The ESRC National Centre for e-Social Science

Peter Halfpenny and Rob Procter
National Centre for e-Social Science

Introduction

The purpose of this paper is to briefly sketch the background to the establishment of the UK National Centre for e-Social Science (NCeSS) and its current operation before describing its plans to facilitate the adoption of e-science tools and services within the social science research community. The bulk of the paper provides a roadmap of future e-Infrastructure developments that have the potential to overcome a variety of key social science research challenges.

NCeSS

NCeSS was established by the Economic and Social Research Council (ESRC) in 2004 as its key contribution to the UK e-Science programme. The Centre forms part of the ESRC's broader strategy to develop leading-edge methodological tools and techniques within the social sciences to enhance the capacity to collect, link, access, share and analyse both quantitative and qualitative data resources.[1] The Council views such methodological development as a key priority as data resources become increasingly multifaceted and multi-layered and the task of analysing them ever more complex.

The Centre's objective is to enable social scientists to make best use of emerging e-Science technologies in order to address the key challenges in their substantive research fields in new ways. In pursuit of this, NCeSS aims to stimulate the uptake and use across the social science research community of distributed computational resources, data infrastructures and collaboration mechanisms by co-ordinating a programme of e-Social Science research, making available information, training, advice and support to the social research community, and leading the development of an e-Infrastructure for the Social Sciences that will provide new resources and tools for social research. The Centre is also responsible for providing advice to the ESRC on the future strategic direction of e-Social Science.

NCeSS has a distributed structure, with a co-ordinating Hub at the University of Manchester, now funded until 2012, and seven three-year research Nodes and twelve smaller one-year projects at universities across the UK. A new round of Node commissioning, open to existing Nodes and new applicants, is currently under way, with the aim of funding eight three-year Nodes from 2008. A further round of smaller projects will follow.

Research Roadmap

The research roadmap is organised around two main strands, the ***applications strand***, which seeks to apply e-Science to the benefit of substantive social science and ***the social shaping strand***, which involves social studies of science and technology approaches to e-Science.

The ***applications strand*** is aimed at stimulating the uptake and use by social scientists of e-Infrastructure in order to make advances in quantitative, qualitative and mixed-methods economic and social research. This strand draws upon unfolding developments

[1] http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/about/strategicplan/

in technologies, tools and services from sources within the UK e-Science programme,[2] and applies them to the particular needs of the social science research community in order to generate new solutions to substantive social science research problems. The long term aim of the applications strand is to build an e-Infrastructure on the UK National Grid Service (NGS)[3] providing Grid-enabled research resources to the wider social science research community. Such resources will include datasets, analysis tools and services, and virtual research environments providing integrated access to them.

The roadmap for the future development of the applications strand is organised around the broad priority areas of ***data infrastructure***, ***data analysis*** and ***collaboration***. In formulating the applications roadmap, the Hub drew heavily on both the Hub's own internal review process and on two external reviews of the NCeSS programme first phase.

Priority 1: Data infrastructure. The ESRC National Data Strategy[4] has identified the provision of a world class social science *data infrastructure* as essential both to improving the re-use of existing data collections and to meeting the challenges of the 'data deluge' arising from a profusion of new, 'naturally occurring' sources of social data. Providing such a data infrastructure requires implementation of a technology strategy that will make available to researchers better tools for describing, locating and accessing data, cleaning it, maintaining its confidentiality, combining datasets, and facilitating secondary analysis. Issues that require research include:

- The proliferation of different user and technical interfaces makes it difficult for social scientists to retrieve and combine **datasets from multiple sources** when seeking to undertake more complex forms of analysis. e-Science technologies offer opportunities to overcome these problems but they face numerous technical, legal, ethical and methodological challenges.

- Many of the **new sources of social data** are distinctive in that they are 'born digital' and continuously updated by people's everyday activities. Research is needed to explore how to realise a 'population observatory' in which social scientists can discover, access and use these new forms of data, while remaining sensitive to ethical issues relating to privacy, confidentiality and access.

- As sources of social data grow and proliferate, the problem of **resource description** becomes more acute and yet remains critical to data discovery and use. Mechanisms are needed to automate adding metadata to datasets.

- Secure procedures for accessing **confidential data** are an overarching requirement for increasing the value to research of existing and potential new sources of social data. Current provision is based on 'safe settings', physically secure locations where researchers must go if they wish to access confidential data. Research is needed to determine whether e-Infrastructure offers opportunities for secure, yet more flexible, access to confidential data. If data subjects are to be persuaded to allow their information to be accessed in this way, they must have trust in the reliability of both the security mechanisms and in the ways that they are put into practice.

---

[2] http://www.epsrc.ac.uk/ResearchFunding/Programmes/e-Science/default.htm
[3] http://www.grid-support.ac.uk/
[4] http://www2.warwick.ac.uk/fac/soc/nds/

Priority 2: Data analysis. Easy-to-use but powerful computational analysis tools will be essential if researchers are to harness the mass of varied digital data that is becoming increasingly available and analyse it in ways that provide a better understanding of complex and dynamic social and economic processes in finer detail and with greater precision. A wide range of issues remain for further investigation to establish where e-Infrastructure tools can offer novel solutions.

- Analyses that require **multiple datasets** pose well-known problems of data harmonisation that are compounded when the datasets are collected by different agencies for different purposes. Another level of complication is added when different data types are brought together, for example, quantitative and qualitative or sound and images or fieldnotes and photographs. Conducting real time analyses adds to the complexity. Can e-Infrastructure tools and services address these issues?

- **Multi-level modelling**, including that which combines physical, biological and socio-economic phenomena, is becoming an increasingly important tool for exploring complex, inter-connected systems such as climate change and disease. Real-time modelling driven by data from sensor grids and from population observatories is increasingly possible. Can e-Infrastructure tools facilitate modelling, encourage the sharing of models, and promote collaboration in model-building?

- The extension of **text mining** techniques to social and economic data will be critical in exploiting new large-scale data resources, with a wide range of potential applications. e-Infrastructure tools bring opportunities to combine text mining with **social network analysis** to extract knowledge from large-scale datasets such as the web. The related technique of **data mining** has yet to find wide application in academic social science research, in part because it is computationally demanding. e-Infrastructure provides the opportunity to overcome this barrier.

- **Visualisation** tools have the promise to make results of complex analyses of multi-dimensional data more intuitively intelligible. Further developments are needed to extend the functionality of visualisation methodologies to capture spatial patterning and network dynamics.

- Solving complex problems often involves **multiple steps** where the output of one step is used as the input in the next. e-Infrastructure tools can expedite execution and management of the analysis process, enable the synthesis of more complex, multi-stage analyses and save them for re-use and sharing with collaborators.

- **Virtual research environments** offer the potential to provide integrated support for the complete research lifecycle, beginning with literature searches and reviews, through analysis, discussing results, to writing papers.

- The capability to gather **qualitative data** now vastly outstrips the capacity to analyse it. One solution would be to develop new kinds of computer-aided analysis techniques. Powerful computational techniques for data mining, text mining and visual data analysis are now available for data exploration, extraction and summarisation. The possibility of deploying such techniques in qualitative social research raises questions of what impact they might have on methods and practices.

- Quantitative research methods are formally powerful but often weak in terms of the understanding they bring while qualitative research methods are descriptively rich but often limited in scope. **Mixed methods** offer an escape from this dichotomy and e-Infrastructure, with its capacity to provide scalable, on demand computing resources and to integrate copious volumes of data, suggests new ways of combining methods.

Priority 3: Collaboration. Meeting the ESRC's strategic research priorities will require more collaborative and inter-disciplinary approaches, and the infrastructure and tools to support them. Continued refinement of these and other types of collaborative tools, guided by a better understanding of a wide range of issues relating to collaborative research, are crucial elements of the e-Social Science roadmap.

- The increasing availability of **collaboration tools** facilitate research collaborations extending over distance and time. Frameworks for building virtual research environments – persistent digital spaces where distributed research teams can share data and tools – are now available.[5] However, much remains to be done to improve the usability and interoperability of collaborative tools, plus their support for – and integration within – the overall research lifecycle.

- **Inter-disciplinary collaborations** involving the social, medical and natural sciences will be essential for success in tackling the ESRC's strategic priorities such as migration, childhood development, aging population and obesity. New collaborative resources, such as 'population laboratories' providing secure access to medical and social data, and new powerful tools for data linking and analysis are required to support such collaborations.

The ***social shaping strand*** is aimed at understanding how e-Science, including e-Social Science, is being developed, how it is being used and what its implications are for scientific practices and research outcomes. 'Social shaping' is defined very broadly to include all social, economic and other influences on the genesis, implementation, use, usability, immediate effects and longer-term impacts of the new technologies. Despite the very substantial current investment in the e-Infrastructure in the UK and elsewhere, 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.

Very broadly, the social shaping roadmap agenda falls into four themes:

- The genesis of e-Infrastructure, including historical comparisons with the development of other communications technologies and a consideration of the broader institutional and political contexts;
- Social, economic and other determinants of the design, uptake, use and sustainability of e-Infrastructures;
- The implications for the nature and practice of science, including social science, and for the character and direction of knowledge production, validation and use;
- International comparisons, examining how different national science policies and legal frameworks influence the funding and organisations of these developments.

[5] One example is Sakai. See http://sakaiproject.org/

The need for further social shaping research is urgent because the initial investments in e-Science are establishing systems that will gradually constrain the direction of developments in the future. While e-Science remains fluid at present, it is beginning to take shape in ways that need to be understood.

The question of **sustainability** presents a key challenge if, as predicted, the adoption of e-Infrastructure promotes an increase in both the numbers and types of research resources available. The accumulation, sharing and re-use of resources lies at the heart of the e-Research vision, however, it seems that responses to the support and financial issues this raises have yet to be factored into planning for e-Infrastructure sustainability. A fundamental question that needs to be addressed is how resources originating in time-limited projects can be curated and managed so that they remain viable for re-use in the long term. In particular, where, in a landscape of multiplying, diverse and distributed resources, will the necessary effort and expertise will come from, and what funding models are most appropriate to pay for it.

The social shaping strand offers the opportunity not only to study the emergence of e-Science but also to intervene in its development trajectory by, for example, contributing social science expertise to the development of e-Infrastructure, the engineering of user interfaces, the development of training and information services, and the raising of public awareness and understanding of ethical issues. More generally, the results of social scientific investigations of e-Science can feed back into e-Science and e-Social Science, for example, by providing an understanding of why e-Infrastructure is being adopted in certain circumstances and not in others, enabling the e-Scientists and e-Social Scientists to remove barriers to adoption. In other words, the outputs of the social shaping research can help monitor and modify e-Science, for the benefit of the applications strand.