The e-Social Science research agenda

Peter Halfpenny and Rob Procter
Manchester eResearch Centre, University of Manchester

Introduction
The aim of this paper is to review the lessons learned from the first five years of the UK Economic and Social Research Council’s (ESRC’s) National Centre for e-Social Science (NCeSS) and outline the implications for the future development of the e-social science research agenda. The first section briefly describes the structure and activities of NCeSS. The second notes the growing diversification of e-social science and how this relates to e-science and e-humanities. The third section comments briefly on the changing technical environment over the past five years, and the fourth section identifies issues that have arisen around funding. The fifth section considers the social science research community’s responses to the e-social science programme. The sixth section concludes with some pointers towards a more grounded programme unfolding over the next few years, while noting the danger that e-social science’s transformational opportunities could yet be missed because of barriers emanating from the wider context within which the initiative operates.

The National Centre for e-Social Science
Over the period 2001 to 2006 the UK Government invested £213M in an e-Science programme. The funding was distributed to the Research Councils, with the ESRC’s allocation being £13.6M over the five years. The ESRC’s programme began cautiously with the commissioning of four scoping studies (Cole et al., 2003; Fielding, 2003; Anderson, 2003; Woolgar, 2003), followed by £500K to fund a set of eleven pilot demonstrator projects. These were intended to test the social science research community’s interest in e-social science by enabling ‘early adopters’ to generate exemplars that would demonstrate the potential of Grid technologies to advance social science and encourage others to adopt the emerging technologies. Subsequently, the ESRC commissioned NCeSS with £6M for the first phase from 2004 to 2007 and, following a positive review, £8M for a second phase from 2008 to 2012. The Centre has a distributed structure, with a coordinating Hub responsible for managing the programme and a small number of large three-year projects or ‘Nodes’ devoted to developing innovative tools and services.

The Scope of e-Social Science
The initial emphasis of the e-science programme was on the Grid, that is, the hardware, software and standards necessary to co-ordinate geographically distributed compute and data resources and deliver them to researchers regardless of location. The ambition was to facilitate bigger and faster science, with collaborators world-wide addressing key challenges in new ways. This model was particularly appropriate to particle physics, and such challenges as weather predictions and earthquake modelling. It was less matched to those disciplines subsequently encouraged to join the e-science bandwagon, including the social sciences, where a mixture of numerous quantitative and qualitative methods is used to pursue relatively small scale issues, with few generic problems requiring complex middleware to coordinate huge distributed compute and data resources.

Accordingly, as the NCeSS research programme unfolded, e-social science broadened out to include the use of digital data harvested from the Web to capture people’s views and map their individual behaviours and their networks; the exploration of new forms of digital data, such as mobile phone logs and GPS (Crabtree et al., 2006); the creation and exploitation of metadata to facilitate the sharing and reuse of data (Hielkema et al., 2007); linking data about individuals and the confidentiality and ethical issues that this raises; webometrics; mapping geo-referenced data (Hudson-Smith et al., 2009); large-scale social simulations of various sorts (Townend et al., 2007); parallelisation of statistical routines; collaborative markup of video data about social interaction (Fraser et al., 2006); text and data mining (Ananiadou et al., 2007; Gibson et al., 2007); and tools for delivering behavioural interventions over the internet.

In the humanities, scholars have been particularly enthusiastic about the digitisation of their data resources and advanced search facilities because in combination these dramatically increase the efficiency with which original sources can be discovered, interrogated, and combined (Goldberg, 2009).

The Changing Technical Environment
The notion of the Grid being at the core of e-science has gradually given way to an emphasis on e-infrastructure (cyberinfrastructure in the US). This is not simply a synonym for the Grid but marks a shift of attention to a broad range of computing tools and services that support the everyday work of scientists, including those that are loosely collected together under the title of Web 2.0. While these are technically less powerful than Grid-based solutions, their relative simplicity – both in terms of implementation effort and ease of use – has made them attractive to users who do not need more sophisticated tools and services, and who are deterred from using Grid services by their complexity and the perceived barriers to access. Many of these tools and services are freely
available on the Web (usually in exchange for accepting advertisements) or are open source, and some have very active and technically adept support groups available. These features considerably reduce the barriers to uptake.

**Funding**

As the NCeSS Nodes have accumulated demonstrators that illustrate the potential of advanced computing technologies to support social science research in a variety of ways, it has become apparent that the innovation pipeline from research to stable and supported tools includes stages that are difficult to support in a research project based funding regime (see Figure 1).

![Figure 1: e-Social Science Innovation Pipeline](image)

Computer scientists’ careers advance through their innovating to the level of proof-of-concept or prototype. The subsequent work necessary to make tools stable and platform independent does not attract rewards in academia, and there is no career structure within UK higher education for the software developers needed to productionise research outputs. A third stage is the provision of support for end-users, including documentation, bug-fixes, and training. Again, this type of activity is rare in universities and carries few academic rewards in research-led institutions.

Just as the second two phases of the innovation pipeline are not promoted by the academic reward system, nor are they supported by the Research Councils’ project-oriented funding. What are needed are new business models that provide a route from research to fully productionised and supported tools, and provide some prospect of long-term sustainability. These business models will need to be agreed by a wide variety of agencies, from the Treasury and government departments (especially the Department of Business, Innovation and Skills, now responsible for university funding), the Funding Councils, the Research Councils, JISC, universities themselves and the commercial sector.

**Understanding the Social Research Community**

‘Early adopters’ are, by definition, keen to experiment with innovations and to take risks. Adoption of e-Infrastructure by the wider social science research community is handicapped (as it is for e-Research as a whole: see Voss et al., 2008) by a complex of factors: lack of awareness of the opportunities e-Infrastructure provides; problems in translating innovations in one field into one’s own research; risk aversion; and IT support that is often dictated by institutional policies and priorities rather than individual researcher needs. Late adopters are resistant to training and require shallow learning curves if they are to adopt new ways of working. They may feel they can achieve their goals – publications and promotions – using the tools with which they became familiar as graduate students. This environment is not conducive to the wide uptake of innovative tools and services.

**e-Social Science of the Future**

The e-social science programme has become highly disparate. From the initial sharp focus on the Grid, which has had very little take-up in the social sciences, it has expanded to embrace an increasing range of innovations in digital technologies. On the one hand this willingness to embrace a broad range of emerging digital technologies is positive, insofar as it reveals that a modest input of technical support can ease the research process and open up new avenues of inquiry, provided that there is very close engagement between developers and users to track and respond to changing requirements so that practices and tools can co-evolve, and provided too that effective local support structures are established (Voss et al. 2008).
On the other hand, radical ambitions for transforming everyday social science have been tempered in the light of growing evidence about the very real barriers slowing widespread adoption of advanced tools and services across the social science research community. Researchers’ restricting themselves to the sorts of social science that can be achieved through an unsystematic mix of existing technologies is severely limiting. Given the huge volume of digital data we generate as citizens in a digital world and the scale of compute resources available to make sense of it, there are real opportunities to conduct a sort of social science that was impossible a decade ago. Routinely recorded transaction and administrative data frees us from reliance on small sample surveys that ask people what they do; instead we have enormous bodies of digital data about what people actually do. Similarly, multi-function mobile phones already act as real-time sensors of their owners’ activities and their scope will continue to expand as they are used to deliver more services. Savage and Burrows (2007) argue that only by embracing the opportunities that new forms of social data make possible can academic sociology be rescued from irrelevance.

Research challenges such as understanding individual behaviour; population change; energy, environment and climate change require more collaborative and inter-disciplinary approaches. Easy-to-use, powerful tools are essential to harness the mass of varied digital data and analyse it in ways that provide a better understanding of complex, dynamic social and economic processes in finer detail, at different scales and with greater precision. Researchers need better and more integrated tools to manage the research lifecycle, from literature review and hypothesis formation, on to data discovery, collection and management, data analysis, through to publication.

Necessarily, however, using these new sources of data to their full potential demands that social scientists embrace innovative research methods and this they have shown themselves reluctant to do. Unfortunately, the incentives that might persuade them lie outside the control of the e-social science programme. These include revisions to the academic reward system to promote data sharing and collaboration, a break from the short-termism of project grants and instead a willingness to offer secure funding for cumulative research over the long term, coordinated implementation by a broad range of stake-holders of a mix of business models that can drive innovations along the innovation pipeline towards sustainability, and much more technical support at the elbow of social science researchers. Without an overhaul of the context within which e-social science operates, the programme may well dissipate into a few isolated remnants, like so many initiatives in the past.

Bibliography


