An architecture for privacy-enabled user profile portability on the Web of Data*

Benjamin Heitmann and Conor Hayes

firstname.lastname@deri.org
Digital Enterprise Research Institute
National University of Ireland, Galway
Galway, Ireland

Abstract Personalisation on the Web is becoming a commodity, yet privacy and personalisation are at odds. We propose an architecture based on Linked Data, FOAF and WebIDs. It combines the privacy-enabling properties of current centralised personalisation architectures with the portability and decentralisation of the emerging Web of Data. It has the potential to enable a universal "private by default" ecosystem which can provide incentives for more users to open up their profile data for personalisation services.

1 Introduction

Personalised recommendations have proven themselves to greatly enhance the user experience of searching, exploring and finding new and interesting content [2], e.g. on social websites like Facebook and Last.fm.

However, privacy and personalisation are currently at odds [5]: In order to provide a personalised experience it is necessary for a website to have access to data about that same user. This requires the user to trust the personalisation service not to misuse or trade his data.

This issue gets aggravated by a paradigm shift which is happening right now in the field of personalisation: User profiles are being collected beyond the context of a single service. This is demonstrated by the Facebook Open Graph protocol[1], which provides the architecture for an ecosystem in which profile data can be aggregated and shared between services. This requires the user to place his trust in a single, centralised storage site for all of his profile data, which then controls access to the profile data by all other services in the ecosystem.

In this paper we propose an alternative: Instead of creating ecosystems around closed data storage silos, we propose to create a universal ecosystem around portable user profiles. These user profiles can be moved between social services or they can be hosted by the user themselves.

2 Centralised personalisation architectures

While most personalised services collect profile data from their own users, a paradigm shift is happening towards ecosystems which allow the user to share his profile data with services belonging to the ecosystem, after explicitly giving his authorisation. In such an ecosystem one site typically has the role of the *hub site*, which provides the main entry point for the whole ecosystem and stores the user profile data. Prominent social networking sites like Facebook and Twitter are such central hub sites. *Third party services* can provide value-added and personalised services for the user of an ecosystem. Examples include TweetMeme.com which shows the most popular links on Twitter, and the Flickr integration for Facebook which posts pictures uploaded on Flickr to the user's Facebook activity stream. *Users* stay in control of their profile data, as their profile is stored on the central hub site and the user can specify which services can access their profile data. If a service e.g. trades the usage data then the user can revoke access for the service.

If such an ecosystem respects the privacy of the user, it can provide powerful incentives for users to allow the sharing of their profile data between different services. However it also leads to user lock-in and social networking data silos: User profiles are not portable between systems, connecting to users from a different system is not possible and the user can not evade changes to the terms of service.

3 An alternative: a decentralised privacy-enabling architecture

In order to enable users to share their profiles with different ecosystems in a decentralised way, while maintaining their privacy at the same time, it is necessary to define an architecture for privacy-enhanced user profile portability. Building on work by Hollenbach, Presbrey and Berners-Lee [3], we present an architecture which describes how to combine existing infrastructure of the Web of Data and existing standards for decentralised identity management in order to achieve privacy-enabled user profile portability. In this section we describe the foundation standards and the roles of the participants in the decentralised architecture, the properties of the resulting architecture and related standards.

3.1 Foundation standards

The *FOAF vocabulary* allows the description of domain independent user profiles [1]. FOAF provides properties to describe all of the details which are usually contained in a social networking profile or on a personal homepage. In addition a FOAF profile provides a container for other information from different domains. For instance, this information could use the SIOC vocabulary to list the content which the users has generated on his blog, on his twitter stream and the comments on different forums.

* This paper is an abbreviated version of Heitmann et al., "An architecture for privacy-enabled user profile portability", Workshop on Information Heterogeneity and Fusion in Recommender Systems at the ACM Recommender Systems Conference 2010, and funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Líon-2).

[1] http://preview.tinyurl.com/facebook-OGP

WebIDs [4] securely connect a user identity to the information in a user profile and can be used for authenticating a user. A WebID consists of two parts: (1) an SSL certificate which contains a link to (2) the URI from which information about the user can be obtained. The data which is obtained from the URI is associated in return with the SSL certificate, as it lists the cryptographic hash of the private key which is associated with the public key contained in the SSL certificate.

3.2 Roles

The interplay between Linked Data, FOAF and WebIDs requires the participants to perform one of three roles: profile storage services, data consumers and user agents.

The *profile storage services* roughly correspond to the hub sites in current profile sharing ecosystems. They provide the storage for the user profile or parts of it, and they secure the access to the profile data based on the authorisation which the user has given to different data consumers. Profile storage services can be either self hosted by the user or they can be hosted by a social networking site.

Data consumers correspond to any type of third party service which is accessing user profile data in current ecosystems. Each consumer has its own WebID, which identifies the service every time it is accessing profile data from a profile storage service. This allows the storage to determine if the access is granted to the consumer.

User agents manage the different identities of a user. Each identity is represented by a WebID, which is used for authenticating the user towards profile storage service or data consuming services.

3.3 Properties

Protection of identity: Users can choose to use multiple identities, each identity being represented by a unique WebID. Each time a user interacts with a data consuming service his user agent can allow him to choose which WebID to use. In this way pseudonymity, unobservability and deniability of the user identity are supported. None of the identities need to be tied to a real world identity, thus supporting anonymity.

Control over the user data: The user stays in control of his profile data, as the portability of user profiles allows him to move his profile freely between storage services or even to host the storage of his profiles on his own server. Lock-in to a specific ecosystem or to a specific storage service should not be possible, as the open standards of RDF, FOAF and SIOC are used for describing the profile.

Non-functional requirements: The presented architecture allows any user agent, profile storage service or data consumer to participate in one universal ecosystem, as all participants will support the same standards and implement the same communication pattern. The architecture is scalable, as there are no bottlenecks or central points of failure, due to the decentralised nature of the used standards. For profile storage and data consumption existing standards and infrastructure from the World Wide Web and the Web of Data, such as HTTP and RDF are reused, thus making future adoption by service providers easy.

3.4 Related standards

OpenID is a standard for decentralised authentication of a user. It provides a way to prove that an end user owns an identity URL without passing around the password of the user to a third party service. OpenID provides the means to decouple identities from real users, thus enabling pseudonymous personalisation. However OpenID is not well suited for machine agents and it requires a large overhead in terms of the number of HTTP connections which are required to gain access to a secured resource [4].

OAuth specifies a protocol for decentralised authorisation of resource access. Third party services obtain access tokens which are used to access the protected resources. OAuth is used the most popular current social websites, such as Facebook, Twitter, Google and LinkedIn. However in addition to two version of the OAuth standard being used at the same time, each ecosystem requires different client implementations, thus leading to fragmentation. WebIDs and Linked Data can enable a universal privacy-enabled ecosystem.

4 Conclusion

In this paper we addressed the problem of preserving user privacy while seeking to integrate multiple personal information sources. The default architectural solution requires a centralised hub with users reliant upon the good will of the service provider to ‘do no evil’. As an alternative we have presented a decentralised architecture for privacy-enabled user profile portability based on existing standards. It allows users to benefit from the privacy that is provided by centralised and closed social networking ecosystems as well as from the portability that is provided by the decentralised and open Web of Data. This could ultimately enable a universal “private by default” ecosystem for personalised services on the Web.

References

1. U. Bojārs, J. Breslin, V. Peristeras, G. Tummarello, and S. Decker. Interlinking the Social Web with Semantics. *IEEE Intelligent Systems*, 23(3):29–40, 2008.
2. R. Burke. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.
3. J. Hollenbach, J. Presbrey, and T. Berners-Lee. Using RDF Metadata to enable Access Control on the Social Semantic Web. In *Workshop on Collaborative Construction, Management and Linking of Structured Knowledge*, 2009.
4. H. Story, B. Harbulot, I. Jacobi, and M. Jones. FOAF+SSL: RESTful Authentication for the Social Web. In *Workshop on Trust and Privacy on the Social and Semantic Web*, 2009.
5. Y. Wang and A. Kobsa. Technical Solutions for Privacy-Enhanced Personalization. *Intelligent User Interfaces: Adaptation and Personalization Systems and Technologies*, 2009.