
Exploiting special collections: using digital methods to enhance their research and learning potential

Stella Butler
*Deputy University Librarian,
John Rylands University Library, The
University of Manchester*
Tel: 0161 275 3751
E-mail: Stella.V.Butler@manchester.ac.uk

In this article I review how the expanding digital environment is impacting on special collections within academic libraries by exploring some examples of digital projects within the John Rylands University Library of the University of Manchester (JRUL) and also elsewhere. I will consider very briefly what the future might hold for special collections as technology continues to open up opportunities and challenges for collection management and access.

THE JRUL SPECIAL COLLECTIONS

The special collections of the JRUL provide rich resources for researchers in the humanities.¹ They originate from three foundation libraries: the John Rylands Library, Deansgate; the Library of Owens College, later the Victoria University of Manchester; and the Joule Library at UMIST (University of Manchester Institute of Science and Technology). The John Rylands Library was founded in the 1890s by Mrs Henriqueta Rylands, as a memorial to her husband, John, one of the most successful cotton entrepreneurs of the mid-nineteenth century. She purchased the printed book collection of the 2nd Earl Spencer and the collection of manuscripts assembled by the 25th and 26th Earls of Crawford and Balcarres as core collections for the library. She herself added many further fine books and manuscripts. In the 1920s, the library first acquired archive collections in the form of family muniment collections. The library now holds the archives of numerous organisations including

commercial companies, businesses, charities and religious institutions. The University of Manchester was founded as Owens College in 1851. The printed book collection of Richard Copley Christie (with its strength in Italian Renaissance books) together with that of Edward Freeman (rich in material on European medieval history) formed the main foundation for the college library in the 1890s. Strengths in special holdings relating to medicine and science reflect the research orientation of the university over the twentieth century. Similarly, the special collections of the former UMIST library focus on the history of science and technology.

RESOURCE DISCOVERY

One of our core functions is to provide access to these resources for as many researchers as we possibly can. First of all, and most basically, we need to tell our potential readers what we have that is available to them. Increasingly our readers regard the internet rather than the physical library as the source of all collection descriptions. If a record describing a book, manuscript or archive is not easily found, preferably through a Google search, then for many readers that resource simply does not exist.

In the 1990s, we benefited enormously from funding made available through the Higher Education Funding Council for England for retrospective conversion of card catalogues, and as a result most of our printed books have electronic records that can be searched through the JRUL web-based catalogue or through COPAC. However, we still have collections of material for which the principal finding aid is the card catalogue. The largest collection in this category is our collection of medical printed books dated between 1800 and 1950. The origin of this collection lies with the Manchester Medical Society, founded in 1834. This collection contains many rare and important items, including first editions of the works of Xavier Bichat, Claude Bernard and others. The collection is particularly rich in continental editions and features an impressive range of titles in the German language. Many would regard it as the best collection of printed material for the history of medicine between Oxford and Edinburgh. It is currently housed in a closed-access store and use is sadly very low. I suspect many researchers are simply not aware of its existence.

We also have a large collection of printed material associated with the history of Methodism, which is also not described by a comprehensive elec-

tronic catalogue and is thus also very difficult to use. This collection was deposited in the library in 1977 and comprises periodicals, books and pamphlets, including material arguing against the ideas of John and Charles Wesley and other early reformers. It contains primary source material for historians of Protestant Nonconformity and is also heavily used by family historians searching for their Methodist ancestry.

I could go on with lists of other equally significant collections. And I am sure we are not alone in having cataloguing backlogs. It is difficult to develop proposals to address these problems as it has become increasingly difficult to secure funding for projects that focus entirely on either retrospective conversion of card catalogues or book-in-hand cataloguing. Programmes of external funding tend to take a dim view of 'simple' electronic cataloguing programmes and expect applicants to devise ways of adding value to the research material – as if the intrinsic historical value is not enough for the academic community. The consequence is that current externally leveraged investment in the most basic of resource-discovery tools, the electronic catalogue record, is very low indeed. As a result, in a tight internal financial environment, collections that have enormous research potential are simply hidden from readers.

We do promote our resources at collection level also and there are many collection-level descriptions on the web-site.² Could cataloguing for printed books at this level provide a more general solution to the lack of investment in item-level descriptions? The problem with printed books in a closed-access library comes, however, when the readers arrive wanting to know exactly what items are in the collection so that they can decide what to order. We cannot allow readers to browse the shelves and so collection-level descriptions are really inadequate for printed-book resource-discovery.

The 'Access to Archives' project³ and the higher education archives hub⁴ are playing a significant role in making archive and manuscript collections better known and better used, by disseminating electronic archive catalogues. We are grateful to the Wellcome Fund for Resources in Medical History for enabling us to create a number of detailed catalogues for archival material relating to the history of medicine. And indeed we are fortunate in employing archivists within our permanent staff who spend a large proportion of their time cataloguing. However, as with printed books, there is also a significant number of collections for which

there is no electronic finding aid. Collection-level descriptions are publicised and for archives they form an important component of resource-discovery for our readers. It could be argued that a collection-level description with a good box list is sufficient. However, archivists assure me that in an ideal world what researchers want is a detailed catalogue as well, so that they have both the big picture provided by the collection description and the details of each item.

As libraries are forced to consider ways of adding value to resource discovery there is increasing focus on finding ways of uncovering the links between collections, and between individual items within collections, in order to encourage new research questions to be posed and, indeed, answered. The database of Italian academies at the British Library provides an outstanding example of this type of work.⁵ Professor Jane Everson of the Italian studies department at Royal Holloway, University of London, and Denis Reidy at the British Library have together developed a database of information associated with the Italian academies, 1530–1650, incorporating records harvested from the British Library catalogue. The database is searched independently of the main catalogue. Searches result in information about people, networks and publications as well as catalogue records associated with books, manuscripts and archives in the British Library collections. Research that was done in the preparation of the database has been used to enhance the book records, with further details associated with each particular academy. This approach is ideal for a set of loosely linked individuals and institutions but it could also be applied to research themes and provides a transferable method of creating pathways through databases. For example, it could work for the Dissenting academies established from the mid-eighteenth century in England, which provided such an important example for the modernization of higher education later in the nineteenth century.

DIGITAL SURROGATES

Resources such as 'Early English Books On-line' and 'Eighteenth Century Collections On-line' have provided researchers with digital surrogates of vast numbers of printed material from 1473 onwards. Is there anything more for special collections librarians to do, therefore, in terms of digitisation of print resources? Of course: copies of books printed before 1701 should be treated as unique in bibliographical terms. And, generally speaking, the earlier the printing the more indi-

vidual each copy becomes. An enormous amount of information lies in the differences between copies. It therefore follows that many surrogates of copies of significant titles could provide a digitally convenient way to compare and collate and so extract that research information. There can be no more significant title than the Bible printed by Johann Gutenberg in Mainz in 1455, the first book to be produced by moveable metal type. In February 2008, JRUL hosted a team from Keio University's 'Humanities Media Interface' (HUMI) project for three weeks while they created very high-quality images of each page of this sacred text. The JRUL Bible is the final piece in the Gutenberg jigsaw that the research team at Keio, led by Professor Toshiyuki Takamiya, have been piecing together over the past ten years or so.⁶ The team began by producing images of the copy of the Bible purchased by Keio University in the mid-1990s. Since then they have photographed the British Library's two copies and the Bury Bible at Corpus Christi College, Cambridge, as well as the copy at the diocesan library in Pelplin, Poland. The Keio team have shown how software and digital images together can provide easier and more productive ways of collating these copies.

Another example of 'surrogate provision' is the Darwin Online project.⁷ This aims to provide access to Charles Darwin's complete publications, including a surrogate first edition of *The origin of species*, as well as 20,000 private papers, together with numerous examples of supporting material. This massive undertaking involved sending copies of Cambridge University's microfilms of Charles Darwin's private papers to India where 90,000 electronic images were scanned and cropped. Basing the project on microfilm has resulted in black-and-white digital images. This doesn't seem to have adversely affected the response, which has exceeded the expectations of project director John van Wyhe. There was wide coverage of the project's launch in the media but it was the traffic on the web-site that was so impressive. On 17 April 2008, launch day, there were 7 million hits on the web-site. On 18 April, there were 8 million. From 17 to 30 April over 20 terabytes of data were uploaded (1 terabyte is 1000 gigabytes). On 17 April alone 14,000 PDFs of the first edition of *The origin of species* were downloaded. One of the astounding statistics that John van Wyhe recounts is that more copies of Darwin's works were distributed in 48 hours than in the whole of the nineteenth century. There can be no doubt about the power of the internet.

However, Darwin and Gutenberg are historical giants, central to the canon, and as such they are exceptions deserving comprehensive treatment when it comes to digitisation. There are of course others in this category, and at JRUL we can identify collections for which the comprehensive approach is the only way to yield the true research potential of a collection. But is comprehensive digitisation feasible for the majority of collections, which often contain thousands of items in a range of formats in varying condition,⁸ and if we cannot digitise everything how do we make choices?

SURROGATE DELIVERY: THE POWER OF THE IMAGE

The JRUL digital-image collection is delivered over the web-site through the Luna Insight software system. The system allows readers to search on a wide range of parameters and then create their own collections, examine them in a virtual lightbox where more than one image can be compared and create and share learning materials.

Collections of other institutions who also mount their collections within Luna can also be searched and thus the system provides potential for collaboration between repositories and, in particular, the sharing of research information.

Collaboration has been a particularly strong theme in the development of a project to catalogue the Rylands Genizah collection, which contains around 11,000 fragments of documents originally stored in the loft of the Ben Ezra synagogue in Old Cairo. This was used from the ninth to the nineteenth century for discarded books and manuscripts of all types. The collection therefore includes texts covering both secular and religious subjects. The papers tell many human stories going back to biblical times. There are a number of Genizah collections in libraries all over the world. The most substantial is held by Cambridge University Library, which holds about 140,000 items.⁹

The Rylands Genizah fragments are in a range of formats, including individual leaves from printed books and manuscripts, archival material such as letters, official documents and even children's exercise books. Several languages are represented, including Hebrew, Arabic and Judeo-Arabic. A grant of £360,000 from the Arts and Humanities Research Council in 2006 enabled us to set up a project under the leadership of Professor Philip Alexander of the University of Manchester's religions and theology group, to digitise and catalogue the collection.¹⁰ We are well on the way to producing images of all the fragments, with

metadata associated directly with each image. Assistant librarian Carol Burrows worked with research associate Renate Smithuis to devise a metadata profile compliant with all major standards for images.

Our aim was to establish a web presence for the collection as quickly as possible and then to add to the collection as soon as basic metadata was available. The collection is identified as a distinct image collection within the library's Luna software.



Two fragments from a manuscript by Moses Maimonides (1135 - 1240) which have been reunited within the collection held at the Rylands. Genizah A281 and B5756.

Within the collection we knew there was an important group of manuscripts associated with Moses Maimonides (1135–1204), physician to Saladin and perhaps the most important of Jewish philosophers. An image of a fragment of his famous *Guide of the perplexed* in his own hand was added to the image collection. Shortly after, Ben Outhwaite, head of the Cambridge Genizah unit, recognised the section as the other half of a fragment within his own collection. As a result two scraps of paper separated for at least a century, and themselves more than 800 years old, have been virtually re-united and researchers can now see the complete page, even though the sections are separated by hundreds of miles. We hope to go on identifying 'joins' and thus to re-construct the contents of the synagogue storeroom, demonstrating how collaborative digitisation adds value to the individual efforts of Genizah repositories.

There are of course many other examples where collections or even single books have been separated over time. The digital environment offers important opportunities for libraries to bring this material together, re-construct wholes from parts and make resources both visible and convenient

for researchers to use. We hope our current JISC (Joint Information Systems Committee)-funded digitisation project focussing on Middle English manuscripts will result in the two-folio fragment of *The Canterbury tales* (Rylands MS 63) being reunited with the eleven-folio fragment held by the Rosenbach library in Philadelphia.

Although a comprehensive approach to digitisation might be ideal, in practice many archive collections are of such a size that such a policy has so far not been adopted. We have, however, tried to identify material that lends itself to attractive images that might act as 'tasters', attracting readers to use the originals. In all externally funded cataloguing projects we now include varying elements of digitisation. Within the Rylands image collection, therefore, we have several hundred relating to medical themes arising from our Wellcome Trust-funded project mentioned earlier. You can find images of surgical operations alongside trade magazines advertising X-ray equipment. There are also images arising from our collections associated with the novelist George Gissing (1857–1903) and the illustrator and designer Walter Crane (1845–1915).



Tinted drawing of the Miller from Chaucer's *'Canterbury Tales'*, English MS 63

SPECIAL COLLECTIONS BORN DIGITAL

Traditionally we have collected modern books, for example in the field of modern literature, where we have associated archival collections. The books and the personal papers together create a complete collection and provide a 'one-stop shop' for researchers. But what of the future? E-books have

yet to catch the mass market but surely it is only a matter of time? Sales of the Sony e-book will certainly be an interesting indicator. The history of music sales in recent years would indicate that if the right delivery gadget is found, with effective software, then take-up will take off. Although special collections librarians will continue to collect printed (or, strictly speaking, published) content to ensure context for archival collections, as e-books gain ground the relationship between special collections staff and IT colleagues will become more central to collection management. While this seems a fairly obvious statement, there are nevertheless some organisational and technical issues to face.



X-ray of a hand taken by the physicist Arthur Schuster (1851 - 1934) in the late 1890s. Papers of Ian Isherwood.

Archives present a different challenge that those who curate and actively collect personal papers must address, and address urgently. We have been fortunate to have been working since 2005 with the Bodleian library (as lead partner) in a JISC-funded research project, 'Paradigm', which has explored the issues associated with collecting the personal papers of living politicians. A major outcome has been a workbook, authored by the project team (led by Susan Thomas of the Bodleian Library but involving also Renhart Gittens, Oxford software engineer, and Janette Martin and Fran Baker, digital archivists at Manchester).¹¹ Their work has revealed the importance of the open archival information system (OAIS) model as a workflow for digital archives. The project also demonstrated very clearly the importance of working closely with the owner or producer of an archive to ensure material of historical significance is not lost or, conversely, buried electronically. Out of Paradigm came CAIRO (Complex Archive Ingest for Repository Objects), also a collaborative project, including this time the Wellcome Library as well as Oxford and Manchester. The aim of this project was to produce an ingest tool from open-source software and it is hoped that this software will be available shortly.

CONCLUSIONS

If we want to increase the use and research impact of special collections then we need to exploit better the power of the web. If special collections librarians fail to address this issue, much of the material within repositories will become or remain invisible. We need to explore ways in which we can use the digital environment to innovate in resource discovery for both print and archive collections, to break down the traditional boundaries between formats. Researchers want the highest number of 'hits' for the fewest number of searches. We need to encourage collaborative projects that 'join up' collections and themes, working not only with other repositories but also with academic colleagues, allowing them into the world of cataloguing and 'resource discovery'. Finally but perhaps most urgently, we must engage with issues around digital preservation to ensure that we do not lose the historical record relating to the people and institutions of these early decades of the twenty-first century.

ACKNOWLEDGEMENTS

I am very grateful to my colleagues in special collections, Fran Baker, Carol Burrows, Anne-Marie Clarkson and Renate Smithuis, for assistance in the preparation of this article.

REFERENCES

- 1 See *A guide to special collections*, Manchester: John Rylands University Library of Manchester, 1999
- 2 <http://www.library.manchester.ac.uk/specialcollections/>
- 3 <http://www.a2a.org.uk/>
- 4 <http://www.archiveshub.ac.uk/>
- 5 <http://www.bl.uk/catalogues/ItalianAcademies/>
- 6 <http://www.humi.keio.ac.jp/en/index.html>
- 7 <http://darwin-online.org.uk/>
- 8 John van Whye, 'A reception study in the making', *Viewpoint: newsletter of the British Society for the History of Science*, 86, June 2008, p 5
- 9 <http://www.lib.cam.ac.uk/Taylor-Schechter/>

- 10 <http://www.library.manchester.ac.uk/specialcollections/collections/specialprojects/genizah/>
- 11 <http://www.paradigm.ac.uk/workbook/glossary/index.html>