

The National Centre for e-Social Science

Rob Procter¹, Mike Batty², Mark Birkin³, Rob Crouchley⁴, William H. Dutton⁵, Pete Edwards⁶, Mike Fraser⁷, Peter Halfpenny⁸, Yuwei Lin¹ and Tom Rodden⁹

¹National Centre for e-Social Science, University of Manchester

²GeoVUE, Centre for Advanced Spatial Analysis, UCL

³MoSeS, School of Geography, University of Leeds

⁴CQeSS, Centre for e-Science, University of Lancaster

⁵OeSS, Oxford Internet Institute, University of Oxford

⁶PolicyGrid, Department of Computing Science, University of Aberdeen

⁷MiMeG, Department of Computer Science, University of Bristol

⁸National Centre for e-Social Science, University of Manchester

⁹DReSS, Department of Computer Science, Nottingham University

Abstract

This paper outlines the work of the UK National Centre for e-Social Science and its plans for facilitating the take-up of Grid infrastructure and tools within the social science research community. It describes the kinds of social science research challenges to which Grid technologies and tools have been applied, the contribution that NCeSS can make to support the wider take-up of e-Science and, finally, its plans for future work.

1. Introduction

The concept of e-Science refers to the growing use of advanced Internet and Grid applications to support scientific research. The National Centre for e-Social Science (NCeSS) is funded by the Economic and Social Research Council (ESRC) to investigate and promote the use of e-Science to benefit social science research. The ESRC's investment in e-Science began in 2003 with the funding of eleven pilot demonstrator projects (PDPs). The creation of NCeSS followed in 2004. The overall goal of NCeSS is to stimulate the uptake and use of emerging e-Science technologies within the social sciences.

2. NCeSS Overview

The Centre is made up of a co-ordinating Hub based at the University of Manchester with support from the UK Data Archive at the University of Essex, plus seven research Nodes based at institutions throughout the UK. A series of smaller e-Social Science projects have been commissioned under the ESRC small grant scheme.

The Hub co-ordinates the research activities of the Centre as well as providing e-Social Science training, technical support, information services and support to users. The Hub acts as the central resource base for e-Social Science

issues and activities in the UK, integrating them with ESRC research methods initiatives and the existing e-Science core programme.

The majority of the Centre's research is undertaken in seven research Nodes, each of which focus on a different area of e-Social Science and funded for three years. There are also twelve Small Grant projects, each funded for one year.

The NCeSS research programme was conceived from the beginning around two distinct strands. The first, the applications strand, is aimed at stimulating the uptake and use by social scientists of new and emerging Grid computing and data infrastructure in order to make advances in both quantitative and qualitative economic and social research. It seeks to draw upon and further advances generic middleware developments from the e-Science core programme and apply them to the particular needs of the social science research community in order to generate new solutions to social science research problems. The second, the social shaping strand, examines the social and economic influences on the development of e-Science and, conversely, the socio-economic impact of Grid technologies. This strand focuses on the factors influencing the design, uptake and use of Grid technologies, and the conditions determining whether and how their potential is realised. As such, the social shaping strand has

potential relevance well beyond the social sciences and we will return to this later.

2.1 Collaboratory for Quantitative e-Social Science (CQeSS)

The overall aim of CQeSS is to ensure the effective development and use of Grid-enabled quantitative methods. Part of the focus of CQeSS is on developing middleware that allows users to exploit Grid resources such as datasets while continuing to employ their favourite desktop analysis tools [6].

As e-Social Science develops, there will be a growing user base of social researchers who are keen to share resources and applications in order to tackle some of the large-scale research challenges that confront us. They will be aware of the potential of e-Science technology to provide collaborative tools and provide access to distributed computing resources and data. However, social scientists are not ideally catered for by the current Grid middleware and often lack the extensive programming skills to use the current infrastructure to the full and to adapt their existing “heritage” applications.

This problem prompted the Grid Technology Group at CCLRC Daresbury Laboratory to write the prototype GROWL: Grid Resources On Workstation Library toolkit.¹ This toolkit provides an easy-to-use client-side interface. CQeSS is now seeking to apply the GROWL middleware to wrap the statistical modeling methods available in SABRE as well as other computationally-demanding models and to make these developments available in the distributed environment as a componentised R library and as a Stata “plug-in”.

2.2 New Forms of Digital Record for e-Social Science (DReSS)

DReSS seeks to understand how new forms of record may emerge from and for e-Social Science. Specifically, it seeks to explore the development of new tools for capturing, replaying, and analyzing social science records [3,4,5,12]. Social scientists work in close partnership with computer scientists on three Driver Projects to develop e-Social Science applications demonstrating the salience of new forms of digital record.

Development work within the Driver Projects focuses on the assembly of qualitative records, the structuring of assembled records, and the coupling of qualitative and quantitative records. The research is underpinned by three

key themes – record, re-representation, and replay – to ensure the development of services that have some general purchase and utility. Work to date has seen the development of a dedicated ReplayTool² which supports the use of records containing multiple forms data from multiple sources, allows data to be represented and re-represented in different ways to support different kinds of analysis, and enable researchers to replay digital records to support analysis. Current work seeks to implement semantic web ontologies to support structured forms of analysis and is also exploring the potential of vision recognition and text mining techniques to support automatic coding of large datasets. The work of the Node is driven by an iterative user-centred prototyping approach to demonstrate the salience of new forms of digital record to future social science research.

2.3 Modelling and Simulation for e-Social Science (MoSeS)

MoSeS will provide a suite of modeling and simulation tools which will be thoroughly grounded in a series of well-defined policy scenarios. The scenarios will be validated by both social scientists and non-academic users [1]. MoSeS is particularly focused on policy applications within the domains of health care, transport planning and public finance. Social science problems of this type are characterized by a requirement for extensive data integration, the need for multiple iterations of computationally intensive scenarios, and a collaborative approach to problem-solving.

The key objective is to develop a representation of the entire UK population as individuals and households, at present and in the future, together with a package of modeling tools which allows specific research and policy questions to be addressed. The advances it will seek to achieve include the creation of a dynamic, real-time, individually-based demographic forecasting model; for defined policy scenarios, to facilitate integration of data and reporting services, including Geographical Information Systems (GIS), with modeling, forecasting and optimisation tools, based on a secure grid services architecture; to use hybrid agent-based simulations to articulate the connections between individual level and structural change in social systems; and to provide high level capability for the articulation of unique evidence-based user scenarios for social research and policy analysis.

¹ <http://www.growl.org.uk>

² www.ncess.ac.uk/nodes/digitalrecord/replaytool/

2.4 Semantic Grid Tools for Rural Policy Development and Appraisal (PolicyGrid)

PolicyGrid brings together social scientists with interests in rural policy development and appraisal with computer scientists who have experience in Grid and Semantic Web technologies. The core objective is to explore how Semantic Grid tools [10] can support social scientists and policy makers who increasingly use mixed-method approaches combining qualitative and quantitative research techniques (e.g., surveys and interviews, ethnography, case studies, simulations).

Provision of metadata infrastructure in this context (to support annotation and sharing of resources) presents many challenges including the dynamic and contested nature of many concepts within the social sciences, the need to align with existing thesauri (where those exist) and the need to support open, community based efforts. The Node is thus exploring so-called “folkology” solutions which exploit both lightweight ontology and folksonomy (social tagging) based approaches [13]. Another activity is investigating the use of argumentation approaches [14] as a means of facilitating evidence-based policy making; argument structures provide a mechanism for linking qualitative analyses with other resources including simulation experiments, and queries over Grid-enabled data services. Enhanced support for social simulation on the Grid [15] is another focus, with a framework for characterising simulation models under development, as well as workflow support.

2.5 Mixed Media Grid (MiMeG)

MiMeG is developing tools to support distributed, collaborative video analysis [11]. The project arises from converging developments in contemporary social science. Digital video is becoming an invaluable tool for social and cognitive scientists to capture and analyse a wide range of social action and interactions; video-based research is increasingly undertaken by research teams distributed across institutions in the UK, Europe and worldwide; and there is little existing support for remote and real-time discussion and analysis of video data for social scientists. Therefore MiMeG has taken as a central concern how to design tools and technologies to support remote, collaborative and real-time analysis of video materials and related data. In developing these tools the project has the potential to support video-based ‘collaboratories’ amongst research teams

distributed across institutions, disciplines and locations. We are undertaking studies of research and analytic practice to develop requirements for the design of these tools, including detailed qualitative studies of current practice in fields of video-based research in both social scientific and professional communities.

MiMeG has built prototype software that can connect collaborators, allowing them to see, discuss and annotate video together in real-time and from remote locations. This software is now freely available under a GPL license in PC and Mac versions and allows real time analysis of video data between two or more remote sites. The software also enables each site to see the others’ notes, transcripts, drawings etc on the video they are watching and talk to each other at the same time. We are now beginning development of a second version of the software, which will begin to consider more innovative ways of capturing and representing communicative conduct in remote data sessions.

MiMeG has collected a wide-ranging data on existing video-based research practice across the social and cognitive sciences and the work of video analysts working in other occupations. Now that the prototype tools are beginning to be adopted we are undertaking a series of studies of their use by social scientists as part of their everyday research activities.

2.6 Geographic Virtual Urban Environments (GeoVUE)

GeoVUE will provide grid-enabled virtual environments within which users are able to link spatial data to geographic information systems-(GIS)-related software relevant to a variety of scientific and design problems. It will provide decision support for a range of users from academics and professionals involved in furthering our geographic understanding of cities to planners and urban designers who require detailed socio-economic data in plan preparation. At the heart of GeoVUE lies the concept of a VUE which represents a particular way of looking at spatial data with respect to the requirements of different kinds of user, thus defining a virtual organisation relevant to the problem in hand.

GeoVUE is developing three demonstrators of increasing complexity and sophistication which will form part of a structured process of sequential development to the ultimate aim of producing a generic framework.

2.7 Oxford e-Social Science Project (OeSS)

OeSS focuses on the inter-related social, institutional, ethical, legal and other issues surrounding e-Science infrastructures and research practices. The design and use of advanced Internet and Grid technologies in the social, natural and computer sciences are likely to reconfigure not only how researchers get and provide data resources and other information but also what they and the public can access and know; not only how they collaborate, but with whom they collaborate; not only what computer-based services they use, but from whom they obtain services [8]. This reconfiguring of access to a wide variety of resources raises numerous issues, including ethical concerns (e.g., confidentiality, anonymity, informed consent), legal uncertainties (e.g., privacy and data protection, liability), social (ownership, trust, credit), and institutional (IPR and risk in multi-institutional collaboration) issues [7,9,16]. OeSS assembled a multi-disciplinary team to analyze e-Sciences in the UK and globally, focusing on a set of in-depth case studies to uncover the dynamics of these ethical, legal and institutional issues that facilitate or constrain the use of e-Science data, tools and other resources, and shape initiatives to address them.

3. Social Shaping

Social shaping is defined very broadly within the NCESS programme to include all social and economic aspects of the genesis, implementation, use, usability, immediate effects and longer-term impacts of the new technologies. Despite the very substantial current investment in the Grid, little is known about the nature and extent of take-up, about how and why and by whom these new technologies are being adopted, nor what will be their likely effects on the character and conduct of future scientific research, including social scientific research.

While OeSS is the only NCESS node which has social shaping research as its principal aim, a number of them address this as a subsidiary aim and it is also a focus for a number of the Small Grant projects. The 'entangled data' project (Essex University), for example, has been conducting a comparative study to understand how groups of research scientists collaborate using shared digital data sources, developing insights into the likely use and non-use of e-Science technologies, and the social and technological innovations that may be

required as e-Science expands from its early adopters [2].

There are also important questions to be answered about how Grid-based tools might be adapted to make them more usable. Members of NCESS have been active organising workshops on usability issues for e-Science³. Finally, the Hub is working with the UK e-Science Institute to investigate how barriers to the wider take up of Grid technologies can be tackled.

4. Developing e-Social Science

Much of what is presently understood about e-Social Science is an extrapolation of existing practice and is focused around areas where it is believed the Grid can address known limitations of research methods.

Computer-based modeling and simulation is a well established social science research tool. As models get more complex, they need computing power and there are similar benefits to be had from the Grid for statistics-based research generally.

The sharing and re-use of data is already well established in the social sciences, but heterogeneity in formats means that linking different datasets together can still prove very difficult. In partnership with UK data centres, NCESS has begun pilot studies of 'grid enabling' selected datasets.

The vast amounts of data generated as people go about their daily activities are, as yet, barely exploited for research purposes. For example, use of public services is captured in administrative records; in the private sector, patterns of consumption of goods and services are captured in credit and debit card records; patterns of movement are logged by sensors, such as traffic cameras, satellites and mobile phones; the movement of goods is increasingly tracked by devices such as RFID tags. Exploiting these data sources to their full research potential requires new mechanisms for ensuring secure and confidential access to sensitive data, and new tools for integrating, structuring, visualisation and analysis.

How e-Social Science will develop in the longer term will become clearer as researchers explore and experiment with the opportunities the Grid provides. To facilitate this process, NCESS has been organizing a series of Agenda

³ See, for example, www.nesc.ac.uk/action/esi/contribution.cfm?Title=613 and www.ncess.ac.uk/support/chi/index.php/Main_Page

Setting Workshops (ASWs)⁴ to which social scientists are invited to hear about opportunities for using the Grid in research and to reflect on how these technologies might address the obstacles. Nine ASWs have been held over the past eighteen months and more are planned.

4.1. Widening Engagement

The ASWs have helped to identify new areas for the application of Grid technologies in the social sciences and community needs. For example, a theme emerging from several of the ASWs is that Grid-enabled datasets, services and tools are key enablers for the wider take-up of e-Social Science. They have also enabled NCeSS to profile the social science research constituency in terms of its awareness and readiness to adopt new technologies. We have used these findings to inform NCeSS strategy for widening the community's engagement with e-Social Science.

Our social science researcher profile identifies three distinct communities – the 'early adopters' who are keen to push to the limit of what is possible; the 'interested' who will adopt new research tools and services if they believe these will provide simple ways of advancing their research; and the 'uncommitted' who have yet to appreciate the relevance of Grid technologies for their research.

Those early adopters who have not been recruited into the NCeSS programme will demand research tools they can apply. Tools are also needed to convert the interested into adopters and demonstrators which will convince the uncommitted to join the ranks of the interested.

To address the needs of these communities, NCeSS has begun to build an e-Infrastructure for social sciences. To engage with the uncommitted, NCeSS will deploy a selection of demonstrators from those being developed within the Nodes (e.g., GeoVUE visualisation tools), NCeSS Small Grants, and demonstrators developed by the e-Social Science PDPs. More generally, NCeSS will provide a platform for disseminating the benefits of e-Social Science to the wider research community, leverage existing e-Social Science and e-Science investments, and ensure the usability and sustainability of middleware, services and tools.

4.2 Solving Real Social Science Problems

The diffusion of any innovation, including e-Social Science tools and practices, will be

shaped by the degree it can address the problems of potential users. NCeSS is developing an increasingly concrete understanding of the applications of e-Sciences in the addressing social science problems. The examples illustrate how social scientists can combine its component parts – i.e., datasets, services and tools – in flexible yet powerful ways to overcome various kinds of problems they face in pursuing their research.

For example, solving complex statistical modeling problems often involves multiple steps and iterations where the output of one step is used as the input in the next: 1) select data or subsets, maybe from more than one source; 2) merge, harvest or fuse; 3) input to model and analyse; 4) repeat previous cycle with new or different data; 5) repeat with different model assumptions, parameters and possibly data; 6) synthesise outputs from multiple models or world views; 7) load output into a different analysis or visualisation tool.

Managing these steps manually is potentially difficult, and performing the data integration and modeling using desktop PC tools may be very time consuming. For example, analysis of work histories requires many different sources of data which need to be reconciled and integrated in order to produce a coherent and contextualised life or work history and takes one week on a desk top PC in Stata. To then simultaneously analyse the data could take months on a desktop PC running serial SABRE and over 10 years using Stata. This presents a major obstacle to research. Using Grid-enabled data, analysis and workflow tools, however, the researcher can compose complex sequences of steps using different computational and data resources, and execute them (semi) automatically using powerful computers in a comparatively short time.

Schelling's model of segregation provides a way of exploring how the spatial distribution of society groups ('agents') responds to different assumptions about neighbour preferences. For example, a social scientist interested in racial segregation could use it to explore questions such as:

- Sensitivity of the model to initial arrangements of agents.
- The relationship between ratio of different 'races' and the final segregation ratio.

To do this, the researcher might need to run simulations for 50 different values of the initial race ratio and 100 different initial agents' arrangements. Using a desktop PC, 5000 simulations will take about one week which may well make it impractical. If the researcher

⁴ Funded by the ESRC/JISC ReDReSS project.

has access to Grid computational resources, however, the same number of simulations can

be run in about five minutes, opening up new possibilities for the researcher.

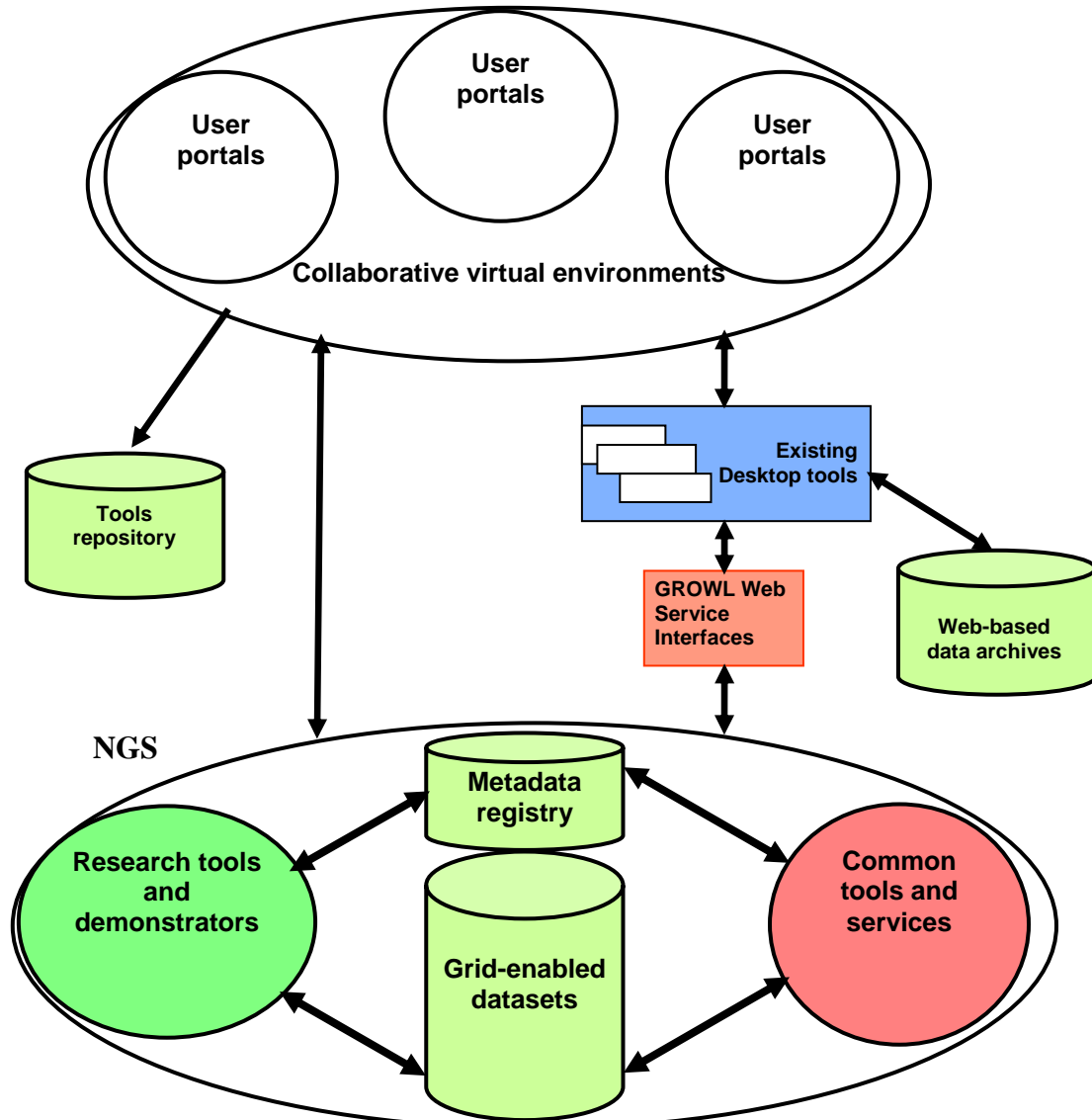


Figure 1: e-Infrastructure for Social Sciences.

5. e-Infrastructure Architecture

An overview of the e-Infrastructure architecture is shown in Figure 1.

Portals provide an integrated, single point of access to e-Infrastructure resources through a familiar and simple-to-use web-style user interface which hides the underlying complexity. Users can authenticate themselves, discover resources (data, tools and services) and create their own 'workflows' from tools and services to carry out analysis. A tool or service is mapped to a Java portlet for insertion into one or more portal frameworks. Each Node can

deploy its own preferred framework and choose from a repository of portlets depending on its user requirements. NCeSS will also host fully-functional portals with both collaboration tools and service portlets for access to Grid-enabled datasets and tools, including tools for generic tasks such as resource discovery, workflow composition, data mining and visualisation. These portals will be designed to be easily usable by a wide range of users. Collaborative virtual environments (such as Sakai) help distant researchers: share in the research tasks, run project meetings, discuss results, and work up presentations and papers.

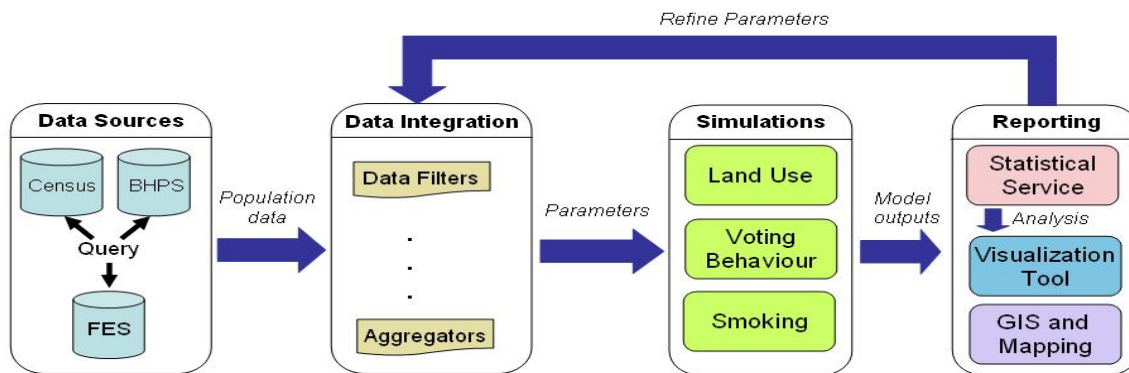


Figure 2: Example simulation workflows.

A range of research tools and demonstrators selected from those being developed within the NCeSS programme will be deployed, with selection criteria informed by ongoing consultations with the wider social science research community. Agenda Setting Workshops have already enabled us to identify simulation and modeling as a priority area. Building on the work of MoSeS and PolicyGrid, simulation and modeling tools will be deployed which will enable researchers to create their own workflows to run their own simulations, visualise and analyse results, and archive them for future comparison and re-use (see Figure 2). This will facilitate development and sharing of social simulation resources within the UK social science community, encourage cooperation between model developers and researchers, and help foster the adoption of simulation as a research method in the social sciences.

A set of common tools and services (e.g., workflow, visualisation) will also be made available. Our approach to e-Infrastructure building does not preclude local community-level deployments of tools and services. Indeed, the NCeSS programme is uncovering requirements for usability, control and trust in qualitative research that may be more appropriately served through localised provision of tools and data. For example, there may be issues with the distribution of sensitive video data to (and via) un-trusted third parties. Development of desktop tools and community-oriented service deployments is part of the ongoing NCeSS e-Infrastructure strategy, particularly in supporting qualitative research. MiMeG and DReSS are building new tools and have a remit within their existing work programmes to capacity build and deploy to interested communities.

A range of Grid-enabled datasets are being made available (quantitative and qualitative), including datasets already Grid-enabled by the PDPs or being Grid-enabled by UK data centres. Others will be selected after consultation with Nodes and Small Grants to identify dataset usage within the NCeSS programme; the major data centres to identify patterns of dataset usage (quantitative and qualitative) within the wider social science community, to understand licensing issues and ensure complementarity with JISC funded Grid-enabling activities (current and planned); and with the social science community to identify research drivers and ensure a fit with the ESRC's future data strategy plans.

6. Benefits for Social Sciences

The e-Infrastructure project will serve a number of important objectives which are relevant to the social science and wider research communities:

- enhance understanding of issues around resource discovery, data access, security and usability by providing a testbed for the development of metadata and service registries, tools for user authorisation and authentication, and user portals;
- lay foundations for an integrated strategy for the future development and support of e-Social Science infrastructure and services;
- leverage the infrastructure investment being made by UK e-Science core programme and JISC for the benefit of the Social Sciences;
- promote synergies across NCeSS and other ESRC investments, co-ordinate activities, encourage mutual support and identify areas in which to promote the benefits of common policies and technology standards.

NCeSS is working closely with the social science community to ensure that the project is driven by research needs and, specifically, to identify the most research-relevant resources, tools and services to incorporate into the e-Infrastructure. It is also working with the UK e-Science core programme (NGS, OMII-UK consortium, NGS, DCC) and JISC to devise an e-Infrastructure development plan which will define technical standards and mechanisms to ensure long term sustainability.

7. Summary and Future Work

In this paper we have presented an overview of how the NCeSS programme is working to develop applications of Grid technologies and tools within the social sciences and to understand the factors which encourage or inhibit their wider diffusion and deployment. As such, NCeSS has lessons for the e-Science community as it begins to grapple with these same problems.

NCeSS will continue to develop its agenda for e-Social Science. Activities focusing on the usability of e-Science are expanding through participation in JISC VRE and EPSRC usability and e-Science programmes. NCeSS is also beginning a series of fieldwork investigations of work practices in developing areas of e-Science so as to better understand the impact of these new technologies and the issues they raise for usability, and the methodologies used for requirements capture, design and development.

Bibliography

1. Birkin, M., Clarke, M., Rees, P., Chen, H., Keen, J. and Xu, J. (2005). MoSeS: Modelling and Simulation for e-Social Science. In Proc 1st Int. Conf. on e-Social Science, Manchester.
2. Carlson, S. and Anderson, B. (2006). e-Nabling Data: Potential impacts on data, methods and expertise. In Proc 2nd Int. Conf. on e-Social Science, Manchester.
3. Crabtree, A. and Rouncefield, M. (2005). Working with text logs: some early experiences of e-SS in the field. In Proc 1st Int. Conf. on e-Social Science, Manchester.
4. Crabtree, A., French, A., Greenhalgh, C., Rodden, T. and Benford, S. (2006). Working with digital records: developing tool support. In Proc 2nd Int. Conf. on e-Social Science, Manchester.
5. Crabtree, A., French, A., Greenhalgh, C., Benford, S., Cheverst, K., Fitton, D., Rouncefield, M. and Graham, C. (2006). Developing digital records: early experiences of record and replay, To appear in Computer Supported Cooperative Work: The Journal of Collaborative Computing, Special Issue on e-Research.
6. Crouchley, R., van Ark, T., Pritchard, J., Grose, D., Kewley, J., Allan, R., Hayes, M. and Morris, L. (2005). Putting Social Science Applications on the Grid. In Proc 1st Int. Conf. on e-Social Science, Manchester.
7. David, P.A. and Spence, M. (2003). Towards institutional infrastructures for e-Science: the scope of the challenge. Research Report No. 2 (Oxford Internet Institute).
8. Dutton, W. (2005). The Internet and Social Transformation: Reconfiguring Access, pp. 375-97 in Dutton, W. H. et al. (eds), Transforming Enterprise. Cambridge, Massachusetts: MIT Press.
9. Dutton, W. D., Jirotko, M. and Schroeder, R. (2006). Ethical, Legal and Institutional Dynamics of the e-Sciences. OeSS Project Draft Working Paper. University of Oxford.
10. Edwards, P., Preece, A., Pignotti, E., Polhill, G. and Gotts, N. (2005). Lessons Learnt from Deployment of a Social Simulation Tool to the Semantic Grid. In Proc 1st Int. Conf. on e-Social Science, Manchester.
11. Fraser, M., Biegel, G., Best, K., Hindmarsh, J., Heath, C., Greenhalgh, C. and Reeves, S. (2005). Distributing Data Sessions: Supporting remote collaboration with video data. In Proc 1st Int. Conf. on e-Social Science, Manchester.
12. French, A., Greenhalgh, C., Crabtree, A., Wright, M., Hampshire, A. and Rodden, T. (2006). Software replay tools for time-based social science data.
13. Guy, M. and Tonkin, E. (2006). Folksonomies: Tidying up Tags? D-Lib Magazine 12(1).
14. Kirschner, P., Buckingham Shum, S. and Carr, C. (2003). Visualising Argumentation: Software Tools for Collaborative and Educational Sense-Making. Springer-Verlag: London.
15. Pignotti, E., Edwards, P., Preece, A., Gotts, M. and Polhill, G. (2005). Semantic Support for Computational Land-Use Modelling. In Proc 5th IEEE International Symposium on Cluster Computing and Grid, Cardiff.
16. Schroeder, R., Caldas, A., Mesch, G. and Dutton, W. (2005). In Proc 1st Int. Conf. on e-Social Science, Manchester.