Using the Semantic Grid to Support Evidence-Based Policy Assessment

Peter Edwards, Alison Chorley, Edoardo Pignotti, Feikje Hielkema

Department of Computing Science
University of Aberdeen
Aberdeen, AB24 3UE
United Kingdom

{pedwards, achorley, epignott, fhielkem}@csd.abdn.ac.uk

Abstract

The PolicyGrid project is exploring the role of Semantic Grid technologies to support eScience for the social sciences, with a particular emphasis on tools to facilitate evidence-based policy making. In this paper we highlight some of the key challenges for the development of *semantic esocial science* and outline our approach for management of evidence. In addition, we discuss issues surrounding creation and presentation of metadata, and explain why there is a need to enhance workflow tools to allow them to operate within our evidence-based framework.

Introduction

The Semantic Grid (De Roure, Jennings, & Shadbolt 2005) is often described 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'; the analogy here being that the Grid and Semantic Grid have a similar relationship to that existing between the Web and Semantic Web (Berners-Lee, Hendler, & Lassila 2001). Semantic Grids thus not only share data and compute resources, but also share and process metadata and knowledge. eScience applications which utilise semantic technologies now exist in areas as diverse as life sciences - myGrid (Stevens, Robinson, & Goble 2003), chemistry - CombeChem (Taylor *et al.* 2006), SmartTea (Frey *et al.* 2004) and earth sciences - SERVOGrid (Aktas, Pierce, & Fox 2004). However, until recently there has been little work exploring the potential of these techniques within the social sciences, arts and humanities. In this paper we describe the activities of the PolicyGrid[1] project, funded under the UK Economic and Social Research Council *eSocial Science* programme. The aims of PolicyGrid are as follows:

- To facilitate evidence-based rural, social, and land-use policy-making through integrated analysis of mixed data types;
- To demonstrate that Semantic Web/Grid solutions can be deployed to support various facets of evidence-based policy-making through the development of appropriate tools;
- To focus on the authoring of relevant ontologies to support rural, social and land-use policy domains;
- To investigate issues surrounding communication of semantic metadata to social scientists and policy practitioners;
- To promote awareness of the Semantic Grid vision and supporting technologies amongst social scientists.

Our activities are framed by a number of methodological issues within contemporary social science research. The first of these is the concept of 'evidence-based policy making' (Bullock, Mountford, & Stanley 2001). This came to the fore in the UK policy environment in response to a perception that government needed to improve the quality of its decision-making processes; it has been argued that in the past policy decisions were too often driven by inertia or by short-term political pressures. Evidence can take many forms: research, analysis of stakeholder opinion, simulation modelling, public perceptions and beliefs, anecdotal evidence, cost/benefit analyses; as well as a judgement of the quality of the methods used to gather and analyse the information. Our collaborators within the social sciences have developed an evidence-based method for the evaluation of policies concerned with accessibility (Farrington *et al.* 2004). We have thus been influenced by their focus on methods and tools for integrated policy assessment, and the associated use of mixed-method approaches (utilising plural types and sources of data).

The remainder of this paper is organised as follows: in the next section we provide an overview of PolicyGrid, highlighting some of the key challenges for the development of Semantic Grid infrastructure and tools to support social scientists; we then discuss the types of evidence which need to be managed to facilitate evidence-based policy assessment, and outline a proposed solution; issues surrounding the creation and presentation of metadata are then described, as is the need to enhance workflow tools to allow them to operate within our evidence-based framework; we conclude with a discussion and a roadmap for future work.

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

PolicyGrid Overview

Our work involves close collaboration with social scientists and other policy practitioners. These interactions have allowed us to explore a range of issues, including: the extent to which these researchers are comfortable with the Grid as a framework for research practice and collaboration; if ontologies are appropriate (and acceptable) to this community as a way of representing concepts to facilitate evidence-based policy making; the utility (or otherwise) of existing metadata frameworks in use by the social sciences; how best to integrate eScience tools and methods into existing working practices.

We will examine one aspect of these discussions in detail here, namely the role of ontologies in facilitating semantic esocial science. From the beginning of our work, user scientists expressed a fear of 'being trapped in the ontology' due to the contested nature of many concepts within the social sciences. Other researchers (Edwards, Aldridge, & Clarke 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). Edwards et al describe a case study in which several drug use ontologies were constructed, some representing the use of concepts in real-world applications (so-called *in vivo* concepts), and some reflecting top-down classification knowledge of the same domain. These 'organising concepts' are used to create mappings between the *in vivo* concepts; for example:

[in vivo]: DopeSmoking *isatypeof*
[organising]:CannabisUse

Within PolicyGrid we are adopting a different approach which supports dynamic, community-driven evolution of metadata (Guy & Tonkin 2006) within a framework provided by a series of utility ontologies. It has recently been argued (Gruber 2006) that the Semantic Web 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 should be integrated with the Semantic Web. 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 interested in the Grid 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. As an illustration, consider a simple OWL-lite ontology based upon the existing UK Social Science Data Archive[2] metadata schema; the ontology[3] defines a number of concepts including `document`, `questionnaire`, `dataset` and a range of object and datatype properties. Permitted values for many of the datatype properties 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.

With this as our overarching approach to 'community-driven semantic escience', we are exploring the following:

- Management of heterogeneous evidence types (including, how to obtain, store and facilitate the use of evidence in policy assessment).
- Support for creation of metadata and access to resources annotated by metadata. As we cannot expect our users to work with complex formalisms such as RDF, a tool is needed that supports metadata creation by other means, such as natural language.
- Support for creation and sharing of rich workflow descriptions to capture process provenance and scientific intent.

Before we discuss these three aspects of PolicyGrid further, it is appropriate to outline our technical approach. We are developing a service-oriented infrastructure (see Figure 1) based upon current Grid and Web service middleware support. The service layers are as follows:

- **Data & Compute** Low-level services for managing data (DBMS), digital objects (e.g. questionnaires, interview transcripts), metadata (ontologies and RDF triples) and services for managing compute intensive social simulation models and tools (e.g. Fearlus, RePast, NetLogo).
- **Metadata** Services which provide different functionality based upon metadata (e.g. provenance management, reasoning service, integration of tags and ontologies.)
- **Presentation** Services which support visualisation of data and compute resources, and different types of metadata (tags, RDF).

These services are used as components by Web or standalone applications (e.g. the metadata presentation service is utilised as part of a portal - see Figure 3) or as activities within a workflow (e.g. a simulation experiment composed from available simulation services - see Figure 5). The PolicyGrid metadata repository service is based upon the Sesame open source RDF framework [4], while the object repository uses a version of the Fedora [5] digital object repository.

Managing Evidence

A researcher conducting a policy assessment exercise will employ some methodology to evaluate the policy's impact (or possible impact) on the community. They may send out questionnaires to members of the public in certain areas of the country to assess public opinion, or organise town meetings and focus groups. They may interview policy makers to gather information about the impact of the policy on the community or other policies. They may perform a cost benefit analysis in order to assess the fiscal impact of the policy. Such an approach is termed 'mixed method' - as the researcher uses a variety of methods and tools to evaluate the

[2]http://www.data-archive.ac.uk/
[3]www.csd.abdn.ac.uk/research/policygrid/ontologies/UKDA/
[4]http://www.openrdf.org/
[5]http://www.fedora.info/

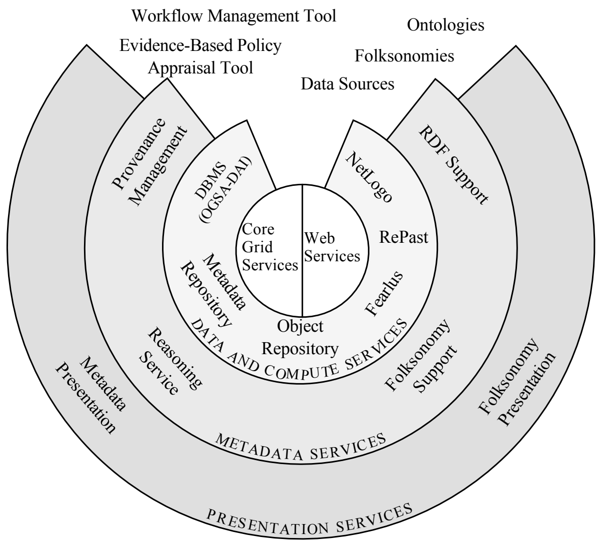

Figure 1: PolicyGrid service architecture.

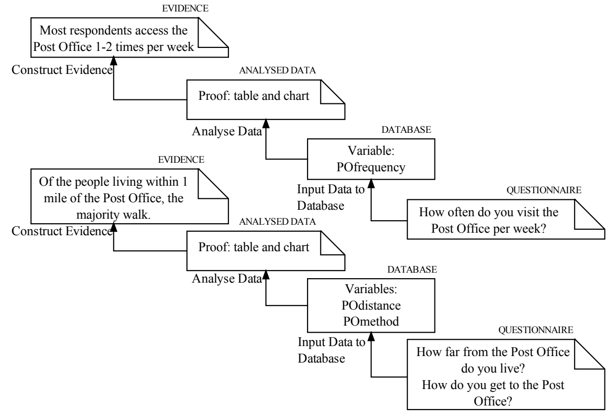

Figure 2: Example evidence pathways (taken from the APAT rural post offices case study).

policy. Quantitative techniques use data obtained from questionnaires and surveys and can be analysed statistically to generate numerical evidence. Qualitative methods use data obtained from interviews, town meetings and focus groups and are usually subject to textual analysis against some conceptual 'coding' framework. Social simulation methods use data obtained from running a model under a specified set of circumstances and then analyse the outcome. APAT (Accessibility Policy Appraisal Tool) (Farrington *et al.* 2004) is an example of a policy evaluation methodology that was designed to examine and evaluate the accessibility impact of policies, using a mixed-method approach. It aims to improve understanding by participants of the accessibility implications of a policy through reflection and analysis and also generates and evaluates alternative policy options.

As we have seen, evidence can take many different forms - quantitative evidence from questionnaires and telephone surveys, qualitative evidence from interviews and focus groups and social simulation results. However, there is another category of evidence that is essential if one is to allow researchers to assess the quality of the methods used to gather and analyse the information - *provenance*. Groth et al. (Groth *et al.* 2006) define the *"provenance of a piece of data as the process that led to that piece of data"*. A suitable provenance architecture for esocial science would allow questions such as 'Where does this evidence come from?' to be answered. Pieces of evidence that form part of a policy appraisal outcome could then be traced back to their source; Figure 2 shows two examples of evidence provenance from an APAT policy assessment case study.

Consider the following example, illustrating the steps involved in producing evidence from quantitative data. The initial resource is the questionnaire sent out to the public to gather their responses; data from the completed questionnaires are gathered and stored in some format (database, text file). The raw data are then analysed using some statistical tool to identify patterns, which are treated as evidential statements in some final report (document). At each of these stages information about the process used should be gathered. For example, when moving from the raw data to the analysed data, the raw data may be sent to an external statistical service which performs the analysis and returns results. Knowing how evidence was derived can prevent problems of misinterpretation and also provides an important audit trail for quality purposes. If a policy maker poses the question 'This evidence seems odd. Show me how it was derived' provenance information could be used to demonstrate that a question used to generate the evidence was misunderstood by subjects completing a survey and they thus gave odd responses, or perhaps a transcription error was introduced.

Our evidence management framework thus comprises the following: metadata support for social science resources (qualitative, quantitative, simulation); a model of provenance (process); and argumentation tools. The resource metadata framework uses an OWL-lite ontology derived from the UK Data Archive schema (itself based upon the Data Documentation Initiative - an international effort to establish a standard for technical documentation describing social science data). Our provenance model for semantic esocial science is based upon the PASOA (Provenance Aware Service Oriented Architecture) project (Groth *et al.* 2006). We are developing an abstract model of provenance which can be instantiated differently for Web/Grid service invocations or human-centred activities (likely to continue to be a key facet of research activity in esocial science). This abstract model of provenance will also be instantiated differently at different stages in the policy assessment process.

To facilitate argumentation we are developing tools which will aid policy stakeholders (researchers, policy makers, others) in using evidence to construct arguments for and against policies. These arguments will be incorporated into an argumentation framework (Dung 1995) which consists of a set of arguments and the attack relations between them. The argumentation framework will allow policy makers to explore a system of conflicting arguments (derived from conflicting evidence) and determine how particular arguments are attacked and defeated (or not) by other arguments.

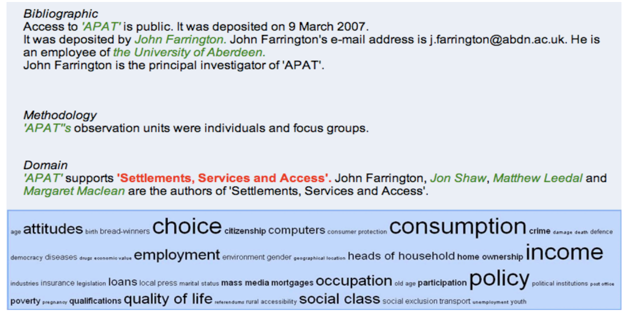

Figure 3: Natural language metadata interface, including tag-cloud.

Facilitating Metadata Creation

Our approach to evidence management requires that the various artifacts associated with a policy assessment exercise have metadata associated with them. Unfortunately, not all metadata can be generated automatically; some types can only be created by the user scientist. A tool is therefore needed that facilitates easy creation of RDF metadata by non-experts, to enable researchers to deposit and describe their own social science resources.

Existing tools are often graphical (Handschuh, Staab, & Maedche 2001). Natural language approaches include GINO (Bernstein & Kaufmann 2006), an ontology editor with an approach reminiscent of natural language menus (Tennant *et al.* 1983), and controlled languages such as PENG-D (Schwitter & Tilbrook 2004). We believe that, for most social scientists, natural language is the best medium to use, as the way they conduct their research and the structure of their documents and data indicate that they are more oriented towards text than graphics.

We also require a tool that is open-ended and flexible. Natural language applications are often domain specific and not very flexible. This makes the open-endedness we need a great challenge. Existing elicitation approaches, such as using controlled languages, restrict in great measure what the user can and cannot say. We believe that to achieve the desired open-endedness and flexibility, the best approach is not based on natural language processing, as it is as yet beyond the state of the art to reliably parse all user utterances, but based on natural language generation. We have chosen an approach (Power, Scott, & Evans 1998) in which the user can specify information by editing a feedback text that is generated by the system, based on a semantic representation of the information that the user has already specified. As the text is generated by the system and does not have to be parsed, we do not have to restrict what can and cannot be said, so the language can retain its expressivity and the user does not need to learn what is acceptable as input. The system is guided by an underlying datastructure, in our case a lightweight ontology plus a series of supporting folksonomies used to suggest feasible tags to influence user-behaviour, without restricting the user to a pre-defined set of concepts.

Figure 3 shows a feedback text (generated by the current system) for the APAT scenario. We are building a tool that elicits metadata from the user, by presenting them with a text containing an expansion point (anchor) for each object that is mentioned, which has a menu with possible properties associated with that object. These objects and properties are defined by an underlying OWL-lite ontology (e.g. the resource metadata ontology mentioned above); however, we intend to ensure that other ontologies can be substituted. Such ontologies should be well-formed, be clear about which objects are permitted in the domain and range of properties, and for the benefit of the generated text should have clear object and property names (e.g. HasAuthor), as these names are used for generation with only some minor adjustments. Each 'string' datatype within the ontology has an associated folksonomy, which is used to generate a tag-cloud that is shown to the user when he/she has to enter a value for that property (see Figure3); the tag cloud gives an overview of tags used by other users, and the frequency with which they have been used (reflected in the relative font size). Use of folksonomies in this way stimulates the emergence of a community set of tags (Guy & Tonkin 2006), prompting the user to use the same values as other users, or to adopt a similar style. It (in part) protects the system from mistakes such as spelling errors, and, when queried, increases the likelihood of a search term being associated with more than one resource. The user however still has complete freedom, as he/she does not have to use the folksonomy values but can still use free text; and every entry the user makes is immediately added to the folksonomy. The folksonomy, then, allows us to subtly guide user behaviour, while being completely unrestrictive.

The current system consists of five components: the semantic graph, the ontology reader, the RDF-creator, the natural language generator (consisting of a text planner and a surface realiser) and the interface.

The **interface** shows the feedback text with anchors indicating expansion points, which contain menus with types of information that can be added. Google Web Toolkit[6] was used to create the prototype interface.

The **semantic graph** stores the information the user is adding. Each node corresponds to an ontology class, each edge to an ontology property. Initially a generic graph (see Figure 4) is created, so an initial feedback text can be produced; the graph is updated each time the user adds information.

The **ontology reader** creates a model of a given OWL-lite ontology, which is consulted throughout the building of the semantic graph and extended with all new properties or objects that the user adds. The ontology specifies the range and domain of the properties; i.e. the properties in each anchor menu, and the (type(s) of) resource that can be selected or added as the range of a selected menu item.

The semantic graph is translated to a list of RDF triples by the **RDF-creator**. These triples are stored, with the relevant

[6]http://code.google.com/webtoolkit/

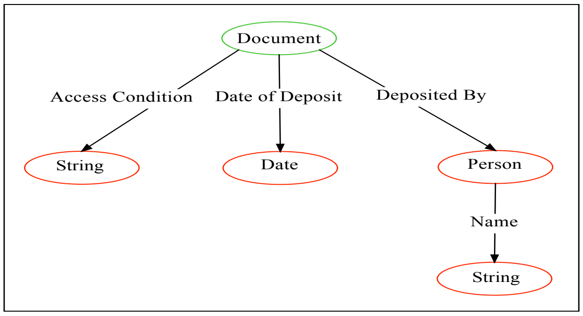

Figure 4: Initial, generic semantic graph.

resource(s), in a Grid-enabled repository of social science resources.

The **natural language generator** maps the semantic graph to (HTML) text that contains anchors. The text is divided in paragraphs to give a clear overview. To keep the text concise some sentences are aggregated (combining two sentences into one, e.g. 'Mary sees Bill' and 'Claire sees Bill' become 'Mary and Claire see Bill') and where possible pronouns (he, she, it) are used.

In the context of our service architecture (see Figure 1), the metadata elicitation tool uses the metadata and the object repositories. The metadata repository contains the underlying ontologies and stores the generated RDF-triples; these are associated with the uploaded resources (datasets, documents, etc.) which are stored in the object repository.

Although our system is driven by an ontology, we have kept this very lightweight (OWL-lite, using only domain and range of properties and cardinality restrictions), and plan to give the user the power to adapt this ontology to his/her own needs. Extending an ontology with new classes and properties is no great problem; however, those properties then need a suitable natural language representation. In our system this means they need an entry in the lexicon that maps them to a tree that represents a sentence in natural language. A straightforward way to obtain such entries is to let a system administrator create them when needed. However, this would cause considerable delays for the user and would appear almost as restrictive as not allowing new property creation at all. Instead, we want to enable the system to create these lexicon entries immediately, so the user can use the new property at once. Using the property name that the user provides, the system can generate possible lexicon entries; the user, who knows what the natural language representation should look like, can choose a suitable sentence from the options offered and perfect it. The resulting lexicon entry would be stored for future use.

We are about to perform a number of user studies to evaluate the metadata elicitation tool. First, we plan to ask users to describe a resource (either their own or a fictional resource we provide) both on paper and using the tool. This will enable us to assess the ease-of-use, reliability and transparency of the tool and the user's learning curve. Comparing descriptions of the same resource in two media, in addition to user feedback, indicates whether the tool provides sufficient support for the description of social science resources. In a second experiment we plan to investigate the property-creation component. Our central question in this case, is whether allowing the user to create new properties contributes significantly to user satisfaction and effectiveness; from an NLG perspective, we are also very curious about the amount of time and effort it takes them to create new properties and whether users find this worthwhile.

Workflow Support

Workflow technologies can assist the scientific process by facilitating the creation and execution of experiments from a pool of available services. By scientific workflow we mean a composition of structured activities (e.g. database queries, simulations, data analysis activities, etc.) that arise in scientific problem-solving (Ludäscher *et al.* 2005). As mentioned above, evidence in policy making/appraisal is not just data (simulations, questionnaires, interviews) but also the process that led to that data. Scientific analysis is a complex process that cannot be captured only in terms of service composition and linear execution but also requires a higher-level description of the scientific process to make hypothesis, experimental conditions and goals of the experiment transparent and therefore challengeable (i.e. can be used as evidence). In this section we discuss an example of social simulation workflow highlighting the limitations of current workflow technologies. We also present an initial design of a framework for capturing scientific intent on top of existing workflow tools such as Kepler (Ludäscher *et al.* 2005) and Taverna (Oinn *et al.* 2004).

We have explored a number of case studies with simulation experts, many of which relate to exploration of parameter spaces and the study of land-use strategies. We now present an example using FEARLUS, a simulation model (Polhill, Gotts, & Law 2001) developed at the Macaulay Research Institute. FEARLUS is an agent-based simulation model of land-use change which aims to study the dynamics of imitative and non-imitative approaches to land use selection change under different circumstances, in the context of environments differing in spatial and temporal heterogeneity. An experiment in FEARLUS might involve studying the differences between imitative and non imitative selection strategies in a specific environment. A scientist wishing to test the hypothesis 'innovators do better than imitators in environment A' might run a set of simulations in which *innovators* compete against *imitators* in that environment using different random seeds. If in this set of simulations, *innovators* outperformed *imitators* in a significant number of simulation runs, the experimental results could be used to support the hypothesis.

Figure 5 shows a workflow built using the Kepler tool that uses available Grid and local services to perform the experiment described above. *Innovators*, *Slow Imitators* and *Fast Imitators* correspond to parameter files defining different strategies where *Model*, *Climate Toggle* and *Economy Toggle* correspond to parameters of the environment represented in the simulation model. The *Parameter Permutation* component is used to create all the possible permutations of two strategies to be tested in the simulation model.

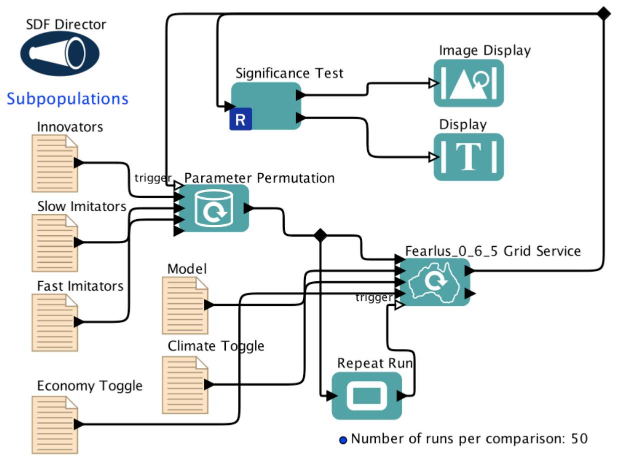

Figure 5: Social simulation workflow example.

The parameters selected are fed to a particular version of the FEARLUS model - *Fearlus_0_6_5 Grid Service*. The *Repeat Run* component is used to trigger different simulation runs based on the same set of parameters (Number of runs per comparison). Results of the simulation runs consist of the number of land parcels associated with a particular strategy at the end of the simulation. These results are fed into the *Significance Test* component which will output the results of the test. The hypothesis is tested by looking at the result of the significance test; if the strategy that we are considering outperforms any other, we can use this result to support our hypothesis.

With existing workflow tools such as Taverna and Kepler scientists can easily define operational workflow in terms of composition of available services. In the context of our evidence-based framework, such workflows are important in order to understand the process that led to a particular resource (the execution details of the workflow act as provenance for a particular piece of evidence). However, the experimental workflow defined in Figure 5 has some limitations. First of all, we don't know *a priori* how many simulation runs per comparison we need to do in order to have a significant number of results. Too few runs will mean that the experiment will return inconclusive data, while too many runs will waste computing resources executing unnecessary simulations. There may also be constraints associated with the workflow (or specific activities within the workflow) depending upon the intent of the scientist. For example, a researcher may be concerned about floating point support on different operating systems; if the simulation model runs on a platform not compatible with IEEE 754 specifications, the results of the simulation could be compromised. We argue that capturing the scientist's intent is thus essential for the transparency of the results generated by a workflow, and propose a framework for capturing such intent which we characterise as a set of goals (or endstate) that the scientist aims to achieve upon completion of the workflow, and a set of constraints that has to be enforced during its execution.

The scientific intent reflected in the FEARLUS example above can be represented as a combination of goals and constraints. **Goal:** run enough simulations to provide valid results to support (*significance test* < *significance level*). **Constraints:** Simulation has to run on a platform compatible with IEEE 754 (*platform = IEEE 754*).

We are developing a model of scientific intent based upon rules[7] which operate on metadata generated by the workflow. Details of the scientific intent are kept separate from the operational workflow as embedding constraints and goals directly into the workflow representation would make it overly complex (e.g. with a large number of conditionals) and would limit potential for sharing and re-use. Such a workflow would be fit for only one purpose and addition of new constraints would require it to be substantially re-engineered. Using the support for scientific intent, a new experiment might be created just by changing the rules but not the underlying operational workflow. We have identified SWRL[8] (Semantic Web Rule Language) as a language for capturing such rules. SWRL enables Horn-like[9] rules to be combined with an OWL knowledge base.

Activities within the workflow must have supporting ontologies and should produce metadata that can be used against scientific intent (rules) to document the execution of the workflow. We have identified the following possible sources of metadata:

- Metadata about the result(s) generated upon completion of the workflow;
- Metadata about the data generated at the end of an activity within the workflow or sub-workflow;
- Metadata about the status of an activity over time, e.g. while the workflow is running.

The kind of metadata generated during the execution of the workflow will depend on the workflow engine used and the metadata support for each activity used in the workflow. The WSRF2OWL-S tool (Babik *et al.* 2005) developed as part of the K-Wf Grid[10] project may prove useful as it allows elements in the Grid service description to be mapped on to ontological concepts. In order to support social simulation activities in our framework we have created an initial social simulation classification ontology[11] capturing the characteristics of a wide range of simulation models, e.g. type of simulation, behaviour, space model, execution type, etc. Collaborators at the Macaulay Institute are continuing work on the development of a simulation modelling ontology to allow a particular piece of modelling software to be described and the structure and context of a particular simulation run to be characterised. We plan to evaluate our approach by assessing the impact of the enhanced workflow representation

[7] http://www.ruleml.org/

[8] http://www.daml.org/2003/11/swrl/

[9] The rules are of the form of an implication between an antecedent (body) and consequent (head). The intended meaning can be read as: whenever the conditions specified in the antecedent hold, then the conditions specified in the consequent must also hold.

[10] http://www.kwfgrid.eu/

[11] www.csd.abdn.ac.uk/research/policygrid/ontologies/Sim/

from two perspectives: (i) the utility of the *intent* construct as additional metadata to facilitate interpretation of experimental results and workflow re-use; (ii) how the real-time monitoring of experiments guided by *intent* affects use of Grid resources.

The K-Wf Grid project is exploring knowledge-based support for workflow construction and execution in a Grid computing environment. Central to their vision is the use of semantic technologies to annotate Grid services (Babik *et al.* 2005). Although the K-Wf Grid project is investigating the role of knowledge in workflows, there is no support for capturing the analytical reasoning (scientist's intent) behind the workflow, which we believe is important to understand how evidence has been constructed.

Discussion

To support semantic esocial science we are developing an evidence-based framework to manage the heterogeneous resources employed during policy appraisal; a hybrid metadata approach combining both lightweight ontologies and folksonomies forms a part of this framework in order to give social scientists the open-ended mechanism they require to annotate resources. The use of tags helps to fill the acceptance/participation gap by allowing individual researchers to describe resources as they deem appropriate, without being constrained by a centrally imposed conceptualisation. To allow users to deposit their resources on the Grid we have developed a tool for metadata elicitation which stores descriptions using RDF-triples; the user has freedom to add values for datatype properties, and is supported by a folksonomy that suggests suitable and popular tags for each property. Although the current version of the tool is driven by one (lightweight) ontology we are exploring ways to enable the user to extend this ontology. As the process used to generate evidence is as important as the evidence itself, we are also embedding support for provenance within our approach. Our abstract provenance model can be instantiated in different ways depending upon the stage in a policy appraisal exercise being recorded and whether the activity was performed by a computational service or a human researcher. We are also investigating how workflow technologies can operate within our evidence-based framework, by enhancing current tools to allow the scientist's analytical reasoning to be made transparent. We are developing a mechanism to capture scientific intent based on the idea of endstate and constraints using a semantic rule language. We are also emphasising the role of ontologies and metadata in workflow activities as a means to extract information during the execution of the workflow so that it can be checked against the scientific intent.

At present our hybrid ontology-folksonomy approach is basic, with many outstanding issues still to be resolved. To date we have employed a standard utility ontology (derived from the DDI standard); this ontology does not attempt to represent concepts which might be imprecise, mutable or contested - as it simply defines standard documentation concepts which then act as containers for tags. We still need to determine what other utility ontologies might be appropriate (and acceptable) for use in semantic esocial science. In our current implementation, tags are simply string tokens with associated frequency counts. What properties (if any) should a tag possess? Should relationships between tags be supported? Can (should?) tags evolve to become ontology concepts?

We hope that the metadata elicitation tool will help us gather user requirements for querying and information presentation using the hybrid ontology/folksonomy metadata representation. Earlier attempts to uncover these failed because users could not articulate the kind of complex queries they might formulate if they had the chance, because at the moment they lack suitable search tools for this, and thus have no relevant experience. We hope that the metadata elicitation tool will clarify for them the types of things that can be said or asked in RDF (presented, of course, in natural language), so they have an easier time speculating about eligible queries. We plan to use the same approach for the querying and information presentation tools as for metadata elicitation. The user should be able to construct a short text describing the type of resource he/she is looking for, which the system should then translate to a SPARQL-query; the answer could be presented in another feedback text in which the anchors would cause related information to be added. We think this approach will be suitable for all three tasks; it also means the user only has to learn to work with one interface which improves usability.

To integrate the various components of our evidence management framework for policy appraisal (and potentially other social science research tasks) we are constructing a community Web portal (*ourSpaces*) which will provide the following functionality: submission of resources, searchable archive(s), enhanced collaboration support, integration with client (desktop) tools. We intend to let users annotate each other's resources and share those annotations (in the manner of Connotea). This approach would especially suit quantitative social science data sets, as they are frequently used for secondary research, and are re-analysed from different perspectives. Users will be able to associate existing resources with new ones, add to an existing description of a resource if they feel the description is missing valuable information, mark resources from one discipline as being relevant to another discipline, etc. We have yet to consider how issues such as trust and reputation will be integrated within our provenance framework; both are clearly important for those involved in policy assessment, and will form part of our plans for future work.

Developing Semantic Grid solutions for social science presents many challenges, some of which we have outlined in this paper. Based upon our experiences to date we are convinced that a solution integrating Grid services with ontologies and community-driven folksonomies is appropriate and will meet the needs of researchers (and others) interested in evidence-based policy development and appraisal.

References

Aktas, M. S.; Pierce, M.; and Fox, G. C. 2004. Designing Ontologies and Distributed Resource Discovery Services for an Earthquake Simulation Grid. In *Proceedings* *of GGF-11 Global Grid Forum Semantic Grid Applications Workshop*.

Babik, M.; Hluchy, L.; Kitowski, J.; and Kryza, B. 2005. WSRF2OWL-S: A framework for generating semantic descriptions of web and grid services. In *Cracow Grid Workshop '05*.

Berners-Lee, T.; Hendler, J.; and Lassila, O. 2001. The Semantic Web. In *Scientific American*.

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

Bullock, H.; Mountford, J.; and Stanley, R. 2001. Better Policy-Making Centre for Management and Policy Studies. Technical report, Cabinet Office.

De Roure, D.; Jennings, N.; and Shadbolt, N. 2005. The Semantic Grid: Past, Present, and Future. *Proceedings of the IEEE* 93(3):669–681.

Dung, P. M. 1995. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence* 77(2):321–358.

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*.

Farrington, J.; Shaw, J.; Leedal, M.; Maclean, M.; Halden, D.; Richardson, T.; and Bristow, G. 2004. *Settlement, Services and Access: The Development of Policies to Promote Accessibility in Rural Areas in Great Britain*. H.M. Treasury, The Countryside Agency, Scottish Executive, Welsh Assembly Government.

Frey, J.; Hughes, G.; Mills, H.; Schraefel, M. C.; Smith, G.; and De Roure, D. 2004. Less is More: Lightweight Ontologies and User Interfaces for Smart Labs. In Cox, S. J., ed., *Proceedings of The UK e-Science All Hands Meeting*.

Groth, P.; Jiang, S.; Miles, S.; Munroe, S.; Tan, V.; Tsasakou, S.; and Moreau, L. 2006. An Architecture for Provenance Systems. ECS, University of Southampton.

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. In *K-CAP '01: Proceedings of the 1st international conference on Knowledge capture*, 76–83. New York, NY, USA: ACM Press.

Ludäscher, B.; Altintas, I.; Berkley, C.; Higgins, D.; Jaeger-Frank; E., Jones, M.; Lee, E.; Tao, J.; and Zhao, Y. 2005. Scientific Workflow Management and the Kepler System. In *Concurrency and Computation: Practice & Experience, Special Issue on Scientific Workflows*.

Oinn, T.; Addis, M.; Ferris, J.; Marvin, D.; Senger, M.; Greenwood, M.; Carver, T.; Glover, K.; Pocock, M.; Wipat, A.; and Li, P. 2004. Taverna: A tool for the composition and enactment of bioinformatics workflows. *Bioinformatics Journal* 20(17):3045–3054.

Polhill, J.; Gotts, N.; and Law, A. 2001. Imitative Versus Non-Imitative Strategies in a Land Use Simulation. In *Cybernetics and Systems*, 285–307.

Power, R.; Scott, D.; and Evans, R. 1998. What You See Is What You Meant: direct knowledge editing with natural language feedback. In *Proceedings of the Thirteenth European Conference on Artificial Intelligence*.

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

Stevens, R.; Robinson, A.; and Goble, C. 2003. myGrid: Personalised Bioinformatics on the Information Grid. *Bioinformatics* 19(1):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):84–101.

Tennant, H.; Ross, K.; Saenz, R.; Thompson, C.; and Miller, J. 1983. Menu-based Natural Language Understanding. In *Proceedings of the Twenty-first Annual Meetings on Association for Computational Linguistics*, 151–158.