Towards Privacy Aware Social Semantic Digital Libraries

Owen Sacco [a] and John Breslin [b]

[a] *University of Malta - owensacco@gmail.com*
[b] *National University of Ireland, Galway - john.breslin@nuigalway.ie*

Digital Library systems provide an effective means to share digital information. These systems provide repositories of digital objects such as catalogues, books, articles, documents, images, video material and audio material. Most digital library systems also provide users to create profiles that include personal information such as their topics of interest, publications, number of citations, subscribed material, history of viewed material, connections with co-authors, and other personal information. Therefore, digital libraries can be seen as implicit social networks since connections amongst users can implicitly be formed from the personal information of users created and gathered within these systems. Moreover, several activities carried out in social networks are also carried out in digital libraries such as social bookmarking. Utilising personal information and the connections between users could enhance personalised services that digital libraries could offer, such as personalised search based on the user's interests and the interests of the user's connections.

Apart from utilising personal information created within digital library systems, these platforms could also be integrated with social network platforms in order to improve personalised services. For instance, recommending or searching digital objects can be based on social network interactions between the user and his/her connections (i.e. "friends"). The digital library platform can recommend articles ranked according to the number of "likes" an article has and whether the topic of the article is similar to the user's interests or similar to the topic of the user's recent publications. Recommending articles can also take into consideration trust measures. For example, the profile similarity between profiles of the user and the user's friends could provide better ranking values especially when comparing the interests between the user's interests and the user's friends interests - those friends that have the same interests as the user and have liked a particular article, then that article will have a higher ranking. Other measures could also be taken into consideration such as the number of interactions amongst users. These trust measures will eventually form a trusted social network which digital libraries could benefit from for providing personalised services. Moreover, digital libraries could be integrated with multiple social network platforms and also with other library platforms in order to take advantage of the personal information about users. This integration will not only enhance searching and recommending articles but also other services such as personal information management within digital libraries.

Digital libraries however pose several challenges: first, most digital libraries are data silos - most platforms structure their data using system specific schema which create a walled garden effect such that digital library data sources are not interoperable with each other and make it hard to link data elements amongst each system. Second, most digital libraries do not make use of current social data residing in social network platforms or in other digital library systems but provide the user to re-create his/her personal data. Digital libraries should utilise social data by aligning and aggregating user's personal information with the user's personal data residing in social network platforms and also residing in other digital library systems. Third, digital libraries lack privacy measures to provide users ways how they want to share their personal information.

The Semantic Web, which evolves from the conventional Web, provides techniques to markup data with meaning which can be processed by machines to offer enhanced services for data sharing and interoperability amongst different data sources. Such structured data is increasing as developers are becoming more aware of the advantages that Semantic Web technologies have to offer. However, the meta-formats which the Semantic Web provides are difficult for non-technical users to grasp in order to structure their data. Therefore, other formats emerged such as microformats[1] which are structured on current standards and provide easy to use formats to markup content with semantics (meaning) in Web documents. All of these formats have an underlying goal: to add structure to Web content in a *graph model*. The aim is to use identifiers, Uniform Resource Identifiers (URIs), to uniquely identify things (also known as resources) such as people, events, blog posts, reviews and tags published in Web documents. Therefore, each resource can link to other resources by referring to the URI of the specific resource to link to. Resources can be depicted as nodes in graphs and the edges between nodes illustrate the links between them. The advantage of linking resources is that different datasets can be linked, and hence create the *Web of Data*.

Content stored in digital libraries and social network platforms can be standardised and represented using various vocabularies such as Friend-of-a-Friend (FOAF)[2] for describing basic personal information, the Relationship Ontology[3] for describing relationship types with other users, the Description-of-a-Career (DOAC)[4] for describing career related information and Semantically Interlinked Online Communities (SIOC)[5] for describing activities. In order to disambiguate terms such as user's interests, DBPedia[6] concepts are used to describe such terms. Standardising how data is represented in this way would enable digital library systems to interoperate amongst each other and also with social networks that would create social semantic digital library platforms. However, this data is easily accessible and open since no access control mechanisms are in place for the Web of Data.

Our work, the *Privacy Preference Framework*, provides an attribute-based access control (ABAC) approach which allows expressing access control restrictions based on attributes which the requester and the restricted data must satisfy. Considering that no vocabulary provides fine-grained access control mechanisms for the *Web of Data*, our

---

[1] microformats – `http://microformats.org/`

[2] FOAF — `http://www.foaf-project.org`

[3] Relationship — `http://vocab.org/relationship/.html`

[4] DOAC — `http://ramonantonio.net/doac/0.1/`

[5] SIOC — `http://sioc-project.org/`

[6] DBPedia — `http://dbpedia.org/`

work provides a vocabulary for describing privacy settings and a manager to filter data based on these settings.

The *Privacy Preference Ontology (PPO)* [2] - `http://vocab.deri.ie/ppo#` - is a light-weight Attribute-based Access Control (ABAC) vocabulary that allows people to describe fine-grained privacy preferences for restricting or granting access to specific Linked Data. Among other use-cases, PPO can be used to restrict part of a user's digital library records only to users that have specific attributes. It provides a machine-readable way to define settings such as "Provide my list of publications only to those who have published articles of the same topic" or "Grant access to my personal contact details only to my co-authors".

As PPO deals with RDF(S)/OWL data, a privacy preference, defines: (1) the resource, statement, named graph, dataset or context it must restrict access to; (2) the conditions refining what to restrict; (3) the access control type; and (4) a SPARQL query, (`AccessSpace`) *i.e.* a graph pattern representing what must be satisfied by the user requesting information.

The *Privacy Preference Manager (PPM)* [1], is a privacy preference manager for the Web of Data. Our aim is to illustrate how PPO can be applied to create privacy preferences and how personal information, specifically digital library data integrated with social data, can be filtered based on those preferences.

The *Privacy Preference Manager* allows users to manage their privacy preferences and also grants or denies access to user's information when requested by others. Using it, users can (1) authenticate to their instance and create privacy preferences for their digital library data and social data; and (2) authenticate to other user's instance and access the filtered digital data of these users. Moreover, the *Privacy Preference Manager* provides an API that can be used by Web systems to take advantage of incorporating privacy preferences enforcement within their system.

In this paper we: (1) Demonstrate how social networks and digital library platforms can be integrated with each other, and how digital libraries can enhance their services using social data; (2) Describe the vocabularies and ontologies required to represent social data and digital library data, and how these can be linked to form a graph of interlinked social data and digital library data; (3) Demonstrate how trust can be calculated on social data to enhance digital library services; (4) Demonstrate how the *Privacy Preference Framework* can be utilised to filter digital library data and social network data within digital library platforms; and (5) Extend the *Privacy Preference Framework* to support semantic digital libraries and also to incorporate digital rights management for restricted access platforms.

References

[1] O. Sacco and A. Passant. A Privacy Preference Manager for the Social Semantic Web. In *Proceedings of the 2nd Workshop on Semantic Personalized Information Management: Retrieval and Recommendation, SPIM2011*, 2011.

[2] O. Sacco and A. Passant. A Privacy Preference Ontology (PPO) for Linked Data. In *Proceedings of the Linked Data on the Web Workshop, LDOW2011*, 2011.