# Personal Semantics: Personal Information Management in the Web with Semantic Technologies

#### Salman Elahi

Knowledge Media Institute (KMi), The Open University, UK s.elahi@open.ac.uk

**Abstract.** Every web user has several online profiles through which personal information is exchanged with many service providers. This exchange of personal information happens at a pace difficult to fully comprehend and manage without a global view and control with obvious consequences on data control, ownership and, of course, privacy. To tackle issues associated with current service-centric approaches, we propose a user-centric architecture where the interaction between a user and other agents is managed based on a global profile for the user, maintained in a profile management system and controlled by the user herself. In this PhD, we will investigate research issues and challenges in realizing such a system based on semantic technologies.

## 1 Research Problem

Web users maintain several online profiles across e-commerce websites, social networks and others, making it difficult for them to realize how much information is exchanged and what happens to that information. In other words, web interactions are happening in a 'one to many' mode, where different agents, with different statuses and relationships to the user take part and receive personal information. We refer to this phenomenon as the *fragmentation of personal data exchange*, where many different 'destinations' of the data receive various fragments of personal information over time, without the user having a global view of her own data.

The problem stems from the model these online interactions are based upon, i.e. a service-centric model, where everything is focused on the needs of a particular organization. Current research suggests a rather contrasting model to address these issues: a user-centric approach [1] where the interaction between a user and other agents is managed based on a global profile for the user, maintained in a profile management system and controlled by the user herself. Parts of this profile can be requested by various agents (websites), with the user being given the possibility to control these accesses and to keep track of them directly within the profile management system. In this PhD, we will investigate research issues and challenges in realizing a system based on user-centric profiles, and show how semantic technologies can support users in making sense, managing and controlling their exchanges of personal data on the Web with such a system.

## 2 State of the Art

Personal information management as defined in [2] is at the core of Semantic Desktops [3], which are concerned with the management of information produced and consumed by desktop applications. While there are significant overlaps, our approach focuses on personal information on the web and on the way this information is exchanged with other agents, rather than on the management of information in the close, local environment of the user.

Identity management shares a thin boundary with the aspects of personal information management we are looking at. The most popular approach for "usercentric" identity management is OpenID<sup>1</sup>. OpenID is a protocol that provides a unique ID to each user to be used over various services. It therefore supports users in reducing the distribution of their identity data. OAuth<sup>2</sup> coupled with OpenID provides secure exchange of data without disclosing users' credentials to third party websites. However, OpenID and OAuth are essentially concerned with the problem of authentication, not the management of personal information. A few initiatives are attempting to realize user-centric identity management frameworks, including Windows CardSpace<sup>3</sup>, LiberryAlliance<sup>4</sup>, Higgins I-Card<sup>5</sup>, etc. These frameworks provide central places for storing and managing personal information (i.e., profiles), to which external websites are granted access. In this sense, they comply with our notion of user-centric personal information management system. While they still suffer from a number of limitations (e.g., in iCard based frameworks, personal data is still fragmented, in the sense that profile information is "boxed" according to the websites requesting it). [4,5] concern flexible user profiles in mash up environments. They focus on domain independent models while [6,7] also discuss user models but with a focus on retail domain. FOAF+SSL [8] is an authentication protocol to provide secure exchange of distributed information. It uses the SSL layer in modern day Web browsers to create a secure conduit of information between two parties. FOAF+SSL is also in an initial phase of development and has not been widely adopted yet.

## 3 Propose Approach and Methodology

A social context provides a very comfortable leverage to someone to represent and manage their persona in different real world interactions as compared to online interactions, where it becomes a tedious task to maintain one's identity (personal information) in one place because of the many different profiles a typical web user usually possesses with each service provider she interacts with. In this scenario, user-centric profiles seem to be a promising solution where a user can define different contexts to disclose certain aspects of her profile depending on what type of agents

<sup>1</sup> http://www.openid.net

<sup>&</sup>lt;sup>2</sup> http://www.oauth.net

<sup>3</sup> http://www.microsoft.com/windows/products/winfamily/cardspace/
default.aspx

<sup>4</sup> http://www.projectliberty.org/liberty/about/

<sup>5</sup> http://www.eclipse.org/higgins/

she is communicating with. However, the realization of a truly user-centric personal profile management system raises complex research issues and challenges, which we intend to tackle in this PhD, partly based on the use of semantic technologies. More precisely, we will focus on the technical and the user aspects [9]:

- The technical aspects include in particular issues related to user modelling in an open, web environment: how to integrate multiple profiles into one global profile? How to manage access policies and access rights on user information, including identifying policy conflicts between consumer and provider and suggesting ways to resolve them? How to make a user's interaction more sophisticated and friendly with new concepts? How to represent behavioural changes in the profile while maintaining consistency?
- The user-related issues regard the way to integrate this user-centric approach in the existing social environment of the user: How much of a profile can be constructed with already available user information? What interaction models are suitable to the management of large, global user profiles by web users? Will users be willing to invest time in such a system? What mechanisms are necessary to ensure trust in the profile providers?

In practice, we envisage the combination of user-centric profile management with semantic technologies as a system where the user is offered the ability to manage one, integrated profile containing information potentially relevant to different agents she is communicating with: a semantic model of her personal information. Such a semantic model of personal information can not only be used to negotiate terms of communications (access policies) with different service providers but can also be reasoned upon, enabling a user to have a global coherent model of her data, helping her in making sense, managing and giving access to them. Through this approach, we intend to answer the following three questions:

- 1. How to model a user? Different users would have different requirements to be able to flexibly structure, organize and represent their arbitrary data.
- 2. How to model semantic access control over semantic data? We will be looking at how access control, and especially definition of access rights in user-centric profile scenarios is realized in a way homogeneous to the description of its semantic data, taking benefit from the flexible data model and inference capabilities provided by the Semantic Web technologies.
- 3. How users would react to such a new approach? We will empirically study the impact of a user-centric profile management system on the practices of data exchange, based on user surveys and on monitoring usages.

The methodology to resolve these questions will consist of the following phases:

1. **Feasibility studies:** The idea of user-centric profile management has not yet been tested in the open environment of the web, where thousands of exchanges happen with hundreds of websites every day. In the next section, we give an overview of a first study where we reconstructed a global user profile from logs of a particular user's activities on the web [10]. Such feasibility studies help us

identifying specific technical issues related to the combination of semantic technologies with user-centric profile management in a web environment.

- 2. Contrasting the user-centric approach with existing models: In this phase, we focus on the aspect of managing the flexible exchange and access to personal information using an ontological model. We are conducting another experiment by creating a semantic representation of the current information and access model of a large organization (The Open University), in order to contrast it with a possible user-centric view of information and access. The Open University represents a perfect example of a large-scale distributed organization with multiple departments, a very large number of distance learning students and off campus employees. In this phase, we intend to demonstrate how a user-centric, semantic model of 'personal information' and 'access policies' provides added value through allowing users to manage, reason upon and control their own information in a more transparent way than in an organization centric model.
- 3. Tackling specific technical issues and prototyping: We propose an architecture consisting of four evolving sub-frameworks. The profile framework will provide intuitive interfaces for distributed management of user-centric profiles. The policy framework acts as a security grid based on the intelligence provided by the Semantic framework, which assesses user behaviour and provide recommendations to evolve the profile and keeps the cycle going. The policy framework derives security policies based on the rich semantic information about the agents the user is interacting with and defines access rights on the contents which are then used in the Content framework when a request comes in for certain information from a consuming agent [9].
- 4. **Empirical study and evaluation:** Along with the evaluation studies of phase 1 and 2 we will be realise empirical studies where the prototype system developed in the previous phase will be used within an organization instead of the usual information and access mechanisms. Through collecting both usage and survey data, we will be able to assess the benefit of the approach, and evaluate specific advantages gained through the use of semantic technologies.

## 4 Initial Results

In this section we discuss initial results obtained (during the first year of this part-time PhD). In [10], our objective was to collect all the fragments of personal information sent by a user during several weeks, and to try to reconstruct from these fragments a coherent global profile. From logs of HTTP traffics through specifically developed tools, the user managed to create 36 profile attributes mapped onto 1,108 data attributes. However, while this small and simple exercise proved to be satisfying the basic hypothesis, we also identified few research issues and challenges which will need to be addressed to create more complex and flexible user profiles [10], including the need for complex and sophisticated profile representations (with temporal and multi-faceted representations), the need to include multiple, external sources of information to enrich the profile (including information about the accessing agents) and the need for appropriate mechanisms for the definition of semantic access control

model over semantic data. On the last point, we are currently investigating a prospective model for access control in a user-centric scenario, and applying it in the scenario of the Open University's information and access model. Through the use of an ontological model, and the possibility of employing inference upon such a model, we expect to obtain results showing how the user-centric approach can outsmart the organization-centric approach, providing better overviews of the access control aspects over the considered data and possibly detecting un-intended behaviours which normally remain hidden in an organization-centric view.

## 5 Future Work

The next major steps for this part-time PhD include completing the current work on Semantic policies and access rights in the next 4 months. The results obtained from this exercise will be used to develop a prototype profile management system based on the user-centric framework discussed earlier. This framework will employ semantic technologies to provide access control through semantic policies with evolvable profiles representing changing needs of a user in today's online interactions. This phase will investigate and try to address the technical issues mentioned above. Social issues will be investigated in the next phase with the deployment of this system in place of the Open University's usual information and access mechanisms to gather empirical evidence of how users interact with such a system and its impact on them. This empirical evidence will be used in the evaluation of the system.

## References

- 1. Iannella, R.: Social Web Profiles. Position paper, SNI (2009)
- 2. Jones, W., Teevan, J.: Personal Information Management (2007); ISBN: 9780295987378
- Sauermann, L., Bernardi, A., Dengel, A.: Overview and outlook on the semantic desktop. In: ISWC (2005)
- 4. Leonardi, E., Houben, G., Sluijs, K., Hidders, J., Herder, E., Abel, F., Krause, D., Heckmann, D.: User Profile Elicitation and Conversion in a Mashup Environment. In: FIWLIW (2009)
- 5. Abel, F., Heckmann, D., Herder, E., Hidders, J., Houben, G., Krause, D., Leonardi, E., Slujis, K.: A Framework for Flexible User Profile Mashups, AP-WEB 2.0 (2009)
- Ghosh, R., Dekhil, M.: Mashups for semantic user profiles. Poster. In: International World Wide Web Conference, WWW (2008), Poster
- Ghosh, R., Dekhil, M.: I, Me and My Phone: Identity and Personalization Using Mobile Devices. In: HPL 2007-184 (2007)
- 8. Story, H., Harbulot, B., Jacobi, I., Jones, M.: FOAF+SSL: RESTful Authentication for the Social Web. In: SPOT (2009)
- 9. Olesen, H., Noll, J., Hoffmann, M. (eds.): User profiles, personalization and privacy. WWRF Outlook series, Wireless World Research Forum (May 2009)
- 10. Elahi, S., d'Aquin, M., Motta, E.: Who Wants a Piece of Me? Reconstructing a User Profile from Personal Web Activity Logs. In: LUPAS, ESWC (2010)