Catching Up or Latecomer Advantage? Lessons from e-Research Strategies in Germany, in the UK and Beyond

R.Schroeder, M. den Besten and J.Fry

Oxford Internet Institute,
1 St.Giles, Oxford University, UK
Email: {ralph.schroeder, jenny.fry}@oii.ox.ac.uk; matthijs.denbesten@oerc.ac.uk
phone: +44- 1865 287224 *fax:* +44-1865-287211

Abstract

This paper will discuss the e-Research strategies in Germany, the UK and elsewhere with a view to identifying potential lessons that can be learned. The UK e-Science initiative has been ongoing for more than five years, as has the cyberinfrastructure initiative in the US. The German D-Grid and other e-Research initiatives are much more recent. Generic lessons identified from our cross-national observations include the need to take usability into consideration at the outset of programmes; 'soft infrastructure' issues, such as legal and ethical considerations, can delay technological developments; 'Grid-based infrastructure' can seem irrelevant to domains outside of large-scale research, complex security systems often lead to 'work-arounds' that undermine the integrity of the system; and uptake beyond the initial development group is highly uncertain.

1 Introduction

This paper will make some comparisons between the German and UK e-Research initiatives in order to assess what lessons can be learned from their respective trajectories. It will outline some of main features of the various D-Grid and related German e-Science projects and of UK e-Science and also put them into the wider context of global e-Research (including the US cyberinfrastructure initiative). One focus of the paper will be on the whether the German e-Science effort can avoid some of the problems that have only emerged over time in the UK. Another focus will be on a number of specific aspects of e-Research, including open access, scale and scope of e-Research, and ethical and legal obstacles. The analysis will contribute to research policy and to identifying the various obstacles and opportunities in strategies of e-Research initiatives.

2 Overview of e-research programmes

Compared with programmes in the US and UK, German e-Research began relatively late: Funding of projects under a dedicated e-Research programme only began in 2005/6 while in the UK and US large-scale initiatives began around 2000. In the United Kingdom, a £250 million, 5-year e-Science programme was initiated in 2001 in order to develop tools, technologies, and infrastructure to support multi-disciplinary and distributed collaborations. The programme tried to address the so-called 'data deluge' (Hey and Trefethen, 2003) in the physical and biomedical sciences. That is,

it tried to address the need for more-and-more computationally intensive simulations, the management of ever-increasing stores of data, and the need for shared access to increasingly expensive instruments (den Besten, Schroeder, and Fry, 2007). The UK e-Science programme was conceived of as a collection of pilot e-science projects in a range of disciplines underpinned by a "core programme" concerned with generic e-science infrastructure (Hey, 2004).

In the US the label 'collaboratories' has been in use for longer than the Cyberinfrastructure programme and continues to have currency (Finholt 2003). The Cyberinfrastructure programme itself took off in 2003 when the Atkins report was published and decision makers were convinced by the potential 'revolution' in science it described and were prepared to allocate the one billion dollars of funding per year it called for. In particular, the report led to a number of initiatives in creating large-scale facilities and projects under several programmes of funding spearheaded by the National Science Foundation (NSF). A central Office of Cyberinfrastructure was created by NSF in August 2005 and the vision of the Atkins report has been extended to the social sciences in a report by Berman and Brady (2005), who have outlined an ambitious series of challenges for various social science disciplines that can be addressed by means of an enhanced 'cyberinfrastructure'.

Despite the relatively late start of e-Research in Germany, there are now seven D-Grid projects as well as a number of additional e-science and e-library projects. Moreover, Germany is poised to expand its e-research programme significantly, with a number of further initiatives under discussion. There are also efforts to extend the sharing of high-performance computing (HPC) resources across a number of centres. Finally, Germany has been a partner in larger European e-science projects for some time, foremost among them the Enabling Grids for E-SciencE (EGEE) project. Some questions about how these efforts can be integrated and coordinated are discussed below.

3 Funding Models and disciplinary representation

A distinctive feature of German e-research is its funding structure. In the UK and US funding has come from the main funding bodies for academic research; the ESRC and EPSRC in the UK and the NSF in the US (see Schroeder and Fry 2007). In contrast, funding for D-Grid did not come from the German Research Foundation (Deutsche Forschungs Gemeinschaft), which is the main the academic research funding body in Germany. Instead, the main funding came from the German Federal Ministry of Education and Research (BMFB), which couples this programme more closely with broader aims of the national system of innovation and training in comparison with the UK and US focus on academic research.

In the UK e-Science "core Programme", while a number of projects have successfully demonstrated grid-based infrastructure and tools as proof of concept, 'roll-out' and sustainability have in the later phases been identified as major challenges. In the D-Grid context, the issue of sustainability and the ensuing debates around continued funding is complicated by the federal nature of (research governance in) Germany, which entails that there is an ongoing discussion about how the individual states (Laender) should contribute when the capacity of HPC resources in one state are tapped by others. This is not an issue in the overall UK e-research programme as funding structures have been developed so that resources, standards and

security are coordinated on a national level. Integration across disciplines may however be easier given that the early phase of D-Grid funding has not differentiated between the natural sciences, engineering, social sciences and humanities, as has been the model in the UK and US.

In theory this has enabled projects to be distributed across the disciplines, but in practice the lack of ring-fenced funding for non-'big science' disciplines has led to an under-representation of certain disciplines. For example amongst the seven D-Grid projects none of them have a social science orientation. As elsewhere, the initial projects in Germany are heavily dominated by the physical science and engineering disciplines, with a focus on physical, biological and chemical data, with the exception of one project based on textual data using computational linguistics approaches and techniques. As the German e-Science effort is still in its early stages, however, it remains to be seen how the composition will shift in terms of focusing on certain domains.

In the UK, changes of emphasis in the e-Science effort can be identified over the course of time. In biotechnology, for example, there was a strong UK focus on tool development early on, but this has gradually shifted towards databases and usability. A National e-Science Centre (NeSC) was established to coordinate e-Science efforts. The subsequent establishment of the National Centre for e-Social Science (NCeSS) by the main funding body for social science research, the ESRC, has ensured that funds are channelled to the 'softer' sciences. There is now also an e-Science support centre for the humanities, the Arts and Humanities Data Service (AHDS), jointly funded by the Arts and Humanities Research Council (AHRC) and the Joint Information Systems Committee (JISC).

4 Sustainability

The "core programme" in the UK is already winding down as a distinct initiative that is separate from mainstream programmes of research funding. This will not mean the end of e-Science, but the integration of e-Science within other initiatives. It also means, however, that unless there are new national programmes, the development of e-Science as a separate research initiative will take on a different guise. In the US, there is likewise increasing concern about how to extend and ensure the future uptake of the cyberinfrastructure initiatives. In both the US and UK, much of the emphasis is thus turning to how the various tools can be put to more widespread use and how the e-Sciences can be sustained in the future. To give just one example, a recent project of the European Commission to enhance uptake of e-Science, AVROSS (Accelerating Transition to Virtual Research Organization in Social Science), drawing particularly on the US and UK experiences, is 'a study on requirements and options for accelerating the transition from traditional research to virtual research organisations through e-infrastructures'.

Unlike in the US and the UK, there has been an early focus on 'sustainability' in the D-Grid funding calls and also in the project goals (for example, a workshop was devoted to this topic, see https://www.d-grid.de/index.php?id=256). Perhaps this kind of foresight is a sign of 'latecomer advantage', where the mistakes of other efforts, like the UK where discussions about sustainability came rather late, can be avoided. There is also a strong emphasis on private sector involvement and on the economics of the projects. German researchers are taking a leading role, for example, in EU projects about the economics of the Grid, including Grid

Econ (Grid Economics and Business Models) and SORMA (Self-Organizing ICT Resource Management). There is also extensive discussion about collaboration with industry, particularly in view of the place of small and medium-size enterprises (SME's) in Germany, which are much more important in Germany than elsewhere (see Keck 1993 for their role in research). These discussions are then framed in terms the needs of these smaller firms to obtain access to HPC resources, which they would otherwise not be able to afford (see, for example, the D-Grid project In-Grid that includes simulations for small firms that focus on specialized casting).

The UK, German and other e-Research efforts can be seen in terms of three layers – with the technical Grid on one side, the different applications or services on the other, and middleware which links the two in between. Much of the recent effort in UK e-Science has been put into creating a repository – the Open Middleware Infrastructure Institute (OMII) – which maintains and makes available the middleware software that has been developed in the UK. This effort is regarded as critical to the sustainability and diffusion of e-Science, but it also faces the problem that some have estimated that it takes 5-10 times as much effort to make middleware available in this way than it took to develop the software in the first place!

In this regard, too, Germany is in a somewhat advantageous position as it is in the process of choosing and combining a number of existing middleware components (Gentzsch 2006). This approach is not without hurdles since the different types of middleware are at different stages of maturity, levels of adoption, interoperability and standards, and in terms of functionality. Apart from middleware, it can be also be asked whether certain tools and datasets arising from e-Science initiatives will become dominant, with implications for the take-up, acceptance and visibility of e-Research.

5 Soft infrastructural issues

The development of e-Research infrastructure involves the co-development of a robust technical system based on technical standards and protocols on the one hand ('hard' infrastructure) and an integrated social system based on the coordination of social standards and protocols on the other ('soft' infrastructure). Soft infrastructural issues that have been illuminated by the UK e-Research programme include security and authentication, anonymization of social science micro-data, usability, uptake and use, contractual arrangements for inter-institutional and cross-sector collaboration, and trust between stakeholders.

There has been minimal attention to ethical and legal issues in German e-Science, even though these have been identified in the UK and in the US as key barriers to further development (David and Spence 2003). The response to these issues in the UK and US has been to prioritise these issues in the second round of funding and to develop some dedicated programme such as the EPSRC's programme on the "Usability Issues to Emerge from e-Science". There is now also work on this topic in the Office of the Cyberinfrastructure of the National Science Foundation (http://www.nsf.gov/dir/index.jsp?org=OCI).

One set of examples here are the legal and contractual issues involved in the 'openness' of e-research (see David, Den Besten and Schroeder 2006). In the UK, as elsewhere, there has been a push on the part of the research councils and other organizations for 'open science' and 'open access'. Yet openness in research, when it is examined more closely, actually consists of

a number of elements in relation to which there are a variety of attitudes and practices: for example, in relation to open source licensing, there are various options, but considerable uncertainty about which of these should be adopted because, among other factors, no dominant solution has yet emerged. Another aspect of 'openness' is documentation of research as an internal process, which has mainly been governed informally. Publication in open access journals has been encouraged by UK research councils, but take-up of this option has been varied as researchers have given priority to publishing in established journals with high status.

6 Towards a sustainable infrastructure

The extent to which e-Science projects are expected to contribute to a greater national infrastructure is another key issue. In the UK, this question has been partly shaped by the combination of organizations (the National Grid Service, the National Centre for e-Science, the National Centre for e-Social Science, the Arts and Humanities e-Science Support Centre, and the Digital Curation Centre (DCC)). Similar institutions will emerge over time on the European level and in Germany. On the European level, there are several programmes in the European Community's current Seventh Framework (FP7 2007-13) that build on previous efforts and seek to strengthen existing and develop new 'e-infrastructures'.

The D-Grid initiative has been planned over several phases. The seven current D-Grid projects that are part of the first (2005-2008) phase are to be supplement in a second (2007-2009) and possible third phase thereafter, with a budget from the BMFB of 100 million Euros (Gentzsch 2006). The second phase is specifically aimed at integrating the Grid-infrastructure horizontally, vertically and by filling gaps, but it remains to be seen how these aims translate into practice. Certainly the strategy of spreading the initiative over several phases can be seen as an attempt to ensure that a longer-term effort which aims at integration is envisioned.

As e-Science reaches maturity, however, two questions arise: one is related to the often stated goal, that e-Science should 'disappear' as it comes into routine use in different applications domains: will this spell the end of a sustained effort to develop e-Science tools and resources and Grid-enable them? A second is whether the national integration of e-Research will continue to be pursued, or if the e-Research effort will devolve to the level of individual discipline- or domain-specific efforts on the one hand, and to a wider collaboration between researchers in different countries sharing tools and resources on the other (in other words, superseding coordinated 'national' e-Research efforts)?

7 Conclusions

It can be seen that the German e-research effort is distinctive and has used the opportunity of being a latecomer in certain respects (sustainability) and is still catching up in others (legal issues). An ideal typical overview of which highlights the main differences is presented in table 1 below. In the UK the e-Science core programme has been a success in terms of demonstrating proof of concept of Grid-based tools, the integration and coordination of goals, approaches and techniques around a central infrastructure has been limited. This outcome has been shaped by 'soft infrastructural' issues as much as, if not more so, than technical development issues. The early

recognition and provision for sustainability in the D-Grid programme may address and forestall some of these issues.

	Germany	UK	US
Main Actors	Federal Ministry,major scientific organizations and academics	DTI, JISC, main Research Councils	HPC and Collaboratory Communities,National Science Foundation
Motivation	Catch-up and integration of efforts within Germany, into Europe and Beyond	Take the lead in emerging technology, develop a UK presence in e-Science	Keeping the lead in developing a national technology infrastructure
Technical Goals	Developing a common Grid to share computing	Address Data Deluge	Share expensive equipment and resources
Timing and Horizon	2005, two or possibly three phases of three years and longer-term aims	Approx. 2001, initially long-term, increasingly short-term funding with longer-term aims	Atkins Report 2003, but 1990s run-up, approx. five year funding cycles
Organizational structure	Federal funding, include projects from all disciplines, computing needs to be shared via Laender	Hub and spokes, with disciplines under their research councils	Individual projects, consortia of HPC centers, all disciplines (except humanities) under NSF
Management	Federal coordination, joint All-Hands workshops	Institutes (NeSC, NCeSS, AHDS), including for middleware, disciplinary All-Hands meetings	National Office, disciplinary workshops

Table 1. Overview of three e-Science Initiatives

In terms of disciplinary representation within e-Research efforts, in the UK programmes in the social sciences and humanities have come considerably later than those in the physical and biomedical sciences. This has led to difficulties in the use and usefulness of existing infrastructure for the type of approaches and tools being developed within the social science and humanities domains. This incremental funding of e-Research has led to

sustainability issues. The funding structure of D-Grid may mean that this can be negated, so long as e-Research priorities look forward to potential applications outside of 'big science'.

Comparing the various e-Research efforts, it can be seen that the aims of national programmes have shifted considerably, but so too have the key policy questions about e-Research programmes.

- How can individual projects be sustained once the project is finished?
- How should individual projects be integrated within a larger whole?
- What kinds of links, to industry and user communities within and outside academia, contribute to sustainability?
- How are e-infrastructures impacted by larger debates about making data and digital resources available in society?

These questions are far from the initial technology-centred questions of e-Research. Yet, as e-Research becomes more part and parcel of the world of research in society-as-whole, the question of its integration in social goals will become more pronounced than the question of catching-up technologically.

Acknowledgements

The work for this paper has been supported by ESRC grant RES-149-25-1022 and is part of the Oxford e-Social Science (OeSS) project.

References

Berman, Francis., & Brady, Henry. 2005. *Final report: NSF SBE-CISE workshop on cyberinfrastructure and the social sciences*. Retrieved September 28, 2006 from www.sdsc.edu/sbe/

David, Paul and Spence, Michael. 2003. Towards institutional infrastructures for e-science. Research Report 2, Oxford Internet Institute, Oxford. [Available at: http://www.oii.ox.ac.uk/research/project.cfm?id=26.]

David, Paul; den Besten, Matthijs and Schroeder, Ralph. 2006. How Open is e-Science?, *Proceedings of the 2nd International Conference on e-Science and Grid Computing*, Amsterdam 4-6 Dec, also available at http://www.oii.ox.ac.uk/microsites/oess/papers.cfm

Den Besten, Matthijs; Schroeder, Ralph and Fry, Jenny. 2007. Biomedical e-Science Collaboration, Journal of Biomedical Discovery and Collaboration, forthcoming.

Finholt, Tom. (2003). 'Collaboratories as a New Form of Scientific Organization.' Economics of Innovation and New Technology, 12, pp. 5-25

Gentzsch, Wolfgang. 2006. National Grid Initiatives: Lessons Learned and Recommendations, keynote at *2nd International Conference on e-Science and Grid Computing*, Amsterdam 4-6 Dec.

Hey, Tony. (2004). Why Engage in E-Science?. CILIP Update, 3(3), pp.25-27.

Hey, Tony. and Trefethen, Anne. 2003. The data deluge: an e-Science perspective, in F. Berman, G. C. Fox and A. Hey (eds.), *Grid Computing: Making the Global Infrastructure a Reality*. Chichester: John Wiley, pp. 809-824.

Keck, Otto. 1993. The National System for Technical Innovation in Germany. In Richard Nelson (ed.), *National Innovation Systems: A Comparative Analysis.* Oxford: Oxford University Press, pp. 115-157.

Schroeder, Ralph and Fry, Jenny. 2007. Social Science Perspectives on e-Sciences, *Journal of Computer-Mediated Communication*, volume 2, issue 2, http://jcmc.indiana.edu/.