Using the Social Semantic Web to Support eSocial Science

P Edwards, R Reid, A H Chorley, F Hielkema, R Mitra and E Pignotti

School of Natural and Computing Sciences
University of Aberdeen
Aberdeen, AB24 3UE
{ p.edwards, r.reid, a.h.chorley, f.hielkema, r.mitra, e.pignotti } @ abdn.ac.uk

Abstract

In this paper we discuss how aspects of the Social Semantic Web can be used in an eScience context to deliver tools to researchers in the social sciences. We explain how user requirements led us to develop a hybrid solution involving use of ontologies and folksonomies in order to document the provenance of research resources, and to situate these within their wider (social) context. A natural language interface able to create, query and browse these descriptions is also briefly described.

Introduction

Scientific research is increasingly carried out by communities that span disciplines, organisations and geographical boundaries. Furthermore, researchers increasingly require access to distributed and heterogeneous data and computational resources. The term *eScience* is often used to describe such activities and has been defined as "global collaboration in key areas of science, and the next generation of infrastructure that will enable it" [1] . Grid technologies have emerged as one of the key components of this infrastructure, by providing middleware support to facilitate access to services and resources. However, to assist collaboration it is also necessary to develop technologies to improve usability, to allow connections between people, ideas and data, and to facilitate the discovery and interpretation of knowledge generated by others (De Roure 2007). De Roure argues that to really bring eScience to the wider research community it is necessary to take into account the "social" aspects of research.

The PolicyGrid[2] project is a collaboration between computer scientists and social scientists, which aims to develop a range of software services and tools to support social science research. In particular, we are developing a software infrastructure that will assist researchers involved in aspects of evidence-based policy research. Evidence is used at various stages of policy making, from the design of new policies to the evaluation and review of existing policy. Support for management of evidence is therefore a significant issue, and there is considerable potential for technology to support such activities. Our experience has shown that the Social Semantic Web (Mikroyannidis 2007) has much to offer.

[1] John Taylor, Director General of UK Research Councils, 1999.

[2] http://www.policygrid.org

Supporting eSocial Science

The ability to capture details about the digital artefacts (questionnaires, interview transcripts, datasets) and activities (surveys, interviews, focus groups) that comprise an evidence-base is crucial. However, any solution must also provide information about the *context* of the data; this could include who collected the data, from whom, when and where and an account of the analytical/ interpretative process including who was involved and any assumptions that were made about the data. Contextual information is particularly useful in collaborative research where it ensures that every member of the team is aware of the various stages and decisions associated with the research project.

Simmham, Plale, and Gannon (2005) define data provenance as "one kind of metadata [which] pertains to the derivation history of a data product starting from its original sources". For our purposes, a provenance framework must also provide support for social and other forms of contextual information. We have developed several OWL ontologies, inspired by ideas from FOAF and the Open Provenance Model (Moreau et al. 2007). Used together, these ontologies allow us to capture relationships between members of a research team, their activities and artefacts; in addition, social connections to others are captured, e.g. the people who participated in an interview activity as the interviewer and the subject. The ontologies are as follows: *Utility* - describes concepts including person, project and organisation; *Resource* - describes the different types of resource used by social scientists (e.g. questionnaire, interview transcript); and *Task* - characterises research activities. While our initial focus has been to support creation of a rich provenance representation for social scientists, all three ontologies have been designed to be easily extensible, allowing new forms of resource or research activity to be added. For further details see: http://www.policygrid.org/ontologies/. The developing evidence base is itself a social construct, produced as a result of the activities of many researchers over time; our framework has thus been designed to allow users to incorporate new resources or activities into the evolving provenance de-

scription through such community action.

While the ontologies provide a means for us to capture the structure of the social provenance graph, they do not attempt to characterise domain concepts that might feature within resource or task descriptions. As social science concepts emerge from debate and are open to indefinite modification through debate, vocabularies tend to be imprecise and mutable - vocabularies tend to change over time to reflect shifts in understanding of social reality (Edwards, Aldridge, and Clarke 2006). Our solution, therefore, is to use ontologies to provide a "conceptual scaffold" for lightweight metadata. In other words, ontologies are used to capture the structure of the social graph, including resources, activities and project descriptions; folksonomies then provide the community-driven mechanism to support the development of domain vocabularies. We believe that this approach provides social scientists with a flexible and open-ended means of describing resources and activities, whilst at the same time providing a context for those assertions through more structured concepts. Permitted values for many of the datatype properties within the ontologies are of type 'string' and it is here that users may enter tag data; as users describe their resources, an underlying folksonomy is constructed which can be used to guide others towards popular tag choices. These folksonomies are kept separate because different values apply to different properties; the property `hasCountry` for example, has rather different tags associated with it than `hasSamplingMethod`. This enables us to generate contextualised tag clouds such as a cloud for a particular property, or for a particular project. Our approach has much in common with Gruber (2006), who has argued that technologies such as OWL and RDF should act as a "substrate for collective intelligence".

The use of such a representation creates a number of user-interface challenges: How will users create RDF when describing an element within the social provenance graph? How should presentation of folksonomies be integrated with tools for RDF creation and query construction? LIBER (Hielkema, Mellish, and Edwards 2008). is an integrated metadata browser, editor and query tool based on the use of natural language generation techniques. It has been designed to enable users with little or no previous experience of semantic technologies to interact with our hybrid metadata representation. The browser presents natural language representations of the provenance graph; the query module enables the user to create SPARQL queries which act upon the graph; and the editing module is used to create metadata. When the user is prompted to provide a value, tag clouds are used to present an overview of the tags popular amongst the user community in that particular context. A feedback text is also generated, so that the user can check whether what was added was what he really meant.

As a platform for our investigations we are developing *ourSpaces*, a virtual research environment (VRE) that allows users to upload, store and annotate scientific resources in a collaborative workspace (Reid and Edwards 2009).

Discussion

Our work to date has shown that there is considerable potential within eScience for the use of a mixed solution combining aspects of the Semantic Web and the Social Web. Our users were comfortable with the use of a formal representation to capture the structure of the evidence base, but desired a lightweight, community-driven approach to provide content. While we have explored some aspects of the Social Semantic Web to date, much still remains to be done. For example, we have yet to implement a solution to deal with the issues of trust and reputation. This clearly plays a vital role in the research process and our social provenance graph should provide a starting point for extensions to document the reputation of an individual or a resource. Extensions to the social provenance graph are also needed to incorporate spatial information and to provide better support for temporal context. Finally, a mechanism to allow the community of researchers to develop new resource and task descriptions is required; the LIBER interface already provides some facilities of this kind, but they are not designed to be driven by a group of users.

Acknowledgements

The authors acknowledge financial support provided by the UK Economic and Social Research Council (ESRC) under the eSocial Science programme; grant reference RES-149-25-1027.

References

De Roure, D. 2007. The New e-Science. IEEE e-Science Conference, Bangalore.

Edwards, P.; Aldridge, J.; and Clarke, K. 2006. A Tree Full of Leaves: Description Logic and Data Documentation. In *Proceedings of the Second International Conference on e-Social Science*.

Gruber, T. 2006. Where the Social Web meets the Semantic Web. In *Proceedings of the Fifth International Semantic Web Conference (ISWC 2006)*, volume 4273, 994. Springer Lecture Notes in Computer Science.

Hielkema, F.; Mellish, C.; and Edwards, P. 2008. Evaluating an Ontology-Driven WYSIWYM Interface. In *Proceedings of the Fifth International Conference on Natural Language Generation*.

Mikroyannidis, A. 2007. Toward a Social Semantic Web. *IEEE Computer* 40:113–115.

Moreau, L.; Freire, J.; Futrelle, J.; McGrath, R.; Myers, J.; and Paulson, P. 2007. The Open Provenance Model. Technical Report, ECS, University of Southampton, http://eprints.ecs.soton.ac.uk/14979/1/opm.pdf.

Reid, R., and Edwards, P. 2009. ourSpaces - A Social Semantic Web Environment for eScience. In *Papers from the AAAI Spring Symposium on Social Semantic Web: Where Web2.0 Meets Web3.0*.

Simmham, Y.; Plale, B.; and Gannon, D. 2005. A Survey of Data Provenance in eScience. *ACM SIGMOD Record* 34(3):31–36.