

CENTRE FOR HISTORY AND ECONOMICS

The Digitization of History

Notes from the talk at Trinity Hall, April 29 2008:

Digitization of the Robert Hart Collection at Queen's University Belfast

In this meeting we welcomed Deirdre Wildy (Senior Subject Librarian, Arts and Humanities, Queen's University Belfast) and Jennifer Regan (Visiting Research Fellow, 2007-08, Queen's University Belfast), who discussed their work in digitizing the rich archive of letters, diaries, photographs and slides of Sir Robert Hart (1835-1911), Inspector General of the Imperial Customs, Peking, 1863-1908. The website, available at the following URL:

<http://digitalcollections.qub.ac.uk/digital-image-gallery/hart/>

provides a valuable demonstration of how smaller-scale (at least, compared with the activities of Google Books, formerly discussed by the group) historical and scholarly project of this type can, through planning and feedback, add immeasurably to the experiences of both the historian and the general public in providing new kinds of access to historical figures and documents. Keyword and full-text searching, (which Dr. Regan has focussed on in her work with one particular diary) quality images allowing close inspection of the images and pages on-site, and lower resolution images available under a liberal license online have sparked a rebirth of interest in the collection.

Several issues emerged from the presentation and discussion that followed:

1. New users, new opportunities:
 - a. Digitization, far from divorcing a collection from its context or allowing historians and students to skip the bricks and mortar archive altogether, can combine the goals of preservation of the materials with increased visibility and use. Prior to the digitization project, Ms. Wildy noted that few scholars had ever made substantive use of the collection. Now, however, both QUB students and outside academics have re-discovered the Hart papers through the website, and use of keyword and full-text searching allows a new kind of interaction with the documents.
 - b. Further, the digital 'home' for the Hart Collection has allowed QUB to link their Hart papers with those held by Harvard, and perhaps, in the future, with other small caches of letters scattered in archives around the globe. Hart's voluminous correspondence can thus become part of this virtual archive, revealing his activities and their impact in a way that the various physical document repositories cannot.
2. Planning:
 - a. Both Ms. Wildy and Dr. Regan highlighted the central importance of planning a digitization project around questions about intended audience, beneficiary of the finished product, and feasibility of maintenance. What distinguishes a research resource from a website aimed at a more general public? Dr. Regan noted that if teachers at all levels wish to provide resources which go beyond the Wikipedia entry, they will need to create a new model of what such a site might be like. Ms. Wildy spoke of the importance of having run a pilot site for the slides and photographs which enabled them to work through their ideas and to design a second-generation site which improved various key features. The

- pilot program also allowed the historians and archivists to see both the potential and drawbacks of the medium.
- b. According to Dr. Regan, it is crucial that academic institutions looking to digitize collections take into account a realistic estimate of the time that it takes to perform the work. The four tasks that were required to complete her project with one Hart diary broke down as follows: creating metadata, one week; transcription, five weeks; adding modified XML code (tagging keywords, formatting the plain searchable text), three weeks; and cataloguing the correspondence and creating an index, three weeks. Yet it is difficult to provide estimates since the researcher's level of familiarity with both the technical and historical aspects of the project are key in determining workload, a point that reinforces the importance of a pilot program. For example, Dr. Regan needed to identify the historical figures mentioned in the letters, often listed only by first name.
3. Trust:
 - a. Ultimately, Ms. Wildy felt that the difference came down to a question of creating a level of trust by both mimicking the face-to-face experience of the archive in two dimensions, and also in providing, through scholarly cataloguing and full-text searching, additional benefits which expand the face-to-face experience.
 - b. The cataloguing of the collection was crucial to the success of the digitization project, but it also demonstrates how digitization can provide an opportunity to engage with a collection in new ways. The collection was estimated at 3,000 items before the project began. In fact, there are in excess of 8,000 items.
 4. Is digitization 'real' scholarly work?
 - a. Dr. Regan spoke about the perception among academics that digitization is not 'scholarly work'. Yet clearly the work of transcribing, editing, and adding meta-data (keywords, shelfmarks, dates etc.) to the Hart diary required a historian well-versed in Hart's life and the events which he recorded. This was a project of editing and interpretation as well as an attempt to mirror the experience of working with the original documents, in which skill and caution was required to avoid passing down 'inherited' errors. Robert Watson noted in the questions that in this sense digitization projects are very like open source software projects in computer science. The computer science community relies upon such projects as a major code base for academic and professional research, but work in open source is often performed by volunteers and does not count toward publication requirements. The liberal license which the project uses, Leigh Denault noted, also echoes those used by open source software.
 5. Digital Rights, Copyright:
 - a. Dr. Steven Murdoch (Computer Laboratory, Cambridge) began a discussion about copyright and digital rights management which further highlighted the extent to which digitization, preservation and access continually come up against the limitations of current copyright law.

The QUB project has benefited from a Centre for Data Digitisation and Analysis that provided the building blocks required by computer scientists, geographers, historians, library scientists and students within the university to rethink digital resources. It seems from this project that a thoughtful interaction with the prospects of digital media offer a real opportunity to change both historians' and archivists' conceptions of how we use, edit, and interpret documents.