The Role of e-Infrastructures in the Transformation of Research Practices and Outcomes

Eric T. Meyer   Ralph Schroeder   William H. Dutton

Oxford Internet Institute, University of Oxford
1 St Giles, Oxford, OX1 3JS, UK
Tel +44 (0) 1865 287210

eric.meyer@oii.ox.ac.uk   ralph.schroeder@oii.ox.ac.uk   william.dutton@oii.ox.ac.uk

ABSTRACT

In this paper, we examine transformations that have taken place in e-Research, and address the potential for additional transformations as e-Research develops and matures. The notion of a transformation in e-Research can operate on many levels: transformations in the tools used to conduct research, transformations in projects that enable new types of e-Research, transformations to ordinary scientific practice, transformations in the types of scientific questions that can be asked and able to be asked, and transformations in the scientific imagination. While much of the current rhetoric implies that e-Research will transform the very nature of science, other types of less-pervasive transformations are more evident at these relatively early stages in the development of e-Research infrastructures, and some evidence supports the idea that continuity has been more common in the ordinary scientific practice of e-Research rather than transformation. The data from this paper draws on the work of the Oxford e-Social Science project (OeSS).

Topics

Information infrastructure development
Information policy, ethics and law

Keywords

e-Research, e-Social Science, scientific transformation, information infrastructure

1. INTRODUCTION

There are major efforts underway in the United States, United Kingdom, Europe, and elsewhere to build scientific infrastructure to enable e-Research. In the U.S., the NSF Office of Cyberinfrastructure is engaged in an ambitious programme aimed at encouraging and developing cyberinfrastructure in a variety of science and social science domains (see, for example, http://www.nsf.gov/od/oci/reports.jsp). In the U.K., the ESRC National Centre for e-Social Science (NCeSS) has recently awarded its second round of funding to projects aimed at understanding the requirements for grid-enabled e-Social Science and developing tools for e-Research [19]. Other specific grid-based computing projects aimed at enabling e-Science and e-Social science projects abound in the U.S., U.K., E.U., and elsewhere.

Much of the language about these new scientific information infrastructures stresses that many funding bodies view e-Research as a way to transform scientific practice. The National Science Foundation (NSF), for instance, has recently revised "the intellectual merit review criterion to specifically include language on potentially transformative research" [20]. Although language indicating the *potential* for e-Research to be transformative abounds, very little work has examined the extent to which this potential is being realized.

The notion of a transformation in e-Research can operate on many levels: transformations in the tools used to conduct research, transformations in projects that enable new types of e-Research, transformations to ordinary scientific practice, transformations in the types of scientific questions that can be asked and able to be asked, and transformations in the scientific imagination. While much of the current rhetoric implies that e-Research will transform the very nature of science, other types of less-pervasive transformations are more evident at these relatively early stages in the development of e-Research infrastructures. In this paper, we examine some of the transformations that have taken place in e-Research, but we also address the potential for various transformations as e-Research develops and matures.

2. SOCIAL SHAPING OF E-RESEARCH

2.1. OeSS: Oxford e-Social Science Project

The Oxford e-Social Science Project (OeSS) is a node of the National Centre for e-Social Science (NCeSS). NCeSS was established with UK Economic and Social Research Council (ESRC) funding in April 2004 "to act as the central resource base for e-social science issues and activities in the UK…[and to] provide a one-stop shop for awareness raising, expertise, training, technical infrastructure, data resources, computer facilities and user-support for e-social science research" [3]. To do this, a central hub was established at the University of Manchester, and eight individual nodes were funded in 2005 for an initial three year period; NCeSS also funded twelve smaller pilot projects that ran for short periods ranging from 9-18 months. In 2008, a second round of three-year funding was announced which included continuing support for the OeSS node through 2011.

Seven of the initial nodes are primarily concerned with creating demonstrator projects which show the potential of e-Social Science in a variety of domains. The OeSS node is somewhat different than the others: rather than creating software tools to enable e-Social Science, the OeSS node studies the social shaping of e-Research. In particular, it focuses on the social, institutional, legal, and ethical contexts that shape the development and adoption of new technologies. It does this work by examining how various e-science and e-social science projects develop and change over time, including cases drawn from the other NCeSS nodes and cases from outside of the NCeSS project.

2.2. OeSS Cases and Themes

The OeSS project is currently in the process of conducting ongoing case studies in several e-Research domains. Each of the case studies involves examining one or more specific projects engaged in developing tools to enable e-Research in a specific domain, and each of these cases also contributes to the investigation of a number of themes in specific areas (including ethics and trust, intellectual property, openness of access, and disciplinarity). An attempt is also made to distribute the cases across a wide range; in social science, this includes qualitative social science, quantitative social science, geospatial modelling and simulation, and web 2.0 collaborative e-Social Science. The methods for this research are a mix of qualitative and quantitative methods, including interviews both with project personnel and the users of their scientific tools, observation of how the tools are implemented, analysis of supporting documentation, and a quantitative survey of U.K. social scientists about their use of e-Social Science tools and their attitudes toward e-Social Science in general.

While these case studies are still ongoing, in this paper we will describe an illustrative case and discuss examples of some of the cross-cutting issues and themes that are emerging across a variety of cases. Finally, we will draw some initial conclusions about how this research helps to illustrate social shaping in e-Research.

2.2.1. Geo-spatial modelling and simulation: GeoVue

GeoVue is an NCeSS node funded for both phase I and phase II of NCeSS. The project is based at the University College of London's Centre for Advanced Spatial Analysis (UCL-CASA), where Mike Batty is the principal investigator. GeoVue's central focus is making tools that allow the fast and easy creation of maps that include data elements for visualizing geo-spatial data. One example of the node's output is the Virtual London demonstrator (see Figure 1), which offered users a virtual fly-through of London using the Google Earth viewer as a platform.

Figure 1: Virtual London (Source: Hudson-Smith [13])

When GeoVue created Virtual London, the target audience they had in mind was primarily urban planners and policy makers. Virtual London was designed not only to create a useful 3-dimension map of London, but also to allow overlaying data. For example, local governments could overlay pollution readings and visualize areas where pollution abatement measures are most needed. Once the application was completed the project attracted the attention of Google, who expressed an interest in including it in the Google Earth viewer so that users anywhere could have access to Virtual London.

It is at this point in the story that institutional and legal arrangements intervened and prevented the public release of Virtual London. The Virtual London application, while it was built in the Google Earth viewer, relied on the Ordnance Survey's *MasterMap* data to build the 3-D projections of buildings. Ordnance Survey is the national agency of the British government that is "responsible for the official, definitive surveying and topographic mapping of Great Britain" [21]. In order to include Virtual London in Google Earth, Google would have had to obtain permission to use the Ordnance Survey data underlying the application. Ordnance Survey data, however, is protected by Crown Copyright, which covers all works "made by an officer or servant of the Crown in the course of his duties" [2]. This is in sharp contrast to the situation in the United States, where "copyright protection … is not available for any work of the United States Government" [1]. Because works created by federal agencies are not subject to copyright protection in the U.S., data created can be used by individuals and by organizations wishing to put it to use. In the U.K., that is not the case – the Ordnance Survey data must be licensed for use.

In the case of Virtual London, Google was willing to pay for a license, and indicated that the actual amount wasn't really too much of a concern. According to a GeoVue staff member, "Google were willing to pay whatever it took, yeah. That was the whole point. They would pay whatever Ordnance Survey said … [But] Ordnance Survey wouldn't budge for any price" [Interview 3 Oct 2007]. The only possible price the Ordnance Survey was able to offer was £1.50 "per click" which was not possible in Google Earth where this type of clicking is not part of the interface. In Google Earth, the user zooms and flies (virtually) through the space. Neither Ordnance Survey nor Google were able to suggest an equitable way of measuring the uses of Virtual London. Even though later discussions took place among GeoVue, Google and the Ordnance Survey, in January 2008 a member of the GeoVue project team indicated that he had given up on holding out much hope for getting Virtual London released and had moved on to other projects [Personal conversation 16 Jan 2008]. It can be added that throughout the duration of the project, there was a high profile campaign led by The Guardian newspaper to open access to the Ordnance Survey and other government data, a campaign which is still ongoing[1]. In other words, the issue of Ordnance Survey access was not limited to the GeoVue project, but part of a wider debate in the UK.

This case is a clear illustration of the legal and institutional issues that can erect barriers to sharing data and implementing e-Research projects. It is often assumed that many projects fail to

[1] See http://www.guardian.co.uk/technology/freeourdata

achieve their goals because of technical barriers to implementation. In the case of GeoVue, however, the GeoVue team had successfully removed the technical barriers and had built a functioning, effective application with the potential to have an impact on public policy and academic research. Legal and institutional arrangements, however, prevented Virtual London from being released widely. The legal issues are based on national copyright requirements, as discussed above.

Institutional arrangements also came into play however. When GeoVue and Google approached the Ordnance Survey for permission to use the Ordnance Survey data that was used to generate Virtual London, they ran into considerable institutional indifference above and beyond the legal barriers. In other words, GeoVue felt that not only was there an issue surrounding copyright and payment for licensing, but also a feeling that Ordnance Survey just didn't think that such an application was of any interest to them. This became even clearer to GeoVue project members after they set up a sample Virtual London in their space on Second Nature Island in the Second Life virtual world. GeoVue felt that the Second Life Virtual London demonstrator would be seen as a feather in the cap of the Ordnance Survey, which was clearly identified as a source of data. The data was not exposed in the demonstrator so it was not making it possible for non-licensed users to obtain protected data, and GeoVue felt Ordnance Survey could use Second Life Virtual London to show people how their data could be used. They were taken aback, then, when the Ordnance Survey response was a demand to immediately remove the demonstrator or risk legal action. They had no choice but to remove it, and replaced Virtual London with another demonstrator, Virtual Phuket.

In this case, several institutional and disciplinary contexts came into conflict. GeoVue is comprised of academic geographers and computer programmers. Like most other academics, the notion of sharing one's results and publishing one's output without expectation of direct financial reward is part of their institutional and cultural expectations. They may also have been somewhat naïve about the legal issues involved, as the team did not include lawyers or copyright experts. There would have been little reason to expect to need to include such experts, when the initial plan was to figure out ways to combine computer programming power with questions of interest to geographers, government planners, and other academics. This lack of expertise, however, resulted in being caught unaware when they came up against the government bureaucrats and lawyers working for Ordnance Survey. Google's institutional culture as an American corporation also had an effect: Google's response was fairly typical of the stereotype of Americans, to simply assume that throwing enough money at the problem would cause it to go away.

While this case illustrates a constraint imposed by legal and institutional considerations, it nevertheless also shows some of the promise of transformation that projects such as GeoVue offer to social scientists. In the past, the use of geospatial information has been relatively limited because of the technical complexity of working with map data and GIS programs. Recently, however, and especially with the advent of the Web, there has been a surge of interest in digital mapping techniques and combining these with different types of data and information. For many social scientists who wanted to include geospatial analysis and data presentation, unless they had a GIS expert on staff or available to them through their organization they would have faced too great a hurdle and often decided to do without the visualization component in their research. GeoVue and another NCeSS node, MoSeS both have been working to make the creation of map data much more transparent to the average scientist; in doing so, they hope to transform the way social scientists see their data. This potential transformation may be quite powerful, and the OeSS project will continue to follow developments in this area[2].

2.3. *Cross-cutting themes and issues*

The case studies in the OeSS project are designed not only to illustrate particular cases of interest, but also to illuminate our understanding of a variety of themes and issues that have emerged during the first phase of the research. These themes and issues are threads that run throughout the work, and help describe how e-Research is undergoing social shaping at this relatively early point in the development of many e-Research projects. Among the important themes are trust, privacy, disciplinarity, and openness. A number of other issues are not properly called themes, per se, but also represent important issues that occur in a variety of projects. These issues include the difficulty of sharing data, the limitations of geography, economic constraints, shifts in funding models, and ways of approaching new scientific questions. In the interests of limited space, we will touch on several of these only briefly here.

2.3.1. Challenges for e-Research

Data sharing

Data sharing has proven to be a significant challenge for many efforts at building collaborative infrastructures for research. Beyond legal and institutional barriers to sharing data, such as the GeoVue case described above, there are a variety of other issues that arise as researchers contemplate sharing data. OeSS project members Carusi and Jirotka, for instance, have published work discussing how a seemingly simple data archive can become an "ethical labrynth" [7]. Digital archives, they argue, can help normalize scientific research by increasing accessibility, uniformity and transparency. Archived data can also help scientists avoid needless repetition, and enable secondary uses of data. However, in practice much social science data resists archiving, particularly qualitative research data. Issues such as guarding subject privacy through anonymization can render otherwise rich data essentially meaningless as important relationships are stripped from sanitized data. Other contextual information may not be apparent in the data, and secondary researchers may mischaracterize data in ways that the original researchers who were embedded in the social relationships under study would not do. Re-using data for purposes beyond those initially stated also has implications not only with regard to informed consent agreements, but also run the risk of jeopardizing the relationship between the original researcher and their subjects if the subjects feel misled or duped by the subsequent uses.

[2] The GeoVue and MoSeS nodes of NCeSS have merged their efforts for phase 2 funding. This will also necessitate some transformations in work practices, as the two nodes are located in different cities and have been targeting somewhat different audiences to this point.

Another major issue standing in the way of widespread data sharing is the unresolved issue of how academic researchers can be assigned credit and rewards for contributing scientific data to a public archive. The reward structure of academia currently favours publication in top peer-reviewed journals, and subsequently having one's work cited by other peer-reviewed articles. Most measures of academic quality offer no credit for having created a dataset that other researchers use, or for creating a software package that enables research. Indexing services such as the Web of Knowledge do not index databases. The promise of e-Research in this area represents a potentially major transformation in scholarly practice: should the publication of scientific data become a standard part of the scientometric measures that measure scientific impact and are often used in the evaluation of researchers, departments, and organizations (see also [5])?

Some efforts are being made to address this issue of academic credit for data sharing. OeSS researcher Meyer has studied an American biomedical project dealing with issues of sharing data, the Genetic Association Information Network (GAIN) [17]. GAIN is a public-private partnership that offered academic researchers high-quality genotyping an order of magnitude more detailed than they had previously used in exchange for releasing their data and blood samples for other researchers and private organizations such as pharmaceutical companies to use. The data distribution for GAIN is being implemented in a system called dbGaP by the National Center for Biotechnology Information (NCBI), a national resource established by the U.S. National Institutes of Health. Among the features of their data distribution site are ways of seeing the number of publications and the authorized data requests that have been made for data (see Figure 2).

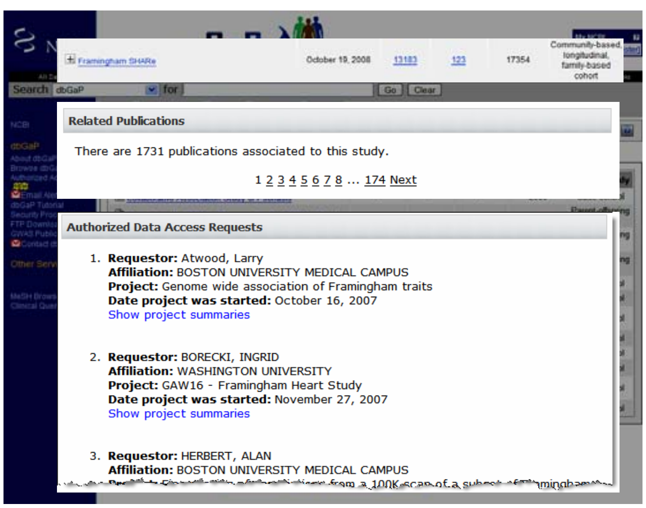

Figure 2: Selections from NCBI dbGap website
(Source: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap)

In Figure 2, we can see parts of the page for the Framingham heart study portion of dbGaP. By identifying the authorized data access requests and the publications associated with the study, the creators of the data begin to have a way to demonstrate the impact of their work creating the dataset. While the well-known Framingham study certainly has many other indicators of success available to them as well, other less well-known studies can track who is using their data, often for the first time. While some of these datasets may have been shared and distributed on an *ad hoc* basis in the past, without a central resource to track data uses, knowing how one's data is used can be difficult or impossible. Furthermore, until resources are made available for professional management of such resources, most datasets will continue to remain scattered across individual researcher's hard drives and local networks.

Again, we see a potential for transformation in scientific practice here, one that is closer to realization through the combination of technological enablers (the dbGaP website), scholarly incentives (access to additional data), and economic incentives (access to additional future funding for scientific work).

e-Science as Open Science?

Another theme of the OeSS project has been the openness -- or otherwise -- of e-Research. One of the promises of e-Research has been that it should enhance open access to knowledge, and thus to "reconfigure access to knowledge" [10] in terms of allowing data and resources from e-Research to be shared and disseminated more freely, with fewer restrictions (for example, by means of copyrights and patents) and perhaps using open source licensing or similar. There has been widespread support from funding bodies, researchers, science policy-making bodies and non-governmental organizations for 'open science' and 'open access' policies and practices, but there has also been evidence of resistance from stakeholders such as publishers who do not want published material made available freely online and researchers who do not want deposit their data immediately to share it with others. On closer inspection then, the openness of e-Research, widely espoused as an aim, needs to be disaggregated and its potential and implementation in practice evaluated in a way that goes beyond the general principle.

To do this, we can take, first, the most macro-level of e-Research, the 'infrastructures' that e-Research is in the process of developing. On this level, an attempt is being made to create shared sets of tools and resources for researchers in and across certain disciplines that are akin to the social infrastructures (transport, communication) that have been created to support social development. This trend is particularly clear in the US, where the term cyberinfrastructure is widely used to describe developments in e-Research. Here the support of science policy makers and funding bodies has been critical, and much of the effort of national and international e-Research programmes has been devoted to infrastructure development. The efforts to make these infrastructures 'open', however, must be put into an even larger context of current conflicts over intellectual property regimes, and the forces that are aligned to change them towards greater openness - or to maintain the existing regimes of restrictions and access rights. These include commercial interests, NGOs, and policy makers, and it remains to be seen how the current interplay of forces will shift the momentum of infrastructural systems in one or other direction [22].

A second level are the policies and practices of individual e-Research projects: whether the materials relating to projects' ongoing research are deposited in a common or openly shared repository; how software tools are licensed (various options of open source licensing); and whether data or publications resulting from the project are made available on project webpages or other shared repositories. A major push, especially from research funding bodies, has been to encourage 'open' practices. For example, an increasing number of funding bodies stipulate that funded projects must make their results freely available online (perhaps after an embargo period with exclusive rights to publish reserved for the contributing researcher). What we found in practice, however, according to in-depth interviews with 12 researchers on UK e-Science projects is that practices vary considerably across projects [11]. Many projects are uncertain about what the policies about copyright and licensing are; they often adopt *ad hoc* policies with regard to which project members have access to ongoing materials created by the project; and finally they are in the early stages of and still unsure about depositing outputs such as data to a common repository or publishing results online (as opposed to via the traditional mechanism of journals).

A third example are the shared 'middleware' software tools, tools which mediate between the shared computing infrastructure and the individual projects or end-user researchers. On this meso-level, we can take case of the Open Middleware Infrastructure Institute (OMII) in the UK as an example. OMII aims to create a suite of widely adopted standard tools which are held and maintained in a repository and made accessible to all researchers. OMII has been promoted by the UK e-Science programme and has also involved the collaboration of commercial software vendors such as Microsoft and IBM. However, despite a push towards open source licensing and standards, the practice has been more complex [9]. It has proved difficult, for example, to achieve standardization, and the efforts to transform the software from the bespoke form that it has within a project into a product that can be shared by many researchers has often proved to be greater than creating the software in the first place.

In short, there have been major efforts to promote the openness of e-Research, and policy debates and debates among researchers [5, 25] continue about the issues that have been described here. It is too early to predict the outcome of a process that is still under way, but by disaggregating 'openness' into a number of levels and component parts, gathering materials on the various levels, and identifying the tensions and forces at work, OeSS will able to analyze this issue, locate it among the other key themes of the projects (there are clearly connections between openness and, for example, data sharing), and assess the implications for scientific practices.

Open science also has the potential to transform research practices on multiple levels. At one level, having research data openly available enables secondary use which potentially allows less advantaged researchers to participate in scientific discovery even if they don't have access to the resources necessary to gather high-quality datasets. Of course, there are other limitations beyond simply access to data when trying to understand why certain scholars have greater scholarly output [18], but lowering the barrier will potentially increase participation. At another level, open data also exposes research data to the same sort of peer review that scholarly publications now undergo. Currently, there is little provision in the practice of science for datasets to undergo peer review beyond a short methodological description in a publication. If data becomes more open and subject to peer review, the quality of the data being used to create scientific knowledge will potentially be transformed for the better.

2.3.2. Constraints on e-Research

A number of the themes and issues identified by the OeSS project pertain to constraints limiting the uptake of e-Research among scholars. **Privacy**, for instance, is a serious concern to all researchers collecting data on human subjects. While quantitative research data is often relatively easy to anonymize and many quantitative researchers routinely create de-identified or limited data sets, qualitative researchers have not generally had to deal with this issue in the past. Since few qualitative data sets were shared outside the original research team, there was little reason to de-identify the data. The challenge in de-identifying qualitative data is non-trivial compared to the relative ease of simply excluding several columns of quantitative data. Video recordings include images of participants, audio recordings can be identified by the sound of a person's voice, and transcribed interview data often includes references to other people by name and other identifying information. It has been suggested that improving the quality of qualitative data is an important goal for social researchers, and that transparency about the data and the research process are key elements to doing so [6]. The question of how to share such rich data while continuing to protect the identity of research subjects, however, is a major barrier to collaboratively sharing qualitative research data. Whether the promise of gains offered by collaboratively sharing qualitative research data is great enough to overcome these constraints remains to be seen; if it is, it would potentially mark a major transformation in the generally solitary and opaque manner in which much qualitative data is collected.

This also relates to another key theme being developed by OeSS researchers Carusi and Jirotka: **trust**. Trust can take many forms, and includes different sorts of trust, such as the trust people place in artefacts, and the trust people have for those who make use of artefacts [8]. Trust is central to human relationships in general and to scientific collaboration in particular; without trust, people are unlikely to share and collaborate. This is an issue for individual researchers and research participants; on a wider level it is also a question about people's trust in government and researchers generally. Even in Sweden for example, which is generally a 'high trust' country in this respect, to continue to maintain this trust and extend it in the face of new e-Research possibilities that are being developed, takes much effort [4]. While it is therefore possible to build information systems that inspire trust because of high levels of perceived security, that trust can be easily destroyed by the most minor of breaches. Trust between individual scientists is equally hard to maintain since it relies mainly on interpersonal characteristics. The challenge for e-Research, then, is to understand which elements of the research process are most reliant on trust, and to ensure that trust is not undermined by the technologies and processes expected of contributors.

Another concern that has been expressed by e-Research projects pertains to the **economic constraints** academic researchers face when compared to industry players. NCeSS nodes are generally small demonstrator projects with little or no support for the eventual development of mature research tools with the extensive user support that would be required if usage were to become widespread. There is a feeling among some in the e-Research community that the most likely outcome of many of these academic demonstrators will be that industry players will see that some of them are viable approaches to a problem, and the industry developers will then create a package which will trickle back down to the research community in several years' time [12]. There is already some evidence for this: recent speculation in *Wired Magazine*'s blog reports rumours that Google is developing a project (code named Palimpset) to host open-source science data in a tool that will include commenting and annotating features [16]. Google's resources are far more extensive than those of academic projects; whether this project develops into something more substantial remains to be seen, but it could become a major player in academic research if the success of Google Scholar is any indication. Should well-funded industry players decide to move into the e-Research arena in a serious way, the entire landscape of e-Research may be transformed.

While many other constraints have limited the early uptake of e-Research, these few examples should give an idea of the types of issues we have identified and the potential for transformation in science if these constraints can be overcome. We next turn to several enablers for e-Research.

2.3.3. Enablers for e-Research

Funding shifts

As indicated in the introduction to this paper, shifts in funding in the U.K., the U.S. and elsewhere are major drivers behind the push to e-Research. The shifts at NSF towards transformative research and the growth in the cyberinfrastructure funding program attract new proposals. Likewise, programs in the U.K. such as NCeSS encourage researchers to focus their attention on developing tools and methods that enable e-Research. If these and other programs in the U.S., the U.K., and elsewhere continue to push new funds to e-Research development, the predictions of the funders that e-Research will become a central focus of science becomes a self-fulfilling prophecy, and the likelihood of transformations in the practice of science become much more likely.

New scientific questions

The ability to potentially answer previously unanswerable scientific questions may be the most important driver behind the push to e-Research. This is particularly true in certain scientific fields where access to massive datasets may be the only way to answer particular scientific questions. In biomedical genetics research, for instance, very few of the disorders for which scientists are trying to identify a genetic basis are triggered by a single gene waiting to be discovered. Instead, many disorders are suspected to be triggered by complex multiple gene interactions, and may also be influenced by interactions between genes and an individual's environment. As such, many genetic studies discuss measuring genetic risk for developing a particular disorder, rather than discovering a gene responsible for the disorder. Because this is much more complex than identifying a gene in the population, genetics researchers are finding that they must rely on ever-larger datasets if they are to have any hope of discovering the genetic alleles that place subjects at risk for disease. One way the scientists are doing this by contributing their collections of subject data and DNA to larger collections, such as the GAIN example described above. By increasing the number of DNA samples available at the same time as increasing the amount of data available from each sample, genetics researchers hope to find better indicators of genetic risk in the populations they study.

Another type of research that can only really be done collaboratively is astronomy. Astronomers have long shared resources such as telescopes and sky data. There are currently several international projects designed to create virtual observatories in the U.K. (AstroGrid), the U.S. (NASA's National VO), Europe (Euro-VO), and elsewhere. These projects have been built to enable astronomers to work with data that is stored distributed across the Grid and analysed using parallel processing techniques. The AstroGrid project, for instance, can find, process, and analyse data from a given patch of sky that has been collected from any instrument to which the project has access to the data on the grid. Only the small portions of the data that are the result of calculations need to be transferred back the astronomer, rather than requiring him or her to download a large dataset and work with the data natively. As these tools become more widely used, astronomers are able to spend less time working with data and more time doing analytic science.

A final example of how these collaborative resources can transform practice in a scientific field comes from the field of oceanography. Lamb [15] has described how oceanographers are now able to use remote sensing to transform the practice of event-driven science. The traditional means of collecting oceanographic data was to send ships out to sea and wait for the types of events the scientists needed data on to occur. In recent years, however, oceanographers have been building extensive networks of remote sensing devices located throughout the oceans. Now, the scientists can monitor the readings from their offices. Also, if an interesting event begins about which they would like to collect additional information on-site, they can be notified of this by the remote sensors and then dispatch a ship to sail to the location of interest. This represents an enormous savings in time and money for the oceanographers, and also allows them to collect much more data than before.

If the necessity of harnessing massive resources is a major enabler for e-Research, this may offer a clue to the relatively slow uptake of the collaborative and Grid-based elements of e-Research among social scientists. Social scientists have traditionally been trained as primarily autonomous researchers, working individually or on small teams. The datasets generated are often small and easily managed and analyzed using desktop computing applications. Unlike running complex astronomical calculations, even the largest social surveys rarely tax the processing capabilities of SPSS or SAS on the desktop. For many social scientists, the Grid doesn't appear to offer compellingly attractive applications at this time. Even for social scientists using quantitative statistics or large-scale Social Network Analysis (SNA), their computational requirements do not generally approach those of scientists modelling complex biological, physical or astronomical systems. That may change if some sort of "killer app" is developed, but even so, there will be disciplinary resistance to a change in the normal practices of science.

2.4. *Future work*

Future work on the OeSS project will move beyond monitoring how various e-Research projects negotiate the constraints with which they are faced -- to include identifying more of the ways in which e-Research realizes the promise of transformation for research, and with what consequences. This will be done through additional case studies, and through additional expansion to a consideration of more international cases, including examination of how national research policies influence the uptake of e-Research techniques.

3. DISCUSSION

While many transformations have taken place in early e-Research projects, a more wide-spread transformation of research practice has remained elusive. The projects have undergone transformation, some notions about the possibility of e-Research have been transformed, and certain scientific fields have seen significant transformations in the way that they work and the types of scientific questions they are able to ask. For many scientists, social scientists and other academic researchers, e-Research is still not central to their work.

One of the challenges in understanding the nature of how e-Research is developing is that many of the projects developing e-Research tools have, until fairly recently, been primarily developing tools that had not yet been released or tested widely in real-world settings. As the efforts have matured, however, it has become feasible to study these projects and to begin developing an understanding of the social shaping that has occurred, and the social, organizational, ethical and legal issues that are associated with the projects. For each of the topics discussed in this paper, the examples are often of projects that have advanced far enough beyond their early stages to have created tools that are being used by scientists. In this way, we are able to examine transformations in tools, projects, scientific practice, and the formulation of scientific questions. In some areas, early results indicate that many transformations have taken place. For instance, at the micro-level, many of the projects have experienced significant transformation from their earliest roots as the initial ideas were shaped by experience. As Schroeder & Fry have observed, “the obstacles to the effectiveness of e-science are not so much technical, as social” [24]. These social barriers are part of the feedback loops that operate as the projects transform over time.

At other levels, however, there is currently less evidence that e-Research tools and e-Infrastructures have yet transformed either meso-level practices of ordinary science or macro-level questions that science is able to imagine and study. One question that remains is whether this lack of fundamental transformation is a temporary problem which will disappear as tools become better developed and more widely used, or whether there are larger issues that make it unlikely that science will be transformed so much as reinforced by e-Research tools. If the former is true, then it becomes just a matter of time and effort to reach the critical mass required for transformation. And if mass adoption of these tools seems an ambitious target, it can be noted that research technologies have, in the past, been drivers for scientific changes because the tools have migrated across different applications and disciplinary domains [23]. If the latter is true, however, then one needs to examine the underlying forces that operate to resist changes in scientific practice and instead co-opt new tools to the service of existing patterns. Indeed, it is very difficult to predict in advance whether new scientific tools are likelier to support normal science or to enable paradigm shifts, to use the language of Kuhn [14].

We are very early in the story of e-Research. Some elements such as electronically accessible journals are well developed, while others such as Grid-enabled applications for social scientists are in their infancy. The extent to which any of these will prove to be transformative remains to be seen, but we would predict that most will instead tend to reinforce existing patterns of scholarship, particularly in the short term. We will continue to research these issues, however, and eagerly await new developments in e-Research as the projects enabling e-Research mature.

4. ACKNOWLEDGMENTS

This research has been supported by NCeSS and the ESRC through grant RES-149-25-1022. The authors wish to thank NCeSS and ESRC for their continuing support, and also wish to acknowledge the contributions of all the people who have worked on the OeSS project, including Annamaria Carusi, Paul David, Matthijs den Besten, Jenny Fry, Marina Jirotka, and other colleagues at the University of Oxford and elsewhere.

5. REFERENCES

[1] Copyright Law of the United States of America. 1976. 17 USC 105.

[2] Copyright, Designs and Patents Act (Great Britain). 1988. Chapter 48 (X) (163).

[3] Allen, R. 2004. NCeSS: the National Centre for e-Social Science. Available online: http://www.grids.ac.uk/twiki/bin/view/ELearning/NCeSS.

[4] Axelsson, A.-S. and Schroeder, R. 2007. Making it Open and Keeping it Safe: e-Enabled Datasharing in Sweden and Related Issues. In Proceedings of the Third International Conference on e-Social Science, (Ann Arbor, Michigan, USA, October 7-9, 2007).

[5] Borgman, C. 2007. Scholarship in the Digital Age. MIT Press, Cambridge, MA.

[6] Brower, R.S. 2000. On improving qualitative methods in public administration research. Admin. Soc. 32 (4), 363-397.

[7] Carusi, A. and Jirotka, M. 2007. From data archive to ethical labrynth. In Proceedings of the 3rd Annual International e-Social Science Conference, (Ann Arbor, Michigan, USA, October 7-9, 2007).

[8] Carusi, A., Jirotka, M. and Parker, M. 2006. Trust and Ethics in e-Science Agenda Setting Workshop Report, Oxford e-Research Centre, Oxford. Available online: http://www.oii.ox.ac.uk/events/details.cfm?id=22.

[9] David, P., den Besten, M. and Schroeder, R. 2006. How Open is e-Science? In Proceedings of the 2nd IEEE International Conference on e-Science and Grid Computing, (Amsterdam, The Netherlands, December 4-6, 2006).

[10] Dutton, W.H. 2007. Reconfiguring Access: A Framework. In Proceedings of the World Wide Science Conference, (Oxford, UK, May 31-June 1, 2007).

[11] Fry, J., den Besten, M. and Schroeder, R. n.d. Open Science in e-Science: Contingency or Policy? J. Doc.

[12] Halfpenny, P. Comments at NCeSS Showcase, Manchester, 2008.
[13] Hudson-Smith, A. 2007. Digital Urban Blog. Available online: http://digitalurban.blogspot.com/2007/08/ordnance-survey-and-google-statements.html.
[14] Kuhn, T. 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago.
[15] Lamb, R. Keynote speech at IFIP-TC9 World Conference on Human Choice and Computers, Maribor, Slovenia, 2006.
[16] Madrigal, A. 2008. Wired Science Blog: Google to Host Terabytes of Open-Source Science Data. Available online: http://blog.wired.com/wiredscience/2008/01/google-to-provi.html.
[17] Meyer, E.T. 2008, Forthcoming. Moving from small science to big science: Social and organizational impediments to large scale data sharing. In Jankowski, N. ed. E-Research: Transformation in Scholarly Practice, Routledge, New York.
[18] Meyer, E.T. and Kling, R. 2002. Leveling the playing field, or expanding the bleachers? Socio-Technical Interaction Networks and arXiv.org (Center for Social Informatics Working Paper Series WP-02-10) Center for Social Informatics Working Paper Series, Center for Social Informatics, Bloomington, IN. Available online: http://www.slis.indiana.edu/CSI/WP/WP02-10B.html.
[19] National Centre for e-Social Science. 2007. NCeSS webpage, Manchester, UK. Available online: http://www.ncess.ac.uk/.
[20] National Science Foundation. 2007. Proposal and Award Policies and Procedures Guide NSF 08-1, January 2008 (Document number nsf081), NSF.gov, Arlington, Virgina. Available online: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp.
[21] Ordnance Survey. 2008. Overview. Available online: http://www.ordnancesurvey.co.uk/oswebsite/aboutus/yourinforights/.
[22] Schroeder, R. 2007. e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production? Prometheus 25 (1), 1-17.
[23] Schroeder, R. 2008. e-Sciences as Research Technologies: Reconfiguring Disciplines, Globalizing Knowledge. Soc. Sc. Inform. 47 (2).
[24] Schroeder, R. and Fry, J. 2007. Social Science Approaches to e-Science: Framing an Agenda. J. Comput.-Mediat. Comm. 12 (2), Article 11.
[25] Willinsky, J. 2005. The Access Principle. MIT Press, Cambridge, MA.