

A Multi-faceted Provenance Solution for Science on the Web

Edoardo Pignotti, Peter Edwards, and Richard Reid

School of Natural and Computing Sciences, University of Aberdeen
{e.pignotti,p.edwards,r.reid}@abdn.ac.uk

Abstract. To support the interface between scientific research and the wider public policy agenda it is essential to make the provenance of research processes and artefacts more transparent and subject to scrutiny. We outline the requirements for a multi-faceted approach to provenance and present a Web-based virtual research environment (ourSpaces) to demonstrate how research artefacts, projects, geographical locations and online communications can be linked in order to facilitate collaborative research.

Keywords: provenance, VRE, collaboration.

1 Introduction

The PolicyGrid project¹ is exploring how novel e-Science technologies can be used to support researchers; in particular, the provision of support for evidence-based policy research. De Roure [1] has argued that in order to assist collaboration among researchers it is necessary to go beyond basic e-Science infrastructure and to develop technologies to facilitate the discovery and interpretation of knowledge generated by others and to allow connections between people, places, organisations, ideas, and data. The Web has drastically improved this exchange. Scientists can now utilise social networking tools as a way to convey ideas much like a person may want to “Tweet” about his or her day. Similarly, virtual research environments enable users to share scientific resources in much the same way as Facebook might be used to share photos. However, provenance information is essential to enable researchers to assess the accuracy, timeliness, reliability, and trustworthiness of information available on the Web. To support science on the Web we require a representational framework for provenance which goes beyond simple metadata descriptions of artefacts and processes. Based upon interactions over a number of years with several research groups and communities, we have identified the following requirements for such a provenance fabric:

1. It should describe and uniquely identify a range of entities: artefacts (digital & physical); processes (services & human activities); people; organisational structures/membership; social networks.

¹ <http://www.policygrid.org>

2. It should situate entities in time and space.
3. It should incorporate online communication (e.g. instant messaging, blog entries, email) into the provenance record.
4. It should allow relationships (e.g. causal, social, organisational) to be defined between entities.
5. It should make explicit goals and constraints associated with processes and associated artefacts, in order to capture the ‘why?’ aspect of provenance.
6. It should facilitate reasoning about access control; documentation policies; completeness of the provenance record; trust and reputation.

While many of the existing provenance solutions [4,5] have focused on specific technologies to support narrow scientific domains, some recent research has focused on interoperability of provenance information across different systems. Most notably the Open Provenance Model was developed to address issues in managing provenance information in science, independent from technology and domain. The aim of OPM is to provide a technology-agnostic model supporting the digital representation of provenance describing any “thing” that is produced by a computer system (or not). OPM is based on three primary entities namely *Artefact*, *Process* and *Agent* and associated causal relationships namely *used*, *wasGeneratedBy*, *wasTriggeredBy* and *wasControlledBy*. OPM also defines a core set of rules that specify inferences that can be made on a provenance record (e.g. *wasDerivedFrom*, *wasTriggeredBy*).

In order to meet the much broader requirements of our provenance fabric an approach is required which integrates organisations, people, domains, technologies, systems and the physical and digital worlds. The Open Provenance Model has made an important step towards realising this vision by allowing the provenance of individual systems to be expressed in a coherent fashion. However, in order to realise a true *provenance fabric* we need to go beyond just descriptions of agents, artefacts and processes.

2 ourSpaces - Supporting Provenance on the Web

ourSpaces (www.ourspaces.net) [6] has been developed as a working realisation of various elements of the *provenance fabric*. Built using a number of Semantic Web technologies [2], users are able to perform various activities such as uploading and describing digital artefacts, maintaining personal profiles, initiating instant messaging (IM) conversations, creating blog posts and calendar events, tagging and commenting on other resources and forming groups (in the form of projects) with other researchers. At the heart of ourSpaces is an OWL representation of the Open Provenance Model [3], used to express metadata regarding digital artefacts and processes. However, additional ontologies (including FOAF SIOC, GeoNames)², are used to capture information regarding people, organisations, social networks, geographical context and online communications.

Through a system demonstration, we will present the following: (a) Resource management including resource upload and description; (b) How OPM can be enriched by social context; (c) Project creation and management features; (d) Use of maps; (e) Multiple approaches to metadata querying and browsing.

² <http://xmlns.com/foaf/spec/>, <http://rdfs.org/sioc/spec/>, <http://www.geonames.org/ontology/>

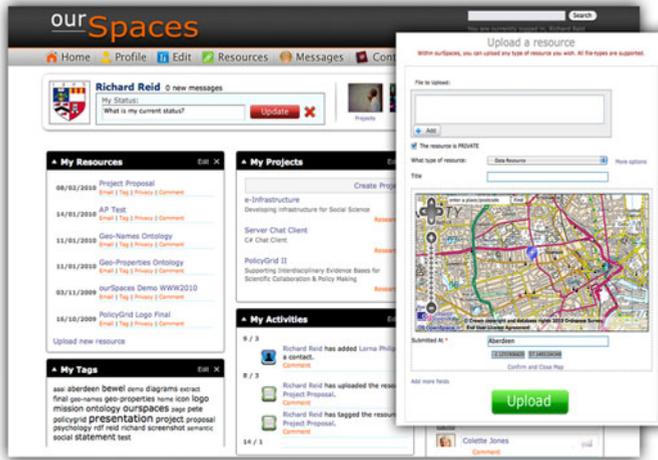


Fig. 1. ourSpaces VRE home page

References

1. De Roure, D.: The new e-Science presentation (2008), <http://www.slideshare.net/dder/the-new-science-bangalore-edution>
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American* 284(5), 34–43 (2001)
3. Moreau, L., Clifford, B., Freire, J., Gil, Y., Groth, P., Futrelle, J., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Simmhan, Y., Stephan, E., Van den Bussche, J.: The Open Provenance Model - Core Specification (v1.1). In: *Future Generation Computer Systems* (2009) (submitted)
4. Simmhan, Y., Pale, B., Gannon, D.: A Survey of Data Provenance in e-Science. *SIGMOD Record* 34(3), 31–36 (2005)
5. Bose, R., Foster, I., Moreau, L.: Report on the International Provenance and Annotation Workshop (IPAW 2006). *Sigmod Records* (September 2006)
6. Reid, R., Pignotti, E., Edwards, P., Laing, A.: ourSpaces: Linking Provenance and Social Data in a Virtual Research Environment. In: *Proceedings of the World Wide Conference* (2010)