ESRC National Centre for e-Social Science

Call for Research Nodes
Deadline: 4pm on Tuesday 3 July 2007

1. Background

1.1. UK e-Science Programme

In 2001, the Office of Science and Technology[1] launched the UK e-Science programme, an initiative involving all the research councils[2]. John Taylor, then Director General of the Research Councils, explained that: 'e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will support it'. This infrastructure, known as 'the Grid' or increasingly commonly as 'e-Infrastructure', comprises networked, interoperable, scalable computational tools and services that make it possible to locate, access, share, aggregate and manipulate digitised data seamlessly across the Internet on a hitherto unrealisable scale, thereby enabling advances in scientific research that would not otherwise have been possible.

1.2. ESRC Strategy

In April 2004, the National Centre for e-Social Science (NCeSS) was established by the ESRC's as its key contribution to the 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.[3] 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.

Working alongside NCeSS is the National Centre for Research Methods (NCRM) which also forms an integral part of the Council's strategy to develop and mainstream innovative research methods. The Council views the infrastructure provided by these two sister Centres as the essential spine through which it will continue to enhance the general skills base in research methods. In this context the Council expects the Centres to build new research capacity and – via their distributed structure (see below) – act as regional incubators, working in collaboration with other institutions to raise methodological standards.

2. NCeSS

2.1. Aims and Objectives

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

---

[1] Now the Office of Science and Innovation
[2] Except the AHRC, which was not in existence at the time, but which has since established the Arts and Humanities e-Science Support Centre: http://www.ahessc.ac.uk/.
[3] http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/about/strategicplan/

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.

2.2 Structure

NCeSS has a distributed structure, with a co-ordinating Hub at the University of Manchester and a set of research Nodes (currently seven) and smaller projects (currently twelve) at universities across the UK. The existing Nodes were commissioned in phases to develop and pursue the NCeSS research programme and raise awareness of e-Social Science. Each Node has a life-time of three years, and their start dates ranged from December 2004 to February 2006. The current phase of Node commissioning is intended to build on the achievements of the NCeSS research programme to date. It is anticipated that a call for a further round of smaller projects will be issued in late 2007, after the Node commissioning is complete. A list of the current Nodes and smaller projects can be found in Appendices 1 and 2 respectively. Further details and useful links can be found at the NCeSS and ESRC web sites.[4]

2.3. Management

The Hub takes responsibility for managing the Centre. The Executive Director chairs a Centre Strategy Board made up of the Hub Directors and the Director and a second member of each Node, together with a representative of NCRM. The Strategy Board meets three times per annum to review the overall progress and development of the Centre.

The Centre has a single external Advisory Committee, with membership drawn from related initiatives: its composition is published on the NCeSS website.[5] It advises on all aspects of the Centre's work. Nodes are not expected to have their own Advisory Committees.

The Centre's Executive Director is responsible to the ESRC's Research Resources Board, and meets bi-annually with a designated member of the Board and ESRC case officer to review progress.

2.4. Research Programme

The Nodes (along with the smaller projects) collectively undertake a programme of research within the ESRC's overall e-Science strategy. This strategy has two main strands, one which seeks to apply e-Science to the benefit of substantive social science and other 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 in technologies, tools and services from sources such as

[4] http://www.ncess.ac.uk and http://www.esrc.ac.uk
[5] http://www.ncess.ac.uk/about/advisory/

the UK e-Science core programme,[6] the JISC e-Infrastructure programme[7] and JISC Virtual Research Environment programme[8] 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 ***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.

3. Node Proposals

3.1. Outline

This call invites proposals within both of the two research strands. Proposals are welcome from existing Nodes and new applicants, including those unsuccessful in previous calls.

There is no constraint on the substantive topics of proposals in this call other than that they fall within ESRC's remit, as articulated in its Strategic Plan 2005-2010.[9] However, it is essential that Node proposals demonstrate collaboration between social scientists and one or more groups from amongst computer scientists, information scientists, data service providers or resource centres. This complementary expertise is crucial because it will enable e-Science technologies and tools to be fully exploited within the context of well-grounded social and economic research problems. It will also ensure that the Nodes contribute to the development of common technical standards with a broad range of potential applications.

Node applicants are strongly advised to consider how their proposals will complement and build on the work of the current NCeSS research programme and the parallel NCRM research programme. Applicants are recommended to discuss their ideas for collaboration with relevant programme members before submitting their proposals: contact details can be found on the NCeSS and NCRM websites.[10] Applicants should address their proposals' relationship to, and potential synergies with, existing NCeSS and NCRM research programme activities in terms of, for example, substantive research topic, key research questions, methodologies, technology platforms and tools. Proposals should also address plans and prospects for international collaboration where appropriate and relevant.

The ESRC Commissioning Panel in reviewing the NCeSS Node applications will take into account the overall balance of the NCeSS research programme, both within and between the two strands. Applicants should note that NCRM will also be issuing an open call for Nodes with a similar timetable to the NCeSS call. The ESRC may seek to promote synergies between the two Centres' research programmes.

---

[6] http://www.epsrc.ac.uk/ResearchFunding/Programmes/e-Science/default.htm
[7] http://www.jisc.ac.uk/whatwedo/programmes/programme_einfrastructure.aspx
[8] http://www.jisc.ac.uk/whatwedo/programmes/programme_vre.aspx
[9] http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Strategic_Plan_2005-10_tcm6-12995.pdf
[10] http://www.ncess.ac.uk and http://www.ncrm.ac.uk

3.2. Applications strand

3.2.1. Key aims of the applications strand

The **first key aim** of the applications strand is to develop e-Social Science tools and techniques that enhance the ability of the social science community to address their substantive research problems. The goal is to produce advances in social science research that could not have been achieved without the application of e-Science technologies. Applicants should note that, within the applications strand, preference will be given to proposals that investigate research problems of major significance in their disciplines – the 'grand challenges' of the social sciences. As e-Social Science is at a relatively early stage, its developmental path is still fluid and applicants are encouraged to take advantage of this by submitting innovative proposals. Nevertheless, there are three priorities for the applications strand of the NCeSS research programme and these are described in section 3.2.2 below.

The **second key aim** of the applications strand is capacity building. This ranges from dissemination of information about how new e-Social Science tools and techniques are able to advance substantive social science research, through encouragement and support for the take-up of e-Infrastructure among the wider social research community, to direct training in the application of e-Social Science within substantive research projects. In pursuit of this last, applicants are encouraged to link studentship and postdoctoral fellowships to their research programme. They must clearly demonstrate that such a studentship or postdoctoral fellowship award will represent a distinct and independent piece of work, and explain how it will add value to the overall objectives of their proposed Node research programme. Detailed guidance covering grant linked studentships can be found on the ESRC website.[11] Whether or not proposals include a studentship, applicants must explain how the tools and services developed by their Node will be promulgated through awareness-raising and other events, and how they will contribute to the body of online and face-to-face training materials designed to increase e-Social Science capacity within the wider social science research community. Applicants should also explain how they will make use of opportunities for collaboration in capacity building with the Hub and other Nodes, and with other ESRC, JISC and UK e-Science initiatives.

The **third key aim** of the applications strand is to build an e-Infrastructure on the UK National Grid Service (NGS)[12] that provides Grid-enabled research resources to the wider social science research community. Such resources will include datasets, analysis tools and services, and easy-to-use virtual research environments providing integrated access to them. Application strand Nodes will be expected to contribute to this activity as part of their research programmes, for example, by creating new Grid-enabled datasets, tools and services to be integrated into the e-Infrastructure. Applicants must therefore include information about how their Node deliverables will be made available to the wider community through the NCeSS e-Infrastructure for the Social Sciences project[13] and how they will be sustainable beyond the lifetime of their Node.

---

[11] http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/opportunities/current_funding_opportunities/grant_linked.aspx

[12] http://www.grid-support.ac.uk/

[13] http://www.ncess.ac.uk/research/hub/einfrastructure/

3.2.2. Priorities within the applications strand

The Hub's review of the Centre's current research programme indicates that, in order to maintain alignment with the ESRC's strategic objectives, the next stage of the programme should be organised around the broad priority areas of ***data infrastructure***, ***data analysis*** and ***collaboration***. Accordingly, proposals within the applications strand must address one or more issues within these three broad areas.

A preliminary list of issues relevant to these broad areas is presented below, though this is not exhaustive and it is not intended to constrain applicants. Provided that they fall within the three broad areas, innovative proposals are welcome, including ones that address issues outside those listed below, or adopt new approaches that address several of the issues simultaneously. It should be remembered, however, that proposals must involve the application of e-Science technologies within the context of undertaking research on substantive areas of social science.

Priority 1: Data infrastructure. With support from JISC, the ESRC has invested substantially in a distributed social science data infrastructure, most notably ESDS[14] which provides access to numerous datasets for the social sciences. There is also an extensive range of information resources available to social scientists through the JISC Information Environment.[15] In addition, there are large bodies of non-ESRC/JISC-funded data resources. Collectively, these resources play a vital role in providing the evidence base for social science research. The ESRC National Data Strategy[16] 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 the 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.

The NCeSS research programme has made significant progress in identifying and delivering some elements of this strategy. For example, the Digital Records for e-Social Science (DReSS) Node has developed new tools through which multi-media social data (video, audio, transcripts, system logs, etc.) can be collected, collated, described and shared. Other elements of this strategy remain as yet less well defined and understood. Some of the issues that require research are as follows:

- 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. The multiple interfaces also act as a barrier to interdisciplinary collaborative working, and to automating complex analyses that must be rerun periodically (for example, whenever a particular dataset is updated). The integration of data from different sources is hampered by lack of harmonisation among data collection instruments and raises a host of issues about the quality and comparability of different research study designs. When successful, combining data increases the risk of statistical disclosure. e-Science technologies offer opportunities to overcome these problems but they face numerous technical, legal,

[14] http://www.esds.ac.uk
[15] http://www.jisc.ac.uk/index.cfm?name=coll
[16] http://www2.warwick.ac.uk/fac/soc/nds/

ethical and methodological challenges. MIMAS and EDINA have recently completed pilot projects to investigate how the technical challenges can be addressed[17] and these need to be extended and further developed.

- Many of the **new sources of social data** are distinctive in that they are 'born digital' and continuously updated by people's everyday activities. Examples include administrative data such as employment and education records, transactional data such as purchases, logs of mobile phone messages and email, and online material from websites, blogs and forums. The progressive 'instrumentation' of social environments through, for example, mobile and ubiquitous computing devices, raises the prospect of complementing (or even replacing) survey data with the real-time continuous 'live' data gathering for tracking patterns of actions and events. 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. Possible issues here are how public sector and commercial datasets might be shared and linked, and how ethical issues relating to privacy, confidentiality and access could be addressed.

- 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 the process of adding metadata to datasets to describe their features and content, and these will have to address the issue of rival metadata standards. Similarly, mechanisms for extracting or constructing appropriate representations of knowledge for social science domains are required, as are tools to translate between different representations. The National Centre for Text Mining (NaCTeM)[18] is already working to adapt existing text-mining tools to the needs of social sciences, but more research is required to identify requirements and to implement practical, usable tools and services for metadata and ontology generation.

- Within social sciences, where meanings and interpretations are diverse and often contested, the application of **formal knowledge representation models** – such as ontologies – remains problematic. The PolicyGrid Node has developed tools which allow social scientists to annotate data using both ontology-based (closed, top-down, expert constructed) vocabularies and folksonomy-based (open, bottom-up, user constructed) vocabularies. One important question is whether this combination can deliver a rich vocabulary that is broadly shared and comprehensible by the user base, and that has the capacity to respond to language change – without the errors that inevitably arise in naive, unsupervised folksonomies. A recent Agenda Setting Workshop has mapped out some areas for further exploration.[19]

- 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. For example, ONS has established a Virtual Micro-data Laboratory (VML) at several sites within the UK and the ESRC plans shortly to announce a call for the development of a secure data service within an HEI location. Nodes addressing issues around confidential data will be required to collaborate with this ESRC-funded service to investigate, for example,

[17] URLs for MIMAS and EDINA projects.
[18] http://www.nactem.ac.uk
[19] http://www.ncess.ac.uk/events/ASW/ontologies/

whether it – and VMLs – impose constraints on how data is accessed that inhibit the development of new research. The e-Infrastructure development community has invested substantial effort in devising mechanisms and models for secure authentication, authorisation, access control and communication. Research is needed to determine whether these mechanisms create opportunities for secure, yet more flexible, access to confidential data. At the same time, if data subjects are to be persuaded to allow their information to be accessed in this way, they need to be confident that the privacy of their data is maintained. To achieve this, data subjects must have trust in the reliability of both the security mechanisms themselves and in the ways that they are put into practice. Protocols need to be established which take account of possible different legal frameworks and ethical issues associated with different data sources, and that are compliant with local legislation, such as the UK Data Protection Act. Appropriate mechanisms for accountability need to be established for credible research governance. Linking of multiple data sources carries the risk that individuals could be identified by accident or design. This risk needs to be minimised in a manner that does not significantly reduce data quality.

- The enhanced means that e-Infrastructure can provide for **documenting and managing project archives** – their datasets, data collection instruments, analysis routines, and published and unpublished results – represents a powerful and novel resource for the dissemination of research findings. For example, it is now technically feasible to document and publish the entire research process, linking data sources, analysis workflows, results and papers as a re-usable research audit trail or provenance. Already, in some fields, publication is conditional on researchers' compliance with agreed provenance standards.[20] Possible issues here concern how the social science research community might take advantage of and adapt these technologies, and what their impact might on analytical practices.

- The question of **sustainability** presents a key challenge if, as predicted (and desired), the adoption of e-Infrastructure promotes an increase in both the numbers and types of research resources available. There will be a cost to realising their value and sustainability strategies based on centralised funding and management may prove inadequate to the task. Issues which need to be investigated include how resources generated in time-limited projects can be curated, managed and maintained so that they remain available for and re-usable by their user communities in the long term, where the necessary effort and expertise will come from, and what funding models are most appropriate in a landscape of multiplying, diverse and distributed resources.

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. Examples of work in this area are provided by the MoSeS Node, which has developed large scale simulation tools to support policy-makers in strategic priority areas such as population change, and the PolicyGrid Node, which has been exploring how to support the use of mixed methods and enable interdisciplinary research collaborations

[20] MIAME – the Minimum Information About a Microarray Experiment– defines what is needed to enable the unambiguous interpretation of array based gene expression monitoring experiments and for other researchers to reproduce them.

between environmental and social scientists. 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?

- **Simulation** is a powerful method for exploring complex, dynamic social systems. Can e-Infrastructure foster the adoption of simulation and enable it to be undertaken on sufficient scale and draw on sufficient cross-disciplinary (economic, social, environmental, medical) data to guide policy-makers in areas such as population change, housing, health and transport planning?

- **High performance computing**: analysis techniques where optimisation is involved enable distributed computing to be employed by breaking down the problem into separate pieces for parallel processing by multiple processors. This suggests that present techniques that use, for example, Markov chain Monte Carlo methods or statistical bootstrapping will have a natural home on e-Infrastructure. The availability of scalable processing will also make some non-parametric techniques more feasible, and enable more exhaustive, automated, model specification searches to be performed. Parallel processing can also be used with less structured data, and could be employed, for example, to enhance or even automate content analysis and linguistic research of very large natural language corpora. Analyses of visual images will also offer opportunities to show the utility of employing high performance computing.

- **Ease of use**: in order to lower barriers to the adoption of e-Infrastructure, it is essential that access to computational and data resources should be as transparent as possible to the user.[21] In pursuit of this, the CQeSS Node has developed mechanisms that enable researchers to access Grid-enabled data and computational resources from some common desktop statistical packages, and the possibility of extending this approach needs to be explored. Work is also needed to develop and refine user environments based on familiar browser-style interfaces where users can gain access through a single authentication process, then discover resources (data, tools and services) and create their own set of linked procedures (known as 'workflows') to carry out their analyses.

[21] This is discussed in http://www.nuff.ox.ac.uk/economics/papers/2001/w22/hpc20013.pdf, in relation to Ox.

- **Visualisation** tools have the promise to make the results of complex analyses of multi-dimensional data more intuitively intelligible to researchers and policy makers. For example, the GeoVue Node has developed a number of GIS-based visualisation tools, ranging from a simple-to-use tool based on Google Maps for displaying social-economic data and to more sophisticated tools for overlaying real-time data (such as air pollution) on 3-D simulations of urban environments. Because of the immediacy of visualisations, and the opportunity they provide for interactivity with data, such tools have the potential to encouraging greater public engagement in policy making and research. A further development could be to extend the functionality of visualisation methodologies to capture spatial patterning and network dynamics.

- 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, including the identification of social trends and the extraction of opinions and market sentiments. 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** is widely used in private sector research for extracting knowledge from structured socio-economic databases but 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.

- Solving complex problems often involves **multiple steps** and iterations where the output of one step is used as the input in the next. Managing these steps manually is potentially difficult, and performing the data integration and modelling using desktop PC tools may be very time consuming. e-Infrastructure tools enable researchers to compose research workflows as sequences of analytical steps using different data sources and tools, and execute them (semi-)automatically.[22] They offer the potential to 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 also offer the potential to provide integrated support for the complete research lifecycle, beginning with literature searches and reviews, through analysis, discussing results, to writing and publishing papers.

- The capability to gather **qualitative data** now vastly outstrips the capacity to analyse it. One solution to this problem would be to develop new kinds of computer-aided analysis techniques and tools. For example, 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.

- The wide uptake by organisations and the public of the internet and mobile devices has lead to the emergence of **novel forms of social interaction** and community such as informal peer-to-peer networks, empathetic communities, virtual teams and virtual organisations. Their study poses novel methodological challenges. Possible issues here concern what kinds of so-called 'virtual methods' are needed to explore and understand these new realms of social interaction. Virtual communities might

[22] See, for example, the P-GRADE portal: http://www.lpds.sztaki.hu/pgportal/

provide a promising site for combining quantitative and qualitative research methods. Virtual research teams might be one solution to the problems of studying distributed, virtual communities. One possibility would be for researchers to use e-Infrastructure to perform social science experiments or observational studies over distributed sites. Research is needed to investigate what tools are needed to support such a development.

- 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. Agent-based modelling, drawing upon qualitative research to enrich formalized descriptions of the real world and capture its complexity, is one example of how this might be achieved.

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. Good progress has already been achieved by the NCeSS programme and the JISC Virtual Research Environment programme in developing collaborative tools and environments. For example, the Access Grid – a high quality, multi-site video-conferencing system – is now used routinely to facilitate meetings between distributed participants. A variety of tools, albeit some only in pilot form, exist to augment the capabilities of the Access Grid and provide effective meeting support.[23] 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, will be fundamental to the success of e-Social Science.

- The ease of sharing data within e-Infrastructure and the increasing availability of **collaboration tools** facilitate research collaborations extending over distance and time. Some of these tools, like wikis, are generic in their applicability while others are more specialised. As an example the latter, the MiMeG Node has developed a suite of tools to support the collaborative analysis of video data by distributed research groups. Frameworks for building virtual research environments – persistent digital spaces where distributed research teams can share data and tools – are now available.[24] In general, 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. Similarly, progress in understanding diseases such as diabetes, which the NHS has identified as the major health challenge of the early 21st Century, will only come through collaboration between medical and social sciences. 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.

---

[23] One example is memetic: See http://www.memetic-vre.net/

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

3.3. Social shaping strand

This strand of the NCeSS research programme draws on social science studies of science and technology. The aim is to understand how e-Science, including e-Social Science, is being developed, the barriers to its adoption, how it is being used, what its impacts and implications are, and how it will be sustained in the medium and long term. 'Social shaping' is defined very broadly here to include all social and economic 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. When investigating the factors that lead potential users to be favourably disposed or otherwise towards adopting new technologies, it is necessary to consider variations across different constituencies, since different areas of science are using and responding to e-Science technologies in quite different ways. For example, the value of e-Infrastructure for sharing data and facilitating collaboration is likely to be of more immediate interest to social scientists than access to scalable computational resources, which is attractive to natural scientists wanting to process very large amounts of data.

The sorts of research issues that might be addressed include: what social circumstances and institutional arrangements encourage or inhibit data sharing and collaborative working? How will the implementation of new e-Infrastructure affect these? Are existing patterns of informal and formal communication between scientists likely to be significantly affected by the adoption and use of e-Infrastructure? Will the enhanced technical capacity for communication lead to more concerns about privacy, intellectual property and research ethics? Will dramatically increased knowledge sharing produce changes in views about trust, reciprocity and disclosure? An Agenda Setting Workshop on the topic of trust and ethics has mapped out the issues where further research is needed if e-Science and e-Social Science are to realise their potential, especially with respect to secondary uses of administrative, transactional and healthcare data. Specific areas highlighted included trust and ethics in the context of the researcher-researcher relationship, the researcher-data subject relationship and the researcher-public relationship. It will be particularly important for the future of the social sciences to understand public attitudes to the re-use of confidential administrative and healthcare data, and to devise strategies for educating the public about the risks and the benefits.

Other research issues include: what social and economic factors affect the uptake and use of e-Infrastructure in different organisational and institutional settings? Are changing scientific reward and recognition practices affecting communication practices, and vice versa? Does the emergence of e-Science have any relation to the commercialisation of knowledge? What procedures and mechanisms are needed to assist in the co-evolution of technologies and research practices? What roles might the e-Social Science user community play in the definition and choice of technical standards for e-Infrastructure components and services? How can the usability of the new technologies be investigated and addressed? What models of technology supply are beginning to emerge and are they likely to be trusted for ensuring the future sustainability of supply?

Another direction emphasises the differences within the social sciences. Given the longstanding disputes between intellectual traditions which favour codification (quantitative data) and those that favour interpretation (qualitative information), will e-Infrastructure been seen to support one tradition over the other? Or do the new technologies provide opportunities to bridge the two traditions in novel ways that support mixed-method studies?

The research questions set out above are far from exhaustive but indicate how the social shaping issues are specific instances or more general questions about virtual communities, communication practices and social networking that exercised social scientists before the computer age. In what ways will the arrival of e-Science modify practices? Do these modifications merely continue the changes that have resulted from the spread of IT or are radically new transformations underway?

Very broadly, the social shaping research 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.

The Oxford e-Social Science Node (OeSS)[25] is addressing a selection of legal, ethical and social issues within these themes, relating to confidentiality, privacy, data protection, intellectual property rights, accountability, and trust and risk in distributed collaborations. Nevertheless, the need for further, complementary work 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. In this respect, NCeSS itself might be the object of inquiry within the social shaping strand, the Hub or Nodes providing case studies of the development of e-Social Science, its determinants and effects, perhaps to be compared with parallel e-Science centres.

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.

[25] http://www.ncess.ac.uk/research/nodes/OeSS/

3.4. Collaboration between Hub and Nodes

Given that the aim of NCeSS is to pursue a collaborative and integrated programme of e-social science research and development, and capacity building, in all cases successful applicants will be required to agree to a governance framework setting out the responsibilities of the grant-holders and the NCeSS Hub. Included in this framework are:

- contribution to the Centre's planning, co-ordination and strategic development through participation by the Node Director and one other member in the Centre's Strategy Board and any Task Forces that it might establish to pursue particular developments;
- the preparation of quarterly and annual reports according to requirements specified by ESRC in consultation with the Hub, and their submission to the Hub, which incorporates them into overall NCeSS reports for submission to the ESRC Research Resources Board;
- contribution of publicity and training materials to the Hub on request, including materials for publication on the central NCeSS website, which operates as the single front-door to UK e-Social Science information and activities. All such materials will be covered by an agreed IPR framework;
- presentation of papers about the Nodes' work at the annual NCeSS International Conference on e-Social Science organised by the Hub. The venue for this alternates between Manchester (2008, 2010, 2012) and an overseas location (2009, 2011). Some financial support for travel, accommodation and registration is provided by the Hub;
- contribution to the annual NCeSS Training School or Showcase;
- outreach activities, including contributions to professional associations' and other conferences, in collaboration with the Hub and other Nodes;
- pursuit of research synergies, for example, with other NCeSS Nodes or with NCRM Nodes, or other appropriate initiatives.

In order to facilitate collaboration between Nodes and the Hub, proposals must include both sufficient staff time and the necessary budget for the above activities, in particular:

- attendance of the Node PI and one other at NCeSS Strategy Board meetings in Manchester three times per year (travel and subsistence costs);
- attendance of all Node PIs, research and administrative staff at the annual two-day gathering of NCeSS, known as the Jamboree and held in Manchester (travel costs – subsistence costs covered by the Hub);

4. Funding

4.1. Budget

The overall budget for the 2008-2011 NCeSS research programme is £10 million of which the ESRC will pay £8 million (80% of the full economic cost). The majority of the budget is divided between two streams, one supporting a set of research Nodes and the other funding a small grants scheme. The budget available for the NCeSS Nodes that are the subject of this call is £8.5 million of which the ESRC will meet£6.8 million,

representing 80% of the full economic cost. The call for small grant projects will be issued at a later date.

4.2. Node awards

It is anticipated that up to eight Nodes will be funded, each with a broad research programme. The average budget per Node is expected to be of the order of £1,060k, of which the ESRC will meet £850k, 80% of the full economic cost, spread over three years, though there is scope for variation around this figure. Applications for small Nodes, with a focussed project and a smaller budget than the average are welcome. Exceptionally proposals for Nodes with broad research programmes requiring larger budgets will be considered, where a strong and compelling case is presented. In all cases, the budget must be fully justified in terms of the scope of the research programme and its contribution to NCeSS's and ESRC's overall strategies. As with all ESRC investments, value for money will be taken into consideration by the Commissioning Panel.

The Nodes could be located in one institution or, given the emphasis within e-Science on collaboration, have a distributed structure involving partnerships across two or more institutions.

Proposals within the applications strand must demonstrate how e-Science technologies will be applied to substantive economic and social research problems in order to develop new solutions. Proposals that fall entirely within the domain of computer sciences, on the one hand, or which do not embrace e-Science technologies as an essential component of social science, on the other, will be rejected. Proposals within the social shaping strand must give a central role to e-Infrastructure developments. Proposals that investigate ICTs in general will be rejected.

e-Social Science will advance through collaborations between social scientists and computer scientists, yet research groups in one of these areas might not have established relations with groups in the other. Similarly, groups with interesting ideas for applying e-Science within their substantive area of research might lack established connections with appropriate data providers. In either case, applicants are invited to contact the NCeSS Hub well before the submission date for advice, although it must be noted that provision of such advice is without prejudice to the recommendations subsequently made by the ESRC Commissioning Panel.

Applicants will be required to provide information about their exit strategy. In particular, in the applications strand they must specify how they will ensure that project outputs made available to the wider community, such as Grid-enabled data and analysis resources, are sustainable beyond the life of the Node. In the social shaping strand, applicants should set out how the developing e-Social Science research programme can benefit from their findings about the influences on such developments.

5. Application process

There will be a single stage application process. Full proposals must be submitted by 4pm on Tuesday 3rd July 2007 through the Research Council's Joint Electronic Submissions (Je-S) process. The ESRC's NCeSS Node Commissioning Panel will consider the proposals in October 2007 and recommendations for funding will go to the

Research Resources Board for approval in November 2007. Decisions will be announced shortly after. Funded Nodes will commence work between January and September 2008.

There will be a **Briefing Meeting** for potential applicants, at which the Hub will present information about NCeSS and the call, with ample opportunity for questions. This meeting will take place from 2pm to 4.30pm on Friday 25 May in the Koln Centre at the Royal Society, 6-9 Carlton House Terrace, London, SW1Y 5AG. The meeting is free but attendees must register beforehand. The registration form on the NCeSS events page (http://www.ncess.ac.uk/events/) will open in early May.

It is emphasised that this call is for proposals that apply e-Science technologies to substantive social science research. Free-standing e-Science proposals that do not engage with the social sciences and social science proposals to which e-Science is not critical are excluded.

See the Guidance Notes for Applicants for further details. These are available on the ESRC website http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/opportunities/

Applicants for funding can contact the Hub team for advice by email addressed to **consultation@ncess.ac.uk.**

Appendix 1

e-Social Science Nodes

Applications strand

- Collaboratory for Quantitative e-Social Science (CQeSS). Director: Professor Rob Crouchley, University of Lancaster.

- Modelling and Simulation for e-Social Science (MoSeS). Director: Dr Mark Birkin, University of Leeds.

- Understanding New Forms of Digital Record for e-Social Science (DReSS). Director: Professor Tom Rodden, University of Nottingham.

- Mixed Media Grid (MiMeG). Director: Dr Mike Fraser, University of Bristol.

- Geographical Urban Environments (GeoVUE). Director: Professor Mike Batty, University College London.

- Semantic Grid Tools for Rural Policy Development and Appraisal (PolicyGrid), Director: Dr Pete Edwards, University of Aberdeen.

Social shaping strand

- Oxford e-Social Science Project (OeSS). Director: Professor Bill Dutton, Oxford University.

Appendix 2

e-Social Science Small Grant Projects

Applications strand

- The use of Grid computing to facilitate disclosure risk assessment. Principal Investigator: Professor John Gurd, University of Manchester.

- New technologies, new applications: using Access Grid Nodes in field research and training. Principal Investigator: Professor Nigel Fielding, University of Surrey.

- Collaborative e-Science for spatial decision-making in distributed environments (CESDEMIDE). Principal Investigator: Dr A Berardi, Open University.

- HeadTalk: understanding the nature and role of gestures in 1-to-1 conversation. Principal Investigator: Professor Ronald Carter, University of Nottingham.

- Grid-enabled occupational data environment (GEODE). Principal Investigator: Dr Paul Lambert, University of Stirling.

- The data chronicles: a repository for social science metadata. Principal Investigator: Dr Karen Clarke, University of Manchester.

- Grid-enabled data collection and analysis: semantic annotation in skills-based learning. Principal Investigator: Professor David de Roure, University of Southampton.

- Intelligent data-driven simulation for policy decision support in the social sciences. Principal Investigator: Dr Georgious Theodoropoulos, University of Birmingham.

- Integrating field and systemic data in a visualisation for collaboration. Principal Investigator: Dr Matthew Chalmers, University of Glasgow.

- Grid-enabled spatial regression models (with application to deprivation indices). Principal Investigator: Dr Richard Harris, University of Bristol.

Social shaping strand

- Learning disabilities data and information infrastructure. Principal Investigator: Dr Udo Kruschwitz, University of Essex.

- Entangled data: knowledge and community making in e-(Social) Science. Principal Investigator: Dr Ben Anderson, University of Essex.