manager is the service that manages the functions of the PDS on PD (such as, storage). This paper does not describe this service in detail; for specific insights, the interested reader is referred to [31].
4. The Data Space, the database where the PD collected by the PDS owner, including the MSPD, are securely stored in a structured way.
5. The Policy Enforcement Point (PEP) is the component embedded in the PDS that intercepts all the requests concerning the storage or the accesses to the MSPD. The PEP invokes the Personal Data Privacy Service to perform the decision process, and enforces the decision. The PEP must be tamper proof and non by-passable, *i.e.,* all the attempt to access the PDS are intercepted by the PEP and forwarded to the Personal Data Privacy Service.
6. The Personal Data Privacy Services in the PDS includes several components:
   - The Mapper from Preferences to executable policies. The privacy preferences are mapped to an executable format, such as, *e.g.,* XACML, the well-known policy language constituting the "de facto" standard for defining access control rules [33]. The Mapper enables automatic translation between users preferences, edited in natural language, and the executable policies.

- The User Registry (UR) that pairs the unique ID of each user with all the PIs related to him (phone numbers, email addresses, and so on).
- The Context Handler (CH) receives the storage (or access) request from the PEP, retrieves the policies concerning the subjects that have some rights on the record that is being stored, along with other subjects' attributes that could be evaluated by the PDP and asks the PDP to perform the decision process evaluating the access request with the selected policies.
- The Policy Decision Point (PDP) performs the decision process by evaluating a set of privacy policies to decide whether the storage (or access) request should be granted or not. The PDP response could also include some obligation that must be performed by the PEP as the result of the decision process *e.g.,* request to pseudo-anonymize, reduce the precision scale or aggregate data before storing or using it.
- The Conflict Detector and Solver (CDS) component determines when two (or more) policies applicable to the same storage request returns conflicting results (*i.e.,* one policy allows the storage of a record while the other denies it) and decides the final result, *i.e.,* the decision that will be enforced by the PEP. The XACML authorization framework comes with a native conflict detector. For solving the detected conflicts, the XACML comes with a set of native *combining algorithms* that define the strategy for solving conflict. Usually, it adopts standard rules, such as, Deny-Override, Permit-Overrides, First-Applicable, and Only-One-Applicable. Other approaches have been proposed in the literature, see, *e.g.,* [26,18,1,16,27,38,29,15]. In particular, here we propose the approach we have designed and implemented in [26,18]. It is based on a multi-criteria decision process that allows to prioritize the conflicting policies by considering the degree of specificity of the elements constituting each conflicting policy. The elements of a policy are the subject, the object, the action, and the environmental conditions. Each of the elements can be characterized by several attributes. As an example, the attributes "category" and "Identification Number" can characterized the element "object" (in our context, the MSPD). To solve a conflict, we evaluate the specificity of the policy attributes. As an example, in [26,18], the category of the MSPD has been considered to be less specific than the Identification Number of the MSPD. Thus, the strategy that ranks the conflicting policies privileges that policy having specified the MSPD Identification Number rather than the one having specified the MSPD category.

Components 1 and 2 are deployed on users' devices, while the other components are generally deployed on servers in the network or "in the cloud".

3.2 Description of Logical Workflow

We concentrate on the logic workflow for storage operations. Similar workflow could be applied for authorizing other operations, such as the MSPD elaboration by user applications or the disclosure of MSPD with 3rd parties. It is important to remark that the storage operation is as representative as elaboration and disclosure to 3rd parties..

Overall, we assume that the MSPD right-holders have composed their privacy preferences through the authoring tools installed in the PDS App on their mobile devices (as indicated by arrow 1 in Fig. 1). The privacy preferences are automatically

translated to executable policies by the mapper (arrow 2, Fig. 1). Then, the logic workflow concerning the storage of a new record in a PDS is described hereafter.

a. The PDS owner requests to store a new record in his PDS. Let us suppose that this record represents a phone call performed by the PDS owner. The PDS owner sends the storage request to the PDS (arrow 3, Fig. 1) toward his PDS App instance.
b. The PEP installed in the PDS intercepts the incoming request, and it creates a storage request message that is sent to the CH (4). This message includes both the data extracted from the incoming request and other data that are collected by the PEP because they are required to perform the decision process.
c. The CH retrieves the relevant privacy policies from the policy repository. In particular, it exploits the User Registry to identify the IDs of the MSPD right-holders from the PIs stored in the record (5). Policies are supposed to be indexed in the repository by the unique ID of the MSPD right-holders. Recalling the CDR running example, suppose that the caller would like to store the record of the call. He is not the only MSPD right-holder, since the callee identifier is owned by the callee herself. Thus, the CH retrieves from the repository both the policies defined by the caller and by the callee (6). Overall, we assume that a default policy exists that allows PDS owners to store data they have rights on in the PDS they own. However, the PDS owners can change this default policy whenever desired.
d. The PDP evaluates the privacy policies selected by the CH to decide whether the storage request can be executed or not. If more than one policy is applicable to the storage request, the Conflict Detector finds out whether the related results are conflicting, and the Conflict Solver is invoked to determine the final response. In our example, a conflict would exist, *e.g.,* if the caller authorizes himself to store the CDR, while the callee denies that storage. The conflict solver will take its final decision based on, *e.g.,* a strategy considering the level of specificity of each conflicting policy, as modeled in [30].
e. The PDP sends the final response to the CH (7), which forwards it to the PEP (8).
f. The PEP enforces a positive (respectively, negative) response by performing the storage request, (respectively, by skipping the request and sending back a notification to the PDS owner) (33,26,18).

4 Related Work

The existing literature refers to “multi owner” data whenever data can be exchanged among several entities that can perform some kind of action on them, with a particular eye to cloud storage. In [21], the authors deal with the sharing of data by considering the untrusted relation between the user and the data center provider. Due to this fact, they mostly rely on the CP-ABE schema. In [25], the authors propose a secure multi-owner data-sharing scheme able to efficiently support dynamic groups and guarantee privacy and anonymity to users. The approach is focused on a cryptographic model for sharing data of a data set among users belonging to a group.

Both these works do not refer directly to PD and PDS but treat the challenge of managing the access to data shared among different entities. Furthermore they propose solutions specific for some aspects related to the sharing of data. Here, instead,

we presented an architecture that aims at being general enough to sustain any specific implementation of the single components, where the PDS is one of the main ones.

For what concerns Personal Data Stores, several platforms and solutions are already implementing PDS-based services. Most of them provide features for enabling the "owner" of a PDS account to control how the stored data can be disclosed or shared with 3rd parties. Here below we review the more relevant ones. None of these PDS platforms, however, considers the MSPD: the owner of the PDS is the only who has the right to control which PD are stored, in which way they are used or disclosed. They do not implement mechanisms to prevent and avoid possible abuses performed by the "PDS-owner" person possibly damaging the other individuals and organizations referred in the stored data.

The open source project *Danube* (*http://projectdanube.org/*) adopts XRI and XDI technology for controlling data access: relationships with individuals and organizations are defined by using XDI (XRI Data Interchange), through which a user can define rules for sharing, linking, importing and synchronizing data.

Higgins (*http://eclipse.org/higgins/*), another PDS-related open source project, gives user control over the information stored in her account, by allowing her to share selected subsets of it with 3rd parties. Relationships with external parties are established as bi-directional data flows to share/synchronize a set of attributes.

The PDS open source platform *ZXID* (*www.zxid.org*), developed inside the IST TAS3 Project (*http://www.tas3.eu*), extends the specification defined by Liberty Alliance related to the access of users' data attributes in the context of an Identity Management framework. It introduces a policy management architecture to make authorization decisions regarding data accesses according to the users' defined policies. The policy enforcement function is enhanced with a notification mechanism used to inform a user about the accesses to the data stored in her PDS.

The *OpenPDS* developed at MIT, instead, provides mechanisms to protect users' privacy by providing a query-based interface so as to enable only the sharing of anonymous and aggregated data (according to users' choices), and not of raw data [12].

A very preliminary approach to MSPD is implemented in the current setting of the PDS developed by Telecom Italia in the context of the *Mobile Territorial Lab (MTL) project* (*www.mobileterritoriallab.eu*), and exploited in its experimentations. The MTL's PDS implements features empowering people with full control over the lifecycle of their PD, from the data collection to the deletion of single/bunches of PD. In particular, a user of this PDS can choose whether to disclose or not the data of a specific type with other users or 3rd parties, and with which level of detail (*e.g.,* in an anonymous or "nominal" way) [40]. In order to avoid the privacy issues arising from MSPD, the *MTL*'s PDS does not include in its records information directly referring to other individuals different from the specific PDS owner.

Some companies, moreover, are starting providing commercial PDS-like services. For example *Personal* (*www.personal.com*) offers a "vault", where a person can store the "details of her life". Data are stored encrypted through a key under the control of the user, therefore, they cannot be accessed by the provider. *Personal* provides features to control the sharing of the stored information, and to improve the user experience in filling web forms, through data stored in user's vault. Analogously the platform developed by *Mydex* (*mydex.org*) implements features to enable users to control which data can be disclosed to another person or accessed by an application.

As previously mentioned in the introduction, user-centric identity management approaches exists [24]. These solutions aim at placing administration and control of identity information directly in the hands of individuals. In this way people have the control on the (certified) attributes to disclose to a provider when they are accessing a service, so as to fulfill the data minimization requirements. Examples of solutions are those based on attribute-based credential technologies [9]. Even if these solutions share with PDS-based approaches the same objective of give more control in the hands of individuals, the addressed scenarios differ: in fact, these solutions aim at performing secure transactions in the digital world, where strong authentication and according authorization based on certified attributes of the requester is paramount for protecting critical information and infrastructures online. Moreover, users' identity is mainly abstracted as a set of (certified) attributes to be passed in a privacy-preserving way to the service providers. Instead PDS-based solutions aims at offering to individuals an environment for the controlled collection, management, exploitation, and disclosure of the PD produced by them or about them.

5 Discussion

As discussed in [23], "there are many requirements for achieving the privacy needs as expressed as law. Currently there is no commonly accepted technical approach for meeting these privacy requirements".

For example, international regulations, such as the European Directive 95/46/EC [13] and its recent reform, give a definition of personal data and attempt to clarify how their privacy should be addressed. However, at the lower level of the single countries, both definitions and methodologies enabling a privacy-aware data management are often in contrast one with each other. To cite a singular example, ``pseudo-anonymity" is a different concept from the 95/46/EC Recital 26 and the UK/IE recommendations points of view. As an attempt to solve contradictions at various country level, the Article 29 Working Party has produced a set of Opinions and Recommendations concerning data protection, with an effort to shed light on how to deploy and implement effective solutions compliant with regulatory normative. In particular, Opinion 04/2007 [2] clarifies the definition of personal data, while Opinions 01/2012 [3] and 08/2012 [5] provides guidelines on their protection.

In this section, we briefly discuss some open issues deserving more investigation, in order to fill the gap between technological solutions and regulatory directives, and achieve a common vision for preserving privacy of shared personal data. This paper has focused on a user-centric model based on Personal Data Stores (PDS) platforms [12], enriched with a privacy policy-based architecture, in part covering some legal issues. However, there is still several questions worth to be addressed. We list them hereafter, and we leave a deeper investigation for future work.

In line with the Directive 2009/136/EC of the European Parliament and of the Council (stating that personal data must be protected against unauthorized management because personal data breach could have very dangerous consequence for data subjects, such as the identity theft [14]), we focus on controlled storage, use and disclosure of PD and MSPD. One possible way to foster the user-centric paradigm is to enable individuals to have a copy of PD (and MSPD). This is claimed to be sufficient

to “create a liquid, dynamic new asset class” [34]. Individuals also achieve the opportunity to combine the data with information from other sources and to set permissions about how others can use data [42]. However, to have a copy of their PD is not enough to create value for people, if not combined with quality services for their collection, control (*e.g.,* on disclosure) and exploitation (*e.g.,* through an ecosystem of applications). A PDS platform provides a person with a data space, where she can collect her PD and access a set of services enabling her to manage and use her PD according to her wills and needs. In some cases, PDSs are built on top of innovative Identity Management platforms [24, 17] and their model supports the guidelines on the minimization of asserted/certified attributes necessary to access digital services [20], enhancing them with new application and business scenarios [32]. Actually, the deployment of PDS-based approach would enable new business opportunities with several advantages to all the actors involved in a PD ecosystem [32].

In the organization-centric model, organizations collect and process the data related to their customers/users according to the terms and conditions agreed with them. There are laws and recommendations that determine “guidelines” on the definition of these conditions and on how the users should express their consent on their application (*e.g.,* the rules on the informed consent) [41]. Unfortunately, there are not clear rules for a user-centric model. In fact, even if the PD Regulation should not apply to exclusively personal or domestic processing of PD (related to other data subjects) by a natural person, the exemption does not apply to actors which provide the means for processing PD for such personal or domestic activities. Moreover, analyses on the impacts of cloud-based services on PD treatment mainly addresses cloud services offered to enterprises and not to individuals [4]. Therefore, a regulation on PD in the context of personal cloud services, such as the PDS-based ones, it seems still missing.

In this paper we tackle with the privacy protection of the PDS owner PD against other PDS owners. We propose a solution based on privacy policies, whose management infrastructure is provided by the PDS manager. However, enabling inexperienced users with even an appropriate technology could not be sufficient. Indeed, especially in non trivial user-centric solutions, the probable low level of users’ expertise may prevent individuals to manage (*e.g.*, edit and analyze) complex privacy policies to define fine-grained access rights or to frequently update these policies to fit new needs. Also, a noticeable study in [28] shows “privacy policies are hard to read, read infrequently, and do not support rational decision making”. This makes worth to better investigate the comprehensibility of the kind of policies individuals are willing to accept. Also, an interesting study in [6] reveals “how technologies that make individuals feel more in control over the release of personal information may have the unintended consequence of eliciting greater disclosure of sensitive information”. This paves the way for further investigation towards benefits and drawbacks of the adoption of privacy-enhancing technologies to protect PD and MSPD.

Finally, we are aware that other critical aspects to be dealt with are: how to protect PD from 1) the PDS manager itself, maliciously acting, *e.g.,* to sell PD of their PDS customers to third organizations; 2) the so called “malicious insider” attacker, *e.g.,* an employee at the PDS manager provider that could access PD of the PDS customers for activity of doubtful legality, and 3) a totally external attacker, able to break the security measures of the PDS manager and accessing in such a way to PD of PDS customers. In the literature some partial solutions able to guarantee privacy properties

between the PDS owner and PDS service providers exist. They are mostly based on cryptographic protocols such as blind signatures [10] and 0-knowledge protocols [39]. However, they do not exhaustively accomplish with all the issues we have listed above leaving space for further investigations.

6 Conclusion and Future Work

This paper describes a technical approach to regulate the storage of MSPD within a user-centric PD management model. Even if we concentrate on the logic workflow required for storage operations, similar workflows can be easily derived for authorizing other PDS operations, such as the MSPD elaboration by personal applications or the disclosure of MSPD with 3rd parties (either other people or organizations).

An area for future work is to extend the solution in order to deal with multiple PDS managers. This is a fundamental requirement in order to enable a person to freely choose her preferred provider. We are considering several options on how the functions in the proposed architecture can be invoked in a multi-PDS context. Moreover we are investigating on how to transform the interfaces internal to the proposed architecture into open protocols, which, in the future, could be object of a standardization process. The integration of the components of our architecture is an ongoing work.

References

1. Al-Shaer, E.S., Hamed, H.H.: Firewall policy advisor for anomaly discovery and rule editing. In: IFIP/IEEE Integrated Network Management, pp. 17–30 (2003)
2. ARTICLE 29 DATA PROTECTION WP136, Opinion 04/2007 on the concept of Personal Data, `http://goo.gl/8hO9m` (last checked February 21, 2014)
3. ARTICLE 29 WP191, Opinion 01/2012 on data protection reform proposals (2012), `http://goo.gl/9tMKa` (last checked February 21, 2014)
4. ARTICLE 29 WP196, Opinion 05/2012 on Cloud Computing (2012), `http://goo.gl/tvKNG` (last checked February 21, 2014)
5. ARTICLE 29 WP199, Opinion 08/2012 providing further input on the data protection reform discussion (2012), `http://goo.gl/1AJXB` (last checked February 21, 2014)
6. Brandimarte, L., Acquisti, A., Loewenstein, G., Babcock, L.: Privacy concerns and information disclosure: An illusion of control hypothesis. In: CIST (2010)
7. Brodie, C., et al.: An Empirical Study of Natural Language Parsingof Privacy Policy Rules using the SPARCLE Policy Workbench. In: SOUPS. ACM (2006)
8. Brodie, C., et al.: The Coalition Policy Management Portal for PolicyAuthoring, Verification, and Deployment. In: POLICY, pp. 247–249 (2008)
9. Camenisch, J., Dubovitskaya, M., Lehmann, A., Neven, G., Paquin, C., Preiss, F.-S.: Concepts and Languages for Privacy-Preserving Attribute-Based Authentication. In: Fischer-Hübner, S., de Leeuw, E., Mitchell, C. (eds.) IDMAN 2013. IFIP AICT, vol. 396, pp. 34–52. Springer, Heidelberg (2013)
10. Chaum, D.: Blind signatures for untraceable payments. Advances in Cryptology Proceedings of Crypto 82(3), 199–203 (1983)
11. Conti, R., Matteucci, I., Mori, P., Petrocchi M.: An Expertise-driven Authoring Tool of Privacy Policies for e-Health. Technical Report IIT TR-02/2014

12. de Montjoye, Y.A., Wang, S.S., Pentland, A.: On the trusted use of large-scale personal data. IEEE Data Eng. Bull. 35(4), 4, 5–8
13. Directive 95/46/EC of the European Parliament and of Council, Official Journal of the European Union, L281/31 (November 23, 1995)
14. Directive 2009/136/EC of the European Parliament and of the Council. Official Journal of the European Union, L337/11 (November 25, 2009)
15. Dunlop, N., et al.: Methods for conflict resolution in policy-based management systems. In: IEEE Enterprise Distributed Object Computing, pp. 98–109 (2003)
16. Hall-May, M., Kelly, T.: Towards conflict detection and resolution of safety policies. In: Intl. System Safety Conf. (2006)
17. Hardjono, T., Greenwood, D., Pentland, A.: Towards a Trustworthy Digital Infrastructure for Core Identities and Personal Data Stores. Global Forum on Identity (2013)
18. Jin, J., Ahn, G.-J., Hu, H., Covington, M.J., Zhang, X.: Patient-centric authorization framework for electronic healthcare services. Computers & Security 30(2-3), 116–127
19. Johnson, M., et al.: Optimizing a policy authoring framework for security and privacy policies. In: SOUPS, pp. 8:1–8:9. ACM (2010)
20. Jøsang, A., Pope, S.: User centric identity management. In: AusCERT Asia Pacific Information Technology Security Conference (2005)
21. Kan, Y., Jia, X., Ren, K.: DAC-MACS: Effective Data Access Control for Multi-Authority Cloud Storage Systems. IACR Cryptology ePrint Archive, 419 (2012)
22. Karat, J., Karat, C.-M., Brodie, C., Feng, J.: Designing Natural Language and Structured Entry Methods for Privacy Policy Authoring. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 671–684. Springer, Heidelberg (2005)
23. Korba, L., Kenny, S.: Towards Meeting the Privacy Challenge: Adapting DRM. In: Feigenbaum, J. (ed.) DRM 2002. LNCS, vol. 2696, pp. 118–136. Springer, Heidelberg (2003)
24. Leenes, R., Schallaböck, J., Hansen, M.: PRIME White Paper, Version 3. PRIME Project (2008)
25. Liu, X., Zhang, Y., Wang, B., Yan, J.: Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud. IEEE Trans. Parallel Distrib. Syst. 24(6), 1182–1191
26. Lunardelli, A., Matteucci, I., Mori, P., Petrocchi, M.: A Prototype for Solving Conflicts in XACML-based e-Health Policies. In: Proc. 26th IEEE International Symposium on Computer-Based Medical Systems, pp. 449–452 (2013)
27. Lupu, E.C., Sloman, M.: Conflicts in policy-based distributed systems management. IEEE Trans. Softw. Eng. 25(6), 852–869 (1999)
28. McDonald, A., Cranor, L.: The cost of reading privacy policies. ISJLP 4, 543 (2008)
29. Masoumzadeh, A., Amini, M., Jalili, R.: Conflict detection and resolution in context-aware authorization. In: IEEE SNDS, pp. 505–511 (2007)
30. Matteucci, I., Mori, P., Petrocchi, M.: Prioritized Execution of Privacy Policies. In: Di Pietro, R., Herranz, J., Damiani, E., State, R. (eds.) DPM 2012 and SETOP 2012. LNCS, vol. 7731, pp. 133–145. Springer, Heidelberg (2013)
31. Moiso, C., Antonelli, F., Vescovi, M.: How do I manage my Personal Data? – A Telco-perspective. In: Proc. Data 2012, pp. 123–128 (2012)
32. Moiso, C., Minerva, R.: Towards a User-Centric Personal Data Ecosystem – The Role of the Bank of Individuals' Data. In: Intelligence in Next Generation Networks (2012)
33. OASIS, eXtensible Access Control Markup Language (XACML) Ver. 3.0 (January 2013)
34. Pentland, A.: Society's Nervous System: Building Effective Government, Energy, and Public Health Systems. IEEE Computer 45(1), 31–38

35. Reeder, R.W., Karat, C.-M., Karat, J., Brodie, C.: Usability challenges in security and privacy policy-authoring interfaces. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 141–155. Springer, Heidelberg (2007)
36. Reuters.com, WhatsApp violates privacy laws over phone numbers: report, http://goo.gl/9tJzF (last checked February 21, 2014)
37. Roussopoulos, M., et al.: Technology-induced challenges in Privacy & Data Protection in Europe. A report by the ENISA Ad Hoc Working Group on Privacy & Technology (2008)
38. Syukur, E.: Methods for policy conflict detection and resolution in pervasive computing environments. In: Policy Management for Web (WWW 2005), pp. 10–14. ACM (2005)
39. Uriel, F., et al.: Zero-knowledge proofs of identity. Journal of Cryptology 1(2), 77–94 (1988)
40. Vescovi, M., Moiso, C., Antonelli, F., Lepri, B., Clippinger, J.-H.: Toward Personal Big Data passing through User Transparency, Control and Awareness: A Living-Lab experience. In: Proc. European Data Forum (to appear, 2014)
41. Whitley, E.: Towards effective, consent based control of Personal Data. In: Hildebrandt, M., O'Hare, K., Waidner, M. (eds.) The Value of Personal Data, pp. 165–176 (2013)
42. World Economic Forum, Rethinking Personal Data: Strengthening Trust (2012), http://www.weforum.org/reports/rethinking-personal-data-strengthening-trust
43. World Economic Forum, Unlocking the Value of Personal Data: From Collection to Usage (2013), http://www.weforum.org/reports/unlocking-value-personal-data-collection-usage