

E-science Challenges in the World of Historical Studies

This document has been prepared to facilitate discussion at the e-Science Scoping Study Expert Seminar for History. It is written by a research practitioner in early-modern history, and based on my own experiences and viewpoints. It cannot claim to be representative of all aspects of these disciplines. Nor does it aim to be comprehensive. Feel free to disagree with all or any of what it contains. It is simply offered as a delineation of some of the key issues on which we shall need to focus in our expert seminar. I would like our discussions to be broad-ranging. But I would also like us to emerge with some practical recommendations and ways forward. The seminar offers us an opportune moment and a unique occasion to look at how best we can use E-science to our advantage in our particular discipline. But, in order to do so, we need to eliminate those, potentially alluring ‘yellow-brick roads’ that will not help us realize our individual and collective research objectives. I start with trying to encapsulate some of the distinctive features in the way in which historical studies are conducted. Then I examine the challenges and opportunities presented by ICT, outlining the areas where it has been of particular importance to us. Finally, I offer a series of headings for us to discuss the e-Science agenda, suggesting where we may most profitably take it forward in our own area.

1. Understanding the World of Historical Studies

There is not much which is unique to the methods and approaches of the historical disciplines. They are, to varying extents, shared with arts and humanities more broadly – and not least since every arts and humanities discipline has its historical research practitioners. But the nature of historical documentation leads to some relevant distinctiveness. Here are 10 points for starters:-

- there is no such thing as a ‘canon’ of historical research materials. There is no primacy of period or geographical place. Historians are necessarily interested in comparative dimensions, both temporal and spatial
- there is no innate primacy of one kind of material about the past, textual or non-textual, over any other kind of material. Just as the past is an exponentially expanding universe, so is its material
- all historical research conclusions are essentially contingent. In historical studies, there is no ‘last word’
- there is no accepted historical ontology for historical studies
- the nature of the historical research process involves sustained engagement with the records of the past. ‘Historicity’ is achieved through successive iterations of critical interrogation of the evidence
- because (in part) there is no ‘canon’ and no accepted ‘ontology’, collaborative historical enquiry has a relatively small basis on which to grow. However (especially where historical documentation requires specialist skills for its collection/collation/interpretation), collaborative work has proved very successful
- the ‘natural partners’ of historians are archivists, librarians, museum-keepers, holders of historical record of one sort or another. These are typically not in academic institutions. They are often in institutions which are relatively poorly

endowed, and sometimes poorly equipped for ICT developments. So, along with research fabric of the AHDS, TNA, BL and IHR, there are numerous smaller repositories on which historians regularly rely for their raw materials. Much historical raw material was made available in published textual editions that have been in the public domain for over a century, and whose viability remains generally solid

- this does not preclude profitable engagement with other academic disciplines, both within the arts and humanities and beyond. But these are often ‘conceptual’ engagements, in which each participant seeks to learn something from the approach of the other, rather than being necessarily involved in research ‘collaboration’ as such
- historical research is not undertaken purely by academic historians. There is an important penumbra of research activity in particular areas undertaken by interested amateurs, on whose work historians often rely for some of their conclusions
- historical research has an important public dimension. Historians are stewards of the past, and that stewardship implies a responsibility for the way contemporary society understands and interprets its past. Because that past is about the lives (and deaths) of human beings, this responsibility is a moral one. It cannot be served in a medium which removes, or immolates, historians from the public sphere
- there is no accreditation structure or overall ‘governing body’ of historians. The edifice of historical learned societies in the UK and abroad is national and local, sometimes specific subject-based. Although the Royal Historical Society, Historical Association, and Society of Antiquaries often speak on behalf of the profession to government, they do not claim to represent this more complex world of learning.

2. Challenges and Opportunities of ICT in Historical Research

I concentrate on the following areas where ICT has proved, or is proving, significant in historical research over the past five years or so. I have divided them (somewhat arbitrarily) into six areas:

1. Information search, retrieval and validation
2. Availability of historical raw materials in facsimile forms
3. Publication and accessibility of textual and non-textual historical outputs
4. Processing/collation of large volumes of data
5. Maintaining formal, institutional and semi-institutional contacts in the world of learning
6. Fostering informal research contacts around particular research themes and objectives

1. Information search, retrieval and validation

This is perhaps the most striking and immediate impact of ICT upon research practitioners in history. As in the wider non-academic community, the current preferred instrument of general internet search is Google. Historians are not particularly worried about the inevitable redundancy in the search results; nor about the non-hierarchical and flat nature of the results. They are relatively comfortable with their capacity to evaluate the reliability of a particular hit. The skills required replicate those that historians use in the evaluation of their historical evidence as a matter of course. But library and archive catalogue searches may well rival Google searches for many practising historians. Union catalogues, or their equivalent in the archive domain (COPAC, A2A, national library catalogues abroad) are essential resources in some historical fields for information search. Historical Abstracts, American History and Life, and the Annual Bibliography of British and Irish History are essential ancillary bibliographic tools to studying history in the relevant subject areas, or within English-language publications. The lack of equivalents in history for other language areas is increasingly crippling to their success in the modern historical research economy. Important, too, are some specialist information providers (e.g. Wellcome Institute for medical history). There is not one information gateway for historians. We are used to a world in which we expect to develop quite sophisticated information literacy about where we are most likely to find relevant information. We are eclectic and philandering, inherently suspicious of monogamous relations with information providers (and tend to equate monogamy with monopoly, or the possibility of withholding information). Information search and retrieval often provides, however, an ‘institutional’ or even ‘received’ wisdom about a subject. Historical research, at its best, is subversion of ‘received’ wisdoms about the past. In that sense, information search in public domains about historical matters is either about confirming details, or (if about larger questions) generally only the beginning of a historical enquiry, and not the end of it.

2. *Availability of historical raw materials in facsimile forms*

The availability of historical materials in facsimile has been often driven by commercial forces, the vagaries of research grant submission success rates, and the activities of pioneers in the field. The result has, by and large, not been ‘strategic’ (but could it have been?). But it has not been bad. The big success story is STConline and its 18thC successor. It is not an exaggeration to say that these instruments have transformed the nature of the questions that historians ask (and can expect an answer) of their materials. They have had a demonstrable ‘cyclonic’ effect on their research areas. ICT has been called an essentially ‘disruptive’ technology. Its effect here, in terms of the ability to get answers to questions very quickly, speeds up the iterative process at the heart of the historical process. It also democratizes the historical processes as well. We might expect the same effects wherever facsimile historical raw materials have been created. In reality, however, the impact has been more variable – registering the great differences in scale, accessibility, searchability, and relevance of historical raw materials to solving one or more kinds of question. We should note that:-

- some worthwhile historical raw materials in ICT facsimile forms are, in reality, only accessible currently with difficulty

- there is a mismatch of provision. The areas where the largest amount of historical raw material is available in traditional forms is in the post-1800 historical domain. This is also where the majority of research practitioners' interests in UK HEI are also located. This is the area, on the other hand, where historical raw materials in facsimile forms are least commonly available. Those that are available (e.g. the Census data) have been created for a number of reasons (public policy; social science research; genealogists' needs, etc), and historians' research interests have not often figured large among them
- in common with other disciplines, historians are faced with a situation where interoperability and cross-searching of one historical dataset with another is becoming a major stumbling block to their being effectively deployed to undertake historical research
- historians, in common with other arts and humanities disciplines, have benefited from the availability of non-textual historical raw materials in facsimile forms. But this raises particular issues of fitness for purpose, and whether a facsimile digital representation really serves as a surrogate for historical research in e.g. art history or material culture
- historians are still undecided about the cost-benefits of full facsimile reproduction over more selective cataloguing and calendaring. The equation is not one of universal applicability, and past experience is no guarantee to likely future need and *praxis*. It is likely that, for the foreseeable future, no matter what the subject domain or the volume and sophistication of the historical materials available online, historical practitioners will still rely on independent field-work in archives, libraries and repositories for their most significant research conclusions

3. *Publication and accessibility of textual and non-textual historical outputs*

In common with other disciplines, historians have begun to take advantage of publishing historical outputs in electronic media. The process has generally been one in which the initiative has been with the information mediators – the traditional publishers making e-books and e-offprints available, journal editors making backruns of journals available through JSTOR, etc. Although the open access movement has provided additional facilities, I am not aware that historians have made extensive use of them. The problems in this area – the need to provide publications that are of high quality, consistently available, methodically catalogued, peer-reviewed, where the access to the material is guaranteed by reliable protocols - are not unique to historical studies. They are none the less relevant to us than to other areas of scientific study. Because of the historians' public role, however, they might be even more relevant, and we should expect electronic publication to be an arena in which there will be more high-profile public contestation over historical issues in the short to medium-term future.

4. *Processing/collation of large volumes of data*

We have already (implicitly) outlined three different layers of historical 'stuff' available in electronic form:-

- information of various kinds, catalogued and uncatalogued, professionally and unprofessionally produced, mediated and unmediated
- historical 'raw' material in facsimile form of various kinds
- works of history available in electronic media

There are probably other layers too.

The reality is that the application of historical intelligence (the creative process at the heart of any historical field of study) requires access to all three concurrently, and in a fashion that cannot be predicted or pre-determined. So the storage of digital information in many different formats, stored and managed in bespoke systems by independent institutions, often does not make the application of historical intelligence through ICT significantly easier than in traditional media. Some significant historical data is only available in particular institutions through commercial site-licences. Other data has to be purchased on an item-by-item basis. Even when available, its manipulation or accessibility can be restricted by the terms of the particular license in question.

Historical material is often rebarbative. It is often not readily capable of being analysed in statistical fashion. The loss of granularity in converting a historical document into a numeric or field-delineated dataset is often a limitation to its historical analysis, and certainly to the subsequent re-usability of the dataset in question. Where statistical methodologies have been adopted, they have often resulted in very sophisticated and important conclusions (e.g. historical demography: social history; crime, etc). But they have been accepted because those involved have the specialist skills both to understand the documentation involved, and to apply the social-science techniques to the analysis of the material.

Historians have been reluctant to develop over-arching ontologies to their subject domain of a kind that would facilitate cross collection searching. That is because historical conceptualization is fluid, often essentially contested. At the level of conceptualization, it would be generally regarded as counter-productive to attempt to build high-level ontologies for searching purposes. Lower-level ontology development is still in its infancy.

Historians are consummate list-makers. Lists constitute the building blocks for our contingent spatial, temporal and nominal historical awareness. But such lists tend to remain private, implicit and subsumed into the historical research process. The historian of eighteenth-century London might, for example, have a working list of goldsmiths at work in the capital in the eighteenth century, and use that in the course of an analysis of its social history, or the operation of credit. But that list might not be explicit in the eventual published work, and it would certainly not be electronically available, even though it might be of considerable research utility to others. There is an issue of what constitutes a 'publishable conclusion' and a 'historical dataset' here which needs to be addressed.

5. *Maintaining formal, institutional and semi-institutional contacts in the world of learning*

Historians have benefited from the large number of institutional structures that exist to support the historical world of learning. AHDS History is strong, responsive to needs, and innovative. The TNA has also been leading the world of archivists, a particularly well-organised group in the application of ICT to records management. The BL is a world-leader in both library and archive ICT applications. The UK has a strong infrastructural provision for historical studies, in comparison with many European neighbours (for example). But there is an issue about coordinating these various institutions, and linking them to the work of historical studies, especially when their mission statements (in some instances) require them to be responsive to public demands in other ways. Although we are not directly involved in the funding of these bodies, we should be aware that their existence is dependant on the use that we make of them. That use implies being able to demonstrate benefit, directly or indirectly from their role.

6. *Fostering informal research contacts around particular research themes and objectives*

The world of historical studies is composed of largely self-reflexive groups of scholars and practitioners, whose activities and interests touch and intersect one another at a variety of points, which constantly change and evolve with the research trajectories and careers of particular individuals. They reflect fashions and trends in historical enquiry, and respond to dominant personalities and the formation of research clusters in particular places. In this respect, historians are no different from other constellations in the republic of science and letters. There are some examples of historical practitioners making use of VRE [Virtual Research Environments]. But I suspect that the concept, and certainly the results of those experiments, is not generally in use as yet in the historical community. This may partly be a result of an important practical constraint on the application of ICT in the historical domain. Most academic departments in history do not have regular access to skilled technician time. History is a popular undergraduate subject with correspondingly high staff-student ratios. There is limited time for individuals to equip themselves with the necessary technical skills – and relevant courses are generally not available for them to do so. What technical proficiency they have acquired so far has been ‘on the fly’. They have become used to that being the norm. Technical developments that are going to assist their research *praxis* have to be weighed against that background.

3. Historians and the E-Science Agenda

E-Science – in this context it might be better to call it E-Research – is flavour of the month. It means all things to all people. I am sticking to the definition of the National E-Science Centre:-

in the future, e-Science will refer to the large scale service that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will

require access to very large data collections, very large scale computing resources and high performance visualization back to the individual user scientists.

[<http://www.nesc.ac.uk/nesc/define.html>]

How, then, should historians react to this possibility?

To answer that question we need to focus on these issues. These emerge from this analysis of how historians work, and how they have reacted so far to ICT opportunities and challenges:-

- We have some fairly large datasets, and some potentially very large demands for fast transfers of data (high resolution of images; video and sound footage of historical materials; national census data, etc). How should we identify where these are, and where their further study and exploration in an e-science Grid environment can be most profitably enhanced?
- We have some potentially very disparate data and some emerging interoperability needs. How can the Grid help us with these? Can we square the challenges of the authentication, licence, copyright and accessibility issues with the opportunities that potentially may lie ahead for us? If so, can we identify a particular domain or area of enquiry where we should concentrate our attentions?
- If we can, what kind of interoperability are we looking to provide? If a high-level ontology of the subject is unlikely to exist in the near future, and perhaps ever, does the provision of lower-level ontologies meet the needs of interoperability? Or are there other ways in which we can develop a Grid-based 'research platform' that would more appropriately provide for historians' needs?
- We have a strong individualist research culture; but also proven evidence of successful collaborative endeavour. Can we base our Grid developments on the latter? If so, what lessons should we draw on so that we can maximise the likely outcome? Are there changes that we need to contemplate in our research culture? Do we need more (or more advanced) research tools? Are these likely to be generic to the arts and humanities, or specific to history?
- We have important infrastructure partners in libraries, archives, etc. The latter, however, are (for the most part) not leading players in the Grid and many lack the necessary technical infrastructure. Is there a potential asymmetry here that we need to address?
- Our informal means of networking are effective in environments where technical support is minimal. How should we approach the move to an environment of higher technical specification for informal networking, with its greater demands? What is the best way of ensuring that historians make effective use of the Access Grid? Of VREs? How should we approach the issues of awareness, training, and support?
- Research collaboration poses fundamental questions about the 'ownership' of research outputs (in a world in which the RAE morcellises those into a return by each individual practitioner). Historians are both very willing to share research results, and quite jealous of their particular 'subject' and sometimes of the 'archive' to which it relates. There is an issue here about the cost-benefit of specialist knowledge of a particular subject in the hands of one individual, and the

collaborative understanding of a historical domain (which, because of the nature of the subject, may never be a completely 'shared' understanding).

Historians, like every other scientific discipline, are faced by an explosion of information. E-science is proposed as a means of managing that data 'bonanza'. The prospect, the overwhelming rationale for this seminar, is that e-Science methodologies help to raise our levels of abstraction (in historical terms, our levels of 'historical intelligence'). If this is so, we should be looking seriously to take advantage of what the e-Science agenda has to offer us.

M. Greengrass

6 June 2006