HAL open science

Privacy for Peer Profiling in Collective Adaptive Systems

Mark Hartswood, Marina Jirotka, Ronald Chenu-Abente, Alethia Hume, Fausto Giunchiglia, Leonardo A. Martucci, Simone Fischer-Hübner

To cite this version:

Mark Hartswood, Marina Jirotka, Ronald Chenu-Abente, Alethia Hume, Fausto Giunchiglia, et al.. Privacy for Peer Profiling in Collective Adaptive Systems. Jan Camenisch; Simone Fischer-Hübner; Marit Hansen. Privacy and Identity Management for the Future Internet in the Age of Globalisation: 9th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6/SIG 9.2.2, International Summer School, Patras, Greece, September 7–12, 2014, AICT-457, Springer, pp.237-252, 2015, IFIP Advances in Information and Communication Technology (TUTORIAL), 978-3-319-18620-7. 10.1007/978-3-319-18621-4_16 . hal-01431584

HAL Id: hal-01431584

https://inria.hal.science/hal-01431584

Submitted on 11 Jan 2017

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

Privacy for Peer Profiling in Collective Adaptive Systems

Mark Hartswood[1], Marina Jirotka[1], Ronald Chenu-Abente[2], Alethia Hume[2], Fausto Giunchiglia[2], Leonardo A. Martucci[3], and Simone Fischer-Hübner[3]*

[1] Oxford University, UK
[firstname.lastname]@cs.ox.ac.uk
[2] Trento University, Italy
[lastname]@disi.unitn.it
[3] Karlstad University, Sweden
[firstname.lastname]@kau.se

Abstract. In this paper, we introduce a privacy-enhanced Peer Manager, which is a fundamental building block for the implementation of a privacy-preserving collective adaptive systems computing platform. The Peer Manager is a user-centered identity management platform that keeps information owned by a user private and is built upon an attribute-based privacy policy. Furthermore, this paper explores the ethical, privacy and social values aspects of collective adaptive systems and their extensive capacity to transform lives. We discuss the privacy, social and ethical issues around profiles and present their legal privacy requirements from the European legislation perspective.

1 Introduction

Smart Society, an EU FP7 FET integrated project, explores how Collective Adaptive System (CAS) comprised of people and machines may help solve problems of urban living. Smart Society aims to capture how contemporary techno-social trends can be harnessed towards solving challenges facing modern society. The Smart alludes to how innovative, social, mobile and sensor based technologies can support powerful collectives of people and machines that are capable of utilizing constrained shared resources, such as, such as transport networks, in more effective and therefore sustainable ways.

Smart Society is partly inspired by the idea of the 'Smart City', a multifaceted concept that recognizes the benefits of urban living but also the strains that are developing on existing infrastructures and resources due to urban growth. According to this vision, cities made 'smart' will be more productive, more sustainable, and pleasanter places to live. One aspect of Smart Cities concerns augmenting service infrastructures, such as transport, energy and health, with sensor-based digital technologies that are able to visualize patterns of service

* This research was funded by SmartSociety, a project of the Seventh Framework Programme for Research of the European Community under grant agreement no. 600854.

delivery and use stretching across space and time and with a high degree of fidelity. The idea is that service operators can utilize this information to make efficiency savings by tailoring provision to match demand, and by shaping demand through use of incentives or other motivating feedback mechanisms. At the same time, users of shared resources can use those resources more effectively if they are aware of the global state of the resource and able to coordinate between themselves about how the resource might be used.

An objective of Smart Society is to identify problems related to CAS and design solutions to preempt them. In the case of protection of personal data, Smart Society identified a set of potential problems in CAS, including: user profiling, 'big data' analysis of personal data acquired from mobile devices, and the use of analytic technologies that reveal user actions, such as activity recognition–all of which resonate strongly contemporary ICT-based privacy challenges. We aim to explore how far state-of-the-art privacy enhancing technologies, such as privacy policy languages, anonymous credentials, and data anonymization techniques, may find an application within Smart Society, focusing particularly on practical implementation issues and how their use trades-off against other important social values, such as accountability and security.

1.1 Smart Society and Collectives

A key principle in Smart Society is founded on the idea of 'collectives'–a collection of humans and / or machines that identify themselves as a group. In Smart Society, collectives are seen as a source of expertise and discoverable via peer profiles. At the same time they are consumers of resources whose patterns of consumption can be shaped by appropriate interventions such as incentives. Diversity within collectives provides a resource pool to enable the development of a range of responses to a situation, but can also be a source of friction and contention. All together, the socio-technical entity powering the Smart Society vision is referred to as a Hybrid and Diversity Aware Collective Adaptive System (HDA-CAS).

Since collective adaptive systems comprise people and machines that seamlessly collaborate to solve problems and execute tasks, a computing platform that supports a CAS has as input two sets of data about:

1. problems and tasks,
2. peers, i.e., people and/or machines.

The objective of a CAS computing platform is to match these two sets in a way that the specific problems are re-directed to the group of peers that are most capable, interested and / or efficient in solving them. Thus, a CAS computing platform needs access to the attributes of all peers stored in peer profiles. However, unlimited access to the peers' attributes in the peer profiles is not desirable because they may include personal data in the case of peers that are people. The collection, storage and processing of personal data is subject to legislation. The CAS computing platform therefore has to accommodate requirements and should be designed upon a privacy-preserving framework.

1.2 Contributions

This paper summarizes the presentations and discussion results of a workshop by the Smart Society project that was held at the IFIP Summer School on Privacy and Identity Management in September 2014. We first discuss the ethical and legal aspects and requirements in regard to peer profiling in Smart Society. Then we outline the first steps for realizing a privacy-preserving CAS computing platform. We present a *semantic schema* for the representation of peers' personal data and *Peer Manager*, which is a distributed database that stores the peers' personal data. Our contribution sets the minimum requirements for the design and implementation of an attribute-based access control privacy policy for a privacy-preserving CAS computing platform.

In the remainder of this paper, we begin by describing the contextual background and ethical and privacy aspects of CAS computing platform that are introduced in Section 2. Section 3 outlines the privacy requirements taken into account in our research work. The semantic schema and the Peer Manager are presented in Section 4. Section 5 summarizes our work and findings as well as the results of the workshop discussions during the summer school.

2 Ethical and Privacy Issues

In this section we explore how the Smart Society project pays attention to issues of privacy, ethics and social values, and expands upon issues associated more generally with 'big data' and profiling driven approaches. In particular, we draw attention to the extensive scope of the Smart Society vision and its extensive capacity to transform our lives to highlight the importance of these issues. To do this we draw on existing literature detailing the ethical challenge that now confront us from increasing levels of digital mediation within our everyday lives.

The reference to 'Society' in the Smart Society name underlines the extensive ambition of the project. Examples and scenarios generated within the project encompass Tourism, Care, Health, Policing and span from grand aims of solving problems of sustainability to assisting the mundane practicalities of finding somewhere to eat in an unfamiliar town. This breadth and depth underlines the vast scope and everyday pervasiveness implied by the ambition of a 'Smart Society' that aims to address 'societal challenges' and operate at 'internet scale' whilst at the same time penetrating into the mundane aspects of many of our everyday routines and activities. The aim is not to leave these activities unaltered, but rather to supercharge them by linking individuals into collectives to access collectives' problem solving and self-organizing abilities, and to draw upon portable devices, sensors, data and algorithms to assist in 'orchestrating' these newly collectivized activities. Part of the motivation for Smart Society stems from the perception that the existing accumulations of digital mediation for everyday activities have until now been technically untidy, due to a lack of appropriate engineering principles, and ethically haphazard, as a consequence of being unplanned and undirected. Hence Smart Society's twin foci on engineering and ethics. Improved engineering is seen to address the problem of ethics by

providing a structured and therefore more considered process for creating such systems. Ethics helps solve engineering problems by guiding the engineers towards solutions that preserve certain important social values, such as privacy.

In the context of an increasingly digitally augmented lives, the performance of everyday activities now involves numerous data streams leading to vast accumulations of data. This data is viewed as a resource towards solving a wide array of social problems, but its misuse is also seen as threatening our privacy and autonomy. Goodman analyzes ethical issues in the era of personal 'big data' draws attention to the following attributes that carry these types of risk: [11]:

- *"The sensor infused world"*. The shear array of sensors, devices and increasingly everyday objects interconnected via online infrastructures passively generating increasing quantities of data from an ever wider range of activities.
- *"Data as commodity"*. As data has become valuable in its own right beyond the services which generated leading to important questions as to who gets to realize value from data and for what purposes? Nissenbaum's concept of 'data integrity' maintains that use of data should be consistent with the values attached to the activities producing the data [17]. The privacy principle of purpose limitation has a role to play here too as an a-priori contractual determination of the sorts of value that may be derived from data - although there are issues with this approach in making it sensitive to context.
- *"Opacity of back-end information exchange"*. The many ways by which data circulates and are traded are hidden from view, as are the ultimate purposes to which those data may be put. So the ways that such data are subsequently used to filter or shape our experience of the world are often concealed.
- *"Mass scale"*. How this is happening on an unprecedented scale and in ways that do not differentiate between diverse cultural expectations about privacy and data use.

Smart Society could be prone to the hazards described by Goodman for social platforms if data is similarly centralized and its stewardship remains in the hands of platform operators. Extending control of data to users and that bind operators to principles of transparency are important challenges for Smart Society - particularly where this creates technical and operational inconveniences. But also, as Goodman points out, 'big data' may be prone to bias and present an unfair representation of a population, which may be further compounded by the opt outs and obscurification of privacy enhancing technologies.

In the era of personal 'big data' it is common to use data to profile individuals and stratify populations as a means to tailoring or individualizing experience, e.g. by targeting advertising, creating recommendations or tailoring services. Profiles are used within Smart Society to involve peers in collectives, perhaps to solve problems, based upon their experience, skills and reputation. We describe below the importance of profiles for Smart Society but first we enumerate some of the hazards of profiling already identified in the literature.

2.1 Social and Ethical Issues of Profiles

- *Social sorting.* Social Sorting refers to how profiling technologies sorts individuals into categories in order to affect their experiences and opportunities. Examples include banks routing calls from wealthier customers deliver speedier service at the expense of less well off customers or internet service providers giving priority to certain traffic or favored customers. Negative effects of sorting include reinforcing existing social divisions and creating new yet invisible hierarchies of access and privilege [16].
- *Autonomy and self-determination.* Often profiles are created and used without our knowledge or consent, and the ways that our experience of services is modified by this is invisible. Where profiles are computed from data about us, then we are subject the values embedded in the algorithms used to sift that information, but not given a voice in the creation of those algorithms. When we create profiles (e.g. on Linkedin or Facebook) then we are still constrained by what we can express and have little control over how much a flat partial identity may be read by others as a literal depiction of who we are.
- *Diminishing diversity.* "With commercial personalization services, the myriad of individual differences is reduced to one or a few consuming categories, on the basis of which their preferences, character, life-style and so forth are determined for a specific context. Because of its tendency to generalize, personalization may lead to diminishing preferences, differences and values..." [12]. A question raised during the Patras workshop underlines this point. The questioner characterized our experience of city life as in turns vivid, serendipitous, frustrating and pleasurable and questioned how city life mediated through Smart Society 'apps' may lead to a dulling, standardisation and impoverishment of these sorts of experience. It is important for Smart Society to retain elements of fun, chance, discovery and provide an experience that is enriching to and complementary to existing beneficial forms of city life, avoiding to frame problems narrowly in terms of optimization.

Profiles are a crucial component within the Smart Society platform and it is the Peer Profile which gives participants their identity within the system and thereby governs the relationship between individuals and the collectives in which they may become involved. The Smart Society Peer Profile codifies the participant's reputation, interests, expertise and actions. The system uses this information to work out if it can recruit the 'peer' to contribute to solving a problem. In this way the peer profile plays a role in determining participants' opportunities within the system. Given the extent of Smart Society vision then this could imply significant advantages or deficits in life-chances where profiles govern participants' access to culture, education, healthcare and their ability to engage in economic activity. In the context of Smart Society then peoples' very participation in civil society may be at stake.

Some risks of privacy profiling within Smart Society are being addressed in Smart Society, as it is presented in the following sections of this paper. In particular, risks to autonomy are addressed by creating mechanisms that give

the participant ownership of their profile by hosting it on their device; allowing the user to edit or amend any aspect; and in specifying policies describing in what circumstances the data may be shared across the platform.

2.2 Privacy Issues of Profiles

Peer profiling may affect privacy in different respects. As the Council of Europe has discussed in its recommendation CM/REC(2010)13 on profiling [4], the collection, linking, calculation, comparison and statistical correction of data with the objective to create profiles may have significant privacy impacts, as profiling enables a person's personality, behavior, interests and habits to be determined, analyzed and/or predicted. Often such profiling is even happening without the knowledge of the individuals concerned. While profiling may offer benefits for users and society at large, e.g. by providing users with targeted and better services addressing personal and societal interests or by permitting an analysis of risks and fraud, profiling techniques can also have the impact on the individuals concerned by placing them in predetermined categories and may unjustifiably deprive them from accessing certain services and by this discriminate individuals [4].

Moreover, profiling techniques do not only allow to analyze data that are actually recorded, but also allow to statistically predict or implicitly derive information from such records. For instance, sensitive data including about political opinions, religious beliefs, intelligence or sexual orientation can be automatically predicted from Facebook Likes (see e.g., [13]).

During the workshop at Patras, the following more specific privacy questions were raised and discussed but not finally answered, which implies that they still largely remain challenges to be addressed within Smart Society:

- How can privacy interests of "collectives" (consisting of several individuals and/or machines) be protected? How can collectives be formed in an anonymous manner, i.e. in a way that it does not relate to any identified or identifiable person? Can privacy policy languages (to be discussed in the next section) be extended to define privacy preferences of Collectives and negotiate privacy policies for Collectives? Is it a challenge to define/jointly agree on privacy preferences for Collectives in regard to personal data that they have in common/share, or can group decisions and crowdsourcing on privacy preference settings enable/motivate users to spend more efforts on privacy preference management?
- In hybrid systems, peer profiles of machines could include personal data of one or even several data subjects. For instance, in the application of Smart Society to a care scenario, sensors may capture data about when and for how long health care professionals and patients have met. This implies that the sensor readings may reveal both personal information about health care professionals and the patients. Under which conditions can data subjects of data relating also to other data subjects can exercise their data subject rights (if for example the data is only intended for the health care professionals to organise their work, the patient (or their relatives) may still have the right

to access data items that relate to them (e.g., to check whether the patient gets the right treatment)?).
- Are anonymous credentials suitable PETs (privacy-enhancing technologies) for enhancing the privacy of passengers and drivers participating in a Smart Society enabled 'ride sharing' platform that is currently under implementation? Both drivers and passengers could be pseudonymously be registered by the platform and prove only certain properties (e.g., passion of driving license for more than five years, reputation scores). Will the use of anonymous credentials in this context be practically feasible and socially accepted?

Privacy-related questions concerning privacy on sensor data collection [3], trust and reputation [15], and provenance [5] were also raised. These topics are being addressed by Smart Society and are not discussed in this paper.

3 Legal Privacy Requirements in Regard to Profiling

In this section we present the requirements for a CAS computing platform that accommodate data protection and it is designed upon a privacy preserving framework. Basic legal privacy principles, especially those enacted by the EU Data Protection Directive 95/46/EC [7], need to be enforced when profiles that include personal data are created and processed.

These basic principles comprise the following:

- **Legitimacy & informed consent:** The collection and processing of personal data in profiles needs to be *legitimate*, which usually implies that the data subjects[4] have given their informed consent (Art. 7).
- **Purpose specification & binding:** Personal data used in the context of profiling must be collected for specified and legitimate purposes and may later only be used for those purposes (Art. 6 Ib).
- **Data minimization:** The amount of personal data and the extent to which they are collected and processed in profiles should be minimized (Art. 6 Ic), i.e., if possible data in profiles should be anonymised or pseudonymised.
- **No sensitive data:** The collection and processing of so-called special categories of data in the context of profiling should in principle be prohibited (Art. 8 I), unless the exceptions of Art. 8 II apply.
- **Transparency & data subject rights:** Data controllers[5] have to provide the data subjects with sufficient privacy policy information pursuant to Art. 10 when personal data are collected in the context of profiling. Data subjects have the right to to obtain information about their personal data, to be informed about the logic underpinning the processing of their data, to correction, deletion and blocking of their data, and not to be subject to a "decision which produces legal effects concerning him or significantly affects

[4] A data subject is a natural person about whom personal data are processed. We use the terms data subjects, users, and individuals concerned interchangeably.

[5] According to EU Directive 95/5/EC, a data controller is an entity that alone or jointly with others determines the purposes and means of personal data processing.

him and which that is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.".

– **Security:** The data controller has to implement proper technical and organizational security measures for the protection of personal profile data.

The Council of Europe proposed more specific privacy principles that should further strengthen the data subject's protection in an appendix to its recommendation CM/REC(2010)13.

In the context of the EU data protection reform, the proposed General EU Data Protection Regulation (GDPR) [8] introduced with its Art. 20 "Measures based on Profiling". This was however criticized by the Art. 29 Data Protection Working Party on focusing merely on the outcome of profiling rather than on the profiling as such [1].

The compromise amendment to the proposed EU Data Protection Regulation [6], which was passed by the LIBE Committee (Committee on Civil Liberties, Justice and Home Affairs) of the European Parliament on October 21, 2013, has taken up this proposal by providing greater transparency and control for data subjects. According to the amended Art. 14 (ga), data controllers should provide "information about the existence of profiling, of measures based on profiling, and the envisaged effects of profiling on the data subject". In addition, the amended proposal includes the right for data subjects to object to profiling (Art. 20 I). Furthermore, pursuant to Art. 20 III, "profiling that has the effect of discriminating against individuals on the basis of race or ethnic origin, political opinions, religion or beliefs, trade union membership, sexual orientation or gender identity, or that results in measures which have such effect, shall be prohibited". Pursuant to Art. 20 V, "Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject shall not be based solely or predominantly on automated processing and shall include human assessment, including an explanation of the decision reached after such an assessment."

The amendment text to the GDPR also introduced in Art. 4 (2a) the concept of "pseudonymous data", which it defines as "personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution". Recital 58a of the amendment, further states that profiling based solely on the processing of pseudonymous data that cannot be attributed to a specific person should be presumed not to significantly affect the interests, rights or freedoms of the data subject.

4 Concept for a Privacy-enhanced Peer Manager

We define *information peers* as equally privileged participants in an information exchange. Subjects (i.e., humans, machines and services) are represented by informational peers, which interact under a common set of rules, and can be both providers and subscribers of different information exchanges at different

times. They are equally privileged with regard to the decision of engaging or not in the information exchange given the set of rules that apply to such interaction.

Personal data protection is embedded in the design of the Peer Manager and its implementation, the Peer Manager Platform. To address the privacy issues raised in previous sections, the Peer Manager follows three core guiding principles for technically protecting privacy:

1. *Well-defined information separation between information owned by different subjects.* For guaranteeing that the subjects control their own data, the Peer Manager creates distributed environments that host the knowledge container of each subject. Therefore, as the knowledge containers can be physically and logically distributed, the personal information of a subject is isolated from the personal information of other subjects.
2. *User-centred identity management.* The Peer Manager gives to each subject the control of the flow over personally identifying information. Furthermore, the identity management system allows subjects the definition and use of pseudonyms for different interactions as part of their partial identities.
3. *A deconstruction and re-imagining of information profiling.* While the high-level objective of the Peer Manager's Profiles is in general the same as the one of traditional profiles (i.e. an information-holding structure that is maintained and updated separately from the subject to which it refers), we have focused on turning around the regular profiling practices by making them transparent and controllable by the subjects that they refer to.

For enhancing management of information and enforcing the privacy requirements of purpose specification and binding, we define the Peer Manager on top of the entity-centric semantic-enhanced model presented in the next subsection.

4.1 Preliminaries

In the Peer Manager, the representation of information related to peers builds upon the notion of a *semantic schema* defining an attribute-based representation of peers' characteristics. The semantic schema we adopt follows an entity-centric approach that uses the notion of entity to refer to a "thing" that exists in the real world. Within the peer manger we formalize this notion of entity and use it, as the basic element representing information about peers. We distinguish between a *schema level* that defines the "format" to represent information and a *instance level* that defines how to instantiate the schema into actual knowledge [10, 14].

A Knowledge Base (KB) stores data from the schema level. The Schema.org[6] initiative defines schemas as "a set of types, each associated with a set of properties and where the types are arranged in a hierarchy". We adopt an approach that is aligned with this idea and allows the definition of templates for each type of entity used in the system. These templates allows to establish restrictions on the set of attributes that can be used to describe a given type of entity. The meaning is further specified by mapping single elements from the schema (i.e.,

[6] `http://schema.org/`

types of entities, the names of attributes and their values) to concepts from the underlying ontology that is also part of the same knowledge base.

- A **concept** is "an idea of what something is or how it works."[7] In the area of knowledge representation, concepts are used to formalize and represent the meaning of words in a language independent manner. Concepts can be mapped to an underlying ontology that greatly helps identifying purposes (for purpose binding) and other hard to manage limitations for the shared information (e.g., alignment between different data sources). It also provides the basis for more accurate access control methodologies as introduced in [2,9].
- An **entity type** ET provides a template for the creation of entities by establishing a set of constraints about the metadata (i.e., attributes) that entities of that type can instantiate. The template for attributes are defined by mean of the so-called attribute definitions. An *attribute definition* AD imposes an explicit constraint about the name and the quantitative or qualitative values of a certain attribute that can be associated to an entity.

An *entity base* (EB) is defined to store information corresponding to the instance level. It includes concrete information about abstract and physical entities that exist in the real world and is represented by the following elements.

- An **entity** (E) is defined as an abstract or physical object, it can be of different types defined at the knowledge level (e.g., person, location, event, etc.) and is described by attributes (e.g., name, birth date, latitude-longitude, size, duration, etc.), which can be different for different types of entities.
- An **attribute** (A) instantiates an attribute definition AD to represent a particular characteristic of the entity. Some attributes may have multiple values, its values may be mapped to a meaning in some knowledge base (i.e., semantic values) or can represent a relation to another entity when the value is a reference to another E (i.e., relational attribute).

4.2 Privacy-enhancing Structures

The three main structures used in the Peer Manager to protect personal data are: the peer, the user, and the profile structures.

Peers as distributed storage providers. Peer structures are units of storage under the control of the Peer Manager Platform and of the subjects that participate in it. The Peer Manager keeps an entity's data and knowledge base. Every user has a its own Peer Manager and defines the access control policies related to their data. An entity's data is kept isolated from the rest, thus helping to promote the privacy principle of *informational self-determination*.

When an entity registers to the platform, it is assigned a peer structure defined as the the tuple $P = \langle ID, KB, EB, ME, \{U\} \rangle$ where:

[7] Merriam-Webster (`http://www.merriam-webster.com/dictionary/concept`).

- ID is a unique identifier and a reference number to a peer;
- KB is the id of the Knowledge Base owned by the peer that will be used to store all of the concepts and Etypes that belong to the peer;
- EB is the id of the Entity Base owned by the peer that will be used to store all of the Entities that belong to the peer;
- ME is the id of the Main Entity of the peer and it is stored in EB. In the case of a person, the main entity is a person entity. There are other types of main entity, such as for service peers and collective peers; and
- $\{U\}$ is a non-empty set of user structures, which is defined below.

The peer structure allows each subject in the system to have its own dedicated storages (to which then they can apply their own policies and AC directives) as shown in Fig. 1.

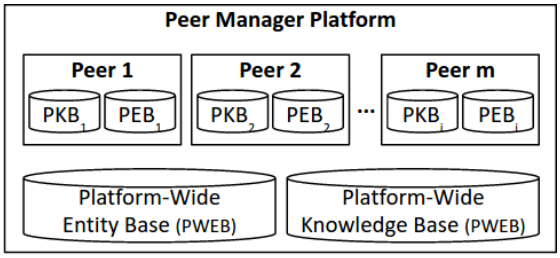

Fig. 1. A Platform handled by the Peer Manager. Each subject its assigned its own peer storage, while the platform itself offers a shared Knowledge and Entity storages for different interactions.

Fig. 1 shows that each Peer has its own PKB (Peer Knowledge Base) and PEB (Peer Entity Base) assigned and clearly separated from the other peers and the platform. The design of the Peer Manager infrastructure also states that only the subject in control of Peer structure has access to it by default and allows for each of these Peers to be stored either in the same server or in different machines altogether. Through this, the platform guarantees that each subject will be always in control of the information stored in his/her assigned peer and that nobody (not even the platform holders) would be able to access this information unless given access by the same subject.

Users structures as subject pseudonyms. When interacting with other peers registered in the platform, entities have the option to control the amount of personal data they reveal. User structures (corresponding to pseudonyms that a subject can act under) are introduced to enhance the privacy of all the subjects that participate in actions/interactions in the Smart Society platform. Entities are able to define N user structures (corresponding to N different pseudonyms) defined as tuples $U = \langle UN, AUTH, P, MPD, \{PD\}\rangle$, where:

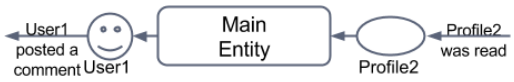

Fig. 2. User structures are used instead of the Main Entity when it acts as a subject and profile structures are used for when the Main Entity is read as an object.

- UN is an alphanumeric string used as the unique identifier to the user structure. This string is a pseudonym for the entity that controls the peer;
- $AUTH$ is an authentication token that is issued by the platform as a proof of peer's identity;
- P is the id of the peer to which a user structure corresponds.
- MPD is the id of the Main profile definition structure that is applied to the peer's ME. It is used to obfuscate (by pseudonymization or anonymization) the link between user structure and the peer that owns it. The resulting Entity Profile is associated to the user structure and (depending on its configuration) may provide none to full identification of the entity that controls it;
- $\{PD\}$ a possibly empty set of profile definitions that subjects create for their entities (e.g. their events, physical and logical objects, and other partial identities) that are linked to the current user structure.

The left side of Fig. 2 shows an user structure being used instead of the ME as the subject of the action "posted a comment". As this example illustrates, the user structure corresponds to a pseudonym for the peer. For achieving a high degree of privacy/unlinkability, different user structures (i.e. different transaction pseudonyms) could be used for different actions.

More in general, Fig. 2 shows how user structures and profiles enhance privacy by providing indirect and partial access to the information from the Peer and its Main Entity, which may contain personal data.

Profiles as object indirections. Instead of directly allowing access to the information contained in the peer's entities, *Profile* structures are created to reply to queries that are sent to the peer (normally only revealing partial or obfuscated information about these entities). The right side of Fig. 2 shows an example where, upon receiving a read query, the peer allows the requester to access the Profile named "Profile2", which contains partial and obfuscated information from ME but not ME, which may contain personal data that the subject may not want to disclose. Profile structures, when they refer to ME, may represent partial identities of the subject controlling the Peer.

The profile structure definition PD is used to define the subset of information to be included in the profile from the ME. A profile definition is defined as the tuple $PD = \langle ID, PE, \{PP\}, \{GP\}, \{NR\}\rangle$ where:

- ID is a numeric unique identifier. It is a reference to the PD;
- PE is the id of the Profiled Entity, the entity to which this profile refers to;

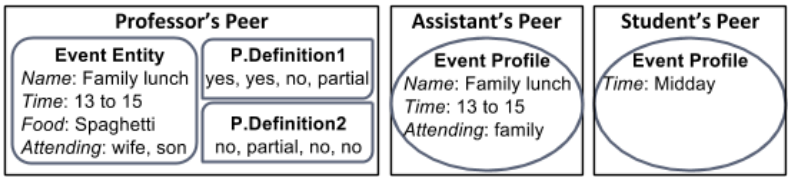

Fig. 3. An example containing Entities, simplified profile definitions and profile materializations for sharing an Event Entity.

- $\{PP\}$ is a non-empty set that specifies the different parameters that feed the algorithms that are to be applied to the profiled entity to obtain the profile, both this set of parameters and the algorithms are entity type-dependent (although future versions may consider its generalization);
- $\{GP\}$ is a possibly empty set that contains the id of all Generated Profiles obtained from the current definition, and;
- $\{NR\}$ represent the Negotiation Requirements that need to be complied by the parties wanting to have access to the information that this profile definition will generate.

Applying a profile definition to the Entity it refers to materializes the actual Profile into an *Entity Profile* structure. An Entity Profile is defined as the tuple $EP = \langle ID, U, S, \{A\}, \{R\}, \{AR\}\rangle$ where:

- ID is a numeric unique identifier
- U is the id of the user structure that was the source of this profile
- $\{A\}$ is a set of attributes defined as before but the specific attribute definitions and values may be different from the ones in the entity.
- $\{R\}$ is a set of relations defined as before but the specific relation definitions and values may be different from the ones in the entity.
- $\{AR\}$ is the possibly empty set of Agreed Requirements set between source user structure of the profile (U) and the owner of this entity instance, this property can be checked to make sure that the terms or agreements are not breached.

The profile definition structure (i.e., the filter to apply to the original information before sharing it) is stored at the source peer's storage while the materialized Profile structure (containing the shared partial and/or obfuscated information) is stored at the destination peer's storage, as shown in the example from Fig. 3.

The left part of Fig. 3 shows an Event (family lunch) that belongs to a professor's Peer. The professor created two profile definition structures, which are presented here in a simplified manner, to define how this event is shared to assistants and students. The rest of the figure shows the peer structures that belong to the subjects with the materialized profiles created from applying the restrictions from the profile definition. These materializations include examples of omitted pieces of information (e.g. the 'food' attribute is not shared in neither profile) and partial/obfuscated information (e.g. the time of the event becomes

'Midday'). It is important to note that the realisation of these materializations requires the formalisation of functions that can provide different levels of abstractions for the original information, subjects should be able then to select a particular level during the creation of a profile definition. However, such formal definition is out of the scope of this paper and is left as part of our future work.

The use of the information in the materialized Profiles is restricted by the Agreed Requirements, which are agreed upon a privacy policy based on PPL [18]. Policy enforcement at the platform-level guarantees that the shared information is only used for the stated purposes. For example, if profiles are used to share contact information, e.g. the professor gives her telephone number to her assistant, the professor may restrict its use to "call over phone only". Therefore, any other operation over the data, such as reading or copying the telephone number is not allowed, i.e., the assistant may call the professor but the actual phone number is not revealed to the assistant.

5 Conclusions

This paper provides an initial discussion of privacy and other ethical issues of peer profiling within the scope of the Smart Society project. It also presents the concept of a privacy-enhanced Peer Manager, which is a fundamental building block for the implementation of a privacy-preserving CAS computing platform. As presented in Section 4, the Peer Manager allows people and other actors, such as sensors and actuators, to store their information in a secure and preserving framework. The design and development of the Peer Manager followed the following three core guiding principles for enforcing legal privacy principles including the principles of data minimisation, purpose binding, transparency and user control:

1. Separation of information among peers.
2. A user-centered identity management.
3. A user-controlled approach to peer profiling.

The information stored by the Peer Manager follows a semantic schema that defines an attribute-based representation of a peer's characteristics. The Peer Manager allows people (i.e. users) to define profiles that contain and reveal only partial or obfuscated information that are used for replying to information requests and is thus enforcing data minimisation. In addition, the purposes and use of personal data are specified and limited by a PPL privacy policy.

The Peer Manager addresses the presented challenges of information profiling, such as the lack of information and feedback about how profile data is collected and traded, by giving transparency and control of this process to the actual subject that the profile refers to. Hence, subjects can create their own profiles with the minimal amount of information needed and share this information with the promise that their requirements will be enforced by the platform and that the shared data will not be misused or traded against their will. Other features of the platform partially reduce the influence of issues like social sorting (profiled users may not consent giving information for these purposes or create a

new pseudonyms to avoid them altogether), diminishing diversity (the greater expressiveness of the semantic schema provides more diversity to data and its representation) and improve self-determination (for the same reasons as the previous two).

It is important to emphasize that the current version of the approach does not completely solve the issues raised in the workshop and that it also introduces potential new issues such information management complexity for users and going against some of the popular business models related to data management.

Management complexity refers to the fact that giving full control of the information could become overwhelming to it owners; specially at the scale in which personal information is produced. For this reason, it is key for the future adoption and practicability of this approach to present itself as an extension of existing identity management systems. While some people may not be interested in fine-grained privacy settings, we plan to provide the possibility to individuals to review and change their privacy settings in a usable way.

It is also worth considering that companies and platforms that base their revenue on the harvesting, processing and reselling customer information may initially be reluctant to give back to their customers more control over their information. There are however, other organizations (e.g. public utilities, health organizations) that value much highly the trust that their customers have in them. Moreover, this higher-degree of customer trust and sense of being in control may also make setting partial limits and transparency settings beneficial to some of the applications that do use this information as source of revenue.

The close attention we pay to privacy and social values for profiles within Smart Society does not mean that all difficulties are circumvented - there are a host of future challenges too. We have to be continually vigilant over whether autonomy and self-determination are indeed preserved by the envisaged technical measures are implementing to protect profiles within Smart Society. For example, despite having the tools to control circulation of their data, a participant may still feel compelled to disclose an item to the system if refusing to do so leaves them materially disadvantaged. This means that we have to keep sight of the wider governance of Smart Society at the same time as we focus at specific technical means that preserve privacy locally.

References

1. Art. 29 Data Protection Working Party: Advise paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation. Available from `http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130513_advice-paper-on-profiling_en.pdf` (13052013)
2. Chenu-Abente, R., Zaihrayeu, I., Giunchiglia, F.: A semantic-enabled engine for mobile social networks. In: Cimiano, P., Fernández, M., Lopez, V., Schlobach, S., Völker, J. (eds.) The Semantic Web: ESWC 2013 Satellite Events, Lecture Notes in Computer Science, vol. 7955, pp. 298–299. Springer Berlin Heidelberg (2013)

3. Christin, D., Roßkopf, C., Hollick, M., Martucci, L.A., Kanhere, S.S.: IncogniSense: An anonymity-preserving reputation framework for participatory sensing applications. Pervasive and Mobile Computing 9(3), 353–371 (2013), http://dx.doi.org/10.1016/j.pmcj.2013.01.003
4. Council of Europe: Recommendation CM/Rec(2010)13 of the Committee of Ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling. Available from https://wcd.coe.int/ViewDoc.jsp?id=1710949 (11 2010)
5. Davidson, S.B., Khanna, S., Roy, S., Stoyanovich, J., Tannen, V., Chen, Y.: On provenance and privacy. In: Proc. of the 14th Int. Conf. on Database Theory - ICDT. pp. 3–10 (2011), http://doi.acm.org/10.1145/1938551.1938554
6. European Commission: Proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011—C7 0025/2012—2012/0011(COD)) Compromise amendments on Articles 1-29. Available from http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/dv/comp_am_art_01-29/comp_am_art_01-29en.pdf (21102013)
7. European Commission: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Available from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML (23111995)
8. European Commission: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final 2012/0011 (COD). Available from http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf (2512012)
9. Giunchiglia, F., Crispo, B., Zhang, R.: Access control via lightweight ontologies. In: Semantic Computing (ICSC), 2011 5th IEEE Int Conf on. pp. 352–355 (Sept 2011)
10. Giunchiglia, F., Dutta, B., Maltese, V.: From knowledge organization to knowledge representation. In: ISKO UK Conference (2013)
11. Goodman, E.: Design and Ethics in the Era of Big Data. interactions 21(3), 22–24 (May 2014), http://doi.acm.org/10.1145/2598902
12. van der Hof, S., , Prins, C.: Personalisation and its influence on Identities, Behaviour and Social Values. chap. 6, pp. 111–127. Springer, New York (2008), http://opac.inria.fr/record=b1126046
13. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 110(15), 5802–5805 (Mar 2013)
14. Maltese, V., Giunchiglia, F., Dutta, B.: Domains and context: first steps towards managing diversity in knowledge. Web Semantics: Science, Services and Agents on the World Wide Web 12(0) (2012)
15. Martucci, L.A., Ries, S., Mühlhäuser, M.: Sybil-Free Pseudonyms, Privacy and Trust: Identity Management in the Internet of Services. Journal of Information Processing 19, 317–331 (Jul 2011)
16. Monahan, T.: Surveillance and inequality. Surveillance & Society 5(3) (2002)
17. Nissenbaum, H.: Privacy as contextual integrity. Wash. L. Rev. 79, 119 (2004)
18. Trabelsi, S., Neven, G., Raggett, D. (eds.): PrimeLife Public Deliverable D5.3.4 – Report on design and implementation (20 May 2011)