Talking 'bout a revolution: Framing e-Research as a computerization movement

Grace de la Flor[1], Eric T. Meyer[2]
[1]Oxford University Computing Laboratory
[2]Oxford Internet Institute
grace.de.la.flor@comlab.ox.ac.uk

Abstract

In this paper we discuss how e-Research initiatives may be framed as a computerization movement (CM) using Kling & Iacono's (2001) conceptual framework. Applying their framework to our case study data has enabled us to explore an alternative analysis of the e-Research vision, which links the uptake and use of Grid technology with a preferred mode of conducting scientific research. We explore the origins of this vision first by tracing the historical trajectory of advances in scientific computing which led to Grid technology. Next, we discuss how the e-Research vision has been communicated through public discourses that present core ideas about the ways in which Grid infrastructure would transform research. Finally, we discuss how the implementation this vision may reconfigure both domain practices and developer-user access to infrastructure. As close participants in the UK e-Research community we are less certain whether it is clear that e-Research currently constitutes a successful computerization movement, or whether it is more accurate to say that some within the community aspire towards building a CM that will in turn have a wider impact on the physical sciences, the social sciences and the humanities.

Introduction

Building a scientific research infrastructure under the banner of e-Research has absorbed the efforts of large numbers of researchers and considerable amounts of research funding over the last decade; the database of the U.S. National Science Foundation's Office of Cyberinfrastructure, for instance, shows 342 awards totalling over $375 million USD since

2000[1]. These programmes go under a variety of names: e-Science (mainly in the UK), e-Social Science (UK), Cyberinfrastructure (US), e-Humanities (UK), and Digital Humanities (US). For this paper, we will use the umbrella term 'e-Research' to reference all of these efforts (Borgman, 2007). We define e-Research broadly, as a new way to undertake research across the physical sciences, social sciences and the humanities through the use of a "distributed infrastructure that supports resource sharing amongst dynamic collections of individuals, institutions and systems" (Jirotka *et al*, 2006).

e-Research aims to embed specific technological artefacts into the routine work practices of scientific research. Such an undertaking is a socio-technical process where multidisciplinary projects, most often including computer scientists and domain scientists from one or more specialist domains, propose changes to the way domain research may be carried out. We examine this process broadly and from a critical perspective through a discussion of the relationship between technological innovation, public policy vision and organizational practices using a CM framework to analyse the activities of e-Research.

Implementing the e-Research vision and much of the language about these efforts promises revolutionary change as a result of widespread adoption including the ability to facilitate long distance collaboration, access to remote instruments, sharing data and computational resources and enabling an orientation toward large-scale research or 'big science' projects (Finholt, 2003). We argue that this last point, enabling an orientation toward big science, has been a key driver of the e-Research vision and can be demonstrated in the much cited quote:

> "e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it." – John Taylor[2]

e-Research has been compared to big science because it shares similarities with it, particularly its provision of complex, large-scale technical infrastructure and its goal to mobilise global collaboration using large research teams, co-located and distributed, to work together on common research problems (Welsh *et al*, 2006).

---

[1] Available online at http://www.nsf.gov/awardsearch/progSearch.do?SearchType=progSearch&page=2&QueryText=&ProgOrganization=OCI&ProgOfficer=&ProgEleCode=&BooleanElement=false&ProgRefCode=&BooleanRef=false&ProgProgram=&ProgFoaCode=&Search=Search#top

[2] Definition from Dr. John Taylor who was the Director General of Research Councils Office of Science and Technology from 1998-2003 and involved at the inception of the UK e-Science programme.

In this paper we discuss how e-Research initiatives may be framed as a computerization movement (CM) using Kling & Iacono's (2001) conceptual framework. Using the CM framework has enabled us to explore the ways in which the e-Research vision has been operationalised in actual practice. We do this first by tracing the historical trajectory of the e-Research vision as it was developed within the scientific computing research community. Next, we discuss how the e-Research vision has been communicated through a government sponsored research agenda. Finally, we discuss how the e-Research vision's implementation may reconfigure both domain practices and developer-user access to infrastructure. We'll begin by first introducing the CM framework and describing its key components.

Framing e-Research as a computerization movement

One way to understand efforts to expand the development and adoption of technologies is through Kling & Iacono’s concept of computerization movements (CMs); this concept was explicated in a series of articles (Kling & Iacono, 1988; Kling & Iacono, 1994; Iacono & Kling, 1996; Iacono & Kling, 2001), and has been followed up with more recent work examining CMs in a variety of contexts (Elliott & Kraemer, 2008). Computerization movements, in Kling & Iacono’s model, "are a kind of movement whose advocates focus on computer-based systems as instruments to bring about a new social order" (Kling & Iacono, 1988: 228). CMs also communicate key ideological beliefs about "what computing is good for" and how project participants "should manage and organize access to computing" (Ibid: 227). Kling & Iacono's main thesis is that "computerization movements communicate key ideological beliefs about the links between computerization and a preferred social order which help legitimize computerization for many potential adopters" (Ibid).

A question we explore is whether the various e-Research initiatives have attempted to communicate a set of ideological beliefs which link the use of Grid technologies with a preferred approach to conducting research, and if so, what might be the motivating factors for this? Others have already framed e-Research as a computerization movement. For instance, Hine's (2006) case study of the UK systematics community outlines tensions which emerged when a report produced by the UK Select Committee on Science and Technology challenged the community to embed new technologies into its research practices. In the paper, Hine describes the strategic manoeuvring sometimes necessary amongst domain scientists when challenged by advocates of e-Research as to why they haven't adopted or implemented its vision in a speedier manner. While proponents of e-Research associated the notion of accessibility to materials with digitization, the complexities of the domain's work practices and their research requirements while in the field generated a "more nuanced understanding of the role of digital resources", as Hine notes:

> "It is recognized that images of specimens can serve a few of the functions of their material counterparts, but most systematic work is still thought to require access to the real thing. Also, an interactive digital identification key may be of little use when circumstances dictate that a more apt solution is a laminated sheet that can be taken into the field and used *in situ* without additional technological assistance" (p. 33).

The paper goes on to describe how visions of technology use are sometimes in conflict with real-world concerns of the domain. In conclusion, Hine suggests that "If much of the potential of a computerization movement is in stimulating creative reactions against it, then more could usefully be made of occasions for expressing doubt and disaffection. This kind of debate should be encouraged…" (p. 45).

Hara & Rosenbaum (2008) also argue that e-Research constitutes a computerization movement describing it as "a fairly constrained CM that encourages discourses among a specialized population…Though the scope of e-science is rather small, it is a CM because technological action frames have arisen around it, it has engendered various types of discourses, and it involves a range of computerization practices within a range of organizations". To clarify, a technological action frame is comprised of a series of arguments developed by proponents of CMs that "form the core ideas about how a technology works and how a future based on its use should be envisioned" (Iacono & Kling 2001: 100). Technological action frames "simplify and condense elements of complex technologies and their potential use...[enabling]...groups of people to interact about what they might mean" (Ibid: 102). They are communicated through public discourses about a technology and reinforced by organizational practices as individuals and organizations "implement and use technologies in their micro-social contexts".

In the next two sections we will analyse three technological action frames used to describe e-Research:

- The positioning of Grid technology as a means to "provide an effective and efficient platform for the empowerment of specific communities of researchers to innovate and eventually revolutionize what they do, how they do it, and who participates" (Atkins *et al,* 2003).

- The proposition of an "imminent data deluge" (Hey & Trefethen, 2003b) which will require "data access, integration and federation capabilities" using Grid technologies.

- Promoting "digital scholarship" through the embedding of Grid technology in the humanities and social sciences will led to "new intellectual products" making it necessary to "give high priority to building tools and collections" (ACLS report, 2006: 7).

The first of these will be addressed in the next section, 'tracing the origins of e-Research', where we will explore the origins of the e-Research vision by tracing the historical trajectory of advances in scientific computing which led to Grid technology. The last two points will be presented in the section, 'a new science and engineering research agenda', where we will discuss how the e-Research vision has been communicated through public discourses that present core ideas about the ways in which Grid infrastructure would transform research.

Tracing the origins of e-Research

e-Research has been characterised as a 'technology-led initiative' (Schroeder & Fry, 2007), indicating an imbalance at the early phases of large-scale projects where "technologies have been developed without taking social aspects into account" (p. 567). We will expand on this characterisation to include technological visions because these are usually developed far in advance of the implementation of large-scale projects. Technological visions give meaning and direction to large-scale projects by providing pre-defined solutions of how technology should operate, which may be used later to justify key decisions about the direction of software development (de la Flor *et al*, 2007). e-Research links the uptake and use of Grid technology with a preferred mode of conducting scientific research and in this section we identify three factors which may have contributed to the development of this vision.

The first begins in the 1950s, with the commencement of projects to address specific research challenges of national concern such as nuclear energy, the space program and weapons development; this period marks the emergence of what has been termed *big science* (Galison, 1992). The second starts with advances in scientific computing research which supplied big science research projects with the most technically advanced machines to support their requirements; the domain of supercomputing (Buzbee & Sharp, 1985). The third begins in the 1990s, as scientific demands for greater performance and storage capacity increased, it galvanised efforts to advance supercomputing even further leading to the design of parallel architectures, initially known as 'metacomputing' (Smarr & Catlett, 1992); this new architecture could harness geographical distributed computational resources as if it were single resource and is known today as the Grid.

The interactions between both big science and computer science have been complementary; with big science providing requirements and research directions to scientific computing research and supercomputing providing technological advances that could be utilised in big science projects. To trace how the e-Research vision may be related to the above we present a broad overview of research efforts in scientific computing.

Supercomputing and big science

The term 'big science' (de Solla Price 1963; Weinberg, 1961) has been used to describe research activities which require large facilities, large budgets and large research teams. Big science projects have typically been high-priority government sponsored research in areas such as nuclear facilities, military projects, particle physics, astronomy and climate research (Galison, 1992).

> "For many the age of Big Science was ushered in by the Manhattan Project during World War II, when the building of the atomic bomb involved the mobilization of much of the U.S. community of physical scientists in an engineering project of unprecedented magnitude. By the early 1960s, with the advent of NASA and the national space program, the term 'Big Science' was firmly affixed as a label for projects that required large-scale organization, massive commitments of funds, and complex technological systems" (Capshew & Rader, 1992: 3-4).

National laboratories in the US, and elsewhere, acquire computing performance through commercially developed supercomputing machines such as Cray, SGI and others (Woodward, 1996). These machines provide the capabilities necessary to support the types of research that big science projects undertake; more specifically, they provide high performance computing and mass data storage capacities. These features are required in big science projects so that researchers can ask questions about phenomena that are difficult to investigate experimentally, for example, the circulation of the oceans and atmosphere, tectonic plate motion and nuclear weapons testing (Buzbee & Sharp, 1985).

Big science research typically requires the use of complex numerical models and other data processing techniques to test predictions or to analyse very large amounts of data. Once a model has processed data, the results are investigated through the use of simulations and visualisation tools[3]. Further experiments can take place through the manipulation of variables available within the model or through modifications to the original. Using this approach, researchers can examine specific variables as they react to any number of simulated circumstances.

Processing *in silico* experiments may take hours or days depending upon the efficiency of the model's algorithm and computational power available. As a response to the increasing computational requirements of scientists, the US National Science Foundation (NSF) established five supercomputing centres in 1985 (Freeman *et al*, 2005); each providing researchers with the most technically advanced supercomputing available[4]. These supercomputing facilities had been utilised successfully for over 20 years. Nevertheless, increased demand for high-performance computing and greater storage capacity prompted several design revolutions in scientific computing research from vector computing in 1967 to "multiple processors arranged in parallel architectures" (Buzbee & Sharp, 1985: 595). It was the design of distributed parallel architectures that led not only to a new way of maximising the technical capabilities of supercomputers, but also to a new way of thinking about the future of networked computing more generally as members of the NSF directorate state in Freeman *et al* (2005):

> "past efforts in supercomputing and high-performance networking are being subsumed into a broader, integrated vision of a more capable, ubiquitous, and accessible cyberinfrastructure" (p. 682).

From supercomputing to cyberinfrastructure; with the introduction of parallel architectures and their ability to harness the computational resources of geographical distributed computers, visions emerged about the ways in which this technical innovation could be implemented more widely.

---

3 This type of research has been referred to as *in silico* research meaning "within silicon".

[4] The centres now organise their resources through TeraGrid, http://www.teragrid.org/about/.

Metacomputing and the Grid

An initiative originally known as 'metacomputing' began as a series of projects that would link the five US supercomputing centres together so that geographical distributed computing devices could be harnessed as a single resource forming "a *network of virtual supercomputers* or *metacomputers*" (Foster & Kesselman, 1997, italics in original). This linking up of resources provided an increase in both the processing power and storage capacity available to researchers. Metacomputing, later known as the Grid (De Roure *et al*, 2003), introduced a new set of protocols to network architecture. De Roure *et al* (2003: 90) explain the ways in which these protocols would transform networked access to information and tools using an analogy; up until now, the Web has enabled the delivery of information using a 'publishing paradigm' by providing a platform in which data can be shared in a distributed manner as images, documents and software tools. Embedding Grid protocols within existing Internet protocols, by comparison, would harness the computational resources of networked machines creating an 'interactional paradigm' so that instruments and facilities at remote locations could be accessed directly.

Implementing these new Grid protocols made possible the virtual consolidation of resources through a network. Originally designed to support big science research through the linking of the five US supercomputing centres, visions to extend its implementation quickly emerged with a goal of eventually creating a network on a global scale (Smarr & Catlett, 1992; De Roure *et al*, 2003). Visions for both its near and longer term, have been elaborated in a series of books (e.g. Berman *et al,* 2003), articles (e.g. Foster, 2008) and government reports (e.g. Atkins *et al*, 2003; Berman & Brady, 2005). The implementation of Grid technology has even been framed as the next generation of protocols for the Internet as Berman *et al* (2003: 40) state "ultimately, one would hope that the Grid will be the operating system of the Internet".

Both the near and longer term visions for Grid technology extend its implementation beyond big science projects and each is communicated through public discourses that present core ideas about the ways in which it would transform scientific research and the Internet[5]. Using Kling & Iacono's framework, we describe these visions as mobilizing ideologies used to establish support, to provide a rationale for the development of government programmes and

---

[5] We briefly explore implications for Internet use in the last section.

to initiate projects that would implement Grid infrastructure and design domain specific tools that could be used within it. In the next section we will explore in greater detail these mobilization efforts.

A new science & engineering research agenda

The technological innovations described in the previous section led to the development of a consensus amongst the scientific computing research community and others that Grid technology should be implemented on a massive scale and that it should even inevitably subsume the Internet. This vision has been communicated in *Revolutionizing Science and Engineering Through Cyberinfrastructure* (Atkins *et al*, 2003) in the US and through *The UK e-Science core programme and the Grid* (Hey & Trefethen, 2002) and *e-Science and its Implications* (Hey & Trefethen, 2003a) in the UK. In 2001, prior to the publication of these vision documents, the UK e-Science programme, in parallel with the US Cyberinfrastructure (CI) initiative, each established a new multidisciplinary science and engineering research programme (Hey & Trefethen, 2005; NSF report, 2007).

The programme's core ideas are described in Atkins *et al* (2003). The report begins by first aligning its vision for the future within the context of larger society:

> "Scientific and engineering research has been crucial in both the *creation* and the advanced *application* of the amazing products of the digital revolution begun some sixty years ago – a revolution that increasingly undergirds our modern world" (p.4).

From this, the report argues for the need to extend technological transformations:

> "indeed a further revolution – in how we create, disseminate, and preserve scientific and engineering knowledge" (p.4).

In the UK, the programme's vision describes the success criteria for embedding Grid technologies in terms of its relevance to research disciplines and ease of use:

> "The Core Programme can assist the e-science projects in building their discipline-centric grids and learn how to interconnect and federate multiple grids in a controlled way. If we can build such grids to production quality, operating 24 hours a day, seven days a week, 52 weeks a year, and that do not require computing experts to use them, then e-science really does have the potential to change the way we do scientific research in all our universities and research institutes" (Hey & Trefethen, 2003a: 1813).

For funded projects, and others considering how e-Research may support their research goals, Grid technologies are presented as a means for fostering distributed research collaboration, the sharing of data and remote instruments, enabling access to heterogeneous datasets and providing high-performance computing power and mass storage capacity[6]. More recently, government programmes have been extended to engage social scientists in the use of Grid technologies. In the UK, the National Centre for e-Social Science (NCeSS) was established in April 2004 and in the US, the social, behavioral and economic (SBE) sciences have been presented with a vision to implement Grid technology (Berman & Brady, 2005). Implementing the vision of a Grid-enabled social sciences is being coordinated in a similar fashion to initiatives in the physical sciences; through the development of programmes that coordinate both the funding and promotion of Grid-based research projects.

We will now explore how two specific visions are used to communicate the ways in which Grid infrastructure would transform research. First, 'managing the data deluge', which has been aimed initially at the physical sciences and second, 'enabling digital scholarship', aimed at social sciences and humanities researchers.

e-Science: managing the data deluge

As discussed previously, while Grid technologies developed and were used within a limited number of big science research projects, consensus grew within the scientific computing research community that these new protocols might benefit a wider range of computer users (De Roure *et al*, 2003: 70). One argument used to frame the need for Grid technologies has been referred to as "the imminent data deluge" (Hey & Trefethen, 2003b: §36.2) which asks us to consider that:

> "data generated from sensors, satellites, high-performance computer simulations, high-throughput devices, scientific images and so on will soon dwarf all of the scientific data collected in the whole history of scientific exploration" (p. 812).

Archiving this type of data would require mass storage capacity. A solution to this challenge is the implementation of Grid technology because it would enable machines to share storage capacity efficiently across a network.

---

[6] These high-level aims are not as straightforward as they may appear, as we discuss in the 'reconfiguring practices' section.

Another argument for the potential benefit of using Grid technologies is its ability to link data stored in distributed heterogeneous datasets, which is thought, are not being utilised to their full potential:

> "It has been argued that alongside the two traditional methodologies of science - theory and experiment - computational science has now emerged as a third methodology. With the advent of scientific data warehouses such as virtual observatories we may be seeing the emergence of a fourth methodology, that of collection-based science" (Hey & Trefethen, 2003a: 1823).

Implementing Grid technology would provide researchers access to large and varied data sets which might allow data to be analysed in new and unpredictable ways. These are the two main arguments used to frame Grid technologies as providing solutions to both data storage and for data sharing.

e-Social science: enabling digital scholarship

As mentioned previously, developing Grid technologies for researchers in the social sciences and humanities is relatively new. The second vision we explore, 'enabling digital scholarship', is used to argue for the implementation of Grid infrastructure for these research communities. The vision for social science use of the Grid is much more general than in the physical sciences; its main funding body in the UK states that it would like to see it implemented to “address the key challenges in their substantive research fields in new ways”[7]. Grid services for the social sciences might include the design of applications that use data mining tools across a diverse range of large datasets or collaboration tools that provide mechanisms for the sharing, annotation and analysis of both quantitative and qualitative data[8].

Digital scholarship has been described as the use of Grid-enabled software tools to enable both geographically distributed and co-located researchers to collaborate digitally (ACLS report, 2006: 7). More specifically, Berman & Brady (2005) envision:

> "the development of more realistic models of complex social phenomena, the production and analysis of larger datasets (such as surveys, censuses, textual corpora, videotapes, cognitive neuroimaging records, and administrative data) that more completely record human behavior, the integration and coordination of disparate datasets to enable deeper investigation, and the collection of better data through experiments and simulations on the Internet" (p. 5).

---

[7] ESRC: http://tinyurl.com/4okh7k [accessed 11 August, 2008]

[8] See http://www.ncess.ac.uk for a list of past and present software development projects

The argument from e-Research proponents is that the social sciences and humanities can benefit from the use of Grid technologies. However, some researchers have discussed challenges to mapping the computational requirements for the physical sciences directly on to the social sciences and humanities (Carusi & Jirotka, 2007; Meyer *et al,* 2008). In the next section we discuss how implementing the vision to embed Grid infrastructure may not be as straightforward as it may appear as both domain practices and developer-user access to infrastructure is reconfigured. We also introduce data from interviews we have conducted with e-Research stakeholders to further inform the discussion.

Reconfiguring practices in domain research & access to infrastructure

The e-Research vision of enabling large-scale, distributed research across a much broader swath of research domains has been likened to big science approach to research "in the sense that there are large teams of people working together in an attempt to solve particular scientific problems, these teams are of necessity international in character and offer a very wide range of expertise" (Welsh *et al*, 2006: 1537). Previous case study research has shown that there are challenges to implementing this vision because of the ways in which Grid infrastructure would reconfigure domain and organisational practices (Jirotka *et al*, 2005; Welsh *et al*, 2006) and developer-user access to infrastructure (de la Flor *et al*, 2007). The reconfiguration of these practices may in turn have an impact on Grid technology uptake and use. This last section will briefly explore these challenges. The interview data presented spans a two-year period where we have had the opportunity to discuss these issues with computer scientists, domain scientists and other e-Research stakeholders.

Reconfiguring domain practices

Understanding the specific scholarly activities and disciplinary requirements of potential users of Grid infrastructure has been recognised as a major research challenge (Proctor *et al*, 2006; Jirotka *et al*, 2006). This section will focus on a sample of some these challenges for the social sciences. Meyer *et al* (2008) summarise reoccurring themes that have appeared across case studies conducted as part of the OeSS[9] project and include:

---

[9] http://www.oii.ox.ac.uk/microsites/oess/

- institutional and legal challenges to data sharing
- issues with reward structures and assigning credit for contributing data to public archives
- taking into consideration the policies and practices of individual e-Research projects
- achieving standardization across heterogeneous data archives
- ensuring privacy when researchers are collecting data on human subjects
- conveying trust in both the expertise of people and the content of artefacts

This list of challenges provides an indication of how data storage and sharing may not be as straightforward as it might appear. Carusi & Jirotka (2007) also discuss a whole range of ethical challenges and requirements that emerge when considering the sharing of qualitative data in a digital environment.

While data sharing is central to most visions of the grid, evidence suggests that most social scientists are not in the habit of using the data of others, whether obtained from a data archive or from a colleague directly. In Dutton and Meyer's survey (Dutton & Meyer 2008; Meyer & Dutton, 2008) of attitudes towards e-Research and research practices (n=526), they found that relatively few researchers, qualitative or quantitative, had either used archived data (22%) or data from a colleague (17%). In addition, their sample suggests a need to support bottom-up research innovation if research infrastructures are to be designed to reflect the ways scientists work: across all categories of software tools, a significant proportion of users in each category (ranging from a low of 15% of database users to a high of 56% of data integrating software users) relied on bespoke in-house software tools to support their research. This even applies in areas such as quantitative research software: while SPSS is dominant (82%), fully 17% report creating their own tools for quantitative analysis. They interpret these findings as evidence that there is still a gap between the bottom-up practices of ordinary social scientists and the top-down visions of Grid advocates.

One of our interviewees describes what may be different requirements for data access between the physical and social sciences:

> "It's easier for you [social scientists] to collect the data that you need for your particular needs. Whereas, in other communities [the physical sciences] their part of a bigger pool of knowledge that one person or one group couldn't possibly replicate on their own" (ER01b-08).

And:

> "There's not that culture of sharing in the same way [as the physical sciences] because they don't have to do it, because maybe it's just one or two people working on something" (ER01c-08).

These comments indicate that the nature of both the data and research goals within the physical and social sciences may be different enough that perhaps different technological solutions are necessary for each. An interviewee confirmed this when stating:

> "I think there are genuine, specific concerns for the social science community, in that the nature of the data is more sensitive" (ER01a-08).

As discussed in Hine (2006), albeit in a different context, the linking of accessibility to materials with digitization may not solve the research challenges of a domain but may instead complicate or constrain research activities because the complexities of the domain may not be conducive, at this point in time, to the digitisation of materials or procedures (Hine, 2006; Jirotka *et al*, 2005). Also, as Woolgar & Coopmans (2006) argue, more case study work is needed to assess the practical application of e-Research's vision of data mobility and reuse through a series of ethnographic accounts that would document the actual, real-world "practical usage and enactment of Grid technologies", which would make available for analysis both "the emergence and impacts of Grid technologies".

This section has briefly discussed some of the challenges that the embedding of Grid infrastructure presents as it may reconfigure domain practices, the next section will extend the discussion to how it may reconfigure developer-user access to infrastructure.

Reconfiguring access to infrastructure

The e-Research vision has linked the use of Grid technology with a preferred approach to conducting research. Because Grid protocols are embedded into current Internet protocols they would define a set of new conventions used to transfer and share data and computational resources across the Internet. The implementation of these new protocols might also reconfigure the activities of developer-users who operate within the Grid. It would be helpful, in this section, for us to present an analysis of infrastructure from a social science perspective.

Star & Ruhleder (1996) state that "*An infrastructure occurs when the tensions between local and global is resolved*. That is, an infrastructure occurs when local practices are afforded by a larger-scale technology, which can be used in a natural, ready-to-hand fashion" (p. 114, italics in original). Along with this definition is a list of nine characteristics of infrastructure (See Star & Ruhleder, 1996: 113; Star, 1999). We cannot address each of those nine characteristics as it relates to the Grid in this paper. However, we can conduct an initial analysis of whether or not Grid infrastructure may be used in a "natural, ready-to-hand fashion" or if this ability is lost, constrained or made too complex for average users when they try to use Grid infrastructure for their own *ad hoc* purposes.

In our interviews with developers-users some described the difficulties of integrating their tools into infrastructural standards and APIs. As they explained, a system's architecture has implications for how applications can be designed, because they must take into account the workflows and constraints built into Grid middleware:

> “I've got a Grid API against which I had to program in order to get images over and so on. Was that usable for me as a user of Grid technologies? Probably not.” (ER02a-06).

Other developer-users are avoiding the use of traditional Grid protocols altogether by designing their own bespoke software; the 'communication application':

> "One of the definitions of what we're doing is setting up a Grid, 'cause we're connecting experimental facilities together over a network. You could call it a Grid...[but]...anything that we do we program ourselves" (ER03a-08).

In this instance, developer-users are avoiding the implementation of traditional Grid protocols in favour of their own bespoke software that connects distributed systems in a peer-to-peer network that is 'good enough' for their intended purposes.

These two examples demonstrate first, the challenges of using a Grid infrastructure to design applications and second, the avoidance of its use altogether. They also imply that the natural, ready-to-hand qualities of the Internet's infrastructure may be lost or constrained with the implementation of Grid protocols. The next section will discuss how the vision of Grid infrastructure proposed in Foster, *et al* (2001) and the actual practice of its implementation in e-Research projects may be two entirely different idealizations.

So what exactly is the anatomy of the Grid?

In Foster, *et al'*s (2001) influential paper, *The anatomy of the Grid: enabling scalable organizations*, the authors define the field of Grid computing as "an extensible and open Grid architecture, in which protocols, services, application programming interfaces, and software development kits are categorized according to their roles in enabling resource sharing". It will be argued in this section that the authors' vision of Grid infrastructure and the practice of its implementation in the real world may not correspond.

Grid protocols not only enable the sharing of computing power and storage capacity across multiple computers, as discussed in previous sections, it also provides two key additional mechanisms: 1) identity management tools, and 2) the ability to define virtual organisations (VOs). These two features of Grid infrastructure may provide greater control over resources through authentication and authorisation, but they also introduce added complexity.

> "our Grid architecture is first and foremost a protocol architecture, with protocols defining the basic mechanisms by which VO users and resources negotiate, establish, manage, and exploit sharing relationships" (Foster et al, 2001: 205).

These two protocols, in addition to protocols that enable computational resource sharing, are core components of Grid infrastructure and define a set of new conventions used to transfer data and share computational resources across the Internet. Both identity management and VO protocols work together within a Grid infrastructure to enable any number of sharing rules between both resource providers and end-users. A VO can include individuals, groups and institutions that agree to share computing resources for specific purposes. These protocols are capable of being highly controlled:

> "The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving and resource-brokering strategies emerging in industry, science, and engineering. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a virtual organization" (Foster et al, 2001: 172).

And highly dynamic:

> "Because of their focus on dynamic, cross-organizational sharing, Grid technologies complement rather than compete with existing distributed computing technologies...Grid technologies can be used to establish dynamic markets for computing and storage resources, hence overcoming the limitations of current static configurations" (Ibid: 202).

The authors describe what appears to be two contradictory objectives; the design of a highly controlled and highly dynamic digital sharing environment. Highly controlled in the sense that specific computational and data resources can be secured for access to designated end-users and highly dynamic in the sense that end-users may have the ability to create and dissolve virtual organisations in an *ad hoc* manner and where resources are shared on an 'as needed' basis.

> "Just as the Web revolutionized information sharing by providing a universal protocol and syntax (HTTP and HTML) for information exchange, so we require standard protocols and syntaxes for general resource sharing" (p. 205).

It is important to note here that the Web revolutionized information sharing because its protocols are easy to use. As discussed previously, one of the success criteria for embedding Grid technologies for the UK e-Research programme has been ease of use so that it can be utilised in a way that it would not "require computing experts to use them". However, we would argue, based upon initial interviews and two years of participant-observation within e-Research projects, that Grid infrastructure is being implemented in a top-down manner and includes the design of highly-specified applications built by teams of computer programmers. This may be especially true of projects aimed at the social sciences and humanities, but it does not necessarily exclude those in the physical sciences.

We do not take issue with this approach to software development in and of itself. We do question, however, if these practices might lead to the black-boxing of Grid infrastructure. As end-users come to rely more and more on highly trained software developers to program applications within Grid infrastructure, the division between those with the knowledge and expertise to assemble VOs and enact the sharing of resources and those who rely on services offered within it increases. We maintain that keeping Grid infrastructure in the hands highly trained software developers may lead to a situation where average users loose the ability to design *ad hoc* solutions, creating what Zittrain (2008) calls a 'sterile technology'. Sterile technologies are characterised as systems that may provide useful applications but they inhibit access to its underlying architecture which is ultimately controlled by its developers, thus making it difficult for end-users to adapt or change the way it operates.

Evidence of this type of top-down development is indicated by one of our interviewees:

> "I think you gotta look at who's actually producing these tools. They're being produced as part of e-social science projects or e-science projects in general, and those research projects are producing tools that are of relevance to their research interests. So, while those tools will be generally, genuinely useful to other people, there may be gaps that just doesn't happen to tie in with anybody's research interest that will never get filled" (ER01c-08).

An exception to this can be seen in development of Grid infrastructure within the domain of particle physics where developer-users are actively involved in the design of both the infrastructure and the applications that run within it (Venters *et al*, 2007).

Elsewhere, Grid infrastructure are being conceptualised as 'service-oriented architectures', which uses a marketplace analogy (De Roure *et al*, 2003: 86) to describe how end-users and service providers would engage with each other in the negotiation of resources. This new conceptualisation of digital sharing may further promote an infrastructure of producers and consumers where networked applications are rigidly defined by their service providers.

In contrast to these developments, for the past 20 years, scientists and other end-users have shared resources on the Internet and Web, according to their *ad hoc* needs through file sharing, application development and service offerings. Web 2.0 developments are somewhat in line with this bottom-up approach, barring the fact that some of these applications may run on proprietary systems. The architecture of the Internet is said to be a 'generative technology' (Zittrain, 2008) in the sense that it doesn't promote any particular purpose, it enables self-programming and activities are not restricted by any one business model.

We offer this discussion of our initial analysis of the Grid's ability to support the "natural, ready-to-handness" and the *ad hoc* assembly of VOs to encourage reflection on the ways in which developer-user access to infrastructure may be reconfigured in a Grid-enabled environment and also to encourage examination of how the vision of Grid infrastructure and the actual practice of its implementation may be at odds. We came away from our reading of the *Anatomy* paper thinking that the original intention of the framers of the Grid vision remains the design of a dynamic environment where end-users create and dissolve virtual organisations and share resources on an 'as needed' basis. With VOs emerging in unpredictable ways and dissolving when their usefulness has expired.

Possible solutions to some of the challenges presented here may be to provide Grid end-users open access to Grid infrastructure. Büscher *et al* (forthcoming) argue, in their case study research of palpable computing, for the design of open architectures which support the "visibility and inspectability of available resources (processing power, available memory, network bandwidth etc.)". This may be especially important when otherwise transparent connections breakdown. Additionally, end-users "should be able to deconstruct an [ambient] assembly of devices and services, both to inspect it for repair and to use its elements for new assemblies". This same approach, applied to the design of Grid infrastructure, would provide both software developers and domain researchers an open, flexible infrastructure that would begin to deliver something more akin to the original vision of a dynamic Grid. Initiating this approach would mean readjusting the e-Research vision and as a consequence, technical effort. Focusing efforts on the development of a transparent architecture would enable a broad base of knowledgeable end-users to dynamically create VOs on an 'as needed basis' and share resources in a similar manner in which they can easily share resources on the Internet today.

Discussion

In this paper we presented a discussion of how e-Research, "a developing and ongoing revolution" (Freeman *et al*, 2005), may be framed as a computerization movement using Kling & Iacono's (2001) conceptual framework. We discussed the historical trajectory of e-Research and have shown how it can be linked to visions of technology use that originated within the scientific computing research community. Next, we described the mobilisation of the movement through government agenda setting workshops and funding programmes. In addition, we discussed how this simple vision of enabling global research through the implementation of 'next generation infrastructure' includes in it a complex set of technical and social challenges, especially as both domain practices and developer-user access to infrastructure are reconfigured. We agree that implementation of technical innovations can provide new conceptualisations and exciting improvements to networked computing. However, we argue that operationalising the e-Research vision may not be a straightforward as it might appear to be on the surface.

Acknowledgments

We would like to thank all participants interviewed and the EPSRC for funding The Embedding e-Science Applications - Designing and Managing for Usability project. Grant No. EP/D049733/1. Also, thanks to Marina Jirotka for valuable suggestions.

References

Atkins, D. et al. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. National Science Foundation Report, http://www.nsf.gov/od/oci/reports/toc.jsp [accessed 01 July, 2008].

American Council of Learned Societies. (2006). *Our Cultural Commonwealth, The Report of the Commission on Cyberinfrastructure for the Humanities and Social Sciences.* Available at: http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf [accessed 30 June, 2008].

Bergman, M., King, J.L. & Lyytinen, K. (2002). Large Scale Requirements Analysis as Heterogeneous Engineering. *Scandinavian Journal of Information Systems,* 14. p. 37-55.

Berman, F. & Brady, H. (2005). *Final report NSF SBE-CISE workshop on cyberinfrastructure and the social sciences.* National Science Foundation. http://vis.sdsc.edu/sbe/reports/SBE-CISE-FINAL.pdf. [accessed 05 September, 2008]

Berman, F., Fox, G. & Hey, A. (2003). The Grid: past, present, future. In Berman, F., Fox, G. & Hey, A. (Eds.) *Grid Computing: Making The Global Infrastructure a Reality*. Wiley, p. 9-50.

Borgman, C. L. (2007) *Scholarship in the Digital Age: Information, Infrastructure, and the Internet,* Cambridge, MA, MIT Press.

Büscher, M., Christensen, M., Hansen, K., Mogensen, P., Shapiro, D. (Forthcoming). Bottom-up, top-down? Connecting software architecture design with use. In Voss, et al (Eds.). *Configuring User-Designer Relations: Interdisciplinary Perspectives*. Springer.

Buzbee, B.L., & Sharp, D.H. (1985). Perspectives in supercomputing. *Science*, 227(4687), p. 591-597.

Capshew, C., Rader, K. (1992). Big science: Price to present. *Oriris*, 2(7), p. 2-25.

Carusi, A. & Jirotka, M. (2007). From data archive to ethical labyrinth. *3rd International Conference on e-Social Science*. http://ess.si.umich.edu/papers/paper171.pdf. [accessed 07 September, 2008]

de la Flor, G., Lloyd, S., Jirotka, M., Warr, A. (2007). Designing software in support of workplace activities – embedding e-science applications. *3rd International Conference on e-Social Science*. http://ess.si.umich.edu/papers/paper189.pdf. [accessed 07 September, 2008]

De Roure, D., et al. (2003). The evolution of the grid. In Berman, F., Fox, G. & Hey, A , eds. *Grid Computing: Making The Global Infrastructure a Reality*. Wiley, p.65-100.

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 [Electronic version available at http://ssrn.com/abstract=1150422]. Paper presented at the 4th International Conference on e-Social Science, Manchester, UK.

Elliott, M. S. & Kraemer, K. L. (Eds.) (2008). *Computerization Movements and Technology Diffusion,* Medford, NJ, Information Today, Inc.

Finholt, T. (2003). Collaboratories as a new form of scientific organization. *Economics of Innovation and New Technology*, 12(1). p. 5-25.

Foster, A. (2008). New Ways to Connect Data, Computers, and People. *The Chronical of Higher Education*. http://chronicle.com/wiredcampus/article/3146/new-ways-to-connect-data-computers-and-people [accessed 06 September, 2008].

Foster, I., Kesselman, C. & Tuecke, S. (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International Journal of Supercomputer Applications*, 15(3), p. 200-222.

Foster, I. & Kesselman, C. (1997). Globus: A metacomputing infrastructure toolkit. *International Journal of High Performance Computing Applications*, 11(2), p.115-128.

Freeman, P. et al. (2005). Cyberinfrastructure for science and engineering: promises and challenges. *Proc. of the IEEE*, 93(3), p. 682-691.

Galison, P. (1992). The many faces of big science. In Galison, P., Hevly, B. (Eds) *Big science: the growth of large-scale research*. Stanford, CA: Stanford University Press. p. 1–17.

Hara, N. & Rosenbaum, H. (2008). Revising the Conceptualization of Computerization Movements. *The Information Society,* 24**,** 229-245.

Hey, T. & Tefethen, A. (2005). The e-Science Challenge: Creating a Reusable e-Infrastructure for Collaborative Multidisciplinary Science. *CTWatch Quarterly*, 1(4). Available at: http://www.ctwatch.org/quarterly/print.php?p=17 [accessed 12 June, 2008].

Hey, T., Tefethen, A. (2003a). e-Science and its Implications. *Philosophical Transactions of the Royal Society,* 361. p. 1809-1825.

Hey, T. & Tefethen, A. (2003b). The data deluge: an e-Science perspective. In Berman, F., Fox, G. & Hey, A. (Eds.) *Grid Computing: Making The Global Infrastructure a Reality*. Wiley, p.809-824.

Hey, T., Trefethen, A. (2002). The UK e-Science core programme and the Grid. *Future Generation Computing Systems,* 18. p. 1017-1031.

Hine, C. (2006). Computerization Movements and Scientific Disciplines: The Reflexive Potential of New Technologies. In Hine, C. (Ed.) *New Infrastructures for Knowledge Production: Understanding e-Science.* London, Information Science Publishing. p. 26-47.

Iacono, S. & Kling, R. (2001). Computerization Movements: The Rise of the Internet and Distant Forms of Work. In Yates, J. A. & Van Maanen, J. (Eds.) *Information Technology and Organizational Transformation: History, Rhetoric and Practice.* Thousand Oaks, CA, Sage Publications.

Iacono, S. & Kling, R. (1996). Computerization Movements and Tales of Technological Utopianism. In Kling, R. (Ed.) *Computerization and Controversy: Value Conflicts and Social Change.* 2nd ed. San Diego, CA, Academic Press.

Jirotka, M., Proctor, R., Rodden, T., Bowker, G. (2006). Special issue: Collaboration in e-Research. *International Journal of Computer Supported Cooperative Work,* 15(4). p. 251-255.

Jirotka, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., Hinds, C. and Voss, A. (2005). Requirements for Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case Study. *International Journal of Computer Supported Cooperative Work,* 14(3). p. 369-398.

Kling, R. & Iacono, S. (1988). The Mobilization of Support for Computerization: The Role of Computerization Movements. *Social Problems,* 35**,** 226-243.

Kling, R. & Iacono, S. (1994). Computerization Movements and the Mobilization of Support for Computerization. In Starr, L. (Ed.) *Ecologies of Knowledge.* SUNY Press.

Lessig, L. (1999). *Code: and other laws of cyberspace*. Basic Books.

Meyer, E. T., & Dutton, W. H. (2008, 8-11 Sept). Top-Down e-Infrastructure Meets Bottom-Up Research Innovation: Fitting e-Social Science Visions to the Realities [Electronic version available at: http://ssrn.com/abstract=1262211]. Paper presented at the UK e-Science All Hands Meeting, Edinburgh, UK.

Meyer, E., Schroeder, R., Dutton, W. (2008). The Role of e-Infrastructures in the Transformation of Research Practices and Outcomes. *iConference 2008*. http://www.ischools.org/oc/conference08/pc/PA10-2_iconf08.pdf. [accessed 5 September, 2008].

National Science Foundation, 2007. *Cyberinfrastructure Vision for 21st Century Discovery*. Available at: http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf [accessed 12 June, 2008].

Price, D.J.d.S. (1963). *Little Science, Big Science*. Columbia University Press.

Procter, R., Borgman, C., Bowker, G., Jirotka, M., Olsen, G., Pancake, C., Rodden, T., Schraefel, M.C. (2006). Usability research challenges for cyberinfrastructure and tools. *Proc. of ACM CHI 2006, Workshop on Human Factors in Computing Systems*, Vol. 2, 1675-1678.

Schroeder, R. & Fry, J. (2007). Social science approaches to e-Science: Framing an agenda. *Journal of Computer-Mediated Communication*, 12. p. 563-582.

Smarr, L. & Catlett, C. (1992). Metacomputing. In Berman, F., Fox, G. & Hey, A . (Eds.) *Grid Computing: Making The Global Infrastructure a Reality*. Wiley, 2003, p.825-835.

Venters, W., Zheng, Y., Cornford, T. (2007). Collaborative Construction of Grid: Usability Within Particle Physics. *Third International Conference on e-Social Science*. October 7-9, 2007, Ann Arbor, Michigan.

Weinberg, A. (1961). Impact of large-scale science on the United States, *Science*, 134. p. 161-164.

Welsh, E., Jirotka, M., Gavagan, D. (2006). Post-genomic science: cross-disciplinary and large-scale collaborative research and its organizational and technological challenges for the scientific research process. *Philosophical Transactions of the Royal Society. A*, 364. p. 1533-1549.

Woodward, P. (1996). Perspectives on supercomputing: three decades of change. *Computer*, 29(10), p. 99-111.

Woolgar, S. & Coopmans, C. (2006). Virtual witnessing in a virtual age: a prospectus for social studies of e-Science. In Hine, C. (Ed.) *New Infrastructures for Knowledge Production: Understanding e-Science.* London, Information Science Publishing. p. 1-25.

Zittrain, J. (2008). *The Future of the Internet--And How to Stop It*. Yale University Press.