eSocial Science and Evidence-Based Policy Assessment: Challenges and Solutions

P Edwards[1], J Farrington[2], C Mellish[1], L Philip[2], A H Chorley[1], F Hielkema[1], E Pignotti[1], R Reid[1], J G Polhill[3] & N M Gotts[3]

[1] School of Natural & Computing Sciences, University of Aberdeen

[2] School of Geosciences, University of Aberdeen

[3] The Macaulay Institute, Craigiebuckler, Aberdeen.

{p.edwards, j.farrington, c.mellish, l.philip, a.h.chorley, f.hielkema, e.pignotti, r.reid} @abdn.ac.uk

{g.polhill, n.gotts}@macaulay.ac.uk

Abstract

The PolicyGrid project is exploring the role of Semantic Grid technologies to support *eSocial Science*, with particular emphasis on tools to facilitate evidence-based policy making. In this paper we discuss the challenges associated with construction of a provenance framework to support evidence-based policy assessment. We then discuss ourSpaces, a virtual research environment for *eSocial Science* that uses the Web2.0 paradigm as well as Semantic Grid technologies, and which provides researchers with facilities for management of digital resources. We describe a novel approach for the creation and presentation of metadata, before exploring how policy-oriented use of social simulation can be integrated with our approach.

Introduction

The UK government has defined policy making as “the process by which governments translate their political vision into programmes and actions to deliver ‘outcomes’ – desired changes in the real world” (UK Cabinet Office, 1999). An important element of twenty-first century policy making, at all levels, is a move away from opinion-based decision making to decision making that is grounded in evidence and subject to rigorous evaluation. Evidence is used at various stages of policy making, from the design of new policies to the evaluation and review of existing policy. An evidence base supports transparency and accountability in the policy decision making process. The evidence used in policy making is derived from a range of sources, and academic researchers from across the social sciences make a notable contribution to the evidence base for many areas of public policy: they produce ‘new’ knowledge by undertaking primary research or conducting further analysis of existing data, they produce reviews and critical appraisals of existing research and are often commissioned to review and evaluate specific policies.

The PolicyGrid project[1], a collaborative venture between computer scientists and social scientists interested in policy-related research, is investigating how social scientists can be supported in their policy assessment activities through the application of *eSocial Science* tools and techniques, in particular the application of Semantic Grid (De Roure *et al*. 2005) technologies. The term *eScience* has come to refer to the use of advanced computing approaches to support scientific research, while *eSocial Science* reflects the extension of these ideas to support social science. The various activities being undertaken in the PolicyGrid project and their application to evidence based policy assessment (hereafter EBPA) will be described in this paper. The remainder of this introductory section of the paper will introduce the concept of EBPA, describe what some of the requirements for evidence to be used in policy making are and, in so doing, identify some of the challenges for *eSocial Science* tools.

Evidence Based Policy Assessment is associated with all aspects of policy-making and is required, in the UK, of national and local government. Best practice principles for *ex ante* **appraisal** of policy and *ex post* **evaluation** of policy are set out in the UK Government's Green Book (HM Treasury, 2003). Our use of the term **assessment** embraces both appraisal and evaluation. The Green Book's primary concern is economic assessment whereby the costs and benefits of policy options are estimated, a process which includes formal consideration of 'non-market' policy impacts. Economic assessment requires that:

- Reports should provide *sufficient evidence* to support their conclusions and recommendations;
- There should be an easy *audit trail* to allow decision makers to understand the assumptions underlying conclusions and recommendations;
- There should be sufficient information to support any *later evaluation*.

(HM Treasury, 2003, para 2:14, our emphasis)

Two generic types of policy evaluation are recognised in the UK Government's performance management system (UK Cabinet Office Strategy Unit, 2003-5 – The Magenta Book – p4; Davies, 2004 p19): *formative* or *process evaluation* (how, why and under what conditions does a policy intervention 'work' or 'fail to work'?) and *summative* or *impact evaluation* (what impact does a policy intervention have?). These types of policy evaluation may overlap and interact and, importantly, recognise that policy evaluation can benefit from using a wide range of evidence, not just (economic) cost-benefit analysis. Useful evidence for policy evaluation can be the data and/or the conclusions from previous research, and new material collected to supplement existing evidence, to fill in gaps in existing evidence, or to create completely new evidence. Policy makers are encouraged to critically appraise the evidence they work with. For example, they should ask:

- Was the evidence sifted and graded for quality?
- Were the inclusions and exclusion criteria explicit?
- Is the evidence easy to understand?
- Has the strength of the evidence been assessed?

---

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

The nature of research questions, the primary data collection methods used, secondary source materials and the various analytical and interpretative techniques employed set limits on the reliability, replicability, generalisability and overall validity of conclusions that can be drawn from a piece of research that academic researchers may produce for use in policy assessment. To illustrate some of these points further, consider a research study which has the goal of identifying the socio-economic, cultural and political changes which could bring about a reduction in carbon-intensive energy demand from the household sector; and the policy instruments which could be employed to facilitate such changes, The study examines the similarities, differences and interactions between cities and their functionally associated rural areas, and a number of emerging initiatives, both grassroots and top-down, from an interdisciplinary social science viewpoint. The project involves both qualitative and quantitative fieldwork, and agent-based simulation modelling. It addresses policy development from the perspective of multiple disciplines, working closely with associated policy-makers and stakeholders. Outputs are expected to be a sequence of policy briefs.

What forms of computational support would assist the researchers involved in the work outlined above? Perhaps the most significant of these is the issue of management of evidence. This refers to both the digital artefacts associated with the project, and their associated *provenance*[2]. Simmham *et al.* (2005, p31) defined data provenance as "one kind of metadata [which] pertains to the derivation history of a data product starting from its original sources". A number of computational provenance models have previously been reported in the *e*Science literature, in a variety of areas including life sciences (e.g. myGrid - Stevens *et al*, 2003), chemistry (e.g. CombeChem - Taylor *et al.* 2006) and High Energy Physics (Branco & Moreau, 2006). What are the particular issues that arise in developing a provenance model for evidence-based policy research? The ability to capture details about the resources (questionnaires, datasets, simulation experiments) and tasks (surveys, interviews, model runs) that comprise an evidence-base is crucial if one wishes to satisfy the requirements set out in the Green Book and the Magenta Book. Additionally, such a framework must provide support for the creation of audit trails to allow evidence to be assessed/validated, and must ultimately include mechanisms for creating policy arguments and options from the evidence base. Such a provenance model would facilitate better communication and collaboration between members of a project team, while at the same time enhancing the 'interface' between the *producers* of evidence (e.g. academic researchers) and the *users* of that evidence (e.g. policy makers).

The existence of a framework for management of evidence of the kind suggested above raises significant challenges in terms of the design of appropriate user-interfaces and tools for researchers. How do we support users who wish to create, query and browse provenance information without them needing to understand the intricacies of specific metadata formats? Any solution should acknowledge (at least in part) the conventions adopted by existing Web-based archives for social science research, such as Intute and the UK Data Archive. *Intute: social sciences*[3] is an online service that provides access to Web resources for education and research, that have been evaluated and selected by subject specialists. Users can browse the database by topic (offering several thesauri for use), search by keyword (and in advanced mode search by resource type, topic and keyword), and suggest new additions. Intute also contains additional services such as a blog, events, timelines, newsround, etc. The UK Data

2 A variety of terms are used in the computer science literature to describe the origins, analysis, transformations, assumptions and conclusions drawn from data, including *lineage*, *pedigree,* and *genealogy*. In this paper, and the wider PolicyGrid project, we are using the term *provenance*.

3 http://www.intute.ac.uk/socialsciences/

Archive[4] offers searches on different criteria (title, geographic coverage, time period, creator, etc) with regular expressions/key words. Users can deposit data by submitting a form, which is then processed by a human agent who interprets the information on the form into a description in a standard format. The existence of an evolving evidence-base should become a powerful tool to support interactions between project co-workers and external stakeholders. How do we develop collaboration tools which support (amongst others): resource discovery, sharing and informed re-use of evidence, tracking of the development of evidence trails and ultimately policy arguments, participation by representatives of multiple disciplines?

The exemplar study above highlights another important challenge if we are to produce support for EBPA activities: How best to integrate a range of evidence types and methodologies, spanning not just qualitative and quantitative research methods, but also the emerging field of social simulation? This will necessitate development of appropriate resource and task descriptions within the provenance model, and in addition, some mechanism for encoding simulation experiment specifications in a form suitable for their re-use.

In the following sections, after briefly introducing the main concepts of the Semantic Grid, we describe how the PolicyGrid project interprets these challenges in terms of technical requirements, and the overall form of the solutions that the project proposes.

The Semantic Grid

Grid computing (Foster *et al.* 2001) plays an important role in delivering many eScience applications. Such technologies facilitate access to computing hardware, storage and application services, and at the same time ensure effective management of quality of service issues, including security and authentication. While Grid technologies provide much of the infrastructure backbone required to deliver *eScience* solutions, there is still a need for enhancements to improve ease-of-use, automate service/data discovery and integration, deliver collaboration tools, and so on. The Semantic Grid was conceived as "an extension of the current Grid in which information and services are given well-defined meaning, better enabling computers and people to work in cooperation." In others words, by describing resources and services available on the Grid through the use of a semantic data model, new forms of software could be created to meet the needs of the end-user (scientific) community. In earlier work (Pignotti *et al.* 2005, Polhill *et al.* 2007) we argued that such an approach has significant potential within *eSocial Science* in helping deliver the vision of a more 'human-centred' Grid (Anderson, 2003) which facilitates tasks such as collaboration, shared experimentation, and annotation of resources. Furthermore, we demonstrated how Semantic Grid techniques could be employed to capture qualitative scientific arguments, supported by a mix of quantitative and qualitative data and results.

The adoption of metadata and ontologies to describe resources is central to the Semantic Grid. Metadata are (computer-readable) data about data - they are used to describe resources (such as documents, datasets or web pages). As such they facilitate the understanding, use and management of data. A frequently used medium to present metadata is RDF[5] (Resource Description Framework), a metadata model used to make *predicate(subject, object)* statements about resources (known as RDF-triples), e.g.:

---

4 http://www.data-archive.ac.uk/

5 http://www.w3.org/RDF/

```
<j.0:Person rdf:ID="F1">
      <j.0:Name>John Smith</j.0:Name>
</j.0:Person>
```

This RDF-fragment states that there is a `Person` whose `Name` is ‘John Smith’. `Person` is a class in an ontology (defined by the namespace ‘j.0’); `Name` is a property of that class. ‘John Smith’ is an individual of this class. Ontologies are thus data models that represent a set of concepts (e.g. `Person`) within a domain and the relationships between those concepts and/or datatypes (such as strings). Such ontologies are commonly described using the Web Ontology Language (OWL[6]), a vocabulary for providing descriptions about classes and properties of RDF resources. The query language SPARQL[7] supports complex RDF queries containing optional patterns, disjunctions, etc.

From the beginning of our interactions, social scientists expressed a fear of ‘being trapped in the ontology' due to the contested nature of many concepts within the social sciences. Others (Edwards *et al*. 2006) have noted that as social science concepts emerge from debate and are open to indefinite modification through debate, vocabularies also tend to be imprecise (e.g. there is no precise definition of ‘anti-social behaviour') and mutable (vocabularies tend to change over time to reflect shifts in understanding of social reality). It has recently been argued (Gruber, 2006) that technologies such as OWL and RDF should act as a ‘substrate for collective intelligence' - in other words that the community-driven approach to creation and management of content now increasingly popular on the Web[8] should be integrated with these technologies. Within PolicyGrid we are adopting an approach which supports dynamic, community-driven evolution of metadata (Guy & Tonkin, 2006) within a framework provided by a series of OWL ontologies. Our approach is similar in form to Gruber's suggestion of integrating unstructured user contributions (tags) into a structured framework (ontology). We believe that it provides social scientists with a flexible and open-ended means of describing resources, whilst at the same time providing a context for those assertions through more structured concepts. Details of the ontologies are given in the next section on provenance. Permitted values for many of the datatype properties within the ontologies are of type ‘string' and it is here that users are permitted to enter tags; 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; a property *`HasCountry`* has rather different tags associated with it than *`HasSamplingMethod`*. When a user selects a property, its folksonomy is presented in the form of a ‘tag cloud’ (see Figure 6 - middle), which contains the most frequently used tags for this property; the user is able to select a term from the emerging community vocabulary or is free to use their own.

Figure 1 provides a schematic overview of the PolicyGrid software architecture. Several aspects of this diagram are discussed elsewhere in this paper, but certain key components will be introduced here. To manage the digital artefacts utilised during the research process, we use a digital object repository; this is based upon the highly scalable Fedora[9] open-source technology. Digital resources stored within Fedora are associated with metadata managed by

---

6 http://www.w3.org/TR/owl-features/

7 http://www.w3.org/TR/rdf-sparql-query/

8 Often referred to as *Web2.0*.

9 http://www.fedora-commons.org/

an instance of the Sesame[10] RDF triple-store framework; this provides support for RDF for storage, inferencing and querying. Together with a relational database (used to manage tag data) these repositories provide the platform for all other PolicyGrid software and services. The box labelled ‘Grid services’ appearing in Figure 1 refers to a set of social simulation models (and other computational services) that have been deployed onto the Grid, and which can be invoked through an appropriate user-interface. The ontologies layer in Figure 1 provides the conceptual framework within which the various software services, tools and user-interfaces interoperate. In the next section we describe three of the ontologies which define our provenance model.

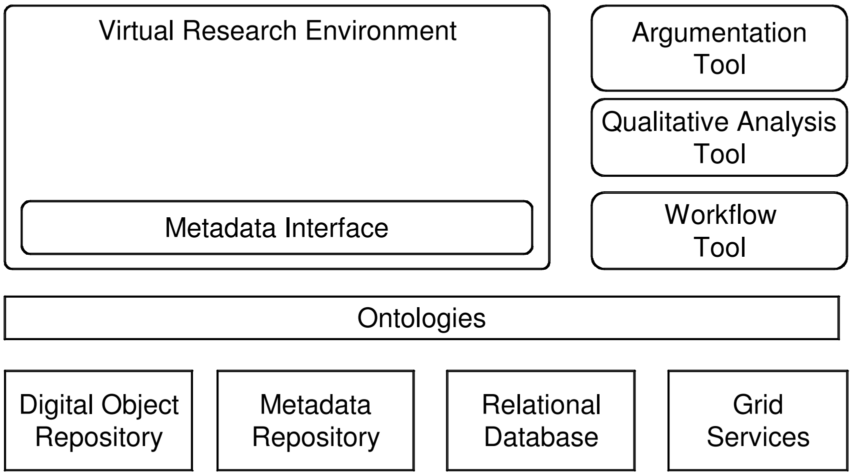


Figure 1: PolicyGrid software architecture.

Provenance

Provenance applied to *eSocial Science* should provide information about how data were created and provide information about the context of the data. Such context could include, for example, an account of the characteristics of secondary data and why it was used; an account of data collection methods, including who collected the data, from whom, when and where; an account of the analytical/interpretative process including who was involved and any assumptions that were made about the data; and a record of what conclusions were drawn, how data were written up and what dissemination has taken place. In EBPA, context is particularly important because of the need to evaluate the quality (and thus the reliability) of data, the robustness of analysis, the generalisability and the validity of findings, conclusions and/or recommendations. Context also helps when data are reused, particularly when that analysis is not performed by those who created the data and or conducted the initial analysis.

Parallels may be drawn between some research activities in the natural and social sciences, notably the application of statistical and modelling techniques to large quantitative data sets. However, qualitative research is arguably as important to the social science community as quantitative research, the policy community has become increasingly receptive to qualitative data in the UK and a ‘mixed methods’ approach is frequently adopted in policy-related research. A provenance model for EBPA in the social sciences should not only draw upon lessons learnt from the development of provenance architectures for *e*Science, it should also be capable of handling quantitative, qualitative and mixed methods approaches.

10 http://www.openrdf.org/

A provenance system that tracks evidence-conclusion chains in any type of social science research potentially raises philosophical, ethical and legal issues. Social scientists working in the policy field must not only conduct their research in an ethically appropriate manner, they should be able to produce, through transparent processes, a robust system for the analysis of data which forms evidence of policy 'success/failure'. Legal issues relating to research materials, for example, provisions under data protection, copyright, intellectual property rights and freedom of information legislation (*c.f.* Bishop, 2005; Parry & Mauthner, 2004; 2005) also apply. For some qualitative researchers each data collection event is unique and unreplicable; the narrative content is singular and not capable of analysis by anyone other than the researcher who collected the data. Others view data as being capable of analysis by others, and also subscribe to *extension through generalisation.* Can, or should, both positions be captured by a provenance model? Further, epistemologically informed positions surrounding the secondary use of qualitative data (Mauthner *et al.* 1998; Fielding, 2004) can affect attitudes towards and the practicalities of archiving data and allowing other researchers access to source materials, regardless of whether those researchers be from the policy (government), academic or consultancy communities. Notwithstanding these concerns, in the context of UK government practice there is a case for developing a provenance model to support the range of data used in social science research.

The challenge, then, is to track the evidence-gathering and evidence-analysis/conclusion process for qualitative and quantitative research in the social sciences, consider the issue of data re-use for EBPA *and* be sensitive to epistemological concerns which express an uneasiness about reusing qualitative data in particular. The PolicyGrid schema for an *e*-based provenance system was first described in Chorley *et al.* (2007). The model defines provenance as a description of how a resource came about, capturing information about the activities performed to create that resource. For example, a quantitative data set is associated with the questionnaire that elicited the data. Contextual information, such as the date that a questionnaire was administered and the response rate, can also be recorded. Formally recording such information allows questions such as 'how was the evidence derived' to be answered, providing evidence that would allow a third party to ascertain the robustness, or *truthfulness* of the data collection process. Information about data analysis and interpretative processes could also be captured because the provenance model allows the researcher to record what they do. For example, if working with interview transcripts a description of the process by which thematic codes were developed could be recorded (in effect, these would be the memos or other analytical notes routinely made during qualitative analysis but not routinely recorded beyond the researcher's own informal notes). If these interview transcripts were re-used a second layer of analytical information could be added if the second researcher drew additional, or different, conclusions from the data. How the provenance information is collated and recorded by the user (researcher) is described in the next section of this paper.

To develop our provenance model, we began by analysing the descriptions (available online) of the various datasets held in the UK Data Archive. The descriptions were used to identify the concepts necessary to characterise a social science resource. The result of this initial exercise was an OWL ontology which modelled objects such as `Document` and `Person`, but avoided abstract domain concepts such as `Rural_accessibility` and `Poverty`. More information about this initial ontology can be found in Chorley *et al.* (2007). Evaluation of this ontology (Hielkema *et al.* 2007) concluded that it did not match the subject's model of resource descriptions closely enough to enable them to easily create satisfactory descriptions. We needed a revised model of provenance that was both more comprehensive and contained less ambiguous or duplicate properties, and which enabled users to describe resources in a manner with which they were comfortable.

We thus set out to develop a revised set of ontologies that could be used by all tools and services that PolicyGrid is developing, while at the same time retaining a degree of compatibility with the UK Data Archive schema. In a series of interviews with social science researchers, we listed the possible resource types that might be deposited, and the information that should be recorded about them including the details of the method(s) by which they were generated/revised/analysed.

The PolicyGrid provenance model comprises three different ontologies: *Utility*[11], *Resource*[12] and *Task*[13]. The *Utility* ontology is used to describe utility items such as projects and persons, while the *Resource* ontology describes resources, including information such as title, author, access rights and dates of creation and/or publication. The *Task* ontology is used to describe, for example, a specific data collection method, including details such as the time and location of an interview and the format the interview took; dissemination tasks, such as conference presentations, the final report and academic papers; or the preparation of a literature review. The three ontologies are separate but compatible, so that the description of a resource and the process through which it was created uses elements from them all. Using the example of the description of an interview transcript, Figure 2 depicts parts of all three ontologies and how they interoperate; the notation used in the figure is as follows: dotted ovals represent a resource, solid ovals represent tasks, and rectangles are properties. An interview transcript is a resource that is associated with a particular interview task, which could itself have produced other resources, e.g. a recording of the interview and interviewer's written notes. These tasks and resources all have properties such as `DateOfDeposit` or `DateOfInterview`. They will also have properties describing roles, such as the interview transcript `was_deposited_by` a person and an interview task `was_conducted_by` a person. `Person` and `Address` are concepts from the *Utility* ontology, and are represented in Figure 2 by an oval with a double outline.

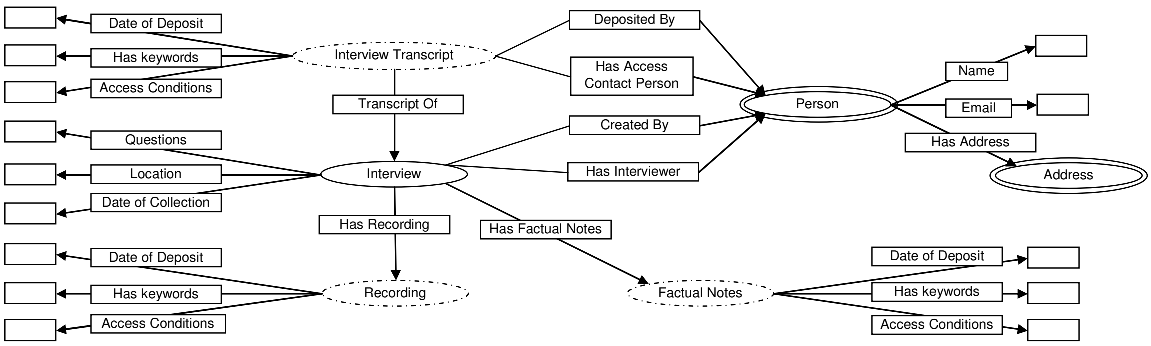


Figure 2: Using the *Resource*, *Task* and *Utility* ontologies to describe an interview transcript.

In order to assess this approach to provenance a series of case study discussions with social scientists were conducted. In each case, interviews were used to elicit full descriptions of a previous (completed) research project, and these descriptions then analysed to determine if the ontologies were capable of capturing the provenance record. During each interview a diagrammatic representation of the project and the relationships between resources and tasks was created on a whiteboard. An example of such a record is reproduced below in Figure 3;

11 http://www.policygrid.org/utility.owl

12 http://www.policygrid.org/resource.owl

13 http://www.policygrid.org/task.owl

this characterises a project which developed a novel approach to contingent valuation - the CV Market Stall (Philip and Macmillan, 2005).

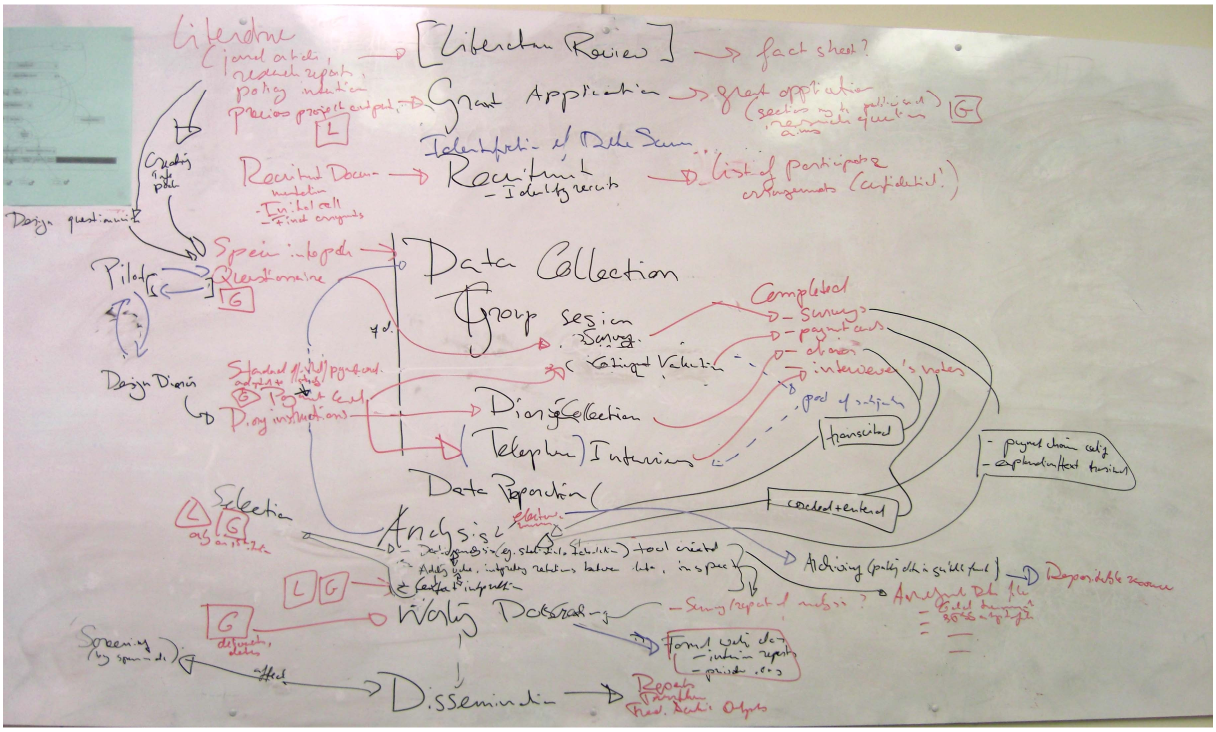


Figure 3: Whiteboard image of description of the CV market stall case study.

Publications and other written documents associated with the case study projects were reviewed after each interview to ensure that all aspects of the project were represented and, along with the interview notes, were used to prepare a diagrammatic representation of the project which was used to elicit feedback. Figure 4 reproduces an extract from such a diagrammatic representation of the CV Market Stall project. Tasks are depicted as boxes (labelled with the tasks details) and resources are depicted as circles (labelled with the resource type).

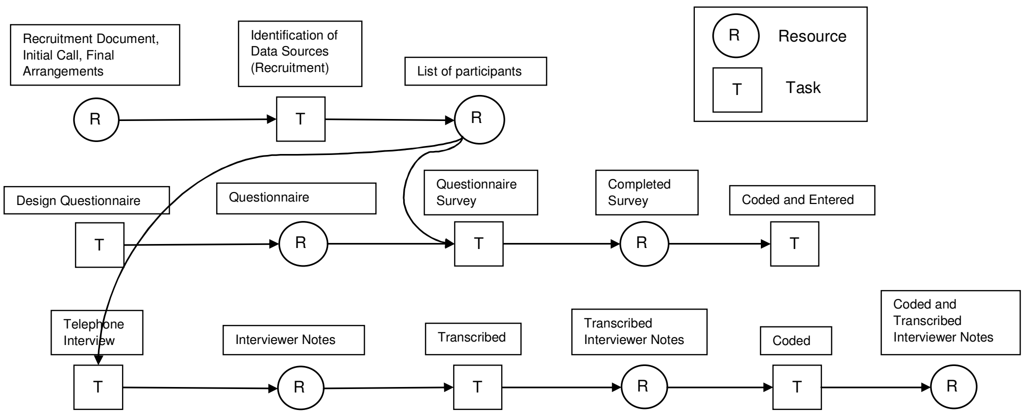


Figure 4: Provenance extract from the CV market stall case study.

Figure 4 demonstrates how various tasks and their associated resources can be captured using the PolicyGrid provenance model. The top row of the diagram shows how participants were recruited. The second row illustrates the various stages associated with the background questionnaire and the third row records the various stages associated with the telephone interviews. The findings from this and the other case studies confirmed that the ontologies could capture the metadata required to document the provenance record and thus provide an audit trail for research activities. In the next section we describe user-interface tools that have been designed to allow social scientists to utilise the provenance model to enhance their research activities.

User-interface Issues

As outlined earlier, there are many challenges to be addressed if we are to build usable *eSocial Science* tools to support those conducting evidence-based policy research. Managing the resources associated with an EBPA activity can be an onerous task, especially when that activity involves a large (possibly multi-disciplinary) effort. While the provenance model described above forms a part of the solution, appropriate user-interface tools are also essential. Semantic Grid technologies are powerful but complex, yet user interfaces must be designed to support researchers to deliver the services they require without them becoming expert in, or even having to be aware of, the underlying infrastructure.

A number of existing Web-based user-interfaces or portals aim to facilitate access to eScience services; these virtual research environments (VREs) include myExperiment (De Roure *et al.* 2007) and SciSpace[14]. myExperiment is a collaborative environment which supports the publishing of experimental workflows and other digital objects. It allows scientists to share, re-use and re-purpose workflows by supporting many of the features seen in social networking sites, including tagging of resources and commentaries. SciSpace, on the other hand, has been designed as a social networking service for scientists, with tools such as blogs, wikis, and tagging. It also ensures that such networks can remain private, allowing small groups of researchers to develop/share ideas. Inspired by these and other developments, we are currently developing a VRE to support social scientists conducting evidence-based policy research; this environment (*ourSpaces*) owes much to social networking sites such as MySpace and Facebook (see Figure 5). Support is provided in the prototype version for the following: social networking (messaging, collaboration invites, etc.); resource management (upload, search, annotation); creation of project 'spaces' (allowing project teams to manage membership, aggregate digital artefacts, organise activities according to project stages); privacy controls; publishing of blogs and wikis; execution/monitoring of simulation workflows. Each of these activities is enabled by the rich and pervasive metadata infrastructure described earlier.

ourSpaces enables users to upload and describe resources themselves, rather than through a third party. How can we elicit metadata from, and present metadata to, social scientists? RDF-triples are difficult to read and write: even experts make mistakes producing them with a text editor. Tools that use graphical representations (e.g. CREAM: Handschuh *et al.* 2001) are more effective, but for users unused to complex graphical presentations or ontologies they can be difficult to interpret. Petre (1995) argues that graphical readership is an acquired skill. She describes experiments into reading comprehension of various graphical and textual representations which showed that graphical representations are significantly slower than text. Novices in particular suffered from mis-readings and confusion. We argue that for many users, viewing natural language descriptions of metadata would most likely be both faster and less error-prone than viewing a graphical representation. Finding and describing resources are tasks peripheral to research, and not all researchers who need to do this are experts in reading and writing metadata.

---

14 http://www.scispace.net/

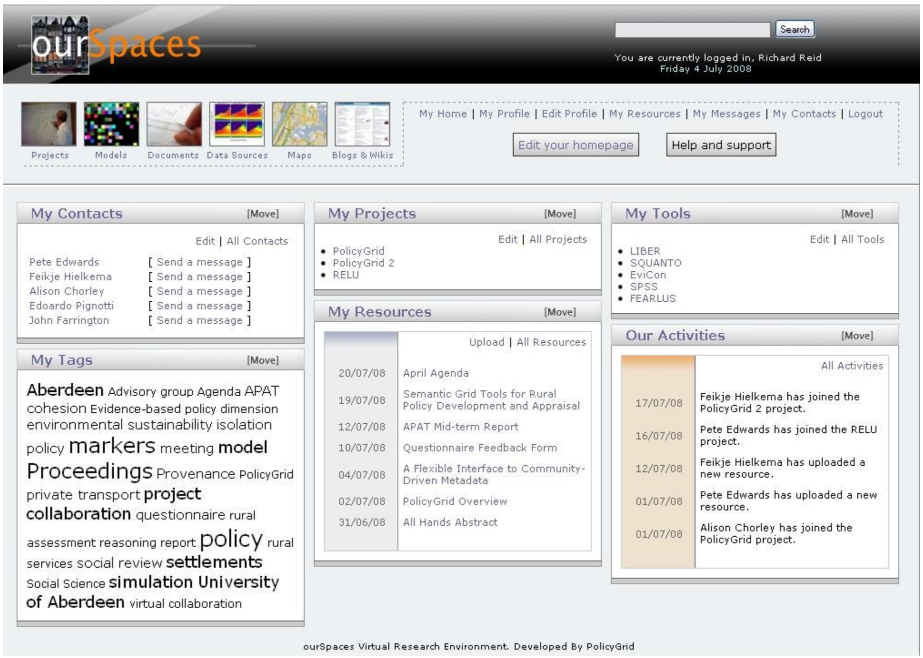


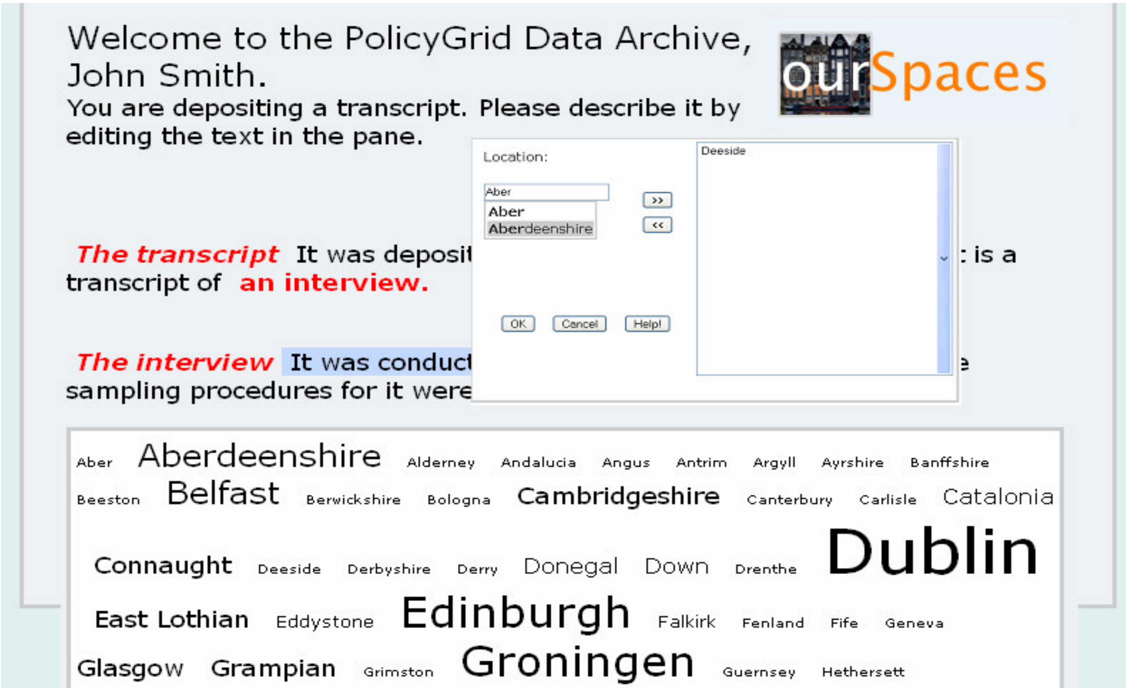
Figure 5: The ourSpaces virtual research environment.

Existing natural language approaches for providing access to databases include menu-techniques in which the user constructs an acceptable sentence by selecting words from a series of menus (Bernstein & Kaufmann, 2006), and controlled languages (Schwitter & Tillbrook, 2004). Such techniques severely restrict what can be said, limiting the language and often making it stilted, so that there is a small learning curve before the user knows which structures are permitted. However, some limitation on what can be expressed is inevitable, because a system that can parse every utterance the user creates is outside the scope of the current state-of-the-art.

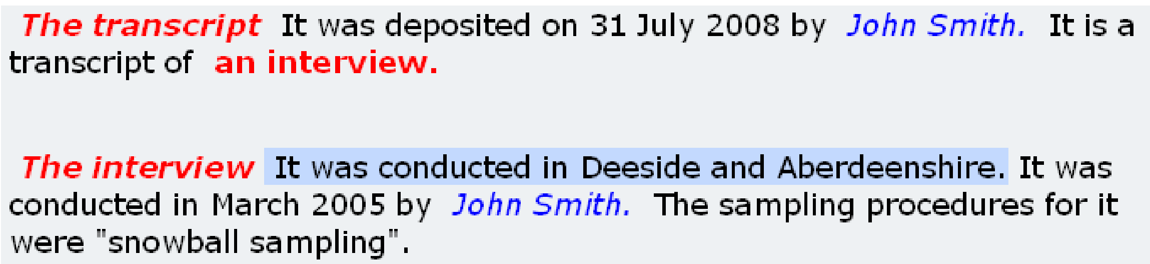
In order to maintain full expressivity and maximise usability, we have elected to use WYSIWYM (What You See Is What You Meant; Power *et al*. 1998). This is a natural language *generation* approach developed for providing access to databases or other knowledge structures, where the system generates a feedback text for the user that is based on a semantic representation. This representation is edited directly by the user by manipulating the feedback text. As the text is generated by the system and does not have to be parsed, we do not have to restrict what can be said, so the language retains its expressivity and the user does not need to learn what is acceptable input. LIBER[15] (Language Interface for Browsing and Editing RDF), a metadata interface for creating RDF (editing), SPARQL (querying) and viewing existing metadata (browsing) is shown in Figure 6.

15 http://roc.csd.abdn.ac.uk:8091/liber/

Top: Selecting a menu item from the 'interview' anchor.



Middle: Providing 'location', prompted by the tag cloud.



Bottom: The updated feedback text.

Figure 6: Creating metadata using the LIBER metadata interface.

We have evaluated LIBER's usability and usefulness through a series of qualitative and quantitative experiments (Hielkema *et al*. 2007; 2008). The most recent experiment was conducted during the Fourth International eSocial Science Conference, and involved eight subjects, most of whom came from a social science background and six of whom had not used ontologies before. After viewing a four-minute introduction video, they were handed a resource description and asked to find the resource in the archive using LIBER's querying and browsing components. Most subjects succeeded in this task despite their unfamiliarity with LIBER, although the experiment emphasized that accessing metadata is a much more complex task than key-word searching, and that for simple queries the current interface is overly complex. An earlier evaluation study of the metadata creation component showed that subjects did find LIBER useful for the more complex task of creating descriptions.

Until now, all evaluation involving LIBER has been with it operating as a stand-alone web service. LIBER is currently being integrated fully into *ourSpaces*, so that information about the user's current activities and other resources contained in *ourSpaces* can be automatically added to the user's descriptions and queries. This should reduce the burden on the user, while improving the quality and completeness of the metadata descriptions. Our aim is to integrate the existing support within *ourSpaces* for simple key word searches through tag clouds, with the much richer querying and metadata creation facilities provided by LIBER.

Integrating Social Simulation

Social simulation has been proposed as a 'third way' for social science, both formal and descriptive in its representations of reality (Moss and Edmonds, 2005). It focuses on explicit representation of individuals and their actions, and so has theoretical roots in complex adaptive systems (Holland and Miller, 1991). Agent-based social simulation (ABSS) modelling, in which decision-makers (individuals, firms, states) are explicitly represented within the simulation model, has considerable potential use in policy-related applications. Nevertheless, actual policy-related uses have been relatively infrequent. There are a number of key challenges concerning how ABSS models are used, which need to be overcome before this can be expected to change.

Calibration, validation, and subsequent policy-oriented use of agent-based models all require the model to be applied to data; often, this will be data of several different types, probably from different sources – and may include both quantitative and qualitative social science data. With respect to all these stages, a considerable number of simulation runs will be required, along with extensive post-processing using statistical and graphical software, and probably pre-processing of data as well. It is therefore vital to keep track of the simulation runs performed, the inputs and outputs, and the purpose of each set of runs. This should be done in a way that makes it easy for either the original investigators, other researchers, or policymakers to check and understand the investigative process. Moreover, the research process is very unlikely to be a linear one: results of simulation runs at any stage may suggest problems with the data or model, leading to revisions, or additional forms of post-simulation statistical and graphical processing. Since social simulation modelling is often only a part of a larger project including empirical social science and theoretical development, the complexity of the relationships between the simulation model and its scientific and policy context can be enormous.

With PolicyGrid we are attempting to address some of these challenges through the adoption of scientific workflow technologies (Pennington, 2007); these provide a user-friendly environment for scientists to create and execute experiments from a pool of available data and computational services. However, one weakness of current workflow languages such as MoML[16] , BPEL[17] and Scufl[18] is that while they use representations which make the steps of the experimental method explicit, they do not provide support for capturing the information about goals and contextual conditions vital in understanding an experiment. We are thus exploring how such workflow languages can be extended, by capturing a higher-level description of the experimental process: the *scientist's intent*. To illustrate, Figure 7 presents a workflow created using the Kepler [19] tool and used for the calibration and validation of a land-use model (Polhill *et al.* 2001). The workflow is itself embedded into a wider research activity, involving use of quantitative data to establish values for model parameters, and qualitative data drawn from semi-structured interviews to influence decision algorithms within the model. The goal and constraint shown on the right hand side of Figure 7 cannot be conveniently embedded in a workflow description. In addition to ensuring that there is support within our provenance model for the resources and tasks associated with simulation experiments, we have developed a rule-based framework (Pignotti *et al*. 2008) for capturing scientist's intent. The rules, written in SWRL[20] (Semantic Web Rule Language), act upon metadata generated from workflow activities (e.g. inputs, outputs, execution). Details of the intent are kept separate from the operational workflow (itself a digital artefact), to avoid embedding and thus obscuring, important provenance information within the workflow representation.

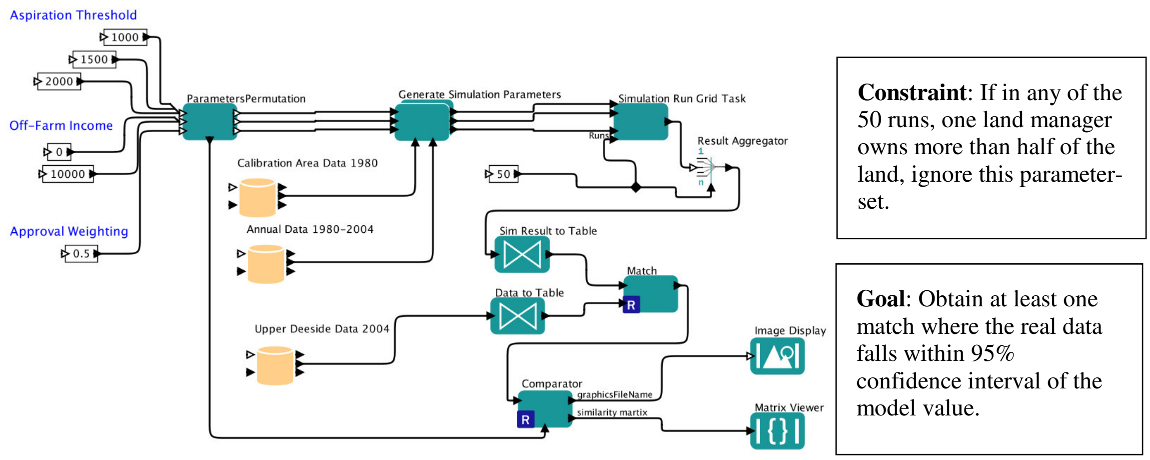


Figure 7: An example workflow with associated intent.

The constraint in Figure 7 (top-right) aims to save computational resources by terminating the exploration of a parameter-set that does not add value to the results. The constraint: 'if in any of the 50 runs, one land manager owns more than half of the land, ignore this parameter-set' can be represented by combining rules with the metadata available from the simulation service. The main challenges we faced were to represent scientist's intent so that:

- It is meaningful to the researcher, e.g. providing information about the context of an experiment so that the results can be interpreted;

16 http://ptolemy.eecs.berkeley.edu/projects/summaries/00/moml.html

17 http://www.ibm.com/developerworks/library/ws-bpel/

18 http://www.cs.man.ac.uk/~witherd5/taverna-site/scufl/index.html

19 http://kepler-project.org/

20 http://www.w3.org/Submission/SWRL/

- It can be reasoned about by a software application, e.g. to use the intent information to control, monitor or annotate workflow execution;

- It can be re-used (the same intent may apply to different workflows);

- It can be used to further enhance the provenance record (documenting the process that led to some result).

Discussion

We have developed a provenance model able to support descriptions of a range of social science resource types, and methods. It satisfies the requirements set out in the UK Government's Green Book, by facilitating the creation of an audit trail, and making explicit the context within which conclusions were drawn. At present, our model only captures those stages of the EBPA process concerned with activities such as planning, data collection, analysis and dissemination of results; it does not yet support the creation of (alternative) policy arguments. We are therefore beginning to turn our attention to the issue of support for construction of policy arguments using the evidence represented using the provenance model. Argument schemes (Bex *et al.* 2003) seem to offer an appropriate method for construction and validation of such arguments.

Our approach to provenance requires rich metadata descriptions to be created about resources, projects, data collection tasks, etc. It is unrealistic to expect a user to create descriptions that are complete in all necessary aspects; such a task would be far too time-consuming. The LIBER metadata interface tool supports creation of RDF using natural language. Although this interface has many positive features, it is clear to us that more needs to be done to exploit the context that the *ourSpaces* VRE can provide to automatically suggest parts of the metadata description. For instance, if a user uploads a transcript into a project space while using *ourSpaces*, the system should infer that the resource was produced by that user during that project, during the data collection stage, and suggest possible authors, interviewers, dates, locations, etc. using the information already held in the metadata repository. In this way the user only has to specify in what way the transcript differs from the other resources in the project, while the other (larger) part of its description is created automatically. Similarly, LIBER may be too complex for many simple queries. Tag clouds and keyword searching may be more appropriate in such situations, with the power of the metadata interface being reserved for complex searches, such as those requiring secondary data from a certain time, place and period.

We envision *ourSpaces* becoming a central workplace for the social scientist, rather than 'yet another tool we must login to'. As many activities as possible should be integrated into this environment, which must be adaptable to the user's own preferences. We have already implemented a means for the user to configure and personalise aspects of their own personal workspace, or to tailor a project space to their needs, e.g. by only displaying relevant resource types and tools. In the coming months we intend to evaluate *ourSpaces* and LIBER further, by observing user behaviour in real-life usage scenarios, rather than in an artificial laboratory setting.

The policy-oriented use of simulation models requires that the experimental methodology be accessible to investigators, other researchers and policymakers. Through our scientist's intent framework we aim to make social simulation experiments more transparent, by providing a closer connection between workflow experiments and the goals and constraints of the researcher. While this approach provides important additional provenance information for the experiment, we argue that its use should also facilitate management of workflow execution. We are currently assessing our approach using a series of case-studies which aim to explore our intent representation in terms of its expressiveness, and its ability to enhance reusability of workflow experiments.

This paper has demonstrated that, drawing upon the experience of *eScience*, there is potential to apply the concept of provenance to *eSocial Science*. A provenance model capable of supporting a range of quantitative and qualitative data types and analytical/interpretative methods has being developed which allows users to record whatever they do: it is thus capable of accommodating the different styles of quantitative and qualitative research (at least in the EBPA context), may be capable of furthering the application of a mixed methods approach in the social sciences because *contextual* information demonstrates how useful combining different types of data can be, and it facilitates the re-use of data. Encouraging social scientists to work with the 'provenance concept' has the potential to support methodological rigour and overall good research practice because it will support and encourage researchers to be reflective and reflexive at all stages of their research. However, challenges remain. How receptive will qualitative researchers be to formally recording the analytical/interpretative process? Will all social scientists be willing to record their data and methods of working in the detail required for reuse? The next stage for the PolicyGrid team is therefore to test the provenance model (using the *ourSpaces* VRE), a process which no doubt will answer some questions and raise more.

The technologies described in this paper were not designed to support interdisciplinary activity; rather, the assumption made throughout has been that we were supporting small, single discipline teams. However, many of the problems facing society (e.g. climate change) will require interdisciplinary collaborations in order to address the societal and scientific implications. How will our provenance model need to change to support interdisciplinary evidence bases? As well as appropriate resource and task metadata for other domains (e.g. ecology) issues surrounding interoperation of social science and natural science evidence types need to be explored. Some of the issues likely to be encountered when integrating such evidence include: how to make explicit constraints on sharing and re-use of evidence, the role of shared terminology within an interdisciplinary project, granularity of evidence types (single sample result vs. macro level survey outcome), temporal scale (data gathered over minutes vs. months), spatial context (data gathered at specific geospatial coordinates vs. a focus group exercise conducted at administrative region level). Investigation of these will no doubt highlight many more challenges.

Acknowledgements

The PolicyGrid project is funded by the UK Economic and Social Research Council, under their *eSocial Science* programme; award reference: RES-149-25-1027. We would like to extend our thanks to all those social scientists that have participated in evaluation of our tools to date, and wish to acknowledge the important contribution they have made to the development of our ideas.

References

Anderson, A.H. (2003): 'Human centred design and grid technologies', *Economic & Social Research Council Discussion Paper*.

Bernstein, A. and Kaufmann, E. (2006): 'Gino - a guided input natural language ontology editor', *International Semantic Web Conference 2006*, pp. 144–157.

Bex, F., Prakken, H., Reed, C. and Walton, D. (2003): 'Towards a formal account of reasoning about evidence: argumentation schemes and generalisations', *Artificial Intelligence & Law*, 11 (2-3).

Bishop, L. (2005): 'Protecting respondents and enabling data sharing: reply to Parry and Mauthner', *Sociology* **39**, pp. 333-336.

Branco, M. and Moreau, L. (2006): 'Enabling provenance on large scale e-Science applications', *Proceedings of the International Provenance and Annotation Workshop (IPAW'06)*, Volume 4145 of *Lecture Notes in Computer Science*, pp.55-63.

Chorley, A., Edwards, P., Preece, A. and Farrington, J. (2007): 'Tools for tracing evidence in social science', *Proceedings of the Third International Conference on eSocial Science*, Ann Arbor, Michigan, USA.

De Roure, D., Jennings, N. and Shadbolt, N. (2005): 'The semantic grid: past, present and future', *Proceedings of the IEEE* **93** (3), pp. 669–681.

De Roure, D., Goble, C. and Stevens, R. (2007): 'Designing the myExperiment virtual research environment for the social sharing of workflows', *e-Science 2007 - Third IEEE International Conference on e-Science and Grid Computing*, Bangalore, India, 10-13 December 2007, pp. 603-610.

Edwards, P., Aldridge, J. and Clarke, K. (2006): 'A tree full of leaves: description logic and data documentation', *Proceedings of the Second International Conference on e- Social Science*.

Fielding, N. (2004): 'Getting the most from archived qualitative data: epistemological, practical and professional obstacles', *International Journal of Social Research Methodology* **7** (1), pp. 97-104.

Foster, I., Kesselman, C. and Tuecke, S. (2001): 'The anatomy of the grid: enabling scalable virtual organizations.' *International J. Supercomputer Applications* **15** (3), pp 200-222.

Gruber, T. (2006): 'Where the social web meets the semantic web'. In Cruz, I.; Decker, S.; Allemang, D.; Preist, C.; Schwabe, D.; Mika, P.; Uschold, M.; and Aroyo, L., eds., *Proceedings of the 5th International Semantic Web Conference*, ISWC 2006, volume 4273, 994. Athens, GA, USA: Springer Lecture Notes in Computer Science.

Guy, M., and Tonkin, E. (2006): 'Folksonomies: tidying up tags?' *D-Lib Magazine* **12** (1).

Handschuh, S., Staab, S. and Maedche, A. (2001): 'CREAM: creating relational metadata with a component-based, ontology-driven annotation framework', *K-CAP '01: Proceedings of the 1st international conference on Knowledge capture*, ACM Press, New York, NY, USA, pp. 76–83.

Hielkema, F., Edwards, P., Mellish, C. and Farrington, J. (2007): 'A Flexible Interface to Community-Driven Metadata', *Proceedings of the Third International Conference on eSocial Science*, Ann Arbor, Michigan, USA.

Hielkema, F., Mellish, C. and Edwards, P. (2008): 'Evaluating an Ontology-Driven WYSIWYM Interface', *Proceedings of the Fifth International Conference on Natural Language Generation* (INLG2008), Columbus, Ohio, USA.

HM Treasury (2003): 'The green book: a guide to appraisal and evaluation', HM Treasury, London.

Holland, J. H. and Miller, J. H. (1991): 'Artificial adaptive agents in economic theory', *The American Economic Review* **81** (2), pp. 365-370.

Moss, S. and Edmonds, B. (2005): 'Towards good social science', *Journal of Artificial Societies and Social Simulation* **8** (4), 13.

Mauthner, N.S., Parry, O. and Backett-Milburn, K. (1998): 'The data are out there, or are they? Implications for archiving and revisiting qualitative data', *Sociology* **32** (4), pp. 733-745.

Parry, O. and Mauthner, N.S. (2004): 'Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data', *Sociology* **38** (1), pp. 139-152.

Parry, O. and Mauthner, N.S. (2005): 'Back to basics: who reuses qualitative data and why?', *Sociology* **39** (2), pp. 337-342.

Pennington, D. (2007): 'Supporting large-scale science with workflows', *Proceedings of the 2nd workshop on Workflows in support of large-scale science*, High Performance Distributed Computing 2007.

Petre, M. (1995): 'Why Looking isn't always Seeing: Readership Skills and Graphical Programming', *Communications of the ACM* **38** (6), pp. 33–44.

Philip, L.J. and Macmillan, D.C. (2005): 'Exploring values, context and perceptions in Contingent Valuation studies: the CV Market Stall Technique and Willingness to Pay for Wildlife Conservation', *Journal of Environmental Planning and Management* **48** (2), pp. 257-274.

Pignotti, E., Edwards, P., Preece, A., Polhill, G., Gotts, N. (2005): 'Semantic support for computational land-use modelling', *Proceedings of the Fifth IEEE International Symposium on Cluster Computing and Grid (CCGrid2005)*, IEEE Press. Volume 2, pp. 840–847.

Pignotti, E., Edwards, P., Preece, A., Polhill, J.G. and Gotts, N. (2008): 'Enhancing workflow descriptions with a semantic description of scientific intent', *Proceedings of Fifth European Semantic Web Conference*, Springer-Verlag, pp. 644-658.

Polhill, J.G., Gotts, N. and Law, A. (2001): 'Imitative versus non-imitative strategies in a land use simulation.' *Cybernetics and Systems,* **32** (1), pp. 285–307.

Polhill, J.G., Pignotti, E., Gotts, N., Edwards, P. and Preece, A. (2007): 'A semantic grid service for experimentation with an agent-based model of land-use change'. *Journal of Artificial Societies and Social Simulation* **10** (2) 2.

Power, R., Scott, D. and Evans, R. (1998): 'What you see is what you meant: direct knowledge editing with natural language feedback', *Proceedings of the Thirteenth European Conference on Artificial Intelligence,* Brighton, UK, 1998.

Schwitter, R. and Tilbrook, M. (2004): 'Controlled natural language meets the semantic web', *Proceedings of the Australasian Language Technology Workshop 2004*.

Simmhan, Y.L., Plale, B. and Gannon, D. (2005): ‘A survey of data provenance in e-Science’, *ACM SIGMOD Record* **34**(3), pp. 31-36.

Stevens, R., Robinson, A., and Goble, C. (2003): ‘my-Grid: Personalised bioinformatics on the information gGrid’, *Bioinformatics* **19** (1), pp. 302–304.

Taylor, K., Essex, J. W., Frey, J. G., Mills, H. R., Hughes, G., and Zaluska, E. J. (2006): ‘The semantic grid and chemistry: experiences with CombeChem’, *Journal of Web Semantics* **4** (2), pp. 84–101.

UK Cabinet Office Strategy Unit (2003-5): ‘The magenta book’, *Government Chief Social Researcher’s Office*, London.

UK Cabinet Office (1999): ‘Modernising government’, *White paper*, HM Stationary Office, Cm 4310, March 1999.