

Talking 'bout a revolution:

Framing e-Research as a computerization movement

e-Research '08 Conference
September 2008

Grace de la Flor
Eric T. Meyer

grace.de.la.flor@comlab.ox.ac.uk
Oxford University Computing Laboratory
Wolfson Building
Oxford OX1 3QD

- Introduction
- Framing e-Research as a computerization movement
 - Kling & Iacono's computerization movement framework:
 - Enables us to explore an alternative analysis of the e-Research vision
- Tracing the origins of e-Research
 - The origins of the vision:
 - Tracing the historical trajectory of advances in scientific computing which led to Grid technology
- A new science & engineering research agenda
 - Public discourses:
 - Core ideas about the ways in which Grid infrastructure would transform research
- Reconfiguring practices in domain research & access to infrastructure
 - Reconfiguring activities:
 - Domain practices and developer-user access to infrastructure

Introduction

- Building a scientific research infrastructure under the banner of e-Research

- e-Research

"...global collaboration in key areas of science, and the next generation of infrastructure that will enable it." – John Taylor

- Socio-technical process

- Multidisciplinary: computer scientists & domain scientists
- Propose changes to the way domain research is conducted
- Technological innovation, public policy vision and organizational practices
 - We examine this process broadly and from a critical perspective

- Revolutionary change?

- Widespread adoption enabling an orientation toward large-scale / 'big science' research

- Big science

- A key driver of the e-Research vision
 - Provision of complex, large-scale technical infrastructure to mobilise global collaboration using large research teams, co-located and distributed, to work together on common research problems (Welsh et al, 2006)

Framing e-Research as a computerization movement

- "A kind of movement whose advocates focus on computer-based systems as instruments to bring about a new social order" (Kling & Iacono, 1988: 228)
- communicate key ideological beliefs about "what computing is good for" and how project participants "should manage and organize access to computing" (Ibid: 227)
- "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)

Technological action frames in e-Research?

"simplify and condense elements of complex technologies and their potential use...[enabling]...groups of people to interact about what they might mean" (Iacono & Kling 2001)

1) 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)

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

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

Tracing the origins of e-Research

- A 'technology-led initiative'

- indicating an imbalance at the early phases of large-scale projects where "technologies have been developed without taking social aspects into account" (Schroeder & Fry, 2007: 567)
- Technological visions developed far in advance of implementation & give meaning and direction 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)

- Three factors that may have contributed to the development of the e-Research vision

- Big science
 - Scientific computing research
 - Metacomputing
- 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.

Supercomputing and big science

- Big science

"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).

- *in silico* experiments

- Complex numerical models to test predictions or to analyse very large amounts of data
- Results are investigated through the use of simulations and visualisation tools

- Supercomputing

- Provide high performance computing and mass data storage capacities
- Supports big science research
- NSF five national supercomputing centers (Princeton, San Diego, Illinois, Cornell, Pittsburgh)

"past efforts in supercomputing and high-performance networking are being subsumed into a broader, integrated vision of a more capable, ubiquitous, and accessible cyberinfrastructure" (Freeman et al, 2005: 682)

Metacomputing and the Grid

- A new set of protocols for network architecture
 - Transform networked access to information and tools; analogy of change (De Roure et al (2003: 90))
- 'publishing paradigm'
 - data can be shared in a distributed manner as images, documents and software tools
- 'interactional paradigm'
 - Instruments and facilities at remote locations could be accessed directly
 - Grid protocols make possible the virtual consolidation of resources through a network

"ultimately, one would hope that the Grid will be the operating system of the Internet" (Berman et al, 2003: 40)

A new science & engineering research agenda

- Vision documents

- **US:** *Revolutionizing Science and Engineering Through Cyberinfrastructure* (Atkins et al, 2003)

- **UK:** *The UK e-Science core programme and the Grid* (Hey & Trefethen, 2002) & *e-Science and its Implications* (Hey & Trefethen, 2003a)

- Establishment of a research programme

"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" (Atkins et al, 2003)

"indeed a further revolution – in how we create, disseminate, and preserve scientific and engineering knowledge"(Atkins et al, 2003)

"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).

e-Science: managing the data deluge

- e-Research: support research goals. Grid technologies are presented as a means for fostering distributed research collaboration, sharing of data and remote instruments, enabling access to heterogeneous datasets and providing high-performance computing power and mass storage capacity

"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" (Hey & Trefethen, 2003b: 812)

"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)

"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" (Berman & Brady, 2005: 5)

- 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 in mapping the computational requirements for the physical sciences directly on to the social sciences and humanities (Carusi & Jirotko, 2007; Meyer et al, 2008)

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)

- Challenges for the social sciences

- 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

"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).

- "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" (Star & Ruhleder, 1996: 114)

- Can Grid infrastructure be used in a **natural & ready-to-hand fashion**?

- Challenges of use:

"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).

- Avoidance of use:

"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).

So what exactly is the anatomy of the Grid?

- 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).

- 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" (p. 202).

"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 & does not "require computing experts to use" (Hey & Trefethen, 2003a: 1813)

From black box to glass box

"visibility and inspectability of available resources (processing power, available memory, network bandwidth etc.)" Büscher et al (forthcoming)

"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" (Ibid)

- 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.

- recalibrating the e-Research vision?

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

e-Research '08 Conference
September 2008

Grace de la Flor
Eric T. Meyer

grace.de.la.flor@comlab.ox.ac.uk
Oxford University Computing Laboratory
Wolfson Building
Oxford OX1 3QD