

# Seminar report: e-Science and Literary and Textual Studies

## Seminar contributors

Peter Robinson [chair] and Barbara Bordalejo (University of Birmingham); Lorna Hughes (ICT Research Methods Network); Tobias Blanke and Stuart Dunn (AHSSC), Luke Blaxill, Sheila Anderson, Sanjay Vivekanandan and Katrin Weidemann (AHDS); Peter Ainsworth (University of Sheffield); David Robey (University of Reading/AHRC); Alan Bowman, Martin Wynne, Charles Crowther, Ralph Schroeder (University of Oxford); Pamela King (University of Bristol)

The seminar was held at the AHDS Executive Offices at 26-29 Drury Lane on 31 May 2006. It was organized by Peter Robinson and Luke Blaxill.

## Summary

Two areas for priority action within e-Science emerged from the discussions of this group.

The first area was specific to literary and textual studies. Priority should be given to the development of *specific tools for collaborative research in literary and textual studies* within the e-Science framework. Particularly, we need tools for collaborative resource creation, comparison (that is: collation and alignment), analysis and annotation, and we need tools that any humanities scholar can use.

A second area is general to all humanities scholarship. It is the need for digital scholarship in the humanities to demonstrate its value. Priority should be given to wider action within e-Science for the development of *proofs of the value of digital scholarship*.

## Challenges, barriers, opportunities

This discussion ranged wide. Seminar participants referred to:

- The difficulties of making high-quality digital scholarship, against a background of competing interests within the worlds of libraries, commercial software and search engine companies, publishers and academic culture. An instance: to make a single digital facsimile of a single manuscript might mean lengthy and complicated negotiations with a library, which is itself wrestling with negotiations with (for example) Microsoft or Google, which might affect the reproduction rights for the manuscript. There is a need for collaborative partnership (or at the very least, dialogue) between scholar, curator, conservator, publisher and digital photography specialist. Such partnerships do exist (the Michigan/Oxford/ProQuest etc Text Creation Partnership comes to mind); we need more such.
- The difficulty of distinguishing what is high-quality digital scholarship from all the materials jostling on the internet, most of dubious intellectual quality. Students and beginning researchers may lack the tools to make this discrimination.

- The need for peer-review systems tuned to the digital world, able to help others make quality judgements about the resources they meet on the net.
- The need for positive, quantitative, evidence to demonstrate the value of digital-tools based research. The AHRC, and other bodies, have spent considerable amounts of money on digital scholarship, and are continually reviewing the effect of their funding. There have been remarkable instances of the success of digital scholarship, based on the application of new methods to old problems. Extraordinary new resources have been made. We believe that digital tools are changing scholarship dramatically, and changing it for the better. If that is true, we should be able to show this, very clearly. But so far we have not done so. We need to offer more than rhetoric and anecdotes.
- There is a gulf between a few who are able to use advanced digital tools to make digital scholarship (usually, with exceptional access to infrastructure such as at Kings, Oxford or Birmingham) and the many scholars whose digital skills extend only to word-processing, email and web browsing. Many more scholars need to be able to produce digital scholarship; too few can. (A later meeting of some seminar participants identified another gap: many libraries in key positions to undertake and advise on digitization lack necessary expert knowledge.)
- An immense investment has been made in two areas: infrastructure (AHDS, etc) and digital resources (e.g AHRC resource enhancement). Yet very little investment has been made in specialist digital tools for humanities scholars. Accordingly, there is a large gap between the tool makers (of whom there are very few) and the tool users (potentially, every humanities scholar). We need tools every humanities scholar can use. We don't have them. Further, good tools need real long-term funding: the short-term research project funding model just does not fit this.
- Interdisciplinarity can really, really work: participants cited the Vindolanda tablets; the use of evolutionary biology methods in the Canterbury Tales manuscripts; the new modes of working induced by models from different disciplines, such as computer science and business studies.
- Digital scholarship can not be done by one scholar alone: it must be done in partnerships, among humanists as well as among humanists and specialists in other disciplines.

Overall: there was a strong sense that extraordinary scholarship has been produced in literary and textual studies using digital methods, and that we are only at the beginning of the transformation that digital methods may bring to us. The discussion identified two barriers as standing in the way of achieving this transformation:

- The lack, as yet, of firm evidence of the transformation of scholarship by digital methods. We have some superb examples and case studies. These do not, as yet, amount to a fully articulated argument. Related to this: the lack, as yet, of accepted and appropriate means of distinguishing good digital scholarship from bad.
- The small number of scholars able to use digital tools to make high-quality scholarship. The term 'coffee-table projects' was used: because so few scholars can do this work, what they do is isolated, and outside the mainstream of scholarship in terms of publication and esteem (which remain resolutely wedded

to print). This feeds back to the first problem: because (as yet) digital scholarship is limited to a very few, it has difficulty achieving due recognition from the many. It also greatly limits the possibilities for collaboration, essential to digital scholarship: if one's collaborators need a high level of skill in digital tools, then one will have very few collaborators.

Finally, one should note the emphasis in the day on collaborative digital scholarship, clearly perceived as an area of high priority and promise, with particular focus on the opportunities offered by digital resources. There was mention, but little discussion, of other areas where over the past decades literary and textual scholars have used digital methods: authorship/stylistic studies; text pattern discovery. The clear consensus of the group was that at this moment in time, digital resources are the key factor in the interaction of digital methods and literary and textual studies.

### **The potential of e-Science**

One may summarize e-Science in terms of three grids: each (potentially) relevant to the account of the challenges facing literary and textual studies presented above:

1. The *computational* grid. This is relevant where very large amounts of data (e.g. weather data) require very rapid processing. The whole set of digital data created by humanities researchers in the last decades is such a large amount of data. Up to now, it has been processed relatively slowly, and in piece-meal project by project fashion. But one can imagine, as individuals increasingly work to common standards and as they increasingly require rapid access to the work of others, that the computational grid might offer the power (for example) to let a large group of researchers continually update the indexes and databases on which they are working together.
2. The *data* grid. This is obviously relevant to our disciplines, and to the collaborative model essential to useful work. We need access to each other's work; the data grid can provide this.
3. The *access* grid. This allows high-quality conferencing for partners working far apart: as close to physical meeting as technology can allow. This can give projects far better discussion than is available via the monochrome of email, far more frequently than can be managed by physical meeting.

Underlying these grids, and so the whole e-Science endeavour, is *collaboration*: not as an occasional necessity, but as the crucial precondition for useful research. To many scholars in Literary and Textual Studies, collaboration is an alien concept: but to those of us who engage in digital scholarship, it is as fundamental as it is to physicists or biologists.

There are several well-advanced instances of digital humanities projects which exemplify e-Science principles. These include:

1. DAM-LR (Distributed access management for language resources): a collaboration of Dutch institutes for language research
2. TextGrid: A German initiative to build a grid-enabled workbench for philological research, with unified access and resource creation tools

### 3. Armadillo: a system for intelligent cross-searching of disparate scholarly resources

The seminar chair presented an outline of projects in which he is working which could benefit from e-Science tools. These projects involve various groups of scholars collaborating on scholarly editions in digital form: making digital transcripts from digital images from manuscripts; collating the digital transcripts; correcting the transcripts and comparisons; analyzing the results of the collations; publishing all this. Because of the scale of the work, there is a need to open it up to others beside the funded academics at the core of the projects, so that (for example) knowledgeable individuals anywhere might contribute transcripts, collations and analyses. But there is a need too to guarantee the quality and integrity of the work we do. This means the projects must be both open and secure: open, when they need to be open to admit the work of others; secure, when they need to be secure, to ensure that work is properly acknowledged and safe from accidental or deliberate defacement.

As an example of this: a library makes available a digital image of a manuscript. I make a transcription of a manuscript, and put it on the web: we need a structured method of ensuring that anyone who wants to, can look at my transcription and the original image. We need too tools which go beyond linking page by page, to linking letter by letter, and even below. Now, scholar B spots mistakes in my transcription. We need tools which enable B to publish his corrections, layered on top of my transcription in such a way so that others can read my transcript with B's corrections – and indeed, make their own corrections, on top of B's corrections. Further: B might be mistaken, so the tools must let us undo what he or she does; and the tools must keep an audit trail, so that at every point it is clear who has done what.

At the same time as this is happening: other people publish other digital images and transcripts of the same work in other manuscripts. Other scholars will wish to compare these various texts of the one work so published with one another, and to publish the comparisons, so that others may annotate, correct and supplement them, and publish these in turn: again, with the tools allowing us to undo or redo this work, and identifying precise responsibilities.

There is an obvious correlation between the collaborative tools needed to enable such work, and the possibilities of e-Science.

#### **Areas of action**

There was a marked shift in the workshop during the day, which can be summarized in the phrase: *from sharing to collaboration*. As the group learnt more about e-Science, it became clear that for e-Science, the sharing of tools and data are not ends in themselves, useful as they may be: they are the key measures for enabling collaboration. Scholarship is an immense collaborative endeavour, and even the classic lone humanities scholar works in a world of libraries and archives, which provide the fundamental materials for the research, and of publishers, which allow others to see the research. That lone scholar's research then becomes part of other's scholarship, in a slow but constant dance of collaboration across generations. Perhaps the greatest promise of e-Science is that it accelerates and intensifies collaboration: exchange of information, modification of

hypotheses, making the new hypotheses available to others, may happen in seconds, not years.

The group agreed, too, that the most fundamental shift in the use of digital methods in the humanities would come when all scholars who have reason to use digital methods (and this might be all scholars) have the ability to use them. Up to now, collaboration in digital scholarship has been limited to a very few scholars, typically those with privileged access to specialist advice and infrastructure. This severely limits the possibilities for useful work, and circumscribes the impact that digital methods could have on the humanities.

There is clear convergence, then, between e-Science and digital scholarship in the literary and textual studies realms. Tools for resource creation, comparison, analysis and annotation are fundamental to our work. But at present, these tools have been designed for the use of single scholars, or of single scholarly groups, working apart from others. We need a new generation of tools designed for collaboration: so that scholars may annotate, supplement or alter the work of others, and so that this new work be available instantly to others.

Accordingly, this group recommends priority be given to the development of *specific tools for collaborative research in literary and textual studies* within the e-Science framework: particularly, tools for collaborative resource creation, comparison (that is: collation and alignment), analysis and annotation. These tools must be so configured that any humanities scholar with reasonable computer skills can use them.

A second priority, emergent from the day's discussions, is the need for digital scholarship in the humanities to demonstrate its value. However, this is a problem general to all humanities digital scholarship, not just to literary and textual studies. We would support wider action within e-Science to develop *proofs of the value of digital scholarship*.