

# FEARLUS-G: A Semantic Grid Service for Land-Use Modelling

E. Pignotti\*, P. Edwards\*, A. Preece\*, N. Gotts†, and G. Polhill†

\*Dept. of Computing Science, University of Aberdeen  
Aberdeen, AB24 5UE, Scotland  
{epignott, pedwards, apreece}@csd.abdn.ac.uk,  
WWW home page: <http://www.csd.abdn.ac.uk/research/fearg>

†The Macaulay Institute  
Craigiebuckler, Aberdeen, AB15 8QH, UK  
{n.gotts, g.polhill}@macaulay.ac.uk  
WWW home page: <http://www.macaulay.ac.uk/fearlus/>

**Abstract.** The vision of e-science is to facilitate scientific activities, including collaboration on a large scale using Grid technologies. In this paper we explore the use of proposed Semantic Grid standards and methodology through deployment of a land use modelling service. The FEARLUS-G service architecture is presented which allows large scale simulation experiments to be distributed over the Grid. We also discuss ontology support for simulation parameters, hypotheses and results that will facilitate sharing and re-use of such resources among land-use scientists. This leads to a description of infrastructure for semantic data management which integrates Jena2 and the ELDAS data access service.

## 1 Introduction

Collaborations between large groups of scientists are increasingly seen as essential to enhance the scientific process. While research has always involved collaboration between individual scientists, there is now even greater necessity for tools to support sharing of knowledge, resources, results and observations. Scientists already rely extensively on computer and communication technologies to bring together different expertise, using the Web as the main vehicle of communication. While the Web does allow access to distributed data, it does not facilitate managed sharing and coordination of computational resources.

For these reasons recent e-science activities [1] have focused on facilitating and promoting collaboration between scientists using advanced distributed information management systems. The vision of e-science is to facilitate large scale science using Grid technologies [2] as a fundamental computing infrastructure to manage distributed computational resources and data. However a major gap exists between current technologies and the vision of e-science. Where Grid technologies overcome some of the limitations of existing Web tools in terms of managing computational expensive tasks, there is still a need for greater ease of use

and seamless automation to support truly flexible collaboration. For these reasons the concept of a Semantic Grid [3] has emerged, which integrates Semantic Web<sup>1</sup> and Grid technologies.

According to the vision of the Semantic Grid community, next generation Grids should include knowledge discovery and knowledge management functionality for applications and system management [3]. An emerging research field known as ‘Grid Intelligence’ studies ways to acquire, integrate, represent and exchange information available in the Grid to produce useful knowledge; such functionality will be encapsulated in so called Grid Intelligence services [3]. Central to the vision of the Semantic Grid is the adoption of metadata and ontologies [4] to describe resources, services and data sources in order to promote enhanced forms of collaboration among the scientific community. Ontologies and metadata facilitate intelligent search mechanisms, one of the key enablers through which such services could be realised.

The FEARLUS-G project, described in this paper, explores the application of emerging Grid and Semantic Grid technologies within the social sciences, through deployment of an existing land-use modelling tool into the Grid context.

The project is one of a number of pilot studies funded by the UK Economic and Social Science Research Council, under their e-social science<sup>2</sup> initiative. FEARLUS [5] is an agent-based model of land-use change developed at the Macaulay Institute in Aberdeen. The system contains objects that represent human decision-makers in the real-world (land managers) and takes into account attributes such as yield from land parcels. Parameters to the modelling environment allow a variety of land-use strategies and their outcomes to be explored. We chose to deploy FEARLUS as a Grid service for a number of reasons: firstly, the Grid infrastructure allows large simulation experiments to be distributed across the grid to make use of unused processing power; in fact, a typical FEARLUS experiment already consists of running a series of simulations, but on a single machine. Secondly, the use of Semantic Web technologies with Grid infrastructure allows us to create a collaboratory where land-use scientists can access the FEARLUS-G service and share and reuse results and observations.

The goals of our work are thus as follows:

- annotate FEARLUS models, experiments and hypotheses with relevant metadata, providing a common vocabulary and shared meaning, so they can be shared;
- describe FEARLUS as a Grid service, so that remote users can discover and invoke it, and obtain interpretable results by making use of the semantic resources;
- manage the Grid service, to control user access and allocation of resources;
- maintain histories of interactions with users, allowing experiments to be replayed and results to be aggregated.

In the remainder of this paper we briefly discuss simulation modelling and in particular the FEARLUS model of land-use simulation. We then describe

<sup>1</sup> <http://www.w3.org/2001/sw/>

<sup>2</sup> <http://www.ncess.org/>

the FEARLUS-G service architecture presenting its various components. We also introduce an ontology to support description of simulations, hypotheses and experiments. As a contribution to infrastructure for the Semantic Grid we present a solution which provides storage, query and retrieval of semantic data to allow a virtual community of land-use scientists to share and reuse FEARLUS-G experimental results and observations.

## 2 Simulation Modelling of Land-Use Change

Using computer technologies to study social and economic phenomena is now commonplace [6]. Such processes are often studied via simulation modelling [7]. In such simulations, the model consists of a representation of the structure and behaviour of a particular real world entity which we wish to study. One of the advantages of modelling is that the results obtained are repeatable. This allows models to be shared among the scientific community in order to be analysed and reused. A simulation consists of running the model under a specific set of circumstances defined by a parameter set and then analysing the outcome. The aim of the simulation is to construct a model where behaviour matches the real entity in at least a significant aspect. By constructing such a model, social scientists aim to develop conclusions that provide insight into the behaviour of real world entities or phenomena; such modelling is often exploratory. Alternatively the simulation may be used to confirm how reliable a predicted behaviour is under certain key conditions which may or may not be under direct control.

There are often many competing models available, and the problem then is how to choose between different types of model in a particular problem context. It may also be difficult to interpret the behaviour from a model even when it has been previously validated experimentally. Ideally a model should be structured in such a way that it is possible to determine if a particular question can be answered using it.

### 2.1 The FEARLUS Model

There has recently been a proliferation of computer models of land-use change and water management; many of these are spatially explicit models, in which a set of distinct localities, and spatial relationships between them, are directly represented [8]. As computational resources, and the availability of machine-readable georeferenced data increase, growth in spatially explicit modelling within these domains is likely to continue. Attention is now turning to how such models can most effectively be used, both in management and policy-related applications, and in social and geographical science.

In FEARLUS Model 0-6-5 parameters to the modelling environment allow a variety of land-use strategies and their outcomes to be explored. FEARLUS aims to improve understanding of land use change, particularly as regards rural Scotland. The agent-based simulation component in FEARLUS is implemented using the Swarm system developed at the Santa Fe Institute [9]. Swarm provides

an environment that can handle large experiments using agent-based simulation models. A FEARLUS simulation might involve for example, studying the dynamics of imitative<sup>3</sup> and non-imitative approaches to land use selection change under different circumstances, in the context of environments differing in spatial and temporal heterogeneity. This involves an initial set of exploratory studies where the model is run in a simulation with different initial parameters and the outcomes observed. The result of the exploratory study is the formulation of a hypothesis regarding patterns of behaviour of specific aspects of the land use model. Figure 1 shows the graphical output of an interactive run of the FEARLUS model. It uses graphical output to render different aspects of the model in order to facilitate exploratory studies. Using experimental validation it is then possible to check whether the model does consistently show the patterns of behaviour suggested by the exploratory studies. Further details of the FEARLUS model can be found here [5].

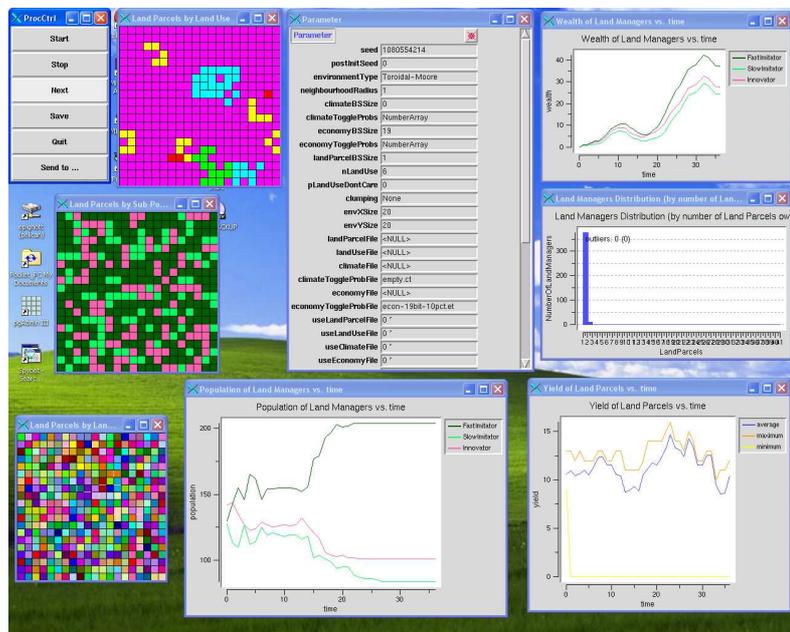


Fig. 1. The FEARLUS Desktop Application.

<sup>3</sup> Imitation is a social phenomenon, e.g. land managers may be influenced by other managers owning a neighbouring land parcel.

### 3 FEARLUS-G Grid Service

The FEARLUS-G Grid service extends the existing desktop application (see Figure 1) by enabling land use simulations to be executed on the Grid. FEARLUS experiments consist of a set of sequential simulations with different initial parameters. Experiments running on a single machine can take a considerable time depending on the number of simulations necessary. FEARLUS-G distributes the load of an experiment across different Grid services depending on the resources available. Furthermore, FEARLUS-G facilitates the reuse of model parameters and experiments using Semantic Web technologies. Ontology support is used in order to maintain histories of interactions with users, allowing experiments to be replayed and results to be aggregated.

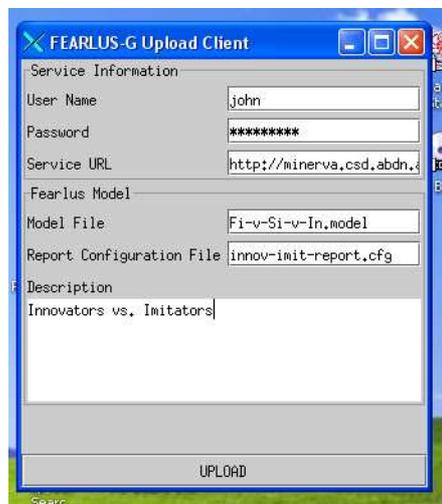
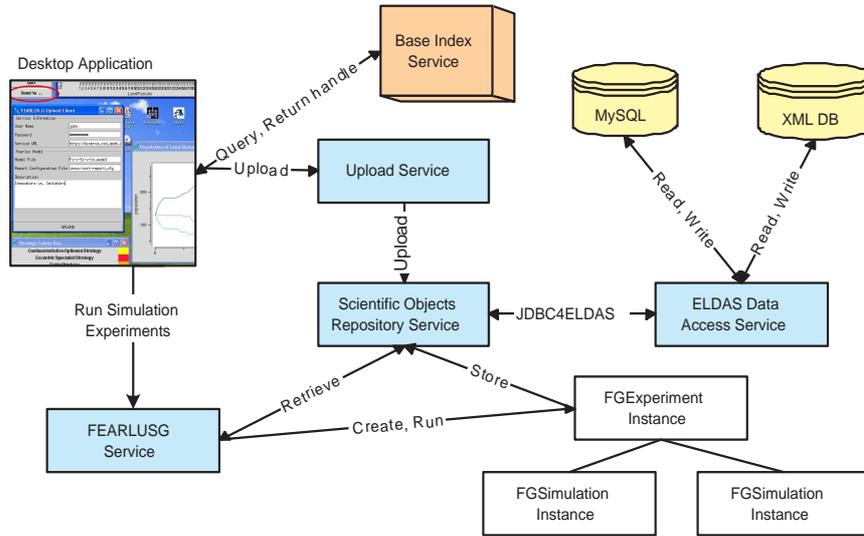


Fig. 2. The FEARLUS-G Desktop Client.

Figure 3 presents an overview of the FEARLUS-G architecture, which is built on top of Globus Toolkit 3 (GTK3). The five core components are as follows:

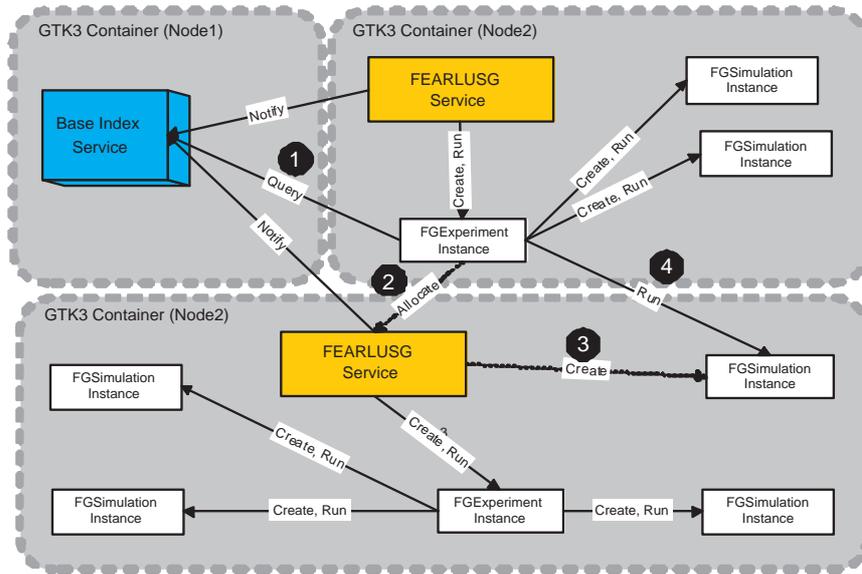
1. **FEARLUS-G Service:** The access point for FEARLUS-G simulations and experiments. This service creates experiment or simulation instances by identifying the components inside an experiment definition.
2. **Upload Service:** Allows the client application to upload Scientific Objects to the Grid application. Scientific Objects are defined in RDF [10] using the ontologies described in Section 4 of this paper. Figure 2 shows an extension of the FEARLUS Desktop application which allows a user to upload a simulation model into the Grid service.
3. **Scientific Objects Repository Service:** Stores and retrieves the model parameters and experiment definitions acquired from the Upload Service



**Fig. 3.** The FEARLUS-G Service Architecture.

- into persistent storage. Integrates the management of data and semantic metadata as described in Section 5.
4. **FEARLUS-G Experiment:** Allocates different simulation instances depending on the characteristics of the experiment. This service also collects the results from the various simulations and stores the results.
  5. **FEARLUS-G Simulation:** Performs the simulation using FEARLUS model 0-6-5 and stores the results. It creates FEARLUS runs, and store the results in the appropriate simulation instance in the Scientific Objects Repository.

FEARLUS-G uses the Globus base index service to identify available services through their service descriptors. Experiments and simulations can be allocated to different nodes depending on the resources available. In our approach, users provide specifications of the experiments and the initial parameters for the simulation. This information is stored using the Scientific Objects Repository. This can be done by a client application integrated within the FEARLUS desktop client (see Figure 1) or through a Web portal designed for FEARLUS-G (see Figure 6). The FEARLUS-G environment contains the middleware to find resources available and the input data required to perform a specific simulation by making use of a service descriptor associated with the FEARLUS-G service. The service descriptor contains the characteristics of the resources available in the node such as the maximum number of simulations and the simulation instances available. In order to make this possible each distributed node with a FEARLUS-G service instance notifies the base index service whenever there is a change in their status. Grid services or user applications that need a FEARLUS-G service can query the base index service to obtain a handle for an appropriate service depending on the size of the experiment they need to perform.

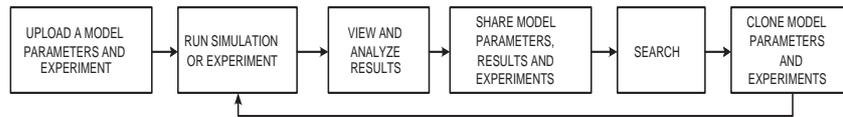


**Fig. 4.** Allocation of a FEARLUS-G Simulation.

Once a request to perform an experiment has been allocated to a FEARLUS-G service, the system allocates simulation instances to perform the various parts of an experiment and collect the results. Figure 4 illustrates how an experiment and its related simulations are allocated in FEARLUS-G:

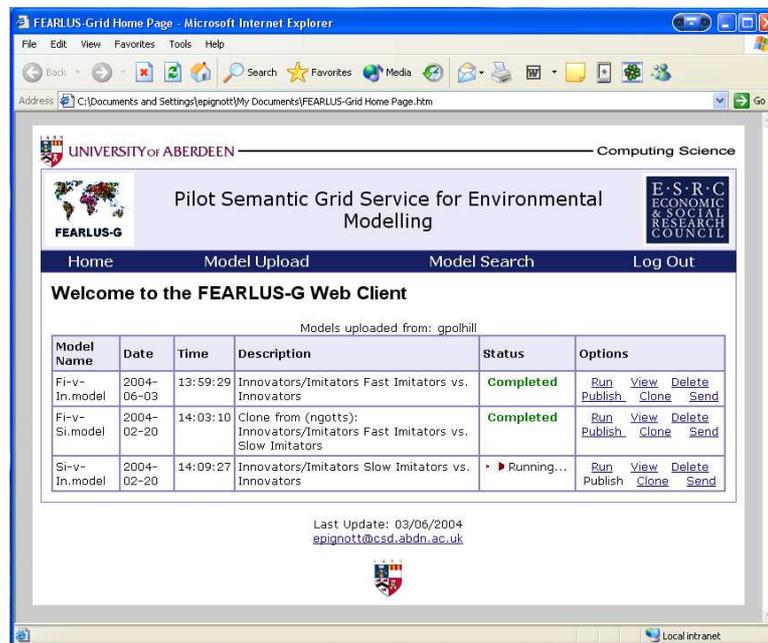
1. The experiment instance queries the base index service to obtain the location of a FEARLUS-G service with available resources. If resources are available on the same node the experiment instance creates instances of a FEARLUS simulation.
2. If the resources available are on a different node the experiment instance sends a request to the remote FEARLUS-G service instance to allocate a simulation instance;
3. The remote FEARLUS-G service creates the simulation instance and returns the handle to the original experiment instance;
4. The experiment instance then uses the newly created remote simulation instance to run part of its experiment.

As explained earlier, FEARLUS-G aims to facilitate sharing and reuse of models and experiments among the land use scientific community. For this reason the Scientific Objects Repository service can be used to search for and clone existing model parameters and experiments. A social scientist can use the cloned model or experiment to further investigate the specific problem addressed within it. For example, they could use an experiment to disprove a previous hypothesis or reuse a model for a different research problem. Figure 5 shows a typical experiment cycle.



**Fig. 5.** The FEARLUS-G Experiment Cycle.

Figure 6 shows the existing FEARLUS-G Web client<sup>4</sup> through which users can perform experiments and share results. The metadata used to describe models, hypotheses, results and experiments within the service environment allows annotations to specify proprieties of models and experiments, including creator, contributors and supporting evidence (publications).



**Fig. 6.** FEARLUS-G Web Client.

## 4 Representing Simulations, Hypotheses and Experiments

As described in Section 2, the classical approach to simulation studies is to create a model of a real entity in order to analyse hypothetical behaviours. Simulations

<sup>4</sup> <http://minerva.csd.abdn.ac.uk:8081/fearg/>

are used to generate numerical results representing future states of the model under specific conditions in order to support or refute a hypothesis. In this section we propose a conceptual layer through the definition of an ontology that captures the concepts and relationships important to scientists performing their research activities.

The key class in our representation is the *Hypothesis*, which we consider to be a scientific concept that has not yet been fully verified. An hypothesis is “a tentative explanation that accounts for a set of facts and can be tested by further investigation”<sup>5</sup>. This implies that a scientific community works to support or refute a hypothesis by contributing publications, experiments and other related hypotheses. Figure 7 shows the core classes and relationships (properties) in our ontology. It includes a collection of generic elements that are intended to be applicable to any e-science application, and also elements that are specific to simulation modelling and FEARLUS-G in particular. In our vision, the classes *Hypothesis*, *Experiment* and *Publication* are subclasses of a generic *ScientificObject* class. We define two important properties **supportedBy** and **refutedBy** which link any given hypothesis to those *ScientificObject* instances which offer support for or refute that hypothesis.

We use the Dublin Core<sup>6</sup> ontology to provide basic annotation of *ScientificObject* instances. For example **dc:creator** defines the creator of an hypothesis or experiment, **dc:contributor** the contributor(s). Figure 8 shows an example. We also aim to provide support for refinement of hypotheses; in particular we use **dc:replaces** to represent that a new hypothesis replaces another.

Another important attribute is **describedIn**. This is used to record that a publication can be used to describe experiments, hypotheses and models. We define *SimulationExperiment* as a subclass of *Experiment* to extend our generic ontology to include a FEARLUS experiment. A typical experiment in FEARLUS compares a number of subpopulations by making them compete in the simulation environment. In our ontology a simulation experiment includes a set of simulations; specifically we define the more specialized class *FEARLUSSimulation* that captures the specific attributes necessary to generate a FEARLUS run.

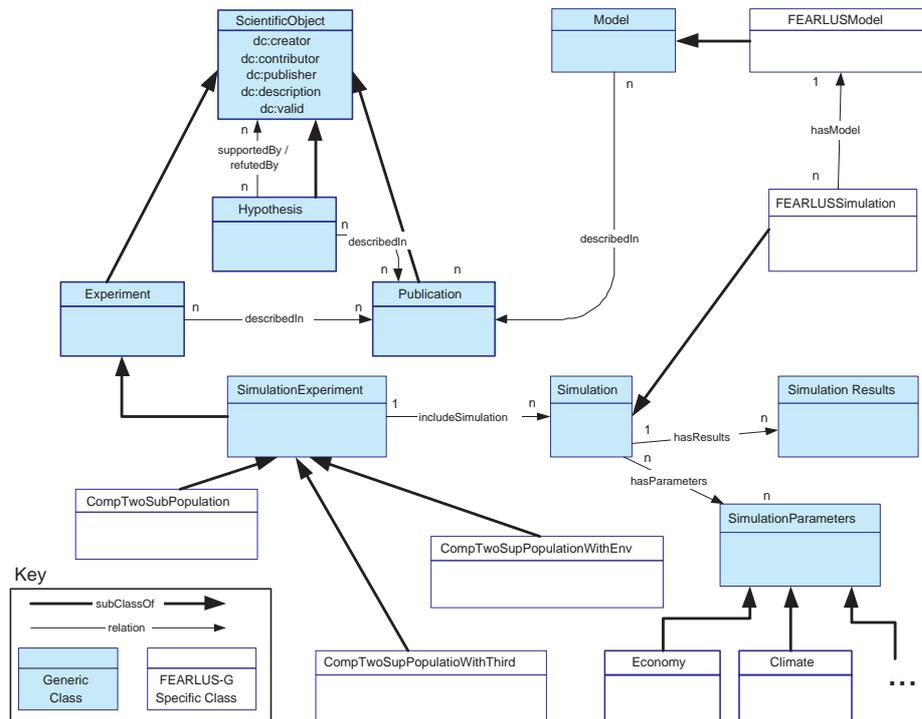
## 5 A Grid Enabled Semantic Data Service

In the previous section we introduced a possible ontology starting point for supporting scientific research via the Semantic Grid. We plan to use this ontology in the Grid context to make use of the potential of both technologies by sharing computational and data resources across different hosts, with the semantic metadata providing support for management of the shared resources.

As part of the FEARLUS-G project we are developing an open source reusable semantic data service based on Jena2 [10] and Globus Toolkit 3 [11] to provide semantic data storage, query and retrieval functions. This service uses ELDAS [12]

<sup>5</sup> [www.pages.drexel.edu/~bcb25/scimeth/vocabulary.htm](http://www.pages.drexel.edu/~bcb25/scimeth/vocabulary.htm)

<sup>6</sup> <http://dublincore.org/>



**Fig. 7.** FEARLUS-G Ontology.

to manage data repositories for RDF models; ELDAS was developed by the UK National e-Science Centre eDIKT project<sup>7</sup>.

Jena2<sup>8</sup> 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 [13]. 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. The following subsections describe these technologies and how we combine them together.

### 5.1 ELDAS Data Access Service

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 [14]. The main

<sup>7</sup> <http://www.edikt.org>

<sup>8</sup> <http://jena.sourceforge.net/>

```

<hyp:Hypothesis rdfs:about="hyp_1">
  <dc:title> Fast imitators do better than innovators </dc:title>
  <dc:creator rdf:resource="http://www.csd.abdn.ac.uk/research/fearg/GaryPolhill"/>
  <hyp:cloneFrom rdf:resource="http://www.csd.abdn.ac.uk/research/fearg/hyp_0"/>
  <dc:contributor rdf:resource="http://www.csd.abdn.ac.uk/research/fearg/NickGotts"/>

  <!-- Innovators/Imitators Slow Imitators vs. Innovators simulation experiment -->
  <hyp:supportedBy rdf:resource="http://www.csd.abdn.ac.uk/research/fearg/exp_10"/>

  <!-- Imitative versus non imitative paper -->
  <hyp:supportedBy rdf:resource="http://www.csd.abdn.ac.uk/research/fearg/pub_15"/>

  <hyp:refutedBy rdf:resource="http://www.csd.abdn.ac.uk/research/fearg/pub_13"/>

  <hyp:describedIn rdf:resource="http://www.csd.abdn.ac.uk/research/fearg/pub_15"/>
</hyp:Hypothesis>

```

**Fig. 8.** Example of an Hypothesis Instance.

interfaces described by GDSS provides methods for accessing the data from a Grid service.

ELDAS [12] 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.

## 5.2 Integrating Jena2 and ELDAS

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.

JDBC<sup>9</sup> is a programming interface which allows external access to databases and query operations using SQL. Moreover JDBC allows the integration of

<sup>9</sup> <http://java.sun.com/products/jdbc/>

database calls with the Java programming environment making database operations simple and intuitive. 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. Figure 9 illustrates the use

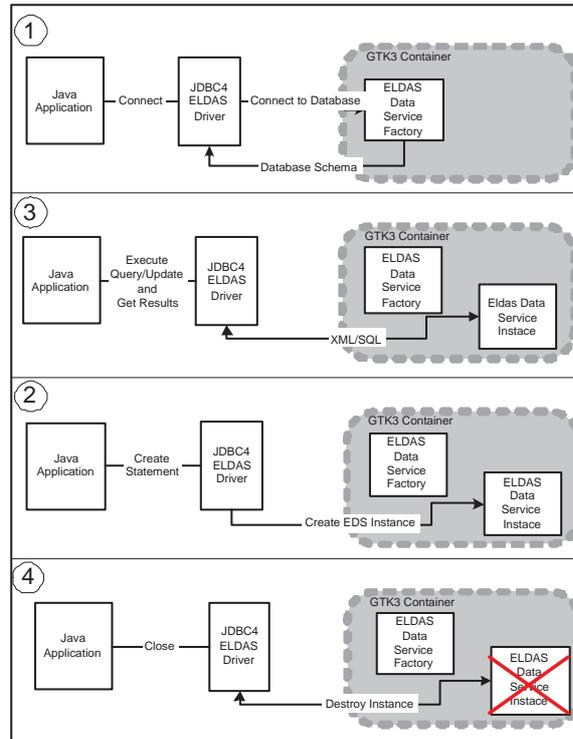


Fig. 9. JDBC4ELDAS Usage Summary.

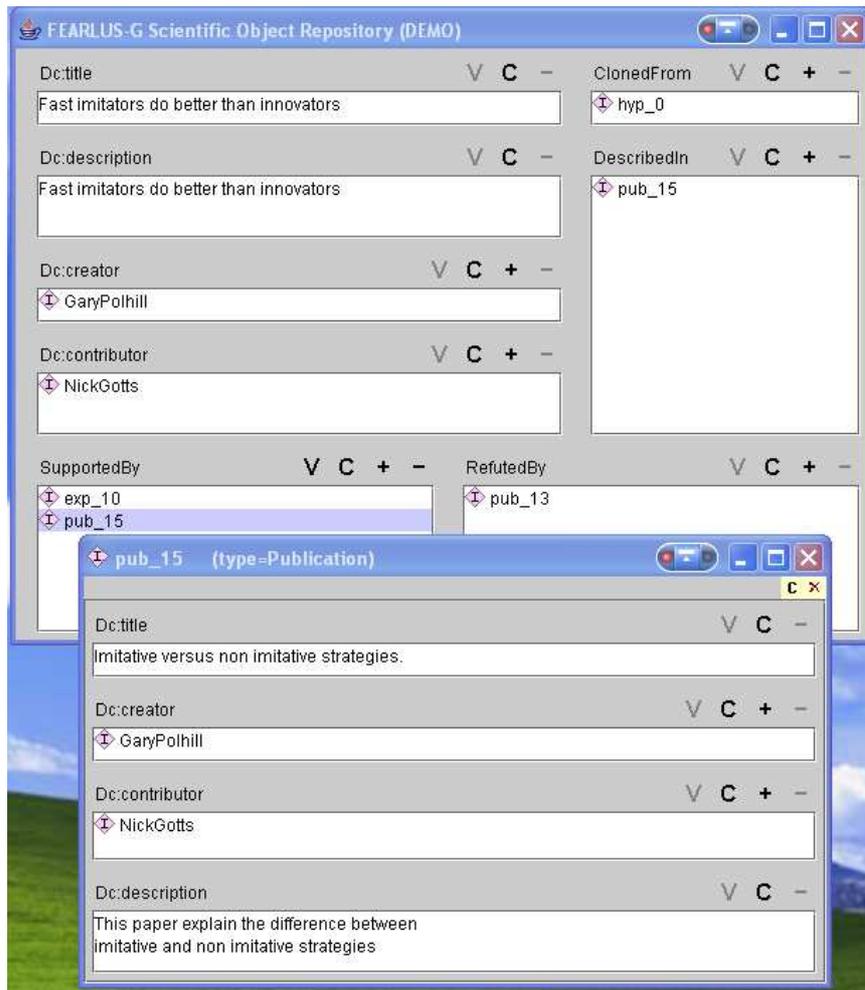
of JDBC4ELDAS.

The JDBC4ELDAS driver<sup>10</sup> 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. We have developed a sample client application using Protégé<sup>11</sup> in order to test the JDBC4ELDAS driver and to explore the potential of Protégé as a visualization/editing tool for the Grid data storage solution. Protégé allows a user to create and edit ontologies and instances

<sup>10</sup> The drivers and associated documentation are available at <http://www.csd.abdn.ac.uk/research/fearg/links.php>

<sup>11</sup> <http://protege.stanford.edu/>

which can be stored in a file or in a database using OWL, RDF or Protégé's own application specific format. The input of new instances is performed through customisable forms automatically generated from the ontology. We configured the existing Protégé database module to use JDBC4ELIAS simply by indicating the name of the driver class in a configuration file; using Protégé in this manner we can easily access our existing ontology and instances. Figure 10 shows an instance of an Hypothesis class (hyp\_1) and an instance of a Publication class (pub\_15) from the FEARLUS-G data repository rendered via a customized Protégé form.



**Fig. 10.** Example of a FEARLUS-G Hypothesis Instance Rendered via Protégé.

## 6 Discussion

In our work to date we have created a set of services to perform land use simulation in the Grid context based on the existing FEARLUS land use model and utilising Globus Toolkit 3. We have developed a Web client application that makes use of the services by uploading and running FEARLUS simulations; the services are also accessible through an extended version of the existing FEARLUS desktop client. There are three main contributions from our work to date. Firstly, our Grid service provides land use scientists with a means to run much larger-scale experiments than previously possible on standalone PCs, and also gives them a Web-based environment in which to share simulation results. Secondly, we have defined an initial collection of ontology elements that describe the scientific objects necessary to enable collaboration between a community of e-social scientists. Our initial ontology represents generic scientific concepts such as hypotheses and experiments, as well as more domain-specific concepts tailored to the use of FEARLUS. The ontology is currently expressed somewhat loosely in RDF, but we plan to tighten the definitions by adding OWL statements in future. Thirdly, we have created a JDBC bridge between Jena2 and ELDAS to enable Grid applications to easily use RDF-based semantic metadata. This bridge also facilitates management and searching of very large collections of RDF resources in the Grid context.

Our future work will focus on refinement and expansion of our initial ontologies, and the further development of our Grid services to promote collaboration among social scientists. We also aim to test and analyse the effectiveness of our technologies with user scientists. We have engaged a group of “assessors” from the international academic communities interested in agent-based social simulation, and in land-use and water management modelling. These assessors have agreed to use the FEARLUS Grid service either in a research capacity (comparing results from their own models with those from FEARLUS, checking results in FEARLUS publications, or exploring FEARLUS’s capabilities and limitations as a stage in the process of designing a new model), or as a teaching tool, or both. Evaluation of the FEARLUS-G service is expected to provide more general insight into the effectiveness of current Semantic Grid technologies and methodologies.

## 7 Acknowledgements

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).

## References

1. Roure, D.D., Jennings, N., Shadbolt, N.: Research agenda for the semantic grid: A future e-science infrastructure. Technical report, UK e-Science Series UKeS-2002-02, National e-Science Centre, Edinburgh, UK. (2001)

2. Foster, I., Kesselman, C., Tuecke, S.: The anatomy of the grid: Enabling scalable virtual organizations. *International J. Supercomputer Applications* **15(3)** (2001)
3. Roure, D.D., Jennings, N., Shadbolt, N.: The semantic grid: A future e-science infrastructure. in *grid computing: Making the global infrastructure a reality*. Anthony J.G. Hey and Geoffrey Fox. John Wiley & Sons Anthony J.G. Hey and Geoffrey Fox. John Wiley & Sons (2003) 437–470
4. Fensel, D.: *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag New York, Inc. (2003)
5. Polhill, J., Gotts, N., Law, A.: Imitative versus non-imitative strategies in a land use simulation. *Cybernetics and Systems* **32 (1)** (2001) 285–307
6. Doran, J.E.: Simulating societies using distributed artificial intelligence. In *Social Science Microsimulation* (eds. Troitzsch K G, Mueller U, Gilbert G N and Doran J E ). Springer: Berlin. (1995) 381–393
7. McHaney, R.: *Computer Simulation A Practical Perspective*. Academic press. (1991)
8. Irwin, E., Geoghegan, J.: Developing spatially explicit economic models of land use change. *Agriculture, Ecosystem and Environment* **85** (2001) 7–23
9. Minar, N., Burkhart, R., Langton, C., Askenazi, M.: The swarm simulation system, a toolkit for building multi-agent simulations. SFI Working Paper 96-06-042, Santa Fe Institute. (1996)
10. McBride, B.: Jena: Implementing the rdf model and syntax specification. Technical report, Hewlett Packard Laboratories (Bristol) (2000)
11. Foster, I., Kesselman, C.: Globus: A toolkit-based grid architecture. In: *The Grid: Blueprint for a Future Computing Infrastructure*. Morgan-Kaufmann (1998) 259–278
12. Baxter, R., Ecklund, D., Fleming, A., Gray, A., Hilld, B., Rutherford, S., Virdee, D.: Designing for broadly available grid data access services. In: *UK e-Science All Hands Meeting (CD-ROM)*. (2003)
13. Seaborne, A.: Rql - a query language for rdf. Technical report, Hewlett Packard Laboratories (2004)
14. Antonetti, M., Krause, A., Hastings, S., Langella, S., Malaika, S., Magowan, J., Laws, S., Paton, N.W.: Grid data service specification: The relational realisation. *Global Grid Forum* 9 (2003)