Digital Humanities Quarterly Special Cluster on Arts and Humanities e-Science

Stuart Dunn <stuart_dot_dunn_at_kcl_dot_ac_dot_uk>, Centre for e-Research, King's College London
Tobias Blanke <tobias_dot_blanke_at_kcl_dot_ac_dot_uk>, Centre for e-Research, King's College London

Abstract

This is the introduction to the special cluster on Arts and Humanities e-Science.

In the UK, most definitions of e-Science begin with that posited in 2000 by Sir John Taylor, then Director General of Research Councils, with the launch of the e-Science Core Programme: "e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it." Concurrently, in the US, similar concepts describing computational mechanisms for dealing with "big data" from "big science" have attracted the term "cyberinfrastructure". The resulting definition of e-Science is generally well understood in the communities concerned with it, even if they also considered it somewhat artificial and "top-down". However, it was not until 2005, when the Arts and Humanities Research Council (AHRC), Joint Information Systems Committee (JISC) and Engineering and Physical Sciences Research Council (EPSRC) launched the Arts and Humanities e-Science Initiative, that a truly domain-specific focus emerged. This background included a major AHRC-funded report, the *Arts and Humanities e-Science Scoping Survey*, which, having conducted detailed research across seven arts and humanities subject domains, defined e-Science as "the development and deployment of a networked infrastructure and culture through which resources — be they processing power, data, expertise, or person power can be shared in a secure environment, in which new forms of collaboration can emerge, and new and advanced methodologies explored" [Anderson 2007]. The arts and humanities e-Science Initiative is deploying a range of activities, and a highly targeted two-stage funding programme, to explore what might be meant by "Arts and Humanities e-Science". This special cluster for DHQ attempts to further that exploration.

Arts and humanities e-Science has proven not only difficult to define, but also a paradoxical coinage. This is rooted in an etymological language distinction. The English language distinguishes between the "humanities" and "science" in a way that reflects a specific treatment of the division of labour in research; while in German both are *Wirssenschaften* , and in French both are *sciences* . In English, science is not just a general state of knowing, like philosophy to the Ancient Greeks; it is linked to specific "scientific" methods, which are mainly found in "hard" or "natural" sciences — a pattern reflected very clearly in the first five years or so of e-Science itself. Such methods are normally *terra incognita* to the communities inhabited by humanists and practice-led researchers in the arts. This makes "e-Science in (or for) Arts and Humanities" a paradox, which is not helped by the general confusion about what "e-Science", in itself, might be. This paradox led to a lively discussion in the early stages of the Initiative as to whether the methods and technologies of e-Science as applied to the arts and humanities should have been repackaged as "e-Research". This term, it was argued, might better reflect the fact that the appeal of those methods and technologies for (e.g.) collaboration, or dealing with fuzzy and/or incomplete data, does *not necessarily* have to extend to dealing with terabyte-scale datasets or massive intercontinental computational processing, as hitherto associated with e-Science and its traditional "big science" constituency. The term "e-Research" was resisted, however, on the grounds that arts and humanities e-Science is about using advanced technology to tackle arts and humanities research questions: technology that originates from far outside the arenas familiar to arts and humanities researchers, even those traditionally involved with ICT (although, as discussed below, a number of UK institutions have had to recognize the difficulties this presents for

interdisciplinary collaboration when setting up dedicated research institutes). The increasing use of so-called Web 2.0 applications in advanced research, and an emerging debate in the broader e-Science world about the role of distributed technology and its relationship with the user, has given extra significance to this question (see also [Dunn 2009]). This significance is further explored in this issue in our interview with David De Roure, one of the key figures in this debate.

The matter of terminology is related to a further major issue, which makes this special cluster for DHQ an important opportunity for intervention: the place of e-Science within existing domain-centred ICT communities of practice. Without addressing the debate about whether or not "Digital Humanities" is a discipline in its own right (see, e.g. [Svensson 2009]), there can still be no doubt that it has a strong intellectual tradition stretching back nearly half a century. It has professional organizations (the Association for Computing in the Humanities, the Association of Literary and Linguistic Computing, the Society for Digital Humanities / Société pour l'étude des médias interactifs); its own journals (e.g. *Literary and Linguistic Computing*, *Digital Studies*, and indeed *Digital Humanities Quarterly*); its own conferences (e.g. Digital Humanities and, in the UK, Digital Resources for the Humanities and Arts) and its own research institutes on both sides of the Atlantic. Despite this, we feel that a high-level methodological conversation between arts and humanities e-Science and Humanities Computing has yet to happen — notwithstanding the essential contribution made at project level by several researchers from humanities computing backgrounds, who have made the leap to become early adopters of arts and humanities e-Science. The problem is that this should be, at most, a step rather than a leap. But there remain major gaps in understanding on both sides. As researchers of the Arts and Humanities e-Science Support Centre, the principal contact point for arts and humanities e-Science in the UK, and as co-editors of this issue, we hope the following articles will facilitate and inform that conversation. We must point out that this problem is not confined to the arts and humanities. A recurring theme of the All Hands Meeting, the UK's principle annual e-Science conference, is how to convince (e.g.) particle physicists who are more interested in particle physics than the infrastructure needed to deal with particle physics data, that engaging intellectually with e-Science is a valuable exercise. The arts and humanities must address this in a way that works for the arts and humanities.

E-Science is often directly associated with specific technologies, namely grids and service-oriented architectures (SOA). The grid is the vision of an integrated network infrastructure for the coordinated sharing of resources and problem solving in distributed environments. Highly flexible resource sharing and computer-supported cooperative work (CSCW) will be established via virtual organizations, which define rules and policies for accessing resources. Generally speaking, there are two types of grids: computational grids and data grids. Computational grids provide the scientists with high-performance nodes that can support computing-intensive tasks such as large simulations with a huge space of unknown conditions. Examples are discussed here in Melissa Terras' discussion of research on census holdings with high-performance computing. Data grids combine distributed data resources to permit researchers to share data and work transparently across data sets, as described in Mark Hedges' article in this cluster. Furthermore, data grid managements systems like Storage Resource Broker (SRB) frequently recur throughout this cluster, as a common example of how to realize data grids for distributed file storage. But, as we argue above, and as several of the contributions reiterate, e-Science is not just about using grids. We have found that a massive range of technologies are used in e-Science projects: frequently, all these have in common is that they share the somewhat loose label of "advanced networking technologies".

In this context, it should be noted that research communities, and in fact industry too, have committed themselves to realizing large e-infrastructures based on grids. IBM, Oracle and Sun have committed to create grid-compliant infrastructures. The Open Grid Service Architecture (OGSA) combines Web services and Grid computing on a large scale. E-Science attempts to create e-Infrastructures and support the work within them by building portals to access grid services or organizing training on e-infrastructure services. These types of activities are confined to experts in the field. Just as universities in the UK started to invest in enhancing their e-learning and e-teaching capabilities, a new type of carrier path is developing. E-Research professionals work on the translation of specific research needs into a research infrastructure to enable these. In the past, they were often based at centres of excellence across the UK, sometimes called regional e-Science centres such as e-Science Central in Newcastle http://www.neresc.ac.uk/ or the Welsh e-Science centre in Cardiff http://www.wesc.ac.uk. A more recent development is the emergence university-based e-Research institutions, such as the Oxford e-Research Centre, or the Centre for e-Research at King's College London.

These centres are called e-Research rather than e-Science institutions to open up towards academic domains such as the arts and humanities, which have not yet been served by developments in e-Science, as generally accepted up until now (see above).

We suspect that arts and humanities research communities themselves will be less interested in the development of professionalized e-Research institutions and groups, which mainly affect information and computing professionals. Arts and humanities researchers are interested in the grassroots activities of other researchers, that attempt to solve similar problems or inspire different approaches to existing research questions. It is such grassroots research in arts and humanities e-Science that is mainly represented in this special edition. In its latest form, e-Science stands for "enabling" science, which can be rightly criticized as relatively unspecific and too broad, lacking clear boundaries. This new definition presents yet another barrier to humanists used to research processes which are centred on the dissection of definitions. However, as often the case in research, when the original definition is rather blurred, this can also be an advantage. It allows for a wide range of grassroots activities to occur. This again is an advantage the arts and humanities, for which research practices in the UK are often only linked by the entirely external reason that they are supported by the same source of funding, usually the AHRC. "Enabling" philosophical debate is completely different from "enabling" (for example) archaeological research, or practice-led research. [6]

This cluster is therefore a collection of ideas, stemming mostly from the early stages of e-Science in the arts and humanities, where comparatively small research grants support ad-hoc experimentation. The first phase of activity to be supported by the UK arts and humanities e-Science programme (many of whose projects are presented in this volume) can be seen as a pioneering phase, in which ad hoc experiments by early adopters have led the way [Blanke et al. 2009]. The challenges of the first phase were to prove that e-Science in the arts and humanities can spark interest among researchers. At the same time, it also confronts established methods in computing with new challenges specific to arts and humanities data, and the research methods needed to analyse it. [7]

The activities within the UK's arts and humanities e-Science community demonstrate the specific needs that have to be addressed to make e-Science work within these disciplines. The early experimentation phase delivered projects that were mostly trying out existing approaches in e-Science. They demonstrated the need for a new methodology to meet the requirements of humanities data: data that is fuzzy and inconsistent, the result of human effort rather than automated production. It is fragile, and its presentation often difficult, as (for example) data objects in the performing arts that only exist as an event. At the same time, wider adoption of computational tools mean that we find more and more data of this kind readily available for computing in digital form. New research questions will develop out of the ability to search archives differently, and to make new connections across data sets. We can see how e-Science in the arts and humanities has matured towards the development of concrete tools that systematically investigate the use of e-Science for research. [8]

Works Cited

Anderson 2007 Anderson, S. *The Arts and Humanities and e-Science: Scoping Survey Report and Findings*. 2007. http://www.ahessc.ac.uk/scoping-survey.

Blanke et al. 2009 Blanke, T., M. Hedges and S. Dunn. "Arts and Humanities e-Science--Current Practices and Future Challenges". *Future Generation Computer Systems* 25: 4 (2009), pp. 474-480.

Dunn 2009 Dunn, Stuart. "Dealing with the Complexity Deluge: VREs in the Arts and Humanities". *Virtual Research Environments: Issues and Opportunities for Librarians* 27: 2 (2009), pp. 205-216.

Svensson 2009 Svensson, P. 2009: "Humanities computing as Digital Humanities." *DHQ*, Summer 2009, Vol 3, Number 3. http://digitalhumanities.org/dhq/vol/3/3/000065/000065.html.