Sustaining an e-Infrastructure for Social Science

Abstract for International Conference on e-Social Science 2007

Michael Daw (michael.daw@manchester.ac.uk)*, Rob Procter*, Yuwei Lin*, Terry Hewitt*, Wei Jie*, Alex Voss*, Kenny Baird*, Andy Turner†, Mark Birkin†, Ken Miller§, William Dutton‡,
Marina Jirotka‡, Ralph Schroeder‡, Grace de la Flor‡, Pete Edwards¥,
Rob Allan**, Rob Crouchley††

* University of Manchester, † University of Leeds, § University of Essex, ‡ University of Oxford, ¥ University of Aberdeen, ** Science & Technology Facilities Council, †† Lancaster University

Introduction

The Economic and Social Research Council (ESRC) has funded a three year project (from January 2007) to begin the building of an infrastructure to provide integrated access to a variety of datasets, tools and services for UK social science research.

This paper describes the project, the aims of which are to: examine the feasibility of and begin to create a coherent e-infrastructure for social science; provide a platform for disseminating the benefits of e-social science; leverage existing e-social science and e-Science investments; and to define a roadmap for the sustainability of middleware, services and tools currently being developed by participants in the National Centre for e-Social Science (NCeSS) programme.

The paper will also focus on what are likely to be the most challenging aspects of the project, namely how to incorporate user requirements so that project outputs are appropriate, useful and highly usable; how to move from research outputs to services for the support of social science research; and how to incorporate this technology within a wider infrastructural framework such as that offered by the UK National Grid Service (NGS).

Project Outputs

The project outputs fall into three broad categories: *research resources*, *support activities* and materials, and *investigations* into the feasibility of technologies, technical standards, sustainability and project evaluation.

Research Resources

Research resources comprise of software, Grid-enabled datasets and the infrastructure on which these will reside. The aim of the software outputs is to provide the core of a UK e-

social science infrastructure, along with sample demonstrator applications and tools presented on the infrastructure.

The user environment is comprised of a portal, a service registry and workflow tools. The portal offers collections of tools and services appropriate to the needs of different social science virtual organisations (VOs). Each VO will have its own "worksite" on the portal that provides easy access to a collection of applications. A number of these tools will be generic to many VOs, such as Wiki's, blogs and videoconference applications, but there will also be available specialised applications or tools to meet the needs of particular VOs. The project is developing these tools to be pluggable into the portal and discoverable via a social science service registry, from which the VO administrator populates the corresponding portal worksite with the tools appropriate to his/her VO.

Such tools include applications that offer simulation models to support land-use and health policy-making; tools to support the more usable use of metadata by social scientists for the annotation, querying and presentation of resources; and development of the Grid Resources On Workstation Library (GROWL), a collection of services that hide the complexity of the Grid from the user and provide a secure means of accessing the Grid from the desktop.

Workflows offer an additional means of employing services and this project is developing sample demonstrations to investigate their efficacy.

The project also aims to determine appropriate security mechanisms to meet the needs of all aspects of this emerging infrastructure and to Grid-enable a number of datasets so that they may be subject to more flexible discovery, interrogation and integration.

Support Activities

In order for the infrastructure to be usable by social scientists, the project is producing a variety of support materials, including documentation, learning objects for inclusion into learning environments, and training materials, which we aim to deliver through workshops to introduce the tools and infrastructure to prospective users.

Investigations

The final category of project output consists of a number of investigations. Work to capture requirements for Grid-enabling datasets will devise criteria for selecting datasets and recommend those datasets that offer the most value to the community were they to be Grid-enabled. This investigation will feed directly into work to Grid-enable the recommended datasets. There will be a study of the feasibility and critical success factors of replacing Safe Settings (which are restricted access secure rooms used to ensure the privacy and anonymity of social survey subjects whose identity is at risk of being revealed) with virtual Safe Settings based on Grid technologies.

The project will conduct a survey to assess the degree of awareness of e-social science to help inform policy and practice within NCeSS as part of an ongoing, continuous process of engagement with social scientists. This survey can act as a baseline to assess the future

effectiveness of the UK e-social science programme. The workings of the project itself will also be subject to ongoing evaluation in order to maximise its effectiveness.

A study to examine technical standards used on the project will facilitate an ongoing discussion of how the disparate technical elements of the social science e-infrastructure can start to converge. The study will take into account and feed into other activities in this area, such as work being done by the e-Research Tools and Resources Interoperability project (eReSS), the Joint Information Systems Committee (JISC) e-Framework, Open Source Software (OSS) Watch and the NGS in order to provide a cohesive approach to standards adoption by this and related communities.

Finally, the project is investigating issues related to sustainability, which will examine the steps needed to take this initiative forward in order to provide an infrastructure that offers highly effective and widespread support for many aspects of social science research.

Challenges

User Requirements

We are aiming for research resources produced by the project to be user-focussed. They should be appropriate to the needs of social scientists, useful in increasing the productivity and effectiveness of social science research, and incorporate highly usable user interfaces to reinforce and increase take-up. The realisation of these aims will enable this project's software outputs to have a significant impact on research practice and begin to realise some of the potential for e-social science. However, whilst desirable, these aims have significant resource and methodological implications that will be hard to meet in the context of this project, which has relatively meagre resources in relation to its ambition.

Although usability focus groups and requirements gathering exercises are a part of the work of this project (for example in the design of the portal user interface and in the assessment exercise of which datasets to Grid-enable), a coherent user-centred design methodology will be challenging given the relatively small parcels of work focussed on developing existing (rather than new) applications. Investigations are underway to ensure the project retains a consistent focus on user needs through the project's internal evaluation exercise, which includes focus groups of project team leaders. The necessity of placing the user at the centre of the development process will also feed into investigations into how to sustain project objectives.

From Research to Service

Much of the project's programme of work involves building on outputs from the first phase of the NCeSS research programme, which yielded many promising e-social science applications. One of the challenges for this project is make the transition for tools that were initially research outputs to prototype services that can support more effective social science research.

This shift in emphasis has implications for academics involved in the original research when engagement in software development to create robust tools for others may not offer the same recognition as more traditional research outputs, such as academic papers.

Additionally, the investment required to develop robust and reliable software is high and may be beyond what this project can achieve. (In this respect, this challenge is similar to the comprehensive user engagement challenge described above.) This is one of the reasons why work to investigate the sustainability of project outputs and objectives is so important to enable this project to become a platform for future work. There must also be management of expectations of what this project can – and cannot – achieve that is commensurate with the resources available.

Embedding within an Infrastructural Framework

The final challenge relates to embedding the infrastructure and associated tools and services within a wider infrastructural framework as offered by the NGS. The NGS's mission statement is to "provide coherent electronic access for UK researchers to all computational and data based resources and facilities required to carry out their research, independent of resource or researcher location". Whilst it is desirable for sustaining the project outcomes to embed the infrastructure within such a centralised national Grid service, the social science community is likely to have particular needs and present particular challenges to a model that has to date largely been focussed on the provision of Grid services for the natural sciences.

The challenge for this project is to articulate social science requirements to the NGS in order to ensure that Grid service provision is appropriate to this community. If successful, we have an opportunity to shape aspects of NGS service provision so that it can offer better support for end-users. This is likely to have ramifications beyond the domain of social science to improve the NGS for all its users.