

Lessons Learnt from Deployment of a Social Simulation Tool to the Semantic Grid

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

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

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

Abstract. In this paper we discuss the lessons learnt from the deployment of FEARLUS, an agent-based model of land-use change, into the Grid Context. We outline the requirements for improving the scientific rigour of agent-based modeling and discuss the solutions adopted. This leads to a description of the infrastructure for managing distributed experiments across the grid and support for semantic data management.

Overview

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. This paper describes work involving social scientists at the Macaulay Institute in Aberdeen investigating land-use change and computer scientists at the University of Aberdeen. Central to this collaboration is FEARLUS (Polhill, J. et. al., 2001), an agent-based model of land-use change which 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.

One of the key justifications for our work is improving the scientific rigour of agent-based modelling. One of the important aspects of science is that work should be repeatable and verifiable. Yet results gathered from possibly hundreds of thousands of simulation runs cannot be reproduced conveniently in a journal publication. Equally, the source code of the simulation model, and full details of the model parameters used are also not journal publication material. We have identified four levels of repeatability that are of relevance to such situations:

- The first level is being able to access the results data itself, to check that the authors' claims (hypotheses) that are based on those results are justifiable.
- The second level is being able to re-run the experiments to check that they produce broadly the same set of results, i.e. that the researchers have not carefully selected runs of their simulation that support their claims and ignored the rest.
- The third level is being able to manipulate the simulation model parameters and re-run the experiments to check that there is no undue sensitivity of the results to certain parameter settings.
- The fourth level is being able to modify the source code and/or re-implement the model to check for what might be termed 'algorithmic sensitivity'.

In this paper we describe lessons learnt from our experience of deploying FEARLUS into the Grid (Foster, I., et. al., 2001) context, and the use of Semantic Web (Berners-Lee, et. al., 2001) technologies to facilitate the scientific process. Where Grid technologies are ideal for social simulation in terms of managing computational tasks, there is still a need for greater ease of use and seamless automation to support truly flexible collaboration. For these reasons we have adopted emerging Semantic Grid (Roure, D. et. al., 2003) technologies, which integrate Semantic Web and Grid technologies. Central to the Semantic Grid approach is the use of meta-data and ontologies¹ to describe resources, services and data sources in order to promote enhanced forms of collaboration among the scientific community.

Simulation Modelling of Land-Use Change

Using computer technologies to study social and economic phenomena is now commonplace (Doran, J.E., 1995). Such processes are often studied via simulation modeling (McHaney, R., 1991). 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 allowing further analysis and reuse. 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.

¹ Ontologies 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.

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 (Irwin, E., Geoghegan, J., 2001). 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 (Minar, N. et. al., 1996). 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 imitative² 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.

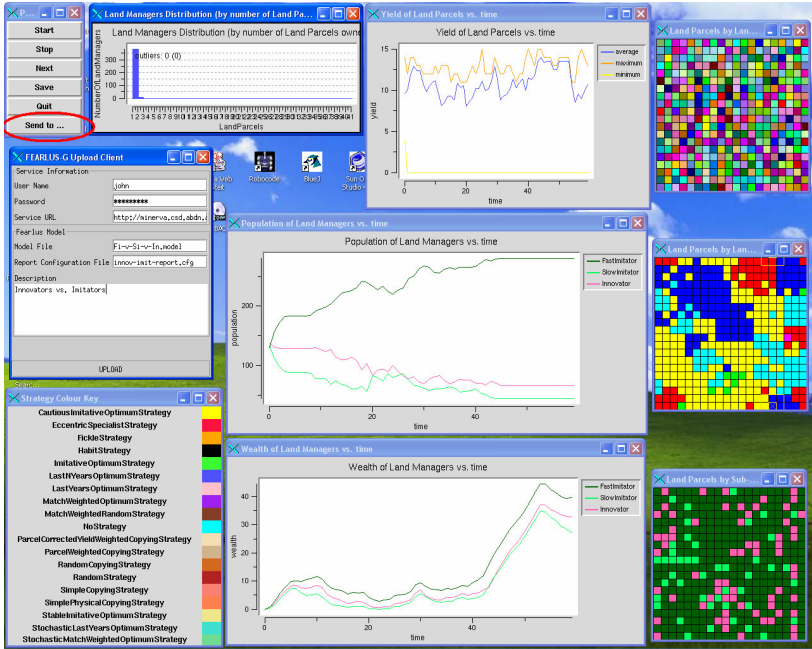


Figure 1. The FEARLUS Desktop Application.

² Imitation is a social phenomenon, e.g. land managers may be influenced by other managers owning a neighbouring land parcel.

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 are available (Polhill, J et. al., 2001).

FEARLUS-G

We have developed FEARLUS-G (Pignotti, E. et. al. 2005a) – a collection of services to perform land use simulation in the Grid context based on the existing FEARLUS land use model and utilising Globus Toolkit 3³ (Foster, I., Kesselman, C., 1998). We have also developed a Web client application (*My Workspace*, see Figure 2) 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.

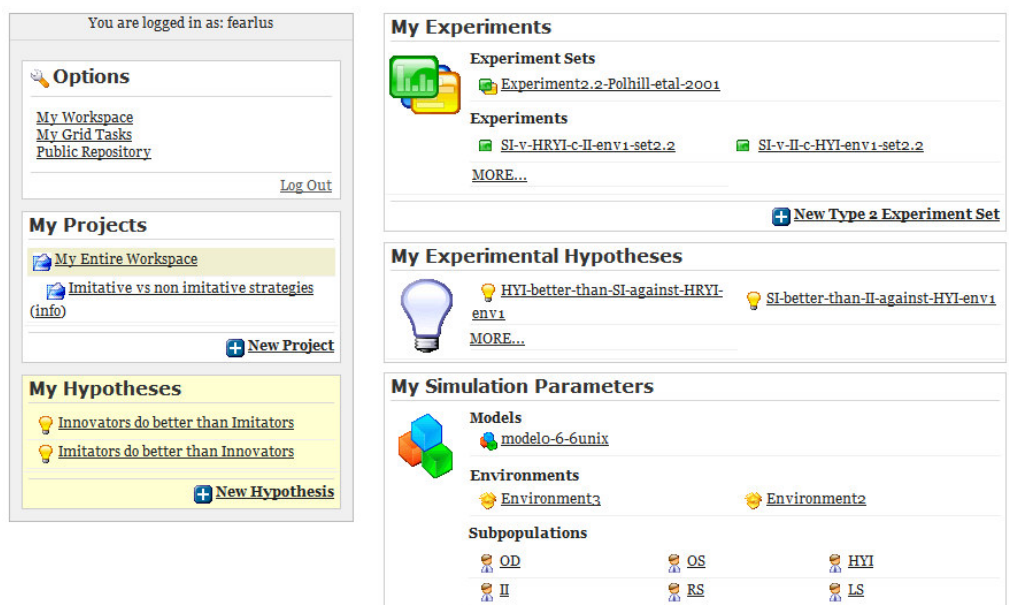


Figure 2. The *My Workspace* interface.

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 involves running a series of simulations, but on a single machine. Secondly, the use of meta-data (Pignotti, E. et. al. 2005b) combined with Grid infrastructure allows us to create a co-laboratory where land-use scientists and others interested in rural policy can access the FEARLUS-G service and share and reuse results and observations.

³ <http://www.globus.org/toolkit/>

Figure 3 presents an overview of the FEARLUS-G system architecture, which is built on top of OGSA 3.2.1 (Globus Toolkit 3) and Web services⁴.

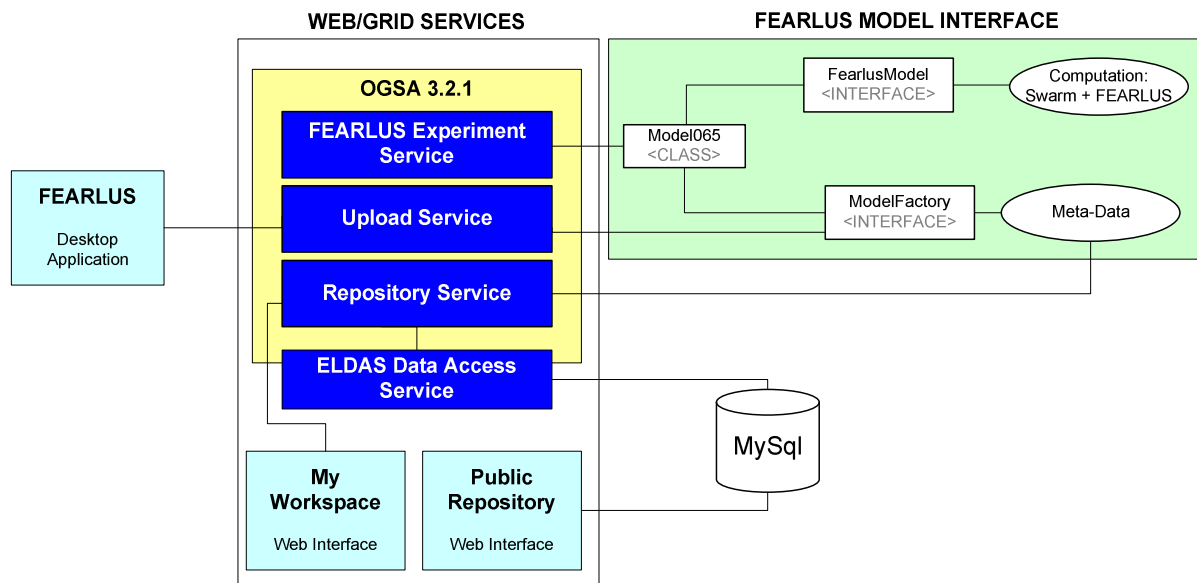


Figure 3. FEARLUS-G System Architecture.

The FEARLUS-G system is composed of two main components: The Web/Grid Services component and the FEARLUS Model Interface component.

The Web/Grid Services component contains the core services that manage access to the computational and data resources. These services are as follows:

- The *FEARLUS Experiment Service* creates simulation runs by identifying the components inside an experiment definition. A simulation run is created by using the FEARLUS Model Interface component which communicates with the FEARLUS agent-based simulation. Moreover, this service distributes the load of an experiment across different Grid services depending on the resources available. Results from the simulation runs are stored using the *Repository Service*.
- The *Upload Service* allows the existing FEARLUS Desktop application to upload model parameters into persistent storage through the *Repository Service*.
- The *Repository Service* allows the other services to store and retrieve meta-data using the *ELDAS Data Access Service* described in Pignotti et. al. 2005b.
- *My Workspace* (Figure 2) is a Web-based interface that allows land-use scientists to manage simulation activities. The resources generated in *My Workspace* are owned by a single scientist and are not visible to other users.
- The *Public Repository* interface is a meta-data browser based on Longwell (Mazzocchi S., 2005), developed by the SIMILE project⁵. Users can browse

⁴ <http://www.w3.org/2002/ws/>

⁵ <http://simile.mit.edu/longwell/>

publicly available scientific resources and clone instances for further study in their own workspace.

The Fearlus Model Interface component contains the interface between the Web/Grid services and the existing FEARLUS agent-based simulation software. This interface serves a dual purpose: firstly, it converts the meta-data from the Grid services into low-level simulation parameters and uses those parameters to perform a simulation run with the FEARLUS agent-based simulation software; secondly, it converts the numerical results from the simulation back into meta-data to be stored in the *Repository Service*.

The meta-data managed in FEARLUS-G is supported by an ontology which captures the concepts and relationships important to scientists performing their research activities. The ontology 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 (see Figure 4). Further details of this ontology and the relevant supporting technologies can be found in Pignotti et. al. 2005b.

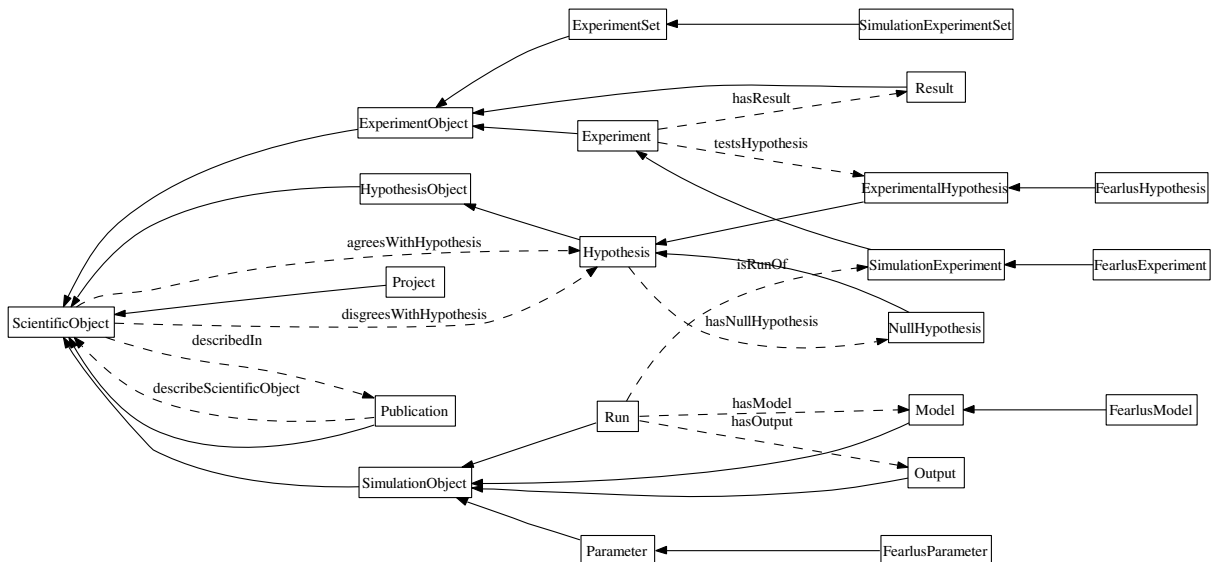


Figure 4. Scientific Object Ontology.

Instances generated from the classes in the ontology are of three types: those which need to be created by the user, e.g. the definition of an experiment; those instances created by the Grid services, e.g. the set of simulation run instances associated with an experiment, and the experiment result instance summarizing the results; the third group are instances created to support scientific argumentation, sharing and reuse. For example, an hypothesis can be associated with related publications and experiments.

The key class in our generic 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". This implies that a scientific community works to support an hypothesis by contributing publications, experiments and other related hypotheses that agree or disagree with the "target"

hypothesis. In our vision, the classes HypothesisObject, ExperimentObject, SimulationObject and Publication are subclasses of a generic ScientificObject class.

We define two properties *agreesWithHypothesis* and *disagreesWithHypothesis* which link any given scientific object to Hypothesis instances. Those properties allow us to support scientific argumentation in the FEARLUS context by linking specific evidence (FearlusExperiment, Run, FearlusResult) to hypotheses. The generic scientific object representation can, of course, be adapted for different scientific domains by extending it with more specific subclasses and properties.

As part of the FEARLUS-G project we have developed an open source reusable semantic data service (Pignotti, E. et. al. 2005b) based on Jena2 (McBride, B., 2000) and Globus Toolkit 3 to provide 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 project⁶.

Jena⁷ 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 (Seaborne, A., 2004). 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.

Related Work

We have identified a number of activities that share important characteristics with FEARLUS-G.

The MyGrid project (Stevens, R., et. al., 2003; Lord, P., et. al., 2004) enables biologists to perform and manage *in silico* experiments by providing a problem-solving workbench based on Semantic Grid technologies. User scientists in MyGrid browse a registry and select a workflow template. The system then asks the user to manually select instances of the services associated with the workflow or automatically selects services based on a user profile (user preferences). The MyGrid services collect and co-ordinate data inputs and outputs for the experiment, as well as provenance information about the performance of experiments. MyGrid develops the use of workflow ontologies to capture web-based procedures and provenance ontologies (Zhao, J., et. al., 2004) to describe how and why results were produced.

Majithia et. al. (2004), propose a system which facilitates automated synthesis of scientific experiment workflows, based on a high-level goal specification. This approach differs from MyGrid as it uses a composition algorithm to adapt available resources on the Grid to high level objectives defined by the user. Moreover, the system makes a distinction between different levels of abstraction of the workflow in order to allow reuse and sharing.

The two systems described above focus on provenance (MyGrid) and workflow abstraction (Majithia) to support sharing and reuse of scientific process information. However, in both systems scientific argumentation is not supported; experiments and results are not associated

⁶ <http://www.edikt.org>

⁷ <http://jena.sourceforge.net/>

with hypotheses. In our opinion the FEARLUS-G argument structure approach is likely to be more intuitive to user scientists, reflecting as it does the evidence-based approach widely used within the scientific community.

SemanticOrganizer (Keller, R.M., 2004) is a knowledge management system designed to support distributed NASA projects, including multidisciplinary teams of scientists, engineers and accident investigators. SemanticOrganizer provides a semantically-structured information repository that serves as a common access point for all work products related to an ongoing project. The repository not only stores files but also metadata describing domain concepts. The system allows accident investigators to collect and manage evidence, perform different type of analyses that generate derivative evidence, connect the evidence together to support or refute accident hypotheses. SemanticOrganizer is implemented using standard Web services for both the repository and the user interface. This project successfully adopts Semantic Web technologies to support construction of an evidence-based argumentation structure. While this approach does have much in common with the FEARLUS-G solution described in this paper, it does not of course provide support for execution of large-scale experiments.

Other relevant Semantic Grid activities include The Smart Tea Project (Hughes, G., 2004) which aims to provide end to end support for data capture and access for experimental scientists by replacing conventional lab notebooks. The Smart Tea ontology supports experimental plans and records, provenance information, and semantic links between data. The myTea Best Practice project (Brostoff, S., 2005) is implementing the lessons learnt from the Smart Tea project within myGrid.

Lessons Learnt

Through our experiences to date we have learnt valuable lessons about the provision of Semantic Grid tools to the computational land-use modelling community; these are now discussed.

Our Grid service provides scientists interested in land-use phenomena 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. The Grid is clearly a valuable tool for managing the computation involved in running large-scale experiments. In fact the Globus environment is ideal for simulation modelling because the cost of a simulation run is less than the cost of the communication necessary to run a distributed experiment and aggregate the results. However, the Globus environment is not well-suited for the fine grained exploratory work that the existing FEARLUS desktop application is able to achieve. Scientists still want to use the patterns of workflow related to the desktop application. For this reason, we have extended the existing desktop application to be able to upload models to the Grid service, creating a link between the exploratory tool and the Grid experiment platform.

We have defined an initial collection of ontology elements that describe the scientific objects necessary to enable collaboration between members of a community of land-use scientists. Our ontology represents generic scientific concepts such as hypotheses and experiments, as well as more domain-specific concepts tailored to the use of FEARLUS. Social scientists are able to run FEARLUS experiments on the Grid and collect the results; a private interface (*My Workspace*) allows them to manage their activities (experiments, investigations, etc.) while their work is in progress. Scientific resources can also be made public, allowing others to investigate them further. We are currently exploring the effectiveness of this approach with a group of assessors taken from the scientific community (see below).

Not all categories of users require access to the same meta-data. For example, we have exposed FEARLUS model parameters to land-use scientists familiar with the existing FEARLUS application in such a way that the complexity of the Grid services is hidden to them. For this reason we provide, together with the desktop application, the *My Workspace* interface which allows users to access and manipulate low-level data. However, we also need to provide high-level data to promote accessibility to the wider scientific community. For this reason we also provide access to our *Public Repository* where resources contain high-level generic properties (e.g. `describedIn`, `agreesWithHypothesis`, `hasDescription`).

Throughout this project we have encouraged computational land-use modellers at the Macaulay Institute to lead the process of engineering the FEARLUS ontology. This important activity forced them to think about what conducting an experiment actually meant. Prior to this project commencing, PERL scripts were used to conduct experiments with FEARLUS and the experimental design was thus hidden. One positive side-effect of the ontology building exercise is that the experimental workflow is now exposed, making its use transparent to other users. Development of the ontology has also encouraged them to begin thinking more generally about meta-data to support agent-based social simulation (Polhill, G. et. al., 2005).

Future Work

As mentioned above, we are currently conducting an evaluation of FEARLUS-G with the help of a group of “assessors” from the international scientific communities interested in agent-based social simulation, land-use and water management modelling. As we write this, the first phase of the evaluation is underway in which assessors are using the FEARLUS-G interface to create and apply meta-data to Grid based land-use simulation experiments. The results of the initial meta-data evaluation will lead to a second evaluation phase in which we will test the wider usability of the FEARLUS-G Grid service, in a research capacity (uploading FEARLUS models, comparing results from assessors’ 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. Evaluation of the FEARLUS-G service is expected to provide more general insight into the effectiveness of current Semantic Grid technologies and methodologies.

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).

References

- 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., Hild, 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.
- Brostoff, S., Cooke, R., Stevens, R., Gibson, A., (2005): 'Transparent Interaction; Dynamic Generation: Context Histories for Shared Science.' In: *Proceedings of 1st International Workshop on Exploiting Context Histories in Smart Environments*, 2005.
- Doran, J.E.(1995): *Simulating Societies using Distributed Artificial Intelligence*. In *Social Science Microsimulation*. Springer: Berlin, 1995, pp. 381–393.
- 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.
- Hughes, G., Mills, H., Roure, D.D., Frey, J.G., Moreau, L., Schraefel, M., Smith, G., Zaluska, E., (2004): 'The Semantic Smart Laboratory: A System for Supporting the Chemical e-scientist.' *Org. Biomol. Chem*, 2004, DOI: 10.1039/b410075a.
- Irwin, E., Geoghegan, J. (2001): 'Developing Spatially Explicit Economic Models of Land Use Change.' *Agriculture, Ecosystem and Environment* , 85, 2001, pp. 7–23.
- Keller, R.M., Berrios, D.C., Carvalho, R.E., Hall, D.R., Rich, S.J., Sturken, I.B., Swanson, K.J., Wolfe, S.R. (2004): 'SemanticOrganizer: A Customizable Semantic Repository for Distributed NASA Project Teams.' In: *Proceedings of Third International Semantic Web Conference*, Hiroshima, Japan, 2004, pp. 767–781.
- Lord, P., Bechhofer, S., Wilkinson, M.D., Schiltz, G., Gessler, D., Hull, D., Goble, C., Stein, L. (2004): 'Applying Semantic Web Services to Bioinformatics: Experiences Gained, Lessons Learnt.' In: *Proceedings of Third International Semantic Web Conference*, Hiroshima, Japan, 2004, pp. 350–364.
- Majithia, S., Walker, D.W., Gray, W.A. (2004): 'Automating Scientific Experiments on the Semantic Grid.' In: *Proceedings of Third International Semantic Web Conference*, Hiroshima, Japan, 2004, pp. 365–379.
- Mazzocchi S., Garland S., R.L. (2005): *Simile: Practical Metadata for the Semantic Web*. XML.com, 2005.
- McBride, B. (2000): 'Jena: Implementing the RDF Model and Syntax Specification.' *Technical Report*, Hewlett Packard Laboratories, 2000.
- McHaney, R. (1991): *Computer Simulation A Practical Perspective*. Academic Press, 1991.
- Minar, N., Burkhart, R., Langton, C., Askenazi, M. (1996): 'The Swarm Simulation System, A Toolkit for Building Multi-Agent Simulations.' *SFI Working Paper 96-06-042*, Santa Fe Institute, 1996.

- Pignotti, E., Edwards P., Preece, A., Polhill, G., Gotts, N. (2005a): 'Semantic Support for Computational Land-Use Modelling.' In: *Proceedings of Cluster Computing and Grid 2005*, IEEE Press, 2005.
- Pignotti, E., Edwards P., Preece, A., Polhill, G., Gotts, N. (2005b): 'Providing Ontology Support for Social Simulation.' In: *Proceedings of First International Conference on e-Social Science*, 22-24 June 2005, Manchester, UK.
- 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.
- Polhill G. J., Pignotti, E., Gotts N. M., Edwards, P., Preece, A., (2005): 'An Ontology for Experimentation with a Social Simulation of Land Use Change' Submitted to: *Third Annual Conference of the European Social Simulation Association (ESSA)*, 2005.
- Roure, D.D., Jennings, N., Shadbolt, N. (2003): '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, 2003, pp. 437–470.
- Seaborne, A. (2004): 'RDQL - A Query Language for RDF'. *Technical Report*, Hewlett Packard Laboratories, 2004.
- Stevens, R., Robinson, A., Goble, C. (2003): 'myGrid: Personalised Bioinformatics on the Information Grid.' In: *Proceedings of 11th International Conference on Intelligent Systems in Molecular Biology*, July 2003, Brisbane, Australia, 2003.
- Zhao, J., Wroe, C., Goble, C., Stevens, R., Quan, D., Greenwood, M. (2004): 'Using Semantic Web Technologies for Representing E-science Provenance.' In: *Proceedings of Third International Semantic Web Conference*, Hiroshima, Japan, 2004, pp. 92–106.