

Reviews in History

Published on *Reviews in History* (<http://www.history.ac.uk/reviews>)

The Virtual Representation of the Past

Review Number:

782

Publish date:

Friday, 31 July, 2009

Editor:

Mark Greengrass

Lorna Hughes

ISBN:

9780754672883

Date of Publication:

2008

Price:

£60.00

Pages:

276pp.

Publisher:

Ashgate

Place of Publication:

Farnham

Reviewer:

Jonathan Blaney

This volume brings together papers given at the Expert Seminar held in Sheffield in April 2006. The seminar allowed historians and archaeologists to share their insights into the use of digital media in their areas of study. Judging from the resulting book, this must have been a stimulating and fruitful occasion.

The volume itself is the first in a series called Digital Research in the Arts and Humanities. Usefully, as an extension to the wealth of web-based resources cited in print, 'a list of ongoing research projects in history and archaeology that are using advanced ICT methods' (p. 3) is being maintained online (www.arts-humanities.net/publications/virtual_representation_past [2]); we are promised that this list will be regularly updated.

Most of these papers take a roundly practical approach to the use of technology in specific projects. Especially so is Meg Twycross's description of her use of Adobe Photoshop and a Video Spectral Comparator (VSC) (1) to examine an overwritten and damaged medieval Memorandum Book in York City Archives. She explains how she used the Photoshop software in some detail, with a number of screenshots, and her step-by-step instructions are explicit enough that anyone with access to this software and high-quality manuscript images should be able to replicate her method.

Along the way Twycross strongly emphasises that she was able to interrogate the manuscript in ways which simply would not be possible with the naked eye. This is not only through advanced techniques (such as the VSC analysis of the ink, which ruled out earlier speculation that some of the overwriting might have been done in the 19th century) but even through an effect as simple as the different appearance that is presented by a back-lit image on screen. The exploration of a palimpsest of this complexity proves an ideal exemplar of the value of such tools for non-invasive analysis: "As we dig down using the electromagnetic spectrum as a spade, we bring the underlying layers to the surface without destroying the original" (p. 47).

It is the strong tendency towards concrete details such as these in the presentation of the technology in research projects that is one of the strengths of this volume. Too often papers of this kind are posed at a level of abstraction which leaves their audience adrift only because they lack a few well-chosen examples as moorings. Donald Spaeth's discussion of the problems of using XML's hierarchical structure to capture historical documents "documents which sometimes inconsiderately refuse to conform to such orderly hierarchies" is absorbing enough to those who like to think about such things; it becomes really vivid when he describes assessors of probate inventories walking around a house and then doubling back on themselves to make trouble for the future XML encoder: "It is not entirely clear why the appraisers returned to the chamber over the kitchen; perhaps a second pass through the chambers to list the bedding led them to discover more objects" (p. 57). With the example of the assessors' footsteps in mind, we are now able to understand the technical XML problem well enough to appreciate and evaluate the three different solutions that Spaeth goes on to present.

Having said that, the most abstract paper in this collection does happen to be one of the most incisive. One of the themes emerging from these essays is the hope that the future will see more convergence of resources. Manfred Thaller argues that a major obstacle is, and will remain, the lack of a standard for the representation of time in databases. In essence this means that it will not be possible to run a successful query across multiple databases which record time in different ways - even if those different ways consist of such humanly similar formats as dividing centuries into thirds versus into quarters. "The different, usually very poorly documented, semantics of existing databases with regard to time is a major obstacle to their future interoperability" (p. 118). Thaller believes that the CIDOC Conceptual Reference Model (2) (which provides a standard for defining concepts and relationships in cultural heritage information) may offer this standard but, he adds with apparent understatement, "It will require, however, a good deal of further development" (p. 124).

In a footnote to his discussion of how a database field might automatically be able to contain a "location in time", Thaller mentions that this concept "has a strong grounding in all information systems which handle cash transactions" (p. 120n). This aside exemplifies a feature of many of the essays: the re-use in humanities computing of software (or software concepts) developed with quite other aims and customers. The VSC software mentioned above is used as a forensic tool by the FBI and other agencies and laboratories; Vincent Gaffney touches upon the notion that gaming technologies may be used to inform development environments; Ian Gregory, when talking about the use of Geographical Information Systems in the Humanities and the drawback of GIS's relatively poor handling of time, mentions ruefully that "GIS software developers do not regard the academic humanities community as a major market" (p. 145).

It is, of course, perfectly natural that software, as with technology generally, will be repurposed for uses unforeseen by its originators. A point that does not emerge from these chapters is that open-source software may offer even greater opportunities for this kind of repurposing for Humanities ICT. Not only is the source

code, by definition, open for modification, but the model of software development encourages the distributed writing of add-on modules by a user community. The Dr Williams's Centre for Dissenting Studies at Queen Mary is about to begin a project, 'Dissenting Academy Libraries and Their Readers', that will ingeniously use KOHA, an open-source Library Management System, in tandem with historic library records (such as handwritten loan registers) to reconstruct the library culture of dissenting academies between 1720 and 1860.(3)

It is clear that, as is frequently mentioned throughout this book, the use of computing in the Humanities lags a long way behind e-Science. Gaffney adds that the use of grid technologies in the Humanities has been less than expected, and mentions some reasons why this should be so, before summing up this technology-envy rather nicely: 'A visit to the live map of the Particle Physics Grid data brokerage site at Imperial College provides an ample demonstration of the scale of such networks, although this is not recommended for arts researchers of a nervous disposition' (p. 126). What is less clear is if there is any practical way to address this disparity, or whether the disparity is, in itself, a problem.

The other comparison sometimes made, for example by Andrew Prescott here, is with projects which have imaged and transcribed literary manuscripts. Again this is an inter-disciplinary difference that seems inevitable. In the case of the Emily Dickinson archive, to take one of Prescott's examples, there are a small number of manuscript pages in one hand; it is, further, relatively easy to make the case for this kind of manuscript archive to be digitised in high-resolution colour. Dickinson is rather unusual in being a much-loved writer whose entire output survives in holograph but who saw almost nothing through to print herself. There could scarcely be a more persuasive case for digitising an archive than this one.

There is apparently an interesting difference of opinion here between Tim Hitchcock and Andrew Prescott. Hitchcock speaks vigorously of the transformative effect archival digitisation could have, if the archive was opened up to keyword searching: 'If historians speak for the archives, their role is largely finished ... If historians no longer 'ventriloquize' on behalf of the archival clerk, then they are free to rethink the nature of social change' (p. 89). Prescott argues for the importance of high-quality colour images of manuscripts. For one the searchable text seems to be the thing, and for the other the manipulable image. Everyone, of course, would really like both, but is this a likelihood in the immediate future?

Hitchcock, writing about 'Plebeian Lives and the Making of Modern London' (4), describes the transcription of 530 reels' worth of manuscript material as 'a relatively small project in comparison to EEBO or ECCO, but it reflects an important shift from the large-scale digitization of printed material to manuscript' (p. 87).

ECCO (Eighteenth Century Online) uses Optical Character Recognition (OCR) software to make searchable an undeniably extensive corpus of 200,000 books, but the accuracy of search results is less impressive: it is fairly easy to find false negatives on the very page on which your hit is highlighted; the same is true of Google Books, which also uses OCR with mixed results. EEBO (Early English Books Online) is not a transcription project at all but is a collection of page images of over 100,000 books, of which only the accompanying catalogue entry is searchable; EEBO-TCP, a separate entity, does transcribe from the page images held in EEBO (the earlier typefaces not being considered suitable for OCR), but has only reached the impressive figure of about 25,000 books completed after nearly a decade of work and a cost of millions of pounds.(5) These are large-scale expensive projects but they are still in the far more manageable medium of print, where OCR or outsourcing of work to keying companies is viable. The only manuscript transcription project I have worked on, to digitise part of the calendar to the Carte Collection in the Bodleian in the Encoded Archival Description format (6), was expensive compared to similar transcriptions from print but, crucially, only consists of one hand throughout.

When Andrew Prescott laments the lack of attention paid to the provision of high-quality colour images compared with transcription on websites such as those of Darwin Online, the Newton Archive, and the Boyle Project, it is noticeable that again we are dealing with one, or predominantly one, hand. Furthermore it is the much-studied hand of a celebrated individual. If a new phase of manuscript digitisation is nearly upon

us then that is entirely to be welcomed, but this book's optimistic picture lacks concrete proposals for how this might be done.

It is often possible to gauge the worth of an online article, such as a blog entry, by seeing if it links to its references; nearly all the good ones do. Yet scholars still seem to prefer to cite print sources where an online version exists (and is conceivably used by those same scholars): 'By ignoring the proximate nature of electronic representations, the impact of new technology has been largely skated over and subtly downplayed' (Hitchcock, p. 86). Indeed digital online resources enforce rigour in ways that can only benefit scholarship: 'Editions like these have also created the obligation to display one's evidence: it is no longer sufficient to state that a reading is there because you have seen it' (Twycross, p. 30). The general point is underlined again by Lorna Hughes in her summary conclusion to this volume 'a volume which should be enough to convince those sceptical of the value of digital resources in the Humanities, were they ever to read it: 'This is an important refutation of the argument that digital surrogates distance the scholar from the original sources. They do not. They give the scholar far greater control over the primary evidence, and therefore allow a previously unimaginable empowerment and democratization of source materials' (p. 192).

It is unfortunate, then, that the conclusion to such an engaging book should, of necessity, end on a downbeat note, as Hughes addresses the closure of the Arts and Humanities Data Service and the ICT Methods Network in 2008. These closures will only make the sustainability of digital resources, and the viability of future projects and collaborations, more uncertain.

August 2009

Notes

1. Twycross recommends this description of the VSC technology: <<http://www.fbi.gov/hq/lab/fsc/backissu/oct1999/mokrzyck.htm> [3]>[accessed 6 May 2009]. [Back to \(1\)](#)
2. <<http://cidoc.ics.forth.gr/> [4]>[accessed 6 May 2009]. [Back to \(2\)](#)
3. At the time of writing the best account available on the web seems to be at <<http://www.english.qmul.ac.uk/drwilliams/research/dissacademies2.html> [5]>[accessed 6 May 2009]. [Back to \(3\)](#)
4. <<http://www.shef.ac.uk/hri/projects/projectpages/plebeianlives.html> [6]>[accessed 6 May 2009]. [Back to \(4\)](#)
5. For ECCO <<http://www.gale.cengage.com/DigitalCollections/products/ecco/index.htm> [7]>[accessed 6 May 2009]; for EEBO <<http://eebo.chadwyck.com/home> [8]>[accessed 6 May 2009] and for EEBO-TCP <<http://www.lib.umich.edu/tcp/eebo/> [9]>[accessed 6 May 2009]. [Back to \(5\)](#)
6. <<http://www.bodley.ox.ac.uk/dept/scwmss/projects/carte/carte.html> [10]>[accessed 6 May 2009]; for EAD see <<http://www.loc.gov/ead/> [11]>[accessed 6 May 2009]. [Back to \(6\)](#)

Other reviews:

oxford journals

<http://ehr.oxfordjournals.org/content/CXXV/512/248.full> [12]

Source URL: <http://www.history.ac.uk/reviews/review/782>

Links:

[1] <http://www.history.ac.uk/reviews/item/3760>

[2] http://www.arts-humanities.net/publications/virtual_representation_past

[3] <http://www.fbi.gov/hq/lab/fsc/backissu/oct1999/mokrzyck.htm>

[4] <http://cidoc.ics.forth.gr/>

[5] <http://www.english.qmul.ac.uk/drwilliams/research/dissacademies2.html>

[6] <http://www.shef.ac.uk/hri/projects/projectpages/plebeianlives.html>

[7] <http://www.gale.cengage.com/DigitalCollections/products/ecco/index.htm>

[8] <http://eebo.chadwyck.com/home>

[9] <http://www.lib.umich.edu/tcp/eebo/>

[10] <http://www.bodley.ox.ac.uk/dept/scwmss/projects/carte/carte.html>

[11] <http://www.loc.gov/ead/>

[12] <http://ehr.oxfordjournals.org/content/CXXV/512/248.full>