

AHRC ICT Methods Network Expert Seminar on Sustainability

PEER REVIEW AND EVALUATION OF DIGITAL RESOURCES FOR THE ARTS AND HUMANITIES

Jane Winters, Institute of Historical Research

1. Context

One of the problems facing those concerned with the sustainability of digital resources is that it is simply not possible to sustain, or even to preserve, everything that has been and will be created by researchers in the arts and humanities. Is all of it even worth sustaining? The starting point for our project was that, at present, we just don't know – we don't have the ability to assess and compare the value and impact of digital resources at a strategic level.

The mechanisms for the peer review and evaluation of the traditional print outputs of scholarly research – monographs, journal articles and the like – are well established, if increasingly under strain. But no equivalent exists for assessing the value of digital resources and of the scholarly work that leads to their creation. A consistently-applied system of peer review and evaluation (of both the intellectual content and the technical architecture of digital resources) would serve a number of purposes.

- First, it would reassure academics and their host institutions of the worth of time spent in the creation of digital resources. It was a common complaint among the resource creators whom we consulted that such activity, unless it is generating really substantial levels of income, is seen as very much subordinate to the production of the requisite four RAE publications.
- Second, it would enable us to establish those types of resource which are of most use and interest to the academic community – of clear importance to both end users and funding bodies.
- Third, it would contribute to the development of common standards and guidelines for accessibility and usability
- And finally, and perhaps of most interest in the context of today's seminar, it would inform proposals to ensure the sustainability and preservation of high-quality scholarly material.

Peer review is fundamental to the academic research process. It underpins traditional scholarly publishing, both monograph and journal, and informs the decision-making mechanisms of the UK research councils. Bodies such as the AHRC and the Economic and Social Research Council (ESRC) have well-established mechanisms for the peer review of research proposals, including those for projects which involve the creation of significant digital outputs.

Evaluation of research output is also of considerable importance to the academy, and again there are robust mechanisms in place. Monographs, and less frequently journal articles, are evaluated by means of the published review. The majority of research projects which receive public or charitable funding are required to produce annual and/or 'end of award' reports outlining their progress, explaining their decision making processes, and addressing any areas in which they have failed to meet their original remit. In some cases, for example ESRC-funded research projects, these reports will themselves be peer reviewed, and they may be made publicly available.

But these mechanisms have not been successfully transferred to the digital environment – or at least not entirely. To take the most obvious example, digital resources are not widely reviewed in scholarly journals,

alongside their print counterparts. Some of the most high profile resources – for example the *Oxford Dictionary of National Biography* or *Early English Books Online* – create a flurry of interest, but such reviews are notable for their very singularity. Even where assessment does take place, a holistic evaluation of digital resources has proven somewhat elusive – a division between the purely ‘technical’ and the purely ‘academic’ persists.

2. The project

So this was what our project sought to address – and we got off to a rather more problematic start than we’d anticipated. The more we discussed it, the more it became apparent that different people understood different things by both ‘peer review’ and ‘evaluation’ – and in the context of digital resources, this was largely related to the point at which the assessment occurred.

Peer review was the simpler of the two concepts – for our purposes, it was understood to mean the formal assessment of *proposed* research. It is undertaken at a sufficiently early stage to influence the course of that research, the nature of its outputs, and ultimately even whether it takes place at all (or is made available to a wider audience). It is usually undertaken by a single academic working in a related field, or by a group of subject experts.

However, we identified two distinct types of evaluation:

- that which takes place during or at the end of a research project as part of a formal process;
- and that which is undertaken by end users, whether informally as feedback or in publicly-available reviews.

In the digital environment, evaluation is most usefully seen as part of an ongoing and iterative process. Digital resources require varying degrees of technical and academic input over time, but few can be said to be ‘complete’ in the way that a book or journal article is complete once published.

3. The survey

Once we had established our terms of reference, we could get to work. The first stage of the project was an online survey, conducted in November and December 2005. The survey questions were designed to elicit opinion as to the usage of digital resources, and no distinction was made between those who are solely users of digital resources and those who are both users and creators. There were 777 respondents to the survey, the majority of whom (56 per cent) identified themselves as being based within UK Higher Education Institutions.

This is not the place to go into the survey in any detail, but it is worth highlighting a few key findings. In response to the question, ‘What is important in determining the value of a particular resource for your own research?’, perhaps unsurprisingly more than three-quarters of respondents indicated content. The next most important factors, in order of popularity, were authority, the lack of availability or inaccessibility of the original analogue material, and comprehensiveness. It is only then that we get to usability, the ability to conduct complex searches and so on – that is, the more technical elements of a resource. One of the most surprising results of this question was the lack of weight accorded to transformative impact on research, with only 23 per cent of respondents regarding this as very important (when the question was reversed, almost as many – 21 per cent – indicated that it was of little or no importance to them). It is a curious response, which suggests that researchers do not always recognise, or articulate, the transformative impact of digital resources on their research practice. It may, however, indicate that, for many, what is important is not innovation for its own sake – they want increased and enhanced access to what they already in some sense have.

This question also highlighted the failure of many researchers to engage with the question of sustainability – only 32 per cent felt that the permanence of a resource was very important to them. Time and again in the focus groups convened by the project team, the sustainability and permanence of digital resources were cited as major concerns for both creators and users – creators, of course, have a vested interest in seeing the outputs of their research maintained, but users also identified the problem of ‘disappearing resources’ as a barrier to take-up. Interestingly, this may be a function of the type of digital resource – there was much greater recognition of the importance of permanence in relation to journal articles published online than in relation to, say, large datasets.

Finally, and most significantly for us, 71 per cent of respondents considered peer evaluation and recommendation to be either important or extremely important in their selection of digital resources for use in their personal research. One academic noted: ‘peer review and provenance are key for me – I can get non-peer reviewed material any time through Google and evaluate its usefulness myself. It is no substitute for the academic resources’.

Other consultation was undertaken throughout the year, with a series of user and focus groups convened to draw out the issues raised in the survey, and a number of interviews were held with key opinion formers. A benchmarking study was also carried out, testing some of the proposals and guidelines that had emerged in the course of discussion. All of this fed into our conclusions.

4. Conclusions

The main conclusions of the report fall under three headings.

The first is cultural change

The need for cultural change was mentioned by many of the participants in the project, by which was meant a change of attitude towards digital resources and their creators, and towards their use for scholarly research. It was thought that this cultural shift was in process – indeed, it was pointed out that many academics already implicitly trust the digital medium, using email as a regular means of academic correspondence, and frequently consulting digital resources such as JSTOR or the Royal Historical Society Bibliography online. It is, of course, the case that for digital resources to become firmly embedded in research culture, there needs to be an accepted mechanism for assessing their value – where we came in! – but cultural change can be driven in other ways.

- First, there needs to be a greater recognition that there is more than one model for research in the arts and humanities. Traditionally, the most valued research outputs have been the work of lone scholars – the creation of digital resources, by contrast, almost always involves collaborative or team working, whether between individual scholars or between researchers and their supporting computing departments. The academy needs to place due value not just on the outputs of collaborative research, but on the work itself. This will also go some way to solving the problem of how to treat the largely unheralded work that is undertaken at the intersection of the technical and the scholarly.
- Second, there needs to be much greater investment in the training of researchers both to use and to evaluate digital resources. The inability of significant sections of the academic community genuinely to comment on and assess the value of digital resources makes any peer review process difficult to manage, and also undermines confidence in its results. Learned societies and subject organisations have a significant role to play in ensuring that their communities engage with this issue, and university libraries and computing centres should be encouraged to provide training to mid and late career academics as well as to new researchers.

- Finally, the editors of scholarly journals can effect change, by routinely commissioning reviews of digital resources and by encouraging their authors to cite digital material where it is available. The creators of digital resources can also help with this last, by providing clear citation guidance on their websites. Some of the reviews commissioned as part of the benchmarking study for this project were published in the IHR's online journal, *Reviews in History*, for example Elisabeth van Houts review of *The Narrative Sources from the Medieval Low Countries* (<http://www.history.ac.uk/reviews/paper/vanhouts.html>). The journal's editorial board have subsequently agreed that reviews of this type should be actively pursued, and it is hoped that others will follow their lead.

Peer review

Many of the project's recommendations concern the mechanics of the peer review process, and specifically as it affects the assessment of research proposals to UK research councils:

- I am sure that most of you are aware of the 'technical appendix' which is required of all applications to the AHRC which have a significant digital element. Almost all of those consulted during the project agreed that this was no longer a reliable indicator of the robustness of methodology or project planning – applicants have simply become too good at filling in the forms, to the point where a 'bad' technical appendix is very rare indeed and a 'good' one no guarantee of successful project delivery.
- Connected with this, bodies such as the AHRC and the ESRC might consider implementing a two-stage application process: an initial summary submission assessed for scholarly value; and a second, more detailed submission, incorporating the information currently relegated to the separate technical appendix. This would encourage both applicants and assessors to view the project as a whole, while retaining emphasis on the importance of content.
- Peer reviewers should be chosen primarily for their subject expertise, but their ability to assess the technical elements of a proposal should also be taken into account. This will both make the process easier to manage – reducing the numbers of academics turning down peer review requests – and make it more robust. Again, learned societies and subject organisations should be prepared to assist in the selection of appropriate reviewers. Bearing in mind the skills gap that I have already highlighted, in the short to medium term it may be necessary to consider review by a subject specialist in conjunction with a humanities and computing practitioner.

Evaluation

The final set of recommendations concerns the procedures for the evaluation of digital resources.

- First, research councils and other large funding bodies should be encouraged to conduct post-completion assessments of those projects which they support financially, with both the evaluation report and any response from the resource creators made publicly available. Any such review should be conducted in a spirit of openness, so that resource creators are encouraged to discuss freely any problems that they have encountered and any innovative solutions that they have adopted, for the benefit of the research community as a whole.
- Both guidelines for potential reviewers and a check-list of basic technical standards would be a useful addition to the process. Our project produced drafts of both, which can be consulted on the IHR's website (http://www.history.ac.uk/digit/peer/Peer_review_report2006.pdf).
- Interestingly, although again perhaps not altogether surprisingly, there was almost complete rejection of any metrics-based approach to assessing the value of digital resources. This was articulated most clearly in connection with usage. While there was acceptance that resources designed

for a wide audience might in some way be deemed to have failed if they were unable to demonstrate high levels of usage, the relative popularity or unpopularity of a resource should, and indeed could not be used as a significant indicator of academic value. The introduction of some system of kite-marking was also felt to be highly undesirable, and many project participants expressed concern that it would lead to over-centralisation and the eventual stifling of innovation. The project concluded that any system of evaluation or review should not adopt a simple 'pass/fail' approach when considering a digital resource in its entirety. Subjectivity was thought to be vital to the assessment process, and should not be masked by any more rigid system of indicating 'approval'.

While there is a role for subject organisations and learned societies in guiding peer review and evaluation, and even recommending or supplying the personnel to undertake such activities, no one body should have the power to say whether or not a resource is 'good' or 'bad'.

5. Wider applications

So, once a structure is in place for assessing the 'value' of digital resources, what are its practical applications?

- It facilitates the assessment of digital resources, and the work which goes into their creation, in formal exercises such as the RAE, with obvious benefits to both resource creators and their host institutions.
- It assists users in making decisions about which digital resources are most appropriate for use in their own research
- It assists librarians in making purchasing decisions.
- And it helps funding bodies to assess whether a particular project should be supported, whether it successfully meets its aims and objectives, and ultimately whether it has in some sense delivered value for money. These criteria in turn inform decisions about sustainability and preservation, the focus of today's seminar.

While sustainability was not explicitly addressed by our project, it was an issue which loomed large in discussions. One focus group, for example, noted 'major anxieties about the sustainability of digital resources and the need to guarantee the great, long-term costs, whether by HEIs or the government'. A participant at another commented that 'It was impossible to produce a trusted site unless there was investment in long-term maintenance'. Scholars are more likely to be unwilling to use a resource and to cite it in their work if they do not believe that it will continue to be available.

Project participants recognised three distinct, if interrelated, elements of sustainability: the financial, the technical and the academic. The issues surrounding the financial sustainability of digital resources largely fell outside the remit of our project, although as I've already noted, rigorous evaluation is an important decision-making tool for funding bodies. Technical and academic sustainability, however, were addressed more directly. The problem of technical sustainability blended into concerns about the long-term preservation of resources, but responsibility for these two elements was assigned differently. The project concluded that, while best practice is continually evolving, it is not unreasonable to expect resource creators to abide by a basic set of common standards agreed, and regularly reviewed, by their peers and implemented through the research councils. Applicants for funding should identify the standards to which they intend to adhere at the technical stage of an application, whatever form that may take. Responsibility for preservation was deemed to reside with bodies such as the Arts and Humanities Data Service.

Academic sustainability was felt to be even more intractable. The issue of updating a resource (both to make corrections and to add new information) after the funding period has ceased is clearly a problem.

Such post-publication activity is likely to be less arduous than the initial creation, but in many cases it is essential to the ongoing relevance and utility of a particular resource. Any solution to this problem is likely to require investment, whether from funders, from libraries or from projects' host institutions. It is possible to envisage a centralised model of maintenance, and despite the costs and organisational challenges that this option would involve, it merits serious consideration. More cost effective would be the informal involvement of the research community, perhaps through the use of the burgeoning wiki technology. In this way, those resources worthy of sustained development might almost become self-selecting.

Research facilitation is central to the Institute of Historical Research's remit, and consequently it tends to be involved in the creation of resources which require financial and technical support, and academic input, over the long term. One of the projects that we host, the Royal Historical Society Bibliography of British and Irish History (<http://www.rhs.ac.uk/bibl/>), illustrates the problems faced by so many of the potentially 'infrastructural' projects run in universities across the country. Let us consider the three elements of sustainability identified here:

- In terms of financial sustainability, the options open to the Bibliography are severely limited by the fact that essentially it provides a route to material offered by other projects and services. It adds enormous value to those resources to which it links, and enhances the user experience significantly, but it is not content based. It also suffers from the 'problem' that, again like so many other projects, it has been funded as a free resource. As resource creators we face the dual problem of user unwillingness to pay for hitherto 'free' material and our own reluctance to remove access from at least part of our constituency by introducing an element of charging.
- As for all digital resources, technical sustainability is of course an issue, all the more so because the Bibliography is to an extent reliant on the technical robustness of those resources to which it links. For us, the role of AHDS History as the ultimate guarantor of the sustainability of the data, if not the interface, is essential. As noted above, like others we consulted during the course of our project, we accept that responsibility for designing a resource with technical sustainability in mind lies with us as data creators – but the sharing of expertise in this area at the project planning stage would lessen the burden.
- Finally, there is the question of academic sustainability, of crucial significance for a project like the Bibliography which will never, in any real sense, be 'complete'. Projects of this type rely for their usefulness in large part on their currency – a bibliography which is not updated will very quickly cease to be of *central* importance for the research community. It will still, of course, have value, but it will need to be supplemented by other resources, and indeed by time-consuming personal research to identify recent publications. In addition, the fact that the Bibliography is continually updated means that it will always require staffing, leading us neatly back to financial concerns.

The IHR's peer review project was not designed to provide answers to these questions, but it at least suggests ways in which we can begin to form judgements about the 'value' of the digital resources which have been created over the last decade and more. This in turn will inform the difficult decisions to be made about sustainability, and perhaps where to target scarce resources – financial, technical and academic.