

Providing Ontology Support for Social Simulation

Edoardo Pignotti¹, Pete Edwards¹, Alun Preece¹,
Gary Polhill², Nick Gotts²

¹ Dept. of Computing Science
University of Aberdeen
Aberdeen, AB24 3UE, Scotland
Email: { epignott, pedwards, apreece }@csd.abdn.ac.uk

²The Macaulay Institute
Craigiebuckler
Aberdeen, AB15 8QH, UK
Email:{g.polhill, n.gotts}@macaulay.ac.uk

Abstract. Semantic Web technologies facilitate tasks such as collaboration, shared experimentation, and annotation of resources and have particular strength in capturing qualitative scientific arguments. In this paper we explore the design of an ontology for a social simulation tool called FEARLUS-G. This lead to a description of the tools used during the ontology engineering, and how the ontology was integrated in the FEARLUS-G service architecture.

Overview

In this paper we explore the use of Semantic Web technologies through the design of an ontology for a social simulation tool (Polhill, J., et. al., 2001). This work was carried out as part of the FEARLUS-G¹ project (Pignotti, E., et. al., 2005) which allows large scale simulation experiments to be distributed over the Grid (Foster, et. al., 2001) FEARLUS-G also includes infrastructure to support management of meta-data; at the core of this is an ontology which facilitates sharing and re-use of resources among land-use scientists. We begin by providing a brief introduction to Semantic Web technologies and their potential role within eSocial Science. Next, we discuss some software tools and outline how these were used during the FEARLUS-G project to construct an ontology. We conclude with an explanation of how this ontology was embedded within the FEARLUS-G service architecture to enhance the scientific process.

The Semantic Web

The Semantic Web (Berners-Lee, T., et. al., 2001) is a vision in which today's Web will be extended with machine readable content, and where every resource will be marked-up using machine readable meta-data; a family of XML-based technologies, most notably the Resource

¹ <http://www.csd.abdn.ac.uk/research/fearg>

Description Framework (RDF)², provide a mechanism for representing that meta-data. Ontologies (Fensel, D., 2003) are used to capture the meaning of meta-data terms and their interrelationships. The main benefit of using ontologies is that they facilitate access to heterogeneous and distributed information sources by defining a machine-processable semantics for those information sources. Important technologies include RDF Schema (RDFS)³ - a vocabulary for describing properties and classes of RDF resources, with a semantics for generalization-hierarchies; and OWL (Web Ontology Language)⁴ - which adds more vocabulary for describing properties and classes, e.g. relations between classes, cardinality, etc.

Figure 1 shows an example of a FEARLUS-G experimental hypothesis represented using RDF. This particular instance has a description “fast imitators do better than innovators” and a number of properties linking to other resources such as: creator, contributor, etc.

```
<ExperimentalHypothesis rdf:ID="hyp_1">
  <hasHypothesisDescription>
    Fast imitators do better than innovators
  </hasHypothesisDescription>
  <dc:creator rdf:resource="#GaryPolhill"/>
  <hyp:replaces rdf:resource="#hyp_0"/>
  <dc:contributor rdf:resource="#NickGotts"/>
  <describedIn rdf:resource="#pub_1"/>
  <hasType rdf:resource="#fearlusType1Hypothesis"/>
  <hasStatus rdf:resource="#hypothesisUnderInvestigation"/>
</ExperimentalHypothesis>
```

Figure 1 RDF Example - ExperimentalHypothesis Instance.

We argue that the Semantic Web approach has significant potential within eSocial Science. Semantic Web technologies can help deliver the vision of a more “human-centred” Grid which facilitates tasks such as collaboration, shared experimentation, and annotation of resources. Furthermore, we argue that these technologies have particular strength in capturing qualitative scientific arguments, supported by a mix of quantitative and qualitative data and results. We believe that the *Scientific Object* ontology presented in the next section is a first step towards this, providing limited support for the creation of argument structures linking simulation experiments, hypotheses and results. We believe that such a mechanism has great potential within eSocial Science.

Ontology Engineering

As part of the FEARLUS-G project we have engineered a Scientific Object ontology, part of which is shown in Figure 2. The ontology includes a collection of generic elements that are intended to be applicable to any e-science application, and elements that are specific to simulation modelling and FEARLUS-G in particular.

² <http://www.w3.org/RDF/>

³ <http://www.w3.org/TR/rdf-schema/>

⁴ <http://www.w3c.org/2004/OWL/>

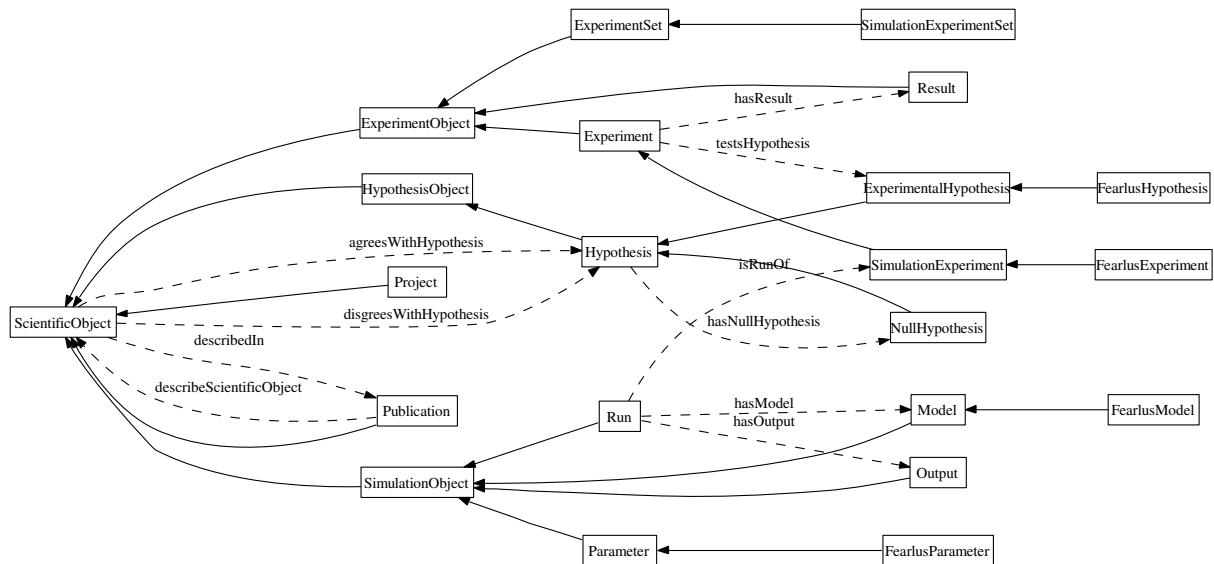


Figure 2 FEARLUS-G Scientific Object Ontology.

The ontology engineering process was initiated by the computer science personnel involved in this collaboration as they were most familiar with Semantic Web tools and languages. A high-level ontology was created initially which captured the concepts and relationships important to scientists performing their research activities. In our ontology, the classes *HypothesisObject*, *ExperimentObject*, *SimulationObject* and *Publication* are subclasses of a generic *ScientificObject* class. To capture the idea that Science is a discourse that does not necessarily present a consistent system of knowledge, *ScientificObjects* may *agreesWithHypothesis* or *disagreesWithHypothesis*. The *agreesWithHypothesis* and *disagreesWithHypothesis* relations are intended to reflect an abstract level of inconsistency between *ScientificObjects*, in contrast to more specific relations that *Hypotheses* have with other subclasses of *ScientificObject*, e.g. *describedIn*.

The Protégé⁵ tool was used during the ontology generation process. Protégé is an open-source tool developed at Stanford Medical Informatics and provides a highly customizable ontology editor that enables conceptual modelling with Semantic Web languages such as OWL and RDFS. The architecture of Protégé is cleanly separated into a “model” part and a “view” part. The *model* is the internal representation of ontologies and knowledge bases. The *view* component provides a user interface to display and manipulate the underlying model. The user interface (see Figure 3) contains a number of useful components: class editor, property editor, instance editor, ontology graph.

An important component of Protégé is the OWL plug-in that allows the ontology (model) to be represented with the OWL syntax. Moreover the OWL plug-in provides an interface for the Dublin Core ontology which has been used to provide basic annotation of *ScientificObject* instances. For example *dc:creator* defines the creator of an hypothesis, publication or experiment, *dc:contributor* the contributor(s).

⁵ <http://protege.stanford.edu/>

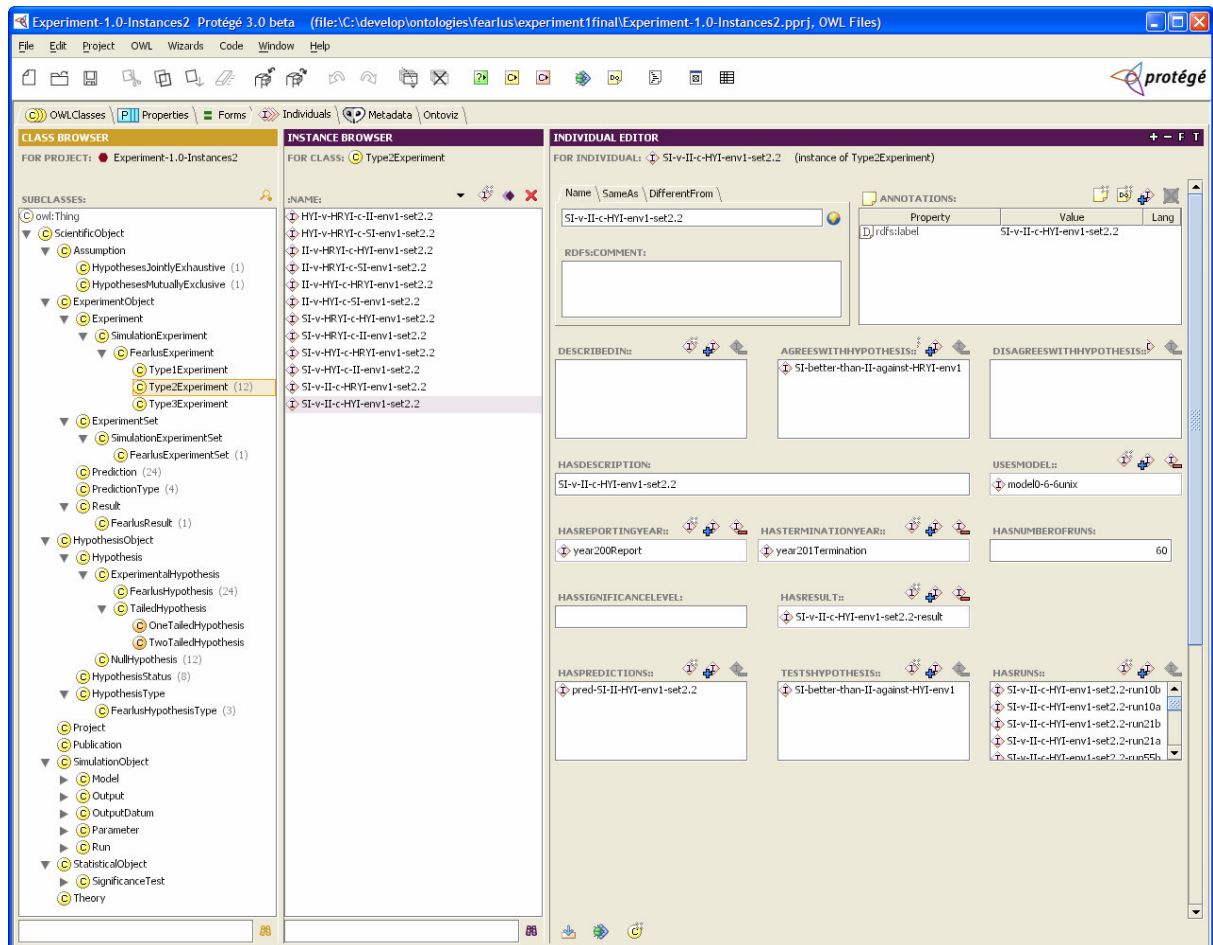


Figure 3 Protégé User Interface.

Once the initial FEARLUS-G ontology had been created, Semantic Web technologies including the Protégé tool were presented to the social simulation scientists involved in the project. This led, in turn, to the process of engineering the specific FEARLUS-G ontology based on the initial high-level ontology. After the first iteration of this process, the resulting ontology contained many low-level, highly specific classes reflecting the parameter space of the FEARLUS-G simulation model. To provide further guidance to the simulation personnel, a tutorial (Horridge M., et. al., 2004) on OWL was introduced. This tutorial focuses on building an OWL ontology and use of an automated reasoner to check the consistency of the ontology and automatically compute the ontology class hierarchy. The tutorial helped the scientists to focus more on aspects of experimental workflow. The resulting ontology (Figure 2) still supports specific FEARLUS-G properties, but also exposes experimental workflow, making its use transparent to other users.

Integrating the Ontology

The ontology described above was integrated into the collection of Grid services which make up FEARLUS-G (Edwards P., et. al., 2005). Together, the full range of technologies allow sharing of computational and data resources across different hosts. The extensive use of RDF meta-data raises human-interface issues, as not all users will be comfortable with mark-up of the kind shown in Figure 1. To address this problem, we have embedded a tool for browsing

meta-data into our Web interface. Longwell⁶ is a suite of web-based RDF browsers that allow complex RDF datasets to be displayed using different styles. The styles can be adapted to any level of user; for example, a user-friendly style where all the complexity of RDF is hidden (see Figure 4) and a RDF-aware style where all the details are shown to the user.

The most important aspect of the Longwell is the ‘faceted browsing’ user-interface paradigm. A facet is a particular meta-data field that is considered important for the dataset that we are browsing. The system extracts a list of facets, their values and the number of times this facet value occurs in the dataset. This is shown on the right hand side of the Longwell screen (see Figure 4), where each facet is grouped independently. By clicking on a facet value, the browser adds a facet restriction and eliminates all items that do not fulfil that restriction.

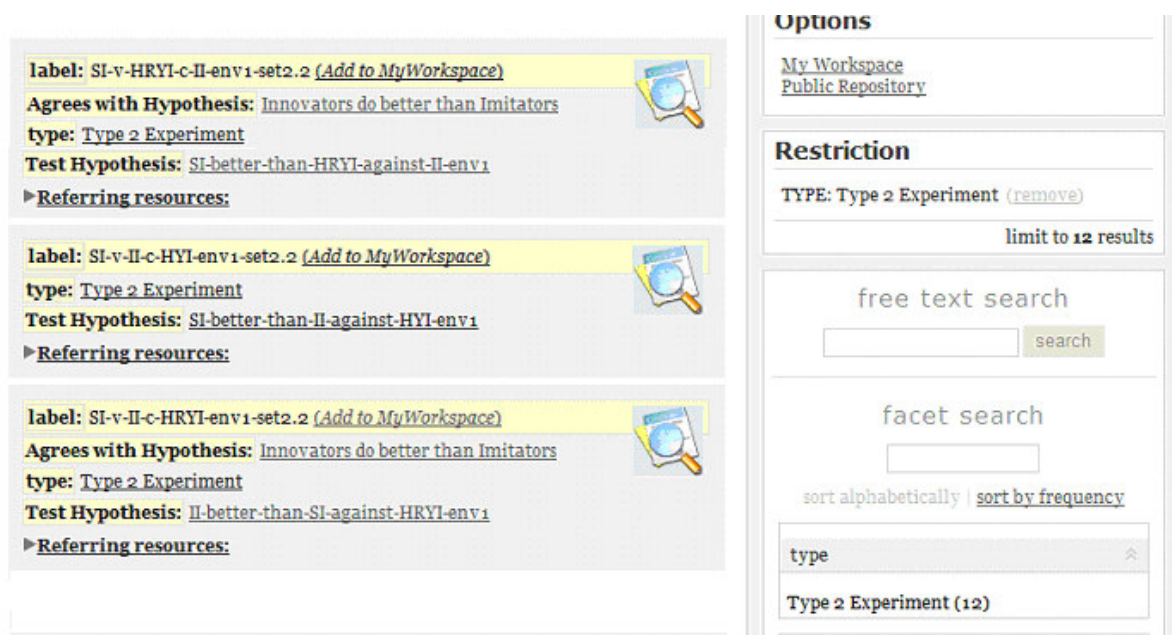


Figure 4. Screenshot Illustrating use of Longwell to Render Experiments.

The facet pane on the right-hand side of the screen is also updated, showing the facets and their values that are available in that sub-set of the data. Clicking on the pane has the effect of ‘zooming in’ on the dataset by locally removing from the view anything that is not of interest at that particular time. The browser also support a ‘free text’ restriction facility; this reduces the browsed dataset to all items that contain the searched strings. By clicking on the icon to the right of each instance, the user can access the full data associated with that instance.

Longwell was a valid tool to support creation of an easy to use meta-data browser however, the layout and rendering templates are not designed to be easily customized. However, we decided to keep the Longwell as a component of our user interface to browse publicly available resources (*Public Repository*).

⁶ <http://simile.mit.edu/longwell/>

You are logged in as: fearlus

Options


[My Workspace](#)
[My Grid Tasks](#)
[Public Repository](#)

[Questionnaire](#)

[Log Out](#)

Scientific Argument

Agrees with Hypothesis

 [Innovators do better than Imitators](#)

Disagrees with Hypothesis

[+ Agrees...](#)
[+ Disagrees...](#)


My Projects


[My Entire Workspace](#)

[Imitative vs non imitative strategies](#)
(info)

[+ New Project](#)



My Hypotheses

 [Innovators do better than Imitators](#)

 [Imitators do better than Innovators](#)

[+ New Hypothesis](#)

Experiment Set

 **Experiment2.2-Polhill-et al-2001** 

Name:	Experiment2.2-Polhill-et al-2001
Description:	Experiment2.2-Polhill-et al-2001
Comment:	
Number of Runs:	60
Significance Level:	0.0010
Single Subject Hypotheses:	true
Experimental Hypotheses:	
Experiments:	II-v-HRYI-c-HYI-env1-set2.2 HYI-v-HRYI-c-II-env1-set2.2 SI-v-HRYI-c-HYI-env1-set2.2 SI-v-II-c-HYI-env1-set2.2 SI-v-II-c-HRYI-env1-set2.2 HYI-v-HRYI-c-SI-env1-set2.2 II-v-HYI-c-SI-env1-set2.2 II-v-HRYI-c-SI-env1-set2.2 SI-v-HYI-c-HRYI-env1-set2.2 SI-v-HYI-c-II-env1-set2.2 SI-v-HRYI-c-II-env1-set2.2 II-v-HYI-c-HRYI-env1-set2.2
Uses Model:	modelo-6-6unix
Termination Year:	year201Termination
Subjects:	II HRYI SI HYI Environment1

[* Remove Instance](#)
[+ Add to Project...](#)

Figure 5. The FEARLUS-G My Workspace Web Interface Rendering an Experiment Set.

In order to enable users to further investigate resources we developed the *My Workspace* Web-based interface (see Figure 5) that allows land-use scientists to manage simulation activities. *My Workspace* has two main components: the side-bar contains an options box, a list of active projects (My Projects) and a list of active hypotheses (My Hypotheses); the body component is used to render selected data resources or to render forms to input new instances. Other context-sensitive options appear in the side-bar when a user navigates to different areas of the workspace. Instances contained in My Workspace are grouped by project and by type of resource (eg. Experiment Set, Hypothesis, Publication). Each resource type has an associated icon used consistently in the workspace environment to facilitate their location. It is possible to browse the properties associated with each instance by clicking on its name (rdfs:label). The properties are rendered based on a template which defines how each resource should be rendered in My Workspace. Templates are encoded in XML defined using a set of Java libraries which contain the code to store, locate and render resources. Templates in My Workspace could be generated automatically by reasoning from the ontology, but to date we have not accomplished this; all existing templates are hand-coded.

New Type 2 Experiment Set

Click on the symbol if you need help with filling in a field.

Name:

Description:

Significance Level: (For example: 0.005)

Subjects of the experiment:

Number of Runs:

Single Subject Hypothesis: true false

Uses Model:

Comment:

```

<rdf:Description
  rdf:about="Experiment2.2-Polhill-etal-2001">
  <hasDescription>
    Testing imitative vs non imitative
    strategies
  </hasDescription>
  <hasSignificanceLevel>
    0.0056
  </hasSignificanceLevel>
  <hasSubjects
    rdf:resource="#OD"/>
  <hasSubjects
    rdf:resource="#OS"/>
  <hasSubjects
    rdf:resource="#HYI"/>
  <hasSubjects
    rdf:resource="#Environment1"/>
  <hasNumberOfRuns>200</hasNumberOfRuns>
  <hasSingleSubjectHypotheses>
    true
  </hasSingleSubjectHypotheses>
  <usesModel
    rdf:resource="#model10-6-6unix"/>
</rdf:Description>

```

Figure 6. Form to Create a New Experiment Set (left); Experiment Set RDF, Instance (right).

Figure 6 shows an example of an input form and the resulting RDF both generated from the defined template. Certain resources in My Workspace (Experiment Set, Hypothesis, Publication) can be used as evidence in order to build a scientific argument. When one of those resources are rendered in the interface, an option is displayed allowing the user to specifying if the resource agrees or disagrees with an existing hypothesis. Figure 7 shows the summary of an hypothesis instance, to which various resources have been linked as evidence for the argument structure.

Hypothesis

Name: Innovators do better than Imitators

Description: Innovators do better than imitators in all environment conditions

Comment:

✖ Remove Instance
+ Add to Project...

Supported By ...

Experiment Sets

[expKMT](#)
 [Experiment2.2-Polhill-etal-2001](#)

Experimental Hypotheses

[expKMT-OS-better-than-OD-against-HYI-in-v-Environment1](#)

Publications

[Imitative versus non-imitative strategies in a land use simulation](#)

Opposed By ...

Hypotheses

[Imitators are better than Innovators](#)

Figure 7. Hypothesis Summary.

Resources appearing in *My Workspace* can be shared with other users by publishing them into the *Public Repository*.

Underlying Technologies

The underlying technologies in order to enable the use of meta-data into the Grid context were crucial for the development of FEARLUS-G. As part of FEARLUS-G Grid services we created a Public Repository service which provides semantic data storage, query and retrieval functions. This service uses ELDAS (Baxter, R., et. al., 2003) to manage data repositories for RDF models; ELDAS was developed by the UK National e-Science Centre eDIKT project. Jena2 is a Java framework for writing Semantic Web applications. It provides programmable access to RDF and OWL sources, ontologies, documents, ontology reasoning and RDF query capabilities. The advantage of enabling Jena2 as a Grid service is that semantic resources can be distributed and used in a dynamic environment. Searching a large collection of RDF resources can be a computationally intensive task; the Grid offers the potential for distributed processing of such queries.

Data access is an important feature in many Grid applications. For this reason the Data Access and Integration Working Group (DAIS-WG) produced the Grid Database Service specification (GDSS). The GDSS presents a specification for a collection of data access interfaces for relational data resources (Antonetti, M. et.al., 2003). The main interfaces described by GDSS provide methods for accessing the data from a Grid service.

ELDAS (Baxter, R. et. al., 2003) is an implementation of the GDSS specifications. ELDAS attempts to overcome some of the issues raised while working with application scientists in several disciplines. The main characteristics of ELDAS are:

- Implementation using J2EE which is machine independent;
- Able to access and integrate data stored in multiple types of data storage system, such as Mysql, Oracle and DB2;
- Accessible as both a Grid service and a Web service.

The main advantage of using ELDAS and the GDSS specification is that they make Grid data services available to the scientific community, resolving issues associated with sharing relational data from different database systems.

There are issues surrounding deployment of Jena2 as a Grid service as it was not designed with this in mind. Jena2 does provide a module which extends the RDF model interface so that it is possible to store and retrieve statements using a database. Although this module supports different database servers such as MySQL, Oracle and PostgreSQL, it needs direct access to the database server and is not designed to operate in a distributed environment such as the Grid.

For these reasons it was necessary to develop a bridge between the Jena2 database module and the ELDAS data access service to allow Jena to communicate with the database across the Grid instead of via a direct connection. JDBC⁷ is a programming interface which allows external access to databases and query operations using SQL. Moreover JDBC allows the integration of database calls with the Java programming environment making database operations simple and intuitive.

⁷ <http://java.sun.com/products/jdbc/>

We have developed a bridging solution, JDBC4ELDAS which allows connections to different databases supported by the ELDAS data access service using standard JDBC library routines. The JDBC4ELDAS driver⁸ enables the Jena2 database module to store models using ELDAS as a Grid data access service. The advantage of using a JDBC bridge driver between Jena and ELDAS is that both applications remain untouched. Moreover the JDBC4ELDAS driver is easily reusable for any Java application that requires Grid data support.

JDBC4ELDAS provides a standard interface for different types of user to access metadata repositories using the most appropriate client interfaces. For example:

- User-scientists can use tailored systems such as our *My Workspace* Web interface to access and manage their meta-data, as discussed in the previous section;
- Computing scientists can use tools such as Protégé to edit ontologies and instance data, as illustrated in Figure 8.

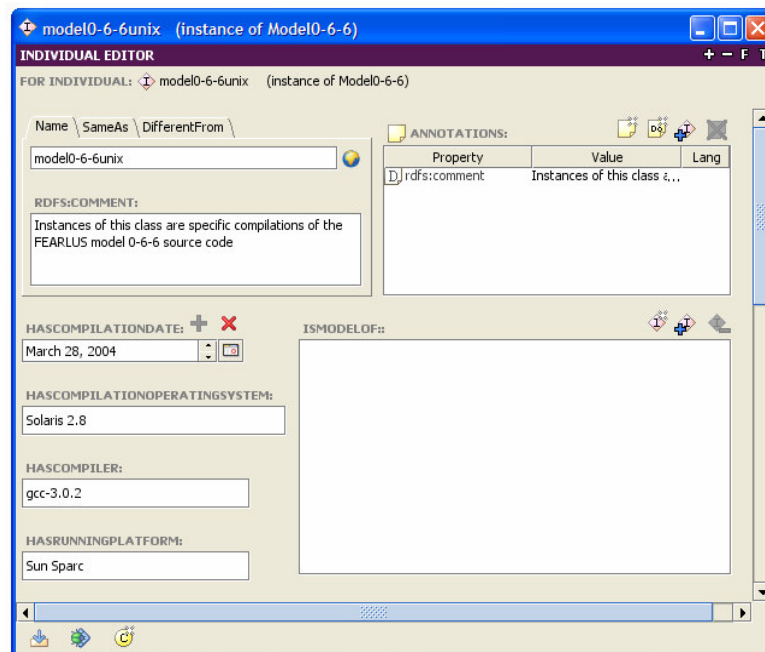


Figure 8. Example FEARLUS-G Instance Rendered via Protégé.

Conclusions

Social scientists must help to shape the Semantic Web and Grid – both in terms of domain ontologies and the new methods of collaboration which these agreed representations will facilitate; Computer Scientists should then provide the tools necessary to exploit the enriched Grid platform. It is unrealistic to expect social scientists and policy practitioners to interact with representations such as RDF or OWL, so mechanisms are needed to “wrap” such formalisms to make them more accessible. In this paper we have discussed a variety of interfaces to allow users to access and manage these kinds of metadata in a structured way. Looking further ahead, we anticipate a growing role for techniques such as natural language generation in enhancing the usability of eSocial science tools.

⁸ The drivers and associated documentation are available at <http://www.csd.abdn.ac.uk/research/fearg/links.php>

Acknowledgment

The project is supported by the UK Economic & Social Research Council (ESRC) under the "Pilot Projects in E-Social Science" programme (Award Reference: RES-149-25-0011).

Bibliography

- Antonetti, M., Krause, A., Hastings, S., Langella, S., Malaika, S., Magowan, J., Laws, S., Paton, N.W. (2003): 'Grid Data Service Specification: The Relational Realisation.' *Global Grid Forum 9*, 2003.
- Baxter, R., Ecklund, D., Fleming, A., Gray, A., Hilld, B., Rutherford, S., Virdee, D. (2003): 'Designing for Broadly Available Grid data Access Services'. In: *UK e-Science All Hands Meeting* (CD-ROM), 2003.
- Berners-Lee, T., Hendler, J., Lassila, O., (2001): 'The Semantic Web'. *Scientific American*, 284, 2001, pp. 28–37.
- Bride, B. (2000): 'Jena: Implementing the RDF Model and Syntax Specification'. *Technical report, Hewlett Packard Laboratories*, 2000.
- Edwards P., Pignotti, E., Preece, A., Polhill, G., Gotts, N. (2005): 'Lessons Learnt from Deployment of a Social Simulation Tool to the Semantic Grid' In: *Proceedings of First International Conference on e-Social Science, 22-24 June 2005, Manchester, UK*,
- Fensel, D. (2003): *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag New York, Inc., 2003.
- Foster, I., Kesselman, C. (1998): 'Globus: A Toolkit-Based Grid Architecture.' In: *The Grid: Blueprint for a Future Computing Infrastructure*. Morgan- Kaufmann, 1998, pp. 259–278.
- Foster, I., Kesselman, C., Tuecke, S. (2001): 'The Anatomy of the Grid: Enabling Scalable Virtual Organizations.' *International J. Supercomputer Applications* ,15(3), 2001.
- Horridge M., Knublauch H., Rector A., Stevens R., Wroe C. (2004): *Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0*. The University Of Manchester, 2004.
- Musen, M.A., Ferguson, R.W., Grosso, W.E., Noy, N.F., Crubzy, M., Gennari, J.H. (2000): 'Component-Based Support for Building Knowledge-Acquisition Systems.' In: *Conference on Intelligent Information Processing (IIP 2000) of the International Federation for Information Processing World Computer Congress (WCC 2000)*, Beijing, 2000.
- Pignotti, E., Edwards P., Preece, A., Polhill, G., Gotts, N. (2005): 'Semantic Support for Computational Land-Use Modelling.' In: *Proceedings of Cluster Computing and Grid 2005*, IEEE Press, 2005.

Polhill, J., Gotts, N., Law, A. (2001): 'Imitative Versus Non-Imitative Strategies in a Land Use Simulation.' *Cybernetics and Systems*, 32 (1), 2001, pp. 285–307.