

## E-Science (E-Research) Expert Seminar: Report on Proceedings

Sheila Anderson, 26<sup>th</sup> May 2004

### Executive Summary

The AHRB e-science policy seminar aimed to identify the role that the outputs from the e-science programme and grid technologies might play in supporting arts and humanities research and teaching. The seminar was held on the 28<sup>th</sup> April at Senate House in London. The seminar comprised a small gathering of scholars from the arts, the humanities, computing science, and information science who are experts in their field, with an excellent understanding of the use and potential of ICT in scholarly research and communication. At least four of the participants are involved in e-science projects and the majority are actively engaged in research and teaching.

This report summarises the issues, challenges and promise of e-science for the arts and humanities as identified during the seminar, and provides a set of recommendations intended to feed into the development of the AHRB ICT strategy and to assist in framing any potential arts and humanities e-science programme activity.

It was agreed that the term e-science was increasingly being replaced by the term e-research, and that this was more appropriate for the arts and humanities. Therefore, the term e-research will be used throughout the remainder of this report.

The seminar reached four broad conclusions:

1. That the e-science programme arose from scientists identifying challenges that could not be dealt with using current technology. In a similar vein the arts and humanities communities need to identify the 'grand challenges' that are particular to arts and humanities research, and to assess how use and development of technology might contribute to solving these challenges.
2. That an arts and humanities e-research agenda should be embedded in research practice and research needs - a research based response to problems which could not be solved with existing computing facilities or technologies.
3. That it was essential that the arts and humanities research community engaged with e-research, and that in the first instance a small selection of demonstrator projects should be identified and taken forward.
4. That understanding of and awareness about e-research generally was low and any e-research agenda for the arts and humanities must address this challenge.

Within this context the group identified a number of recommendations for consideration, both for specific projects, and also as an aide to developing an arts and humanities e-research strategy, and these are summarised at the end of this report.

## Summary Report on the Meeting

After a formal welcome and introductions, David Robey explained the context in which the seminar was taking place, and provided an overview of the AHRB ICT Programme, including details of the Methods Network, and the JISC funded Awareness Programme. He requested two key outcomes from the meeting: first, a set of recommendations for small strategic projects that might be funded through the ICT programme; and second, input into the forthcoming AHRB ICT Strategy and the place that e-research might take.

Sheila Anderson then set the meeting within the context of the wider e-science agenda. She identified the current e-science agenda as being about:

- Using technology to do what couldn't be done otherwise
- Enabling global collaboration
- Grid technologies to provide access to distributed resources of data, computing power, storage space, and applications
- Utility computing to provide integration of remote heterogeneous data, automated capture of metadata, real time data capture
- Dealing with complexity - access and finding tools, semantic web and shared vocabularies, ontologies
- The Access Grid to provide video communications, virtual networks and collaborations

She invited participants to identify if and where this agenda might be mapped onto the arts and humanities, and to seek to identify specific recommendations for a nascent arts and humanities e-research programme.

It was noted that the e-science programme had arisen in response to clearly articulated needs from the science community. The impact of the science 'data deluge', increasing globalisation of research, and increasing costs associated with providing the kind of computing power and access to and management of scientific data had propelled the science community to seek solutions using the power of the Internet and potential of technology. It identified that partnerships between scientists and computer scientists was necessary to achieve this vision and set about creating partnerships that would drive this agenda forward.

The challenge for the arts and humanities research communities is therefore to identify its 'grand challenges' – to ask what researchers want to do that they can't currently do, to ask where the technology fails them, and to articulate how solving these problems would drive forward the research agenda in the arts and humanities. Participants agreed that it was essential that any e-research activity for the domain should be driven by clearly articulated research needs and be seen as a response to these research-driven needs, rather than driven by what the technology might be able to do. Participants also agreed that it was essential to articulate what is unique and problematic for these disciplines, and to identify an arts and humanities agenda separate from that of the scientists.

Fundamental to identifying the 'grand challenges' is the nature and substance of much arts and humanities research. At the heart of this is difference in method – science tends towards testing a hypothesis against a set of data, whereas the arts and humanities are about criticism and meaning, about interpretation, re-interpretation, and extracting meaning. In order to undertake this kind of analysis 'deep' access into the content of resources is required, often at the level of individual words or phrases, or images or sounds. Moreover, this level of access may well be required at the 'locate' stage of the process in order to assess the relevance and usability of a particular resource.

This difference is further exemplified in the nature of the materials with which arts and humanities scholars engage - much of the raw material of research is complex and partial, with the consequent challenge of finding technologies that are able to deal with complexity, fuzziness, and incompleteness. In addition, scholars frequently work with data sources that were created for a different purpose or audience, and over which they have had little or no control in their creation.

The 'grand challenges' are therefore to understand how to locate, access and integrate the content of highly distributed resources that are likely to be incomplete, fuzzy, and complex, and to have been encoded using different standards, described using different standards, and be of variable quality.

Participants argued that there remained a great deal to be done to enable scholars to locate the resources that might be useful for their research, to assess the usefulness and provenance of those resources, and then to be able to download and use them. Technologies existed that enabled some these processes but they lacked sophistication and much further work needs to be done. Issues to be addressed include:

- Integration of disparate and diverse resources
- Provenance
- Common standards
- Shared ontologies
- Common encoding
- Quality assessments and markers
- Security, authorisation, payment etc.

There was some discussion about the need for increased computational power to cope with ever-increasing volumes of data, high resolution images, sound, and moving images. Participants argued that this is likely to prove increasingly important, particularly to enable provisional 'what if' and iterative analyses.

Participants also identified a need for better, shared tools and software to push forward the boundaries of analysis. Researchers, particularly in the arts, often found that existing tools were not sufficient to enable them to follow through their analysis. The development and implementation of tools and software that enabled

interrogation and exploration of resources, and that formed a key part of the research process, rather than as tools to search, sort and retrieve, would be key to the further integration of ICT into arts and humanities research. It was argued that an arts and humanities e-research agenda could play an important part in driving this issue forward because it could concentrate on methods and processes and the technologies required to support these, rather than focusing on research outputs in the more traditional sense.

Following lunch David de Roure summarised the current e-research programme under three main headings:

- Data grid – data management and sharing
- Collaboration – access grid
- Computational grid – software applications

He argued that e-research is about joining things up – not about CPU power and networking. He suggested that the semantic grid developments might be most appropriate for the challenges identified in this seminar, and that semantic web / grid integration holds most promise. This would require community shared vocabularies.

Discussion then took place about the extent to which the arts and humanities might use grid technologies in particular, and technology in general to assist in resolving some of the 'grand challenges'. Discussion ranged across a wide range of topics including the creation of ontologies and grid based language engineering; community building; assigning meaning and whether this is something that might become an automated, or semi-automated process; annotating visual and textual resources; the challenges of providing good metadata; analysis and access issues, and many more.

It became clear that articulating the grand challenges for arts and humanities research, although not easy, was significantly easier than translating those needs into concrete examples of how technology might be used to help resolve those challenges. The attempts to achieve this were also undermined by a lack of meaningful understanding about current e-research projects, and the range and value of tools and other outputs arising from them that might be used by arts and humanities scholars.

## Conclusions

The seminar concluded that drawing up an e-research agenda for the arts and humanities would not be an easy task. Awareness of current e-research projects was low and how it might be applied within the arts and humanities still confused. What might be done and how was still an unknown, and confusion about what was unique or distinctive about e-research was still an issue. A general lack of understanding about what e-research might contribute to arts and humanities research and how it might help push forward the boundaries of research was still uncertain.

However, despite these misgivings participants at the seminar came to the overall conclusion that the arts and humanities must engage with e-research and develop an agenda specific to the domain. The key is to develop an agenda that would be seen to be of real benefit for scholarly research, and to be seen to be driven by research needs rather than the technology. It was strongly recommended that any projects or demonstrators should relate to real-life problems and issues, and be based on use-case scenarios.

Moreover, scholars needed to perceive how the application of advanced ICT could lead to new types of questions, and new forms of knowledge. To do that will require sharing of resources, expertise and knowledge, new forms of collaboration within and across domains, and a secure and reliable infrastructure to underpin new initiatives, new projects, and new forms of working.

## Recommendations

The seminar identified a number of projects that might form a part of an arts and humanities e-research agenda for consideration by the AHRB ICT Programme. It was also suggested that it would be useful to explore collaborative funding models with other research councils, and potentially European Union funding. Projects for consideration include:

1. A Scoping / Mapping Study: in order to overcome the lack of awareness and to start the collaborative process, participants strongly recommended that a scoping/mapping study be carried out to:
  - a. Gather information on e-science projects,
  - b. Identify key features and outputs,
  - c. Match with issues and challenges in arts and humanities
  - d. Dating agency: matching computer scientists with researchers
2. Virtual Network (expanding point d. above):
  - a. Establish forum for bringing together arts and humanities scholars, computing scientists and information professionals, (perhaps through the methods network). This should be used to incubate new projects, new links and alliances, share expertise and knowledge.
  - b. Establish links with e-social science centre to discuss possible link-up – explore the idea of an arts and humanities node?
  - c. Research common issues with qualitative social science
  - d. Use as awareness raising forum
3. Test-bed implementation of grid infrastructure for managing, accessing and delivering widely distributed resources:
  - a. Identify use-case scenarios linked to real research questions
  - b. Investigate use of globus, SRB, grid services
  - c. Test implementations
  - d. Potentially a joint AHDS/Methods Network project
4. Investigate use of the semantic web: perhaps a scoping study that might investigate:

- a. The potential for developing community ontologies
  - b. The use of semantic web technologies to create mappings between ontologies
  - c. Test against real community problems
  - d. Maybe in the area of language/linguistics or archaeology/heritage?
5. Investigate automatic annotation of film, video, and sound
  6. Investigate automatic capture of representation of practice-based arts research process – research process involved in producing a piece of work
  7. Develop a humanities encoding initiative (similar to the TEI for texts) and associated tools for its use.