NSF/ESRC Agenda Setting Workshop on Agent-Based Modeling of Complex Spatial Systems: April 14-16, 2007

Why Agent-Based Modelling of Complex Spatial Systems Needs Cyberinfrastructure.

Edoardo Pignotti
University of Aberdeen
Department of Computing Science
Aberdeen AB24 3UE, Scotland, UK
epignott@csd.abdn.ac.uk

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.

The vision of e-science[1] (or Cyberinfrastructure) is to facilitate large scale science using Grid technologies 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 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[2] has emerged, which integrates Semantic Web[3] and Grid technologies.

Central to the vision of the Semantic Grid is the adoption of metadata and ontologies 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 Semantic Web 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. 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. 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.

Recent activities in the field of social simulation have outlined the need to improve the scientific rigour of agent-based modelling. One of the important characteristics of scientific research 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 several activities that are relevant:

- access to the results data itself, to check that the authors' claims that are based on those results are justifiable.
- an ability to re-run experiments to check that they produce broadly the same set of results.
- manipulation of the simulation model parameters to check that there is no undue sensitivity of the results to certain parameter settings.
- modification of the source code and/or re-implementation of the model to check for what might be termed 'algorithmic sensitivity'.

In a previous project, FEARLUS-G[4], we tried to bring together the needs of agent based modelling with the vision of eScience. This project involved social scientists at the Macaulay Institute in Aberdeen investigating

[1] http://www.rcuk.ac.uk/escience/default.htm
[2] http://www.semanticgrid.org/
[3] http://www.w3.org/2001/sw/
[4] http://www.csd.abdn.ac.uk/research/fearg/

land-use change and computer scientists at the University of Aberdeen. FEARLUS-G aimed to provide scientists interested in land-use phenomena with a means to run much larger-scale experiments than previously possible on standalone PCs, together with a Web-based environment in which to share simulation results. A key facet of this project was the development of an ontology which describes the tasks and entities involved in simulation work, such as experiments, hypotheses, parameters, simulation runs, and statistical procedures.

In the PolicyGrid[5] Project (a UK eScience project funded as part of the Economic & Social Research Council eSocial Science initiative) we are building on the work of FEARLUS-G. The project involves collaboration between computer scientists and social scientists at the University of Aberdeen, the Macaulay Institute (Aberdeen) and elsewhere in the UK. The project aims to support policy-related research activities within social science by developing appropriate Grid middleware tools which meet the requirements of social science practitioners. The project is developing a range of services to support social scientists with mixed-method data analysis (involving both qualitative and quantitative data sources) together with the use of social simulation techniques. Issues surrounding usability of Semantic Grid tools are also a key feature of PolicyGrid, with activities encompassing workflow support and natural language presentation of metadata.

As part as PolicyGrid we are investigating the use of semantic workflow tools to facilitate the design, execution, analysis and interpretation of simulation experiments and exploratory studies, while generating appropriate metadata automatically. We have explored a number of case studies of social simulation research activities. Some of the emerging challenges are:

- Capturing the scientist's goals and constraints associated with a workflow;
- Integrating workflows into a scientific argument structure;
- Improving interoperability and re-use of workflows.

To date, we have created an initial social simulation classification ontology capturing the characteristics of a wide range of simulation models, e.g. type of simulation, behaviour, space model, execution type, etc. Collaborators at the Macaulay Institute are continuing work on the development of a simulation modelling ontology to allow a particular piece of modelling software to be described and the structure and context of a particular simulation run to be characterised.

Questions

- Should there be a more systematic way of sharing and preserving the knowledge produced by the agent-based modelling community?

- Can we increase the scientific value of agent-based models by making them easier to verify and validate?

- Can workflow tools play an important role in capturing agent-based modelling methodology?

- What kind of activities are undertaken in agent-based modelling?
  - What are the relationships between these activities?
  - What kinds of metadata support do they need?

- Is it possible to draw up a general specification of metadata requirements for agent-based modelling, or do agent-based modelling projects differ in too many respects for this to be feasible?

- What software tools are needed (or would be useful) to support the activities listed, and the generation of the required metadata relevant to each activity?

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