Realising the Vision of the Semantic Grid for Computational Land-Use Modelling

P. Edwards* and **E. Pignotti*** and **A. Preece*** and **N. Gotts**† and **G. Polhill**†
Dept. of Computing Science, University of Aberdeen
Aberdeen, AB24 5UE, Scotland
{pedwards, epignott, apreece}@csd.abdn.ac.uk,
†The Macaulay Institute
Craigiebuckler, Aberdeen, AB15 8QH, UK
{n.gotts, g.polhill}@macaulay.ac.uk

Abstract

We explore the vision of the Semantic Grid by discussing deployment of a computational land use modelling service. The FEARLUS-G service is presented which allows large scale simulation experiments to be distributed over the Grid. We also discuss meta-data support for simulation parameters, hypotheses and results that will facilitate sharing and re-use of such resources among land-use scientists, together with the creation of scientific arguments.

1 Introduction

Collaborations between large groups of scientists are now seen as essential to enhance the scientific process. While research has always involved collaboration between individuals, there is now even greater necessity for tools to support sharing of knowledge, resources, results and observations. For these reasons recent e-science activities [Roure *et al.*, 2001] 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 [Foster *et al.*, 2001] 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. The Semantic Grid [Roure *et al.*, 2003] [De Roure, 2005] aims to facilitate ease-of-use of Grid resources and flexible collaboration on a global scale; central to this vision of a 'smart' Grid is the use of Semantic Web [Berners-Lee *et al.*, 2001], software agent and Web services technologies (amongst others).

The Semantic Grid requires adoption of metadata and ontologies [Fensel, 2003] 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 system 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 (Semantic) Grid context.

2 FEARLUS-G Grid Service

FEARLUS [Polhill *et al.*, 2001] 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 have developed FEARLUS-G - 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 3. We chose to deploy FEARLUS as a Grid service for a number of reasons: firstly, the Grid infrastructure would allow large land-use simulation experiments to be distributed across the Grid; a typical FEARLUS experiment already involves running a series of simulations, but on a single machine. Secondly, the use of Semantic Web technologies combined with Grid infrastructure would allow us to create a co-laboratory where land-use scientists can access the FEARLUS-G service and share and reuse results and observations.

3 FEARLUS-G Meta-Data Support

We propose a conceptual layer through the definition of an ontology which captures the concepts and relationships important to scientists in order to improve the scientific rigour of agent-based modelling. Our 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.

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 service, 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"[1]. 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.

We use the Dublin Core[2] ontology 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). Another important attribute is `describedIn`. This is used to record that a publication can be used to describe any given scientific object.

As part of the FEARLUS-G project we have developed an open source reusable semantic data service based on Jena2 [McBride, 2000] and Globus Toolkit 3 [Foster and Kesselman, 1998] to provide semantic data storage, query and retrieval functions. This service uses ELDAS [Baxter *et al.*, 2003] to manage data repositories for RDF models; ELDAS was developed by the UK National e-Science Centre eDIKT project[3]. 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 could offer the potential for distributed processing of such queries.

4 Discussion

The Grid is clearly a valuable tool for managing the computation involved in running such large-scale experiments. In fact the Grid is ideal for simulation modelling because the cost of a simulation run is less that the cost of the communication necessary to run a distributed experiment and aggregate the results. 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. 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. We provide a Web application which allows users to access and manipulate low-level simulation model parameters needed to run a FEARLUS experiment. However, we also need to provide high-level data to promote accessibility to the wider community. 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, 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.

[1] www.pages.drexel.edu/∼bcb25/scimeth/vocabulary.htm

[2] http://dublincore.org/

[3] http://www.edikt.org

Acknowledgments

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

[Baxter *et al.*, 2003] Rob Baxter, Denise Ecklund, Aileen Fleming, Alan Gray, Brian Hilld, Stephen Rutherford, and Davy Virdee. Designing for Broadly Available Grid data Access Services. In *UK e-Science All Hands Meeting (CD-ROM)*, 2003.

[Berners-Lee *et al.*, 2001] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):28–37, May 2001.

[De Roure, 2005] N.R. Shadbolt De Roure, D. Jennings. Semantic grid: Past, present, and future. In *Proceedings of the IEEE, Volume 93, Issue 3*, pages 669–681, 2005.

[Fensel, 2003] Dieter Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag New York, Inc., 2003.

[Foster and Kesselman, 1998] I. Foster and C. Kesselman. Globus: A Toolkit-Based Grid Architecture. In *The Grid: Blueprint for a Future Computing Infrastructure*, pages 259–278. Morgan-Kaufmann, 1998.

[Foster *et al.*, 2001] I. Foster, C. Kesselman, and S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International J. Supercomputer Applications*, 15(3), 2001.

[McBride, 2000] Brian McBride. Jena: Implementing the RDF Model and Syntax Specification. Technical report, Hewlett Packard Laboratories (Brstol), 2000.

[Polhill *et al.*, 2001] J.G. Polhill, N.M. Gotts, and A.N.R. Law. Imitative Versus Non-Imitative Strategies in a Land Use Simulation. *Cybernetics and Systems*, 32 (1):285–307, 2001.

[Roure *et al.*, 2001] D. De Roure, N. Jennings, and N. Shadbolt. 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.

[Roure *et al.*, 2003] D. De Roure, N. Jennings, and N. Shadbolt. 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*, pages 437–470, 2003.