Expert Seminar

What do Humanities Researchers need from ICT?

Chairs: Harold Short and Marilyn Deegan

Introduction

The workshop began with a description by David Robey of the ICT Programme and the proposed ICT Methods Network, then ranged widely across many key issues in the use of ICT by humanities researchers, concluding with a number of recommendations for further work.

1. Database of Research in Progress and Methods Database

The AHDS currently has funding until December 2004 to produce a projects and methods database. This database records detailed information on recent UK-based projects using ICT and on the ICT-related methods they are employing. The database will include information on methods and projects, with details about implementation. The database will relate AHRB-funded research projects to a taxonomy of methods and for a limited number of these projects provide detailed information on the implementation of such methods. As such, it will already constitute a valuable information resource in itself. However, the database has the potential to be of even wider use and value to the community if it could be extended and further developed. The aim is to create a usable, workable and expandable source that people can use as reference to describe their own methods.

There was a great deal of discussion about what should be included in this: some felt that project applications to AHRB that failed should be included as well, since their "technical" failure may not correspond to an intellectual failure (given that there is often insufficient funding for many highly-rated projects). For the moment (1 year), the project will be limited to the UK only, but it may be extended later to other countries. After the first phase, there will be widespread consultation to explore how nbest to keep the resource running. One major issue will be to keep the database constantly up to date.

2. Users

It was clarified that what we mean by users in this context are humanities researchers in HE institutions, and the questions asked here were:

- What the groups of users are doing and NOT doing?
- The range of users is wide, but how sophisticated and how deep is their usage?

While the resources are primarily intended for scholars, there was a view that the results of their research must be made more widely available, even as far as being accessible to the general public as well as to the whole range of 'memory organisations' (libraries, museums, galleries, etc. Also, subjects of research often come from communities outside the academic world who are nevertheless engaged iin academic projects (local historians, for instance). Librarians, archivists and curators need to be involved in the projects as well.

Researchers / User groups

There are different layers of users, some of whom are also creators as well as users.

- What are the different areas of use?
- Who are we trying to help?
- Who are we aiming resources and projects at?
- How do we identify different users?
- What tools/software do users need?

Users were identified as including:

1. Resource creators (developers)
2. Project directors
3. Re-users (using a resource that already exists)
4. Tool users
5. Students (we need to keep in mind their specific needs)
6. Potential users (or current non-users) This may be the biggest category. It is necessary to create a market, to reach people that do not know about ICT methods (and there is a link here to the awareness and education programme).

7. Resource managers. This category includes librarians and information systems managers (who are included in other categories as well).

The three macro-categories that need special attention are the creators, the users/researchers and the managers, that is to say, respectively, the people that “make, use and mediate stuff”.

Use and Re-use

A dichotomy seems to characterise humanities research. Before the widespread use of ICT in the humanities, dissemination and reuse of resources was not an issue for humanities researchers. The primary concern was instead the originality of an unstudied and unpublished area of research. On the contrary, the concepts of re-using and re-packaging involve the sharing of resources. There were a number of comments questioning the real willingness of humanities researchers to share data and results, which was compared to practices in the scientific world.

The view was aired that there seems to be a distinction between projects that create resources for people to use and projects that do not have this aim, but which see traditional publishing as the route to the fulfilment of the researchers’ intentions. There was general agreement that there is a large amount of effort to be put into the process of re-packaging resources for re-use.

The key issues discussed on use and re-use included:

- Standards
- Especially in history and archaeology, there was felt to be a need to find standard ways of identifying persons, institutions, places (geographical entities) and events (times, dates, calendars).
- Metadata
- Inter-operability
- Documentation
- Data Enrichment: base data are distinct from the materials used to enrich the data. The materials used to enrich the data (e.g. linking, coding classification etc.) may be added in order to allow other scholars to use the base data.
- Tools/Software: this category deals with the re-use of technology. The main problem to be faced is the development of specific tools for a whole range of diverse projects.

3. Tools

Tools have two meanings in this context:

- Existing tools (e.g. text analysis tools; tools for using XML, etc.) It was noted that using a tool already available (developed by someone else) has the problematic consequence of lack of control of it. The problem is often the lack of information about existing tools: some studies are needed to identify these.
- Tools to be developed: it might be necessary to set up a collaborative tool development endeavour, analogous to but different from the project hosted by the University of Illinois to investigate and build tools for exploiting digital library content.

Some issues discussed in relation to tools were:

- Commercial tools
- Open source tools (that often need repackaging)
- Open middleware
- Re-packaging of available tools
- Standards and protocols
- Inter-operability
- Cross-domain tool investigation
- Tools/Methods/Projects: the relationship between these needs to be investigated thoroughly, especially for major scholarly projects using ICT. This brought the discussion back to the methods and projects database
- Lateral development/software houses: the games industry was mentioned as having valuable tools that might be further exploited
- The identification of technology deficit: where are the gaps?
- Tools registers: there are existing registers of useful tools that need to be investigated. A good example of this is the TEI which links to XML tools, projects etc.
- Tools and technologies watch: this needs to be undertaken, probably by the Methods Network.
- How does a tool fulfil intellectual goals? It is important to understand the potential use of a tool (issue of the relationship between tools and interfaces).
- Open source, open standards software developments
- Commercial exploitation? To recover some of the costs of development the commercial route could be exploited. Collaboration with the commercial world might provide a solution for funding providing the company does not see this as an exclusive right. The involvement

of Hewlett Packard in the development of MIT's D-Space, which is given away free, is a good example of a possible model.

- Documentation: recording the development process and methods
- Guides to Good Practice
- Formal methods (SSADM): this is part of any engineering project, and provides a check list of the standard principles that need to be followed when embarking on a formal project. It was suggeszted that there should be a research project to investigate these
- Evaluation – Benchmarking - Quality Assurance: there need to be tools for carrying these out. There is the need for tools for projects to evaluate themselves.
- Risk assessment
- There needs to be a research project to evaluate tool building for the humanities, perhaps a scoping study with an international dimension.

In conclusion, it seems that tools for the humanities are likely to be complex and diverse and to derive from many communities: commercial, open source, academic. To what degree should the AHRB be encouraging or funding tool building?

4. Collaboration: Interdisciplinary and Multidisciplinary

A distinction needs to be made between interdisciplinary and multidisciplinary. An example of a multidisciplinary project is the William Blake Archive, where many humanistic disciplines have been involved in the project, together with technical specialists. The surplus added value offered in an interdisciplinary approach consists of a genuine engagement of multiple disciplines on a collaborative basis. A truly interdisciplinary collaboration changes both disciplinary perspectives. AHRB does not have a good strategy for evaluating collaboration.

Other related issues to think about in collaborative contexts are:

1) Academic Recognition
Sometimes, even within the project team itself, people do not understand what is involved in giving true recognition to all contributors.

2) Career development of researchers
The academic recognition question is directly linked to the career development of researchers. What comes after a research project for ther researchers involed? Involvement in an ICT project can seem to be academic suicide, compared to a traditional academic career path. This is also

related to the contract researchers problem, where reesarchers move from one insecure contract to another. The suggestion was made that the AHRB might institute a system of credits, some sort of proof that people have participated in a research project in a defined role. One problem, however, is that the acceptance and accreditation of e-resources are not seen as equivalent to printed publications (though publication in e-journals seems to be no longer problematic).

Two things are needed: some guidelines on how to cite co-authors and co-developers of databases and other ICT products, and a method for the evaluation of such products, with a mapping to conventional research outputs.

3) Regular publications produced by projects

More regular output by the projects should be encouraged, so as to promote good practice in dissemination and to assure that the resources reach a high academic standard.

4) Peer-reviews of databases

Proper editorial control is needed. But it is hard to know who will assess and how (RAE)? The management process is complex for peer review, and in the case of databases it requires technical competence as well as intellectual competence.

5. Training

Institutional provision for postgraduates

For postgraduates, training in methods is fundamental. Currently, training is expected to be institutionalised rather than optional. Certain key issues were discussed:

- Who has the responsibility for training? Who is going to decide what to train?
- Training models for UK: Every institution and every discipline have their own needs. Hence, a distinction has to be recognised between introductory training and more specific or advanced ICT training. The various backgrounds of the students may come into play
- Course credit: when students recognise a certain status to a course, they will think it is worth attending.
- A study of current postgraduate training in ICT research methods in the arts and humanities is needed, and core curricula for training developed.

Training for Researchers and Supervisors

This could be provided via:

- Summer schools (people working intensively)
- Institutions
- Workshops (focused and specialised)

It has been observed that the training of postgraduates is more effective than the training of more established researchers. Researchers claim that they do not have time for taking courses. However, they find time to attend conferences, and so a proposed solution to this problem might be that of taking advantage of conference attendance, and offering courses attached to conferences.

6. Infrastructure

It has been noted that few institutions with a broad expertise in both computing and humanities (humanities in general, not a specific discipline) exist. This urgently needs to be addressed

7. Proposals for urgent funding

1) Database of tools, methods, people, high level equipment
2) Scoping survey of open source tools
3) Some developments of standards for markup and encoding that incorporate but extend beyond TEI. This would include the use of controlled vocabularies etc.
4) Scoping study on the universal identification of people, places, events etc.
5) Investigation into the use of formal methods as used by the software industry, for example the SSADM ((Structured Systems Analysis and Design Methodology) formal system.
6) Investigation into formal methods for knowledge representation, drawing on information science techniques and formalisms.
7) Investigation of formal methods for project analysis, design, and management.
8) Scoping study on user interfaces and user-centred design.
9) Investigations into legal issues of rights, ownership etc.
10) Demonstrator projects on all of the above, perhaps coming from different areas of the humanities (text-based, image-based, artefact-based, time-based).
11) Investigation/scoping study to find out what users are currently doing with ICT and digital resources. This could possibly be a repeat of the study carried out by the Office for Humanities Communication in 1991/2, which was itself a repeat of a survey done ten years previously. This could provide some useful longitudinal information.