

Digital Research and Editing Environments Workshop

Institute of Historical Research, July 7th, 2011

This recent workshop held at the IHR generated illuminating contributions from representatives from a wide variety of historical digitisation projects. Speakers included:

- Jane Winters (IHR) on British History Online and ReScript;
- Philip Schofield (UCL) on the Bentham project;
- Mark Hedges (KCL) on TextVRE;
- Rob Iliffe (Sussex) on the Newton project.

A video of the speakers' presentations may be found at: <http://livestre.am/Rkv7>

Following the speakers' presentations, delegates broke into four groups for an Ideas Cafe, each of which discussed one of the following four themes:

- Ongoing Engagement;
- Skills and Training;
- Generated Content/Authorship; and
- the Research Ecosystem.

These discussions informed the questions raised in the Open Discussion at the end of the workshop.

The feedback reports, written by the moderators of each of the Ideas Cafe tables are included below. Since the discussion included many and various voices, the reports do not represent fully worked-out answers to the questions posed; but the reports do give a useful sense of the opinions of a representative group of scholars.

Generated Content/Authorship

Summary of discussion

When texts can be continually updated and re-edited, what constitutes an “edition”?

- An edition is an entity. The base text has been, or should be, established. There might be many micro-iterations of that entity but defining exactly what constitutes an iteration is difficult. With an edition there is a sense of aiming for a finished product.
- It might be helpful to think of a code analogy: there is a software release, but the ability to rollback in order to see earlier versions/releases.
- There is the question of born-digital content: living authors might deposit material which is unfinished.
- Web archiving can be done as snapshots but there are cost implications, and some authors/institutions may be reluctant to consent to this.
- Online material may be transient: there is a Digital Dark Age, before archiving began.

How do you cite a source that keeps evolving?

- Use persistent IDs, but there has been a low take-up with these; it's not practically possible, either, to apply one for every small change to a source – you have to take an editorial decision. DOIs could be used, there is an encouragement to use DOIs for datasets now.

What issues arise for editors: sustainability:

- Preservation and sustainability are not the same thing. Editors may trust large institutions like the BL, or museums and libraries generally, more than university departments. However many users are simply unaware of the issues in this area and don't think about it.
- Funding councils need to do more here, as do parent institutions, but both of these things incur monetary costs. There also needs to be dedicated support for research IT – too many institutions see support as desktop support alone.
- Perhaps there needs to be a way of finding an electronic parallel to the book: once someone has published 300 copies of a book and had it distributed they don't worry about sustainability.
- There is now talk of “the death of the web”, as people move more and more to using apps. The issue of apps also raises questions of preservation and archiving. Apps and games may be the best way to do effective crowd-sourcing.

What issues arise for editors: protecting the integrity of one's work:

- Peer review and moderation are important. Reputation may be affected by the overall quality of the resource one is contributing to. The question is always, “who polices this?”

- Rating policies may be a good method (as on Ebay reviews or Amazon ratings of other people's reviews) but this requires a critical mass to be reliable.
- There needs to be a culture of web responsibility, moving on from netiquette to the idea of contributing responsibly.

What issues arise for editors: getting credit for one's work

- The sciences use a quantitative citation index; in the humanities there might be quality issues with such a metric.
- For collaborative work it can be hard to determine whose work is whose, or to prove it subsequently. In some forms of crowd-sourcing one can earn "tokens". Student work can earn credit in this way, but how should they be credited on the resource/publication and could this be exploitative?
- However there are other academic outcomes from projects as well.
- It could be included in the REF exercise.
- However there is simply a disparity of esteem. Academic culture just prefers print (at least at the present).

Skills and Training

Questions

What particular skills do researchers need to be able to contribute to online editing?

How does this differ from existing skills?

How much of this gap is due to lack of skills and training and how much with the design of the research tools?

What would be of most benefit to address this gap?

Summary of discussion

- At the moment there remains a great divide between academic and technical skills, with technical skills often denigrated as merely clerical, i.e. as inferior to academic skills. As more digitised resources become available, an integration of scholarly and technical skills will increasingly become necessary. Academics need to move away from a reliance on IT people.
- One reason people may see technical skills as inferior is that when we present our projects, our focus on the ease of use for end-users unwittingly downplays the complexity that goes into creating a resource.
- There is a gap between official, funding-body views of technical skills, and views amongst many academics. While funding bodies increasingly formally recognise the importance of the technical side of projects – and indeed, may insist upon such an element – anecdotal accounts of the experiences of PhD students indicate that often no credit is given for the work that goes into the technical side of such research. This can have a negative impact on whether/to what extent doctoral students make best use of computer-based technologies in their research.
- Economic pressures on traditional publishing models are likely to be an important element in changing attitudes towards the value of technical skills in scholarly research.
- There is often resistance among researchers (students as well as established academics) towards IT training. People tend to think that they already know how to use a computer, so why bother. One way to overcome such resistance is to focus on the practical applications of the training being offered, i.e. when

advertising training courses highlight the research that can be done rather than the technical process itself. In other words, don't call it skills training!

- Academics need more information on what technologies are available and how it can impact on their research. For historians, the IHR could facilitate this by flagging up new developments and linking to appropriate websites.
- Computer technologies have become embedded in everyday life. Learning new computer skills is an incremental process, rather than "big teaching".
- When managing a project, it is dangerous if only one person has a certain set of necessary skills. People change jobs or otherwise move on. Structure your project so that it doesn't collapse if one person is no longer there.
- For crowdsourced contributions, some form of gatekeeping is needed to ensure that contributors have the necessary skills. How this is done will vary between projects. Without such a process, the trustworthiness of the work is compromised.
- Online, evolving editions need to spell out the limitations of the resource to users to encourage them to do further research rather than relying on a particular edition as definitive. While this might at first seem a downside to online, collaborative works, there is actually a lot to be said for maintaining a healthy scepticism when viewing any resource. When a book is published in hard-copy, it feels more definitive and may lead people to accept its contents at face value. Online editions may actually lead to more widespread critical engagement of texts.

Ongoing Engagement

Overview

"This group is going to focus on the *deal* between the editor and the editorial project. In other words, what would constitute a mutually beneficial relationship?"

Questions

What obstacles are there to the adoption of collaborative editing? [*How can these be overcome?*]

What are the benefits of collaborative editing? [*How can these be made clear to potential editors?*]

How do we keep existing editors coming back? [*How do we recruit new editors?*]

Summary

- Considering Wikipedia as an example: Wikipedia is very popular with students, covers topics ignored in other sources, and can be used to make a point about authority. However, citing Wikipedia is unlikely in serious research and researchers will not contribute to Wikipedia until it achieves REF traction. In general, it was felt that the higher the price of the journals which publish your work, the more the recognition, and that Open Access or free work goes unrewarded
- How then to recognise and reward contributors' work ? There was potential to use Print-on-Demand to log/commemorate/acknowledge contributors, but the group saw other difficulties: how convenient is it actually to go around a dozen projects and add up total numbers of contributions? Also, if contributions are of poor quality, should they be acknowledged at all ?
- On the continuing engagement of editors: the group knew of no evidence that users respond to competitive devices such as leader boards. There was also a need to manage fear of change due to complexity of this field.
- A focus on the relationship between the user and the organisation was more important than that between users. The IHR had a strength in event organisation. It was also necessary to create a feedback loop for contributors direct from the organisation. Organisation should broaden the search for funding to include teaching sources, and identify economies of scale and areas of common interest between projects.

The Research Ecosystem

Questions

What are the benefits of having the different online sources linking together, and of more collaborative interaction between people working on different sources?

What issues arise regarding how different sectors and disciplines interact with each other?

What would you like to see the IHR do to help facilitate a more integrated research ecosystem? What can be provided centrally, and what can't?

Summary

- Interoperability was clearly useful, but there was concern to avoid projects having to spend too much time on this up front. It was also noted that there were a number of common standards, authority lists etc. available, but that the problem lay in persuading people to use them. Library of Congress headings etc. were often dismissed as not being specific enough to a particular project or research context. One issue was the relative difficulty of persuading historians to collaborate outside the higher education sector, and particularly with libraries and archives, where there was often a greater awareness of common standards.
- There was also discussion of the number of databases which never make it online, for example those produced by PhD students during the course of the work. The PhD student on our table initially welcomed the idea of putting his personal research database in the public domain, but was subsequently concerned about the amount of work that he would have to do to make it 'presentable'. There was also anxiety about exposing data to other people so that they could re-use it and possibly challenge the initial findings [NB This is interesting in light of Rob Iliffe's 'pyramid of data', as presented in his earlier paper].
- There was some discussion about just how collaborative historians are prepared to be, and precisely what is meant by collaboration. There was general agreement that, especially in the field of editing, interdisciplinary collaboration was beginning to work well within the humanities. It was less clear that this was happening, for example, with the hard sciences, where there was obvious scope for working together in the context of data mining etc.
- Collaboration with libraries and archives was adversely affected by cultural factors (for example a failure to recognise that special collections librarians can be researchers in their own right), by institutional structures, and by the nature of the projects themselves. In the latter instance, libraries are necessarily interested in digitisation and cataloguing, researchers with what happens

subsequently. The Andrew W. Mellon Foundation model of funding that had been used for British History Online was cited as combining these two things very effectively.

- Both of the database projects represented in the group felt that their particular resource could not have been produced using a central service like ReScript. They would, however, have liked to be able to talk to other people about what they were doing, seek advice on best practice etc. before embarking on constructing a database.
- It was suggested that central provision might be most useful in terms of resource discovery (common search etc.) rather than in relation to data input.
- It was noted that there are initiatives underway to create repositories of URLs which support linkage and interoperability by removing the 'dead links' issue. This was felt to be very useful and would greatly assist in encouraging researchers to cite digital materials.
- There was a danger that the multiplicity of generic tools and services being developed (for example there are already several which deal with crowdsourcing) would simply replicate existing problems of reinventing the wheel. However, it was likely that the best would simply rise to the top as they were widely adopted by the research community (in the manner of Zotero).

The ReScript project team

July 2011