The Future of e-Research Infrastructures

Kathryn Eccles[1], Ralph Schroeder[1], Eric T. Meyer[1], Zack Kertcher[2], Franz Barjak[3], Tobias Huesing[4], Simon Robinson[4]

[1] Oxford Internet Institute, University of Oxford, UK
[2] National Opinion Research Center, University of Chicago, US
[3] University of Applied Sciences, Northwestern Switzerland
[4] Empirica, Bonn, Germany

Email address of corresponding author: kathryn.eccles@oii.ox.ac.uk

Abstract. In this paper, we present selected results of a systematic study of different types of e-Research infrastructures. The paper is based on ongoing research to compare a range of e-Infrastructures of broad diversity focusing on: geographical diversity, representing efforts from around the globe; disciplinary diversity, including the natural sciences, social sciences and humanities; organizational diversity, for example, multi-institutional or federated; diverse levels of maturity, from those in the planning stage to those with a well-established user base; and diverse types of target user communities such as specialized niche, discipline-wide, or generic infrastructures. In presenting six initial cases, we discuss some general features that distinguish between different types of infrastructures across different fields of research. Previous analyses of e-Infrastructures have focused on the parallels between these infrastructures and the major infrastructures in society that support national populations. What our cases highlight instead is that e-Infrastructures consist of multiple types of overlapping and intersecting socio-technical configurations that serve quite diverse needs and groups of users. Indeed, the very term 'infrastructures' may be misleading insofar as it connotes support of whole communities of researchers on a large scale, which is currently still premature. The paper derives implications of this heterogeneity for the future outlook on e-Infrastructures.

Introduction

The development of infrastructures for e-Research has reached a point inviting reflection. e-Research infrastructures are beginning to show signs of fulfilling their early promise and are no longer in their infancy, but they are also not yet fully mature. Sizable investments have been made in Europe, the US and elsewhere, and still more investment is being allocated and planned for the future. Yet despite a number of policy and planning studies which point to future options, so far little attention has been paid to developing a systematic understanding of the different types of technological and organizational outcomes in the transition to e-Research infrastructures. Furthermore, although several studies have recently examined a single e-Research infrastructure (Olson, Zimmerman and Bos 2008), much less studied but of considerable empirical and conceptual significance is research that identifies some defining characteristics that distinguish between different types of infrastructures and across different fields of research. In this paper, we present selected results of such a systematic approach and present preliminary data from selected e-Infrastructure case studies and identify some characteristics that can help us to make sense of their trajectories.

e-Research can be defined as ‘the use of networked, distributed and shared digital tools and data for the production of knowledge’ (Schroeder 2008). In the case of e-Research infrastructures (also called e-Infrastructures, or Cyberinfrastructure), we can modify this slightly to ‘networked, distributed and shared digital tools and data that *support* communities of researchers engaged in the production of knowledge,’ since infrastructures are often not aimed at the production of knowledge itself, but at *supporting* the production of knowledge. For example, an e-Infrastructure could provide a generic database and associated tools, whereas e-Research would involve *using* the tools to analyze these data and produce novel results. Since the terms e-Infrastructure and e-Research are nevertheless often mixed in practice, in our research we compare cases from both types of organizations and also reflect, in the conclusion, on the aptness of the term ‘infrastructure’.

Case studies

The following six studies fall into several disciplinary categories: Swiss BioGrid (Biological/Medical Sciences), Swedish National Data Service (Social Sciences, Medical Sciences, and Humanities), DRIVER (e-Infrastructure, ie. generic provision across disciplines), CineGrid (e-Infrastructure/Social Sciences and Humanities), National Virtual Observatory (Physical Sciences and Engineering), and Open Science Grid (e-Infrastructure; again, provision across disciplines). These six case studies, from which we present preliminary data, are drawn from among a larger set (approximately twenty) which are part of an ongoing research project to investigate types of e-Infrastructure, their organizational structures, modes of collaboration and technological development and how they prove most effective in supporting virtual research organizations in different fields. The analysis is based on several interviews conducted in each case, as well as documentary research. The project will yield a roadmap of strategies that will guide public policies and enhance the uptake and use of e-Infrastructures in research.

Swiss BioGrid

Swiss BioGrid (SBG) is an initiative that was launched in 2004 to assess whether Grid computing technologies could be successfully deployed within the life science research community in Switzerland. Several years on, SBG has successfully completed two pilot projects - one for proteomics data analysis and the other for high-throughput molecular docking to find new drugs for neglected diseases. One of the lessons learned in SBG was that a small project-type effort is hard to sustain outside the context of such a larger infrastructural effort (den Besten, Thomas and Schroeder 2009). SBG was composed of six distinct academic groups; The Swiss National Supercomputing Centre (CSCS); Biozentrum, University of Basel, Switzerland; Functional Genomics Centre, Zurich; Friedrich Miescher Institute for Biomedical Research; Novartis, a pharmaceutical company, and the Swiss Institute of Bioinformatics -Vital IT Lab. The project was led by a co-ordinator with the help of a steering committee, in addition to the infrastructure and science personnel.

Unlike other projects which are initiated top-down from national e-science initiatives, demand for SBG was driven by the practitioners themselves. SBG started with a group of researchers at the University of Basel to work on a project called ‘Grid-based Virtual Screening of Dengue Virus Target Proteins’, which employed grid computing to deliver an evaluated list of compounds likely to bind to dengue virus proteins identified by in silico docking. Several academic institutions agreed to free up spare computing resources to enable the grid computing effort. A secondary project, related to proteomics, based at the Swiss Institute for Bioinformatics (SIB) in Geneva aimed at the development of an automated analysis system to identify proteins using data from the virtual screening process. In scientific terms, the virtual screening project has identified c.100 potential drug candidates, a number of which will be put into experimental validation by Novartis. A new drug for Dengue would have a dramatic effect. The proteomics project showed that this infrastructure worked for the specific project that was developed, revealing the potential for similar projects to attempt to employ this kind of solution.

Though primarily aimed at the academic researchers and pharmaceutical industry personnel involved in the project itself, the project was essentially a demonstrator to show what an extensible larger-scale effort might achieve in the future. SBG was wrapped up at the end of 2007, although the science projects are ongoing and continue to use the infrastructure that was set up under SBG. One of the major contributions of this project was to prove that it was possible to build grids organically, and that these did not require top-down governance or dictation of what technology was going to be used. SBG showed that it was possible to build such a project by consensus. Indeed, the main challenge that the project had to overcome (which was successful, but time-consuming) was to put the legal and organizational framework in place for the project to go forward across the boundary of commercial and academic/research institution partners. SBG also had an important impact on national engagement with grid technology. It was one of the factors which led, in 2008, to a new Grid initiative, SwiNG, being set up to coordinate Swiss efforts and represent Switzerland in international Grid efforts. Several of the partners in SBG are now active members of this new structure.

SBG also illustrates a wider problem in the life sciences, which is that rather than becoming integrated around a shared computational infrastructure, e-Science initiatives have resulted in the promulgation of countless heterogeneous resources and efforts (Wooley and Lin 2005). While much of the development of the Grid has been geared towards applications in particle physics, which tend to be centralized and fairly homogeneous, the more heterogeneous requirements of computational biology have been poorly supported by existing Grid solutions, and it is unlikely that this picture will change unless there is a concerted effort in adapting Grid tools and putting them on a permanent footing. This point anticipates our conclusion about this type of project: a small project that arises from a bottom-up effort among researchers can respond to a real need to transform research into a more collaborative way of doing of science that provides advantages over existing ways of working. However, the pay-off is limited unless this kind of limited-duration small-scale effort can be embedded within a larger-scale, long-term institutionalized effort that is made available to a wider user base, one which, moreover, transcends the barriers of existing research practices in the heterogeneous life sciences.

Swedish National Data Service

The Swedish National Data Service (SND) is the national academic data service for Social Sciences, Humanities and parts of Medicine. They are a service organization for all Swedish universities and colleges, whose purpose is to collect, document and disseminate data within the designated areas. SND builds on previous institutions and it is mainly financed by the Swedish Research Council through the Council's Database Infrastructure Committee (DISC). It is also coordinating with European efforts (such as the Council of European Social Science Data Archives (CESSDA). An important purpose of SND is to further international collaboration and the exchange of data by enhancing the research infrastructure. SND also offers professional advice on matters of documenting and archiving data based materials, and is currently expanding to become a much more extensive e-Infrastructure for Sweden and beyond. Approximately 8 full-time staff work at the SND office, but there is a larger network of persons at the research funding councils, Gothenburg University, and associated researchers working on the databases in which SND, which can be seen a data service, is embedded.

SND provides access to uniquely good Swedish datasets and has successfully maintained an unusually high level of trust about the use of sensitive data in Sweden, despite a number of incidents that have challenged this trust and provoked public debate (Axelsson and Schroeder 2009). Sweden is also in a unique position to capitalize on its system of personal identifiers, long-established and comprehensive databases, and trust between researchers and the population – for the coming generation of shared databases in medical, social science and other forms of research. Sweden may provide a model of how this kind of data sharing, especially of sensitive micro-data, can take place, although its unique conditions are unlikely to be found elsewhere.

The Swedish case of an infrastructure for social science, humanities and medicine illustrates the transformation of a well-established resource into the new territory of e-Research. Although the Swedish Data service has had a sizable base of users going back to the 1980s, its recent e-Research incarnation still needs to solve the socio-technical challenges of secure remote access to databases containing sensitive micro-data. The e-Research user base which would make use of this sensitive micro-data remotely and with links between different types of data - again, potentially the most unique resource that distinguishes Sweden from other e-Research efforts - is still largely non-existent, even though ambitious plans are under development. Although well-resourced for the foreseeable short-term future, there are also questions about this service will fare on a longer-term basis.

Digital Repository Infrastructure Vision for European Research

Now in its second phase, the DRIVER project aims to build a Europe and world-wide digital repository infrastructure in order to link users (both academic and public) to 'any form of scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data' (http://www.driver-repository.eu). Building upon the successful worldwide GÉANT network, the first phase of the project, DRIVER, established a network of relevant experts and Open Access repositories. DRIVER-II consolidates these efforts and aims to transform the initial testbed into a fully functional, state-of-the art service, extending the network to a larger confederation of repositories.

DRIVER has 13 partners, of which 10 are universities or university libraries and the rest national representatives of repositories. The University of Athens (NKUA) acts as the project coordinator, maintains the services provided by DRIVER-II and is responsible for scientific, technological and management support, and provides support for enhanced publications and support and training to users. The Consiglio Nazionale Delle Ricerche in Italy is the scientific and technological coordinator of the project, leading a team of technical partners who are dealing with software development.

The motivation behind this project was to bring together scattered scientific information in one place and make it accessible to a wider audience. There are various scientific repositories in Europe in universities, research institutions, and national organisations. At the time DRIVER started, there were scattered national initiatives in several countries (e.g. Netherlands, Germany, UK) and DRIVER set out to bring all this information together and target end users. 225 repository managers have submitted their data and there are currently 10,000 unique hits per month. The project's primary aim was to integrate the data that repositories have across Europe. A secondary aim is to produce a portal that enables scientists to access all kinds of Open Access publications, research materials (non-textual information) and 'enhanced publications', which 'combine interrelated information objects into a logical whole, e.g., publications coupled with relevant presentations and associated datasets.' The DRIVER software is running and can be used to set up similar portals by all kinds of institutions, also to develop new applications on top of the basic services. There is strong community uptake and commitment. Libraries generally have an interest in taking and running the DRIVER service and repositories are willing to conform to the framework.

There are challenges in developing the aims of DRIVER-II and building upon the successes of the first phase of the DRIVER project: open access needs further advocacy; although linking things has been a strong driving force in development in recent years, data and publications are still separated more often than not. Mandates are essential: if research funding hinges on making results available through open access, this will push the idea forward. This promotional effort can also be put into a larger context. The open access 'movement' is a contested movement globally and open publications and repositories play a key role in this contest, the outcome of which is still unclear even if a clear-cut victory for one side or the other seems unlikely (Schroeder 2007). The second context is that there is a plethora of efforts to provide repositories of various types at different levels, from small research groups to universities to national and global levels. If a successful EU infrastructure can be established via DRIVER-II, it remains to be seen how prominent it will be among the various open access initiatives.

CineGrid

CineGrid is a non-profit organisation dedicated to exploring and promoting research, and to the development and deployment of new distributed applications of ultra-high performance digital media (sound and picture) over advanced photonic networks. CineGrid organizes network test beds to host experimental digital media projects and organizes workshops and demonstrations to share results and identify new research areas. CineGrid summarizes its mission as being: 'To build an interdisciplinary community that is focused on the research, development, and demonstration of networked collaborative tools to enable the production, use, and exchange of very-high-quality digital media over photonic networks.' (http://www.cinegrid.org/). Based in the US, CineGrid has an international membership.

CineGrid has been described as a virtual organisation with very few formal requirements out of which dedicated teams are formed to realize joint projects. The focal point for this community, cinegrid.org, is incorporated as a non-profit international membership organisation in California. Still, CineGrid has a formal structure comprising a Board of Directors, an Executive Committee drawn from the Board of Directors and an advisory committee, drawn from the Executive Committee and invited industry leaders. The CineGrid Project Review Committee, drawn from the Executive Committee, evaluates submitted proposals and makes recommendations.

CineGrid currently has around 50 member organizations paying annual membership fees, with 80% of these organisations located in North America, around 10% in Asia (Japan and Korea) and Europe, and significant groups in the Netherlands and the Czech Republic (as of April 2009). CineGrid does not have any dedicated activities or campaigns for enlarging the community or involving potential end-users beyond the CineGrid network. Individually, CineGrid members engage in outreach activities, presenting projects at workshops and events, making demonstrations and performing. It primarily acts as liaison between the members, funding organizations and related public and private organisations. CineGrid members engage in projects to demonstrate the feasibility of digital media applications running over next generation optical networks.

CineGrid has brought together people from different communities and supported their collaboration in joint projects; the growth and maturation of the CineGrid community is considered one of the biggest accomplishments over the past few years. It has created an environment of trust and mutual understanding in which the sharing and joint use of resources is common practice and members help each other with demonstrations and experiments. In addition, CineGrid has raised awareness about new audio-/video-, production and post-production, transmission and display technologies and work-flows among artists, filmmakers and other media professionals; along the same lines, it has raised awareness of using visualization technologies among scientists. With the CineGrid Exchange, CineGrid has started to make media content available for its members thus supporting experimentation, research and development. The Exchange is still considered to be work in progress.

Funding is seen as the biggest challenge to the continuation of the project. CineGrid has had significant problems raising funds for its projects, perhaps because the projects seem less important than those in other domains from a science-funding perspective. A larger funding basis would lead to full-time staff and more communication opportunities among the CineGrid members. Apart from the issue of continued funding, CineGrid is an interesting example of an e-Infrastructure that sits at the intersection of research (high-end distributed visualizations and performances) and the commercial world of distributed video editing and content access. It is also not creating an e-Infrastructure per se (with the possible exception of CineGrid Exchange), but is rather developing a special interest group spanning the disciplines involved in advanced digital media networks. Thus it could be described as a virtual organization which is forming links to push distributed collaboration tools and resources – with a question mark over whether this will gel into sharing these tools and resources with a wider community.

National Virtual Observatory

The NVO is a joint effort of computer scientists and astronomers from 17 US-based institutions to develop an e-Infrastructure for astronomers to identify, retrieve and analyze multiple types of large-volumes of celestial data from disconnected astronomic ground and sky instruments. Working to develop standards and protocols for astronomical data, NVO joins similar efforts in 15 other countries from North America, Europe and Asia, as a part of the International Virtual Observatory Alliance (IVOA). The NVO builds upon a long history of data cataloguing, archiving and sharing in astronomy.

NVO's mandate is to develop and provide the e-Infrastructure and the interface capable to support the integration, or federation, of various astronomical digital data sources from diverse instruments. A major component of this worldwide virtual telescope is to enable efficient processing and visualization of these massive amounts of data. More broadly, the project is spearheading a fundamental change in astronomy. By providing transparent access to supercomputers, high speed networks and federated large volumes of data from different sources—in addition to programs and algorithms to help make sense of this data more efficiently—the NVO is bringing together astronomers from the various wavelength specialties. Visionaries claim this new paradigm will eradicate the artificial barriers that currently separate the different sub-fields of astronomy (Williams 2003).

The NVO involves collaborating participants from 17 US-based institutions, including astronomy data centres, national observatories, supercomputer centres, university departments and computer science specialists. For NVO, the primary users are considered to be research astronomers who want access to comprehensive data sets across wavelength regimes. In the field of astronomy people have tended to brand themselves by wavelength, but lately this has become less clear-cut, and the new research that is being done requires a synthesis of data. In addition, NVO wants to provide access to data for education and public outreach.

Prototypical astronomical research relies on observations of small, carefully selected object samples, within a specific wavelength band. The NVO has advanced an approach of integrating high-quality, homogenous, multi-wavelength data on millions of objects from different observational sources. One way to consider this alternative—and possibly its main contribution— is as supporting an alternative, increasingly popular paradigm for the field of astronomy. Additional discoveries are expected, as multi-wavelength data is necessary for the study of important questions in observational astronomy, such as active galactic nuclei. Further, use of NVO data and tools can allow researchers to check or replicate results more easily, again through the ability to access the necessary data and tools.

NVO makes a practical contribution to astronomers from institutions not directly connected to observatories, and from teaching-oriented institutions, enabling them to continue to engage in research. These individuals are able to download NVO data and could even work on research using a laptop at home. This aspect of the NVO is likely to benefit astronomers in other, less developed countries as they would easily gain access to research resources. Although this ability to access data from many different sources is bound up with the development of the internet, and similar results could be found with something like a web search, the NVO has contributed a clear and cohesive search function of astronomical data in one portal, along with the tools to access and manipulate that data.

A related efficiency is that NVO allows easier collaboration across co-PIs on a project, who in the past may have had to set up a website or to use email to share data with other researchers. With NVO, PIs who have access to pre-release (not yet public) data are able to share this data with one another by providing the private archive key to the data via the NVO. The common standards set by NVO have also created greater efficiency for astronomers, particularly through the sharing of tools and techniques used to translate data coordinates between different coordinate systems. While the NVO—and the IVOA—have made significant advancements in the fundamental computational framework, most astronomers have not yet adopted this technology. In subsequent rounds of funding, NVO leaders aim to push research efforts to focus more closely on considering user practices and soliciting user adoption.

NVO is thus an infrastructure shared within a relatively homogeneous, small and well-bounded disciplinary community. Unlike in other disciplines where there is a major impetus towards data sharing, in the case of NVO (and IVOA) astronomers have come some way in addressing the problem of common standards. Furthermore, there is no problem with the sensitivity of the data (unlike, say, in the case of SND) or its potential commercial value (unlike in SBG). However, this effort will only continue to succeed if the interest of developers can be sustained in providing a shared and useful resource, both nationally within the US and internationally. Interestingly, unlike the cases discussed so far (with the possible exception of DRIVER-II), there is considerable interest from the amateur community and the general public, and this kind of interest can have a role in providing support for funding.

Open Science Grid

Extending early data grid research projects, the Open Science Grid (OSG) is a large, US-based consortium of dozens of academic organizations and national laboratories. While not owning any of the resources, the consortium operates, supports and enables more effective sharing and utilization of available compute cycles and facilitates the use of distributed storage and software through its ‘opportunistic computing’ model. The OSG maintains close ties with other e-Infrastructure providers, developers and resource communities. Like its European counterpart, the Enabling Grids for E-SciencE (EGEE) project, OSG is a central component of the global computational fabric of the Large Hadron Collider. Beyond high-energy physicists, the OSG also caters to researchers from other disciplines, such as astrophysics, climatology, structural biology, nanotechnology and economics.

Uniquely joining three organizational layers, the OSG is a novel type of network organization (Powell 1990): it is a consortium of existing organisations mostly based throughout the US; it is comprised of multiple sites that provide technological capacity; and it is an organizational container and a technological platform upon which various virtual organizations operate and develop. OSG is organized as a consortium that presently consists of 53 academic and research institutions, mostly from the US. The main stakeholders are HEP experiments, major grid technology development projects and national laboratories. Members of the consortium include sites which contribute computational, data and storage services. Sites trust and authorize virtual organizations (VOs). To date, the consortium has served about 2,500 unique users who have moved data or have run computations on the OSG infrastructure. Users are a part of diverse fields of science, such as theoretical physics, industrial engineering, computer science and natural language processing, chemistry, biochemistry, computational biology, genetics, structural biology and economics.

OSG’s aim has been ‘to make collaborative scientific research more effective and widespread, stimulate new and transformational approaches to computationally based scientific discovery, and build intellectual capital for future scientific research relying on distributed cyber-infrastructures.’ Participants in this effort describe their goal as promoting and enabling—by partnerships and interoperation—a truly global grid, a ‘grid of grids.’ Facilitating the vision of e-Infrastructure, members of the OSG envisage such a comprehensive e-Infrastructure fabric to transform the practice of collaborative science, making it more effective and widespread. OSG has established a dedicated function called the ‘engagement team’ that implicitly operates in a unique three-step procedure: user identification, engagement and contagion, the latter being an approach that explicitly emphasizes social networks as conduits of information contagion and as a basis for innovation diffusion within a research community.

A mixture of technological and social challenges may threaten OSG’s sustainability. Firstly, OSG does not anticipate or actively encourage the commercialization of its core technology. In fact, most commercial vendors have recently moved away from grid technology to the development of an alternative technology—which many see as a disruptive, new paradigm to distributed IT: cloud computing. However, the OSG’s flexible economic model, which is based on generalized exchange of opportunistic cycles, does not lock users to a particular technological approach, and bears little direct costs to members.

Secondly, while the computational portion of the e-infrastructure is mature, security mechanisms are not yet at a stage that enables communities to handle secure data—such as human subject data—via remote access (just as in the SND case). Sensitive data are governed by a dense web of regulation and research practices, making their sharing beyond organizational boundaries a considerable challenge. This is a well documented problem of e-Infrastructure research communities (Jirotka et al. 2005; Barjak et al. 2009).

OSG participants have suggested that a grassroots approach is insufficient in the creation of global research communities; it should be supplemented with top-down requirements from funding agencies for collaboration among providers, as well as among research communities.

OSG, like EGEE, is thus a good illustration of infrastructure as a federation of efforts, in this case (again like EGEE) centred on shared high-performance computing. Like all the other projects that we have examined here, a key issue is sustained funding, in this case threatened on one side by new technological models that potentially supersede Grids (cloud computing and web services), and on the other by the need to deepen its socio-technical roots or remain a federation without a driving agenda in substantive research.

Conclusions

The role of research policy is to provide frameworks (including for governance) and guides for the effective support of – in this case e-Infrastructures – such that they enhance the role of research in society. This is a task that a number of bodies have embarked on (see Fry and Schroeder 2009 for an overview) and it should be possible to learn lessons from other infrastructures. In practice, we find that e-Infrastructures are quite diverse and face quite different issues, even before we take into consideration the entwining of policy agendas with different funding programme priorities, the momentum of existing technological and social infrastructures that pre-date e-Research, and how boundaries between research groups and organizations are being redrawn around emerging technological systems.

What we can see even from this small sample of six e-Infrastructures is the diversity of these socio-technical systems, both on the social side where we have top-down discipline-transcending federations (DRIVER-II, OSG, CineGrid) and national efforts (SND) as well as bottom-up efforts (SBG) (see Dutton and Meyer 2008 for the top-down/bottom-up distinction in e-Social Science) and discipline-specific efforts (NVO) – but also on the technical side whereby the main contributions are shared data repositories (NVO, SND, CineGrid), shared hard- and software to digitalize workflows in research (SBG, OSG) and beyond (CineGrid, DRIVER-II).

What is common to these efforts is that none has a hierarchical and centralized structure, and none is developing technology on behalf of a single group. Instead, they are creating longer-term collaborative socio-technical structures. Whether these should be labelled ‘infrastructures’ is questionable: none yet serve the whole research community and none have yet established a user-base that relies on this system. In terms of models of sustainability, some cases examined here are either already (SND) or potentially (SBG) stably embedded within larger established institutions. Others are projects without a future beyond the end of project funding (SBG – except insofar as it finds its way into SWInG). A third type are networks that federate the contributions of members, either on fee-paying basis (CineGrid) or based primarily on research funding institutions (OSG, NVO, DRIVER-II).

The user communities can also be differentiated: although it is envisioned, as per the definition of an infrastructure, that *all* members of a particular academic community will need to use an infrastructure, in practice, this is not so. It may be argued that this limitation is provisional – that ultimately all scholars will come to rely on these infrastructural tools, but this argument only needs to be stated clearly to see that it is misleading: only in certain areas of research will it be essential to use certain infrastructural tools. Thus we can divide the communities into those that consist primarily of early adopters (or are even still limited to developers) as against those which are close to maximizing the relation between potential and actual adopters. For our cases, none can be put into the latter category: SBG and OSG respectively have a small and larger constituency, DRIVER-II and CineGrid have many developer constituencies, and for NVO and SND it depends whether one draws the line around the users of novel uses of data (few) or more long-standing uses of shared data (many in both cases – and it is easiest to see in these two cases how the gap between potential and actual users may one day be closed due to the relative homogeneity of this user base).

The analogy with other infrastructures in history (see Hughes 1994; Jackson et. al. 2007 for e-Infrastructures) can only take us so far. This is partly due to the fact that the cases discussed here are not really infrastructures in the sense that they provide lasting, systemic and essential support to a large body of constituents. Rather, they support larger or smaller specialist groups of researchers. This point also highlights an overlooked feature of infrastructures: that they need to become standard technologies which extend to large user-bases, perhaps becoming monopolies. Traditional infrastructures *are* monopolies (if we think of transport, power and communication) even if attempts have been made to break them up or to allow multiple providers within a single umbrella, and the possibility to cope with multiple technological standards. It is also worth noting that there is a tension in monopolies as they apply to research: innovation is thought to rely on competition, but standards (for example, standard instruments) are sometimes also an essential precondition for advancing knowledge. A related tension is that markets are thought to be effective in driving down costs, but economies of scale and the elimination of duplicate infrastructure efforts should also enhance cost-effectiveness.

Against this background, we can envision what the e-Infrastructures in 2020 and beyond will look like: they will consist of multiple overlapping and intersecting networks rather than monolithic infrastructures; they will be a mixture of monopolies (within certain communities) and duplicate or parallel efforts, of essential tools for everyone within and across certain research communities and tools that are only occasionally used for specific purposes by a narrow subset within or across them, and of permanent but extensible large-scale systems that will provide indefinite and essential support for well-defined large groups of users, but also light-weight tools without any lasting structure and only ad hoc constituencies. This heterogeneity, rather than all-encompassing and overarching infrastructures, will mark the socio-technical systems that support e-Research in the future.

Acknowledgments

Part of the work on this paper (at the Oxford Internet Institute) was supported by ESRC grant RES-149-25-1022 for the Oxford e-Social Science (OeSS) project: Ethical, Legal and Institutional Dynamics of Grid-enabled e-Sciences. The paper is partly supported by the eResearch 2020 project http://www.eresearch2020.eu/ supported by the European Commission.

References

Axelsson, A.-S. and Schroeder, R. (2009).‘Making it Open and Keeping it Safe: e-Enabled Data Sharing in Sweden’, forthcoming, *Acta Sociologic*a.

Barjak, Franz et al. (2009). ‘Case Studies of e-Infrastructure Adoption.’ *Social Science Computer Review*.

Jirotka, M. et al. (2005). “Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case Study.” *Computer Supported Cooperative Work (CSCW)* 14:369-398.

Den Besten, M; Thomas, A. and Schroeder, R. (2009) ‘Life Science Research and Drug Discovery at the Turn of the 21st Century: The Experience of SwissBioGrid’ forthcoming in *Discovery and Collaboration*.

Dutton, W. H., & Meyer, E. T. (2008, 18-20 June). ‘The Diffusion of e-Research: The Use and Non-Use of Advances in Information and Communication Technologies across the Social Sciences’, presented at the *4th International Conference on e-Social Science*, Manchester, UK.

Fry, J. and Schroeder, R. (2009). ‘Towards a Sociology of e-Research: Shaping Practice and Advancing Knowledge’, in N.Jankowski (ed), *e-Research: Transformation in Scholarly Practice,* New York: Routledge.

Hughes, T. P. (1994). ‘Technological momentum’, in L. Marx & M. R. Smith (Eds.), *Does technology drive history? The dilemma of technological determinism* (pp. 101-113). Cambridge, MA: MIT Press.

Jackson, S. J., Edwards, P. N., Bowker, G. C., & Knobel, C. P. (2007). Understanding Infrastructure: History, Heuristics, and Cyberinfrastructure Policy [Electronic Version]. *First Monday*, 12 from http://www.firstmonday.org/issues/issue12_6/jackson/index.html.

Olson, G.M., Zimmerman, A. and Bos, N. (2008). *Scientific Collaboration on the Internet.* Cambridge MA.: MIT Press.

Powell, W. W. (1990). ‘Neither market nor hierarchy: network forms of organization’, *Research in Organizational Behavior* 12:295-336.

Schroeder, R. (2007). ‘e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production?’, in *Prometheus*, 25(1), pp.1-17.

Schroeder, R. (2008). e-Sciences as Research Technologies: Reconfiguring Disciplines, Globalizing Knowledge. *Social Science Information, 47*(2), 131-157.

Williams, R. (2003). ‘Grids and the Virtual Observatory’, in F.Berman, G.Fox and A.Hey (eds.), *Grid Computing: Making the Global Infrastructure a Reality*, pp.837-858.

Wooley, J.C. and Lin H.S. (Eds) (2005). *Catalyzing Inquiry at the Interface of Computing and Biology*. Washington, DC: The National Academies Press