

EEBO, Microfilm, and Umberto Eco: Historical Lessons and Future Directions for Building Electronic Collections

by Shawn Martin

Shawn Martin (shawnmar@umich.edu) is Project Librarian for the Text Creation Partnership at the University of Michigan.

Introduction

In an age of mass digitization with book scanning projects like Google and Microsoft and their open access rival, the Open Archives Initiative, it is easy to forget that this is not the first time such efforts to "organize the world's information and make it universally accessible and useful"¹ have been attempted. In 1926, A. W. Pollard and G. R. Redgrave compiled *A short-title catalogue of books printed in England, Scotland, & Ireland and of English books printed abroad, 1475-1640* which at that time was the most comprehensive bibliography of English printed material in the early modern period. That project later developed into Early English

Books (EEB), a microfilm project started by University Microfilms International (UMI), and an electronic database Early English Books Online (EEBO) produced by ProQuest Information and Learning.

Though current mass digitization projects may hail themselves as the first attempt to organize large amounts of information and make them available, they are not. Certainly they are the first to do so at such a large scale. However, there are lessons that can be drawn from earlier attempts to do the same thing. One could try to do a complete history of information gathering from the time of ancient Egypt. Yet a more useful comparison might be the age of microform. Many of the same

arguments about preservation, greater access, and easier search capability are similar to the arguments about mass microfilming only 50 years ago. What is or is not unique about digitization as opposed to microfilm, and, more importantly, what lessons from mass microfilming can be learned for modern electronic projects? By looking at the history of just one of these mass microfilm/digitization projects, Early English Books Online, it may become possible to discover some of the answers to those questions.

Earliest History: the Short Title Catalog

As early as 1884, efforts had been made to effectively cata-

log English books during the early modern period. A three volume set *Catalogue of books in the Library of the British Museum printed in England, Scotland, and Ireland and of English books printed abroad, to the year 1640* had been published in the nineteenth century. In 1918 Sir William Osler of the Bibliographical Society identified a need to include books not in the British Museum and the need to compile multiple editions of the same work. This was especially important to him given the potential for air raids during World War I and his fear that further wars might jeopardize the collections within Britain. Librarians at Oxford, Cambridge, and other special libraries echoed this concern and believed that they had many copies of books not present in the British Museum and they wanted researchers to be able to locate them. Thus, A. W. Pollard and G. R. Redgrave along with support of the Bibliographical Society compiled a Short Title Catalog meant for scholars to look up titles and see which libraries held them. The first edition of this work was published in 1926.²

Despite the great help that the first Short Title Catalog provided, there were many gaps. Also in the 1920s, many indexers working on linguistic and bibliographic research, particularly in American Universities, noticed significant gaps in the Short Title Catalog and believed that it would be helpful to continue beyond 1640. So, Donald Wing of Yale University with the help of the Index Society started work on a supplement to the original Short Title Catalog in 1939. It

took many years to complete, largely because of the difficulties posed by World War II. However, in 1945, he published *Short-Title Catalogue of Books Printed in England, Scotland, Ireland, Wales, and British America and of English Books Printed in Other Countries 1641–1700*.³

The task of cataloging material from early modern England continues even now. In June of 1976, a meeting of the American Society for Eighteenth Century Studies envisioned a machine readable catalog of all material printed in English speaking countries during the eighteenth century. In 1987 the Bibliographical Society and the Modern Language Association decided to merge that project with the ongoing projects to update the Pollard, Redgrave, and Wing catalogs. Thus the English Short Title Catalog (ESTC) project began. Today the project continues to catalog thousands of works in English from 1470 to 1800 and provides access to this information at the British Library's website.⁴

From Print to Microfilm: Early English Books

In 1938, Eugene Power of Edwards' Brothers Printers left to found his own microfilm company, University Microfilms International. Because of growing fear of a German invasion of the U.K., it was decided to preserve as much of England's cultural heritage as possible. So UMI, with the cooperation of the British government and many other libraries around the country, used Pollard and Redgrave's initial bibliography to film what became the Early English Books

(EEB) microfilm collection. In 1957 he began filming the books contained within the Wing edition.⁵ Later, UMI released supplements to the collection in the form of the Thomason Tracts (a collection of pamphlets from the English Civil War compiled by George Thomason in the seventeenth century and currently held in the British library) and the Early English Books Tract Supplement (a collection of tracts held in the British library). The bulk of this project finished filming in 1988, and it still continues to operate to this day. ProQuest (now owner of UMI) has filmed 82 units of microfilm containing books from Pollard & Redgrave and 132 units from Wing. It releases about 2 additional units each year and anticipates completing this project within the next five to ten years. ProQuest also continues to discover new works and scours the world for rare copies held in obscure libraries (hence the slow progress of the latter stages of the project). Arguably it is still the most important microfilm preservation project in existence.⁶

Originally, this project was envisioned as a preservation project, but soon libraries came to realize how important such a collection would become to scholars on their campuses. The prospect of having copies of nearly every book printed in England made EEB a "must have" collection for campuses around the United States. For over thirty years after the initial filming, university libraries around the world bought EEB, and it became an essential resource for researchers in English literature, history, and other subjects.⁷

Eventually, all of these microfilm reels will be digitized and placed in the electronic project, Early English Books Online (EEBO).

From Microfilm to Electronic: Early English Books Online

Thus, the massive bibliographic projects started by Pollard, Redgrave, and Wing along with supplements from the Thomason Tracts and the Early English Books Tract Supplement form the core of what became the EEBO collection. In 1998 UMI began digitization of the microfilms and by 2003 Chadwyck-Healey had developed an interface for the images.⁸ Within the first year of its release over one hundred and fifty libraries bought EEBO for their libraries. By 2005 over 100,000 of the original 125,000 titles were available in this interface and libraries continue to acquire it in many countries around the world.⁹ Clearly EEBO became an important collection in a relatively short time. In many ways it has even replaced the microfilm collection.

From Image to Text: The Text Creation Partnership

EEBO also spawned an entirely different project. In 2000, seeing that the searchability of EEBO, though great, was not quite utilizing all of the potential that electronic technology had to offer, the University of Michigan and Oxford University started a project to create SGML/XML text that would allow scholars to search individual words within the books themselves

rather than just catalog records. To date this project has created over 10,000 texts and aims to do 25,000 by the end of the project in 2009. Most importantly, all texts that TCP finishes will eventually enter the public domain, thus ensuring that all of these culturally significant works remain publicly available in some form. TCP has also been able to incubate other projects that use its text as a base for further research on topics as diverse as sociolinguistics and Shakespeare studies.¹⁰

EEBO: A Complex History

Therefore, one can see that EEBO has at least five components: bibliography, microfilm, electronic images, e-text, and scholarly projects. All of these components are still going on in some form. The English Short Title Catalog continues to catalog old books. The EEB microfilm project is still searching for books and photographs them. EEBO will digitize those microfilmed books. The TCP is still producing text, and scholarly projects continue to build tools to access those texts. One also notices the disparate timelines required to complete these projects. It took fourteen years to get much of the bibliographic work done. Most of the microfilming was completed within fifty years. The digitization of that microfilm took five years, and the TCP will complete its work in seven years. The scholarly projects involved will finish usually in two to three years. Clearly the microfilm was the foundation and probably most costly step in the entire process. Does this complex history give

any lessons about the future of similar massive digitization projects like Google? Most importantly, what trends can we discern over the ninety or so years of the history of EEBO?

Access, not Preservation

One of the most notable shifts that seemed to happen in the 1950s was the emphasis on getting access to the content within EEB rather than preserving it. Eugene Power originally envisioned his microfilm projects as helping to preserve content of Britain and many other places and distributing that content to libraries around the world so that if any one copy should be destroyed many would still remain.¹¹ That view of microfilm has remained up until the present day. Even into the 1990s this view continued to dominate. If a book was microfilmed, that was the preferred method of preservation of a book. On the other hand, if it had been digitized, that is a method of providing access to that book, not preserving it.¹² More recently the Council on Library and Information Resources noted the interplay between electronic technologies and the emphasis on access rather than preservation.¹³ ProQuest also acknowledges this dichotomy in their own words "digital technology was the 'key' to unlocking greater access to the microfilm images."¹⁴

So, there was a great shift in emphasis from one technology to the other. With microfilm, preservation was originally the primary goal. That changed around the 1950s and access to libraries, particularly small ones without resources to send re-

searchers to the British Library, was equally important. This trend held when the EEB microfilm collection was digitized. Though there is certainly interaction between the two (in order to give access to materials one has to preserve them). The greater emphasis on access rather than preservation is certainly a trend one can see in projects like Google.

Value Added Production

Another trend, particularly in the electronic age has been the weight given to value added access to the collection. EEBO is more convenient than microfilm because users can pull up an individual book rather than locating it on a reel of microfilm. TCP is more convenient than EEBO because one can go directly to a word or concept one is looking for rather than reading the entire book. Scholarly projects are more convenient than TCP because they have particular tags or scholarly apparatus that a particular discipline might want rather than going through the generic TCP or EEBO interfaces. Whereas it was impossible to improve upon the microfilm image, it has become infinitely possible to improve upon the digitized image. Publishers are now focusing increasingly upon adding value to their collections (and charging more for them). Libraries are increasingly required to purchase these improvements to meet the needs of scholars on their campus. When EEB was first produced, libraries were in essence paying for the content itself. With EEBO they were not really pur-

chasing the content as much as they were purchasing more convenient access to that content. The same is true for TCP and all of the additional projects built after it. Dollars are spent now on content the library already owns, but cannot access to the full extent that electronic technology allows.

Replacing the Book

More and more, scholars and librarians alike have feared that students believe the electronic copy in EEBO (which is in fact a copy of a copy of a copy somewhere in a library) is replacing the original book. In some ways this fear is genuine. Many errors were introduced during the microfilm process and were compounded as that process shifted to digital. Diana Kichuk recently identified this process of "remediation" and discussed the problems of using digital facsimiles as replacements of the original book.¹⁵ She notes the problems of attempting to identify context, understanding the dimensions, and replicating the physicality of the book. She also notes the many problems introduced in the microfilming process when books were cropped and parts of pages were lost, pages were distorted in cameras, and many other problems. As scholars and students alike rely more on the digital facsimile, there is a fear among many that important material will be lost. As access becomes more important and ability to travel to the original book becomes less possible, the problems of whether a facsimile found in EEB or in EEBO truly "replaces" the book will be-

come more apparent. The same issues can be found in many of the reviews of the Google Book project and its problems.¹⁶

Infrastructure and Usage Patterns

Libraries have always been the infrastructure for scholarship, particularly in the humanities. Scholars came to a physical place that held collections restricted only to a small number of people. Now, the same researchers can access those collections from home, their offices, or from around the globe. Often the library has become less and less relevant in the eyes of many. Microfilm began that trend. Though many of the books were held in the British Library or other special collections in the UK, researchers no longer had to go there in order to consult the books. They could consult the microfilm images at their own library and then perform the more labor intensive research at the British Library. That trend has held in the electronic world. Richard Ovendon from the Bodleian Library has reported that usage of books in EEBO has dropped, but they have seen a great increase in usage of books not in EEBO and in manuscripts associated with those books.¹⁷ As more and more libraries have greater access to materials in special collections, this trend will likely increase. Also since much of that access will be electronic, libraries will become more defined by the special collections not available online. Additionally the infrastructure of libraries will (and has) changed to provide more value added services to users

rather than access to particular materials.

Where have we been? Where are we going?

Many of these trends started in microfilm and have simply been heightened by access to electronic technology. Therefore, they are not new problems, simply old ones re-emerging. So, what does this mean as we attempt to build collections for the future? First, we have to consider how access and preservation interact. All libraries are interested in digital preservation but few have spent as much time thinking about what that really means. Admittedly, microfilm sits on many library shelves slowly deteriorating from disuse. Similarly publishers give CDs, DVDs, and magnetic tape drives to libraries in order to "preserve" the electronic files. Many of these also sit on shelves slowly disintegrating (and they do so at a faster rate). Providing access is a very short term view designed only for users here and now. Preservation is a longer term strategy that seeks to make sure researchers will always have access to materials. Libraries have always been dedicated to this. Now it is an even more important mandate given the fast pace of change in the electronic world.

Second, with access to the same materials being available virtually at any library in the country, it becomes more important to think about the additional services libraries can provide. Publishers are beginning to see these trends and acting on them. Libraries need to do the same. TCP was designed to

meet a need not available within the EEBO collection (the ability to search text within the book). Other projects using TCP have also identified needs not met by current tools.¹⁸ Users will need to see a reason to go to a particular database in order to use it. They will not go simply because it is provided by a particular publisher or library.

Third, librarians need to think about what role electronic plays vis-à-vis analog books. Clearly as Diana Kichuk has shown, they are not replacements. However, they are not useless. What is the role of the electronic book in a database? What is the role of a print book in a library? What are the possibilities of one and not the other? Electronic books allow much greater searching and allow users to pull out particular bits of information from multiple books in ways that print books do now allow. Print books provide an artifactual context that an electronic book can never provide. Scholars and students, because of the problems EEBO and Google have presented tend to reject the utility of electronic books. Rather, librarians need to think of ways to engage faculty about the uses of different forms of the same content.

Finally, as the infrastructure of scholarship changes, builders of digital collections need to think about how users are accessing content, what they are doing with it, and how to build a system around that. Physical libraries will likely have a place within this infrastructure. Electronic libraries will probably have a greater role. Users will want specific types of services offered to them and particular kinds of

material offered to them in a print environment as opposed to an electronic one. The question remains as to which places require which services.

Conclusion

This debate about preservation, access, and the creation of new knowledge reminded me of a passage I remember from one of my favorite books, *The Name of the Rose* by Umberto Eco. In it the main character William of Baskerville has a debate with Jorge of Burgos, one of the scholars in the Abbey about whether the purpose of the library is to preserve knowledge or to search for new knowledge.¹⁹ In many ways, we are still having this debate. Traditionally libraries have been a place to preserve knowledge; yet in the Middle Ages, the time in which *The Name of the Rose* is set, changes were taking place. Greek books were coming into Europe that had been preserved by the Arabic world. The economy was changing so that fewer people became monks and more people were joining radical movements outside of the Catholic Church. Jorge of Burgos provides a conservative view of librarianship; abbey libraries exist only to preserve the past, not to contribute to the present. William of Baskerville on the other hand suggests that abbey libraries have an important role in defining the future. Microfilm was a unique invention, but it was one that changed from a preservation medium, to an access medium, and finally to an electronic one. If librarians do as Jorge suggests they will simply preserve the content for the

next generation without caring who uses it or why. If they follow William's suggestion they will be important in integrating new knowledge and new ways of thinking into their old systems. That is exactly what needs to be done, and by understanding how large microfilm collections have shaped the current system, we can begin to map out a course for the future.

Endnotes

- ¹ Google Mission Site, <http://www.google.com/intl/en/corporate/index.html> (accessed September 1, 2007).
- ² A. W. Pollard and G. R. Redgrave, *A Short-Title Catalogue of Books Printed in England, Scotland, & Ireland* (Oxford: Oxford University Press, 1926), v – ix.
- ³ Donald Wing, *Short-Title Catalogue of Books Printed in England, Scotland, Ireland, Wales, and British America and of English Books Printed in Other Countries 1641-1700* (New York: Columbia University Press, 1945), v – vii.
- ⁴ British Library Site – ESTC, <http://www.bl.uk/collections/early/estc2.html> (accessed September 1, 2007).
- ⁵ G. W. Cooke, "Eugene B. Power: Father of Preservation Microfilming," *Conservation Administration News*, (1993): 54.
- ⁶ <http://eebo.chadwyck.com/about/about.htm>
- ⁷ Mark Sandler, "Academic and Commercial Roles in Building 'The Digital Library,'" *Collections Management* 28, nos. 1/2 (2003): 107–119.
- ⁸ More information on the building of this collection are available in A. J McLean, "Early British Printing Meets the Electronic Age: A Large-scale Digitization Case Study," *Microform & Imaging Review* 30, no. 4 (2001): 127–134.
- ⁹ Diana Kichuk., "Metamorphosis: Remediation in *Early English Books Online (EEBO)*, *Literary and Linguistic Computing* 22 (September 1, 2007): 291–303.
- ¹⁰ Shawn Martin, "Collaboration in Electronic Scholarly Communication: New Possibilities for Old Books," *Journal of the Association for History and Computing* 9, no. 2 (October, 2006), <http://mcel.pacificu.edu/jahc/jahcix2/articles/martin.htm>.
- ¹¹ G. W. Cooke, "Eugene B. Power: Father of Preservation Microfilming," *Conservation Administration News* (1993): 54.
- ¹² Stephen Chapman, Paul Conway, and Anne Kenney, *Digital Imaging and Preservation Microfilm: the Future of the Hybrid Approach for the Preservation of Brittle Books*. Council On Library and Information Resources (1998), <http://www.clir.org/pubs/archives/hybrid/intro.html> (accessed September 1, 2007).
- ¹³ Oya Reiger, *Preservation in the Age of Large Scale Digitization*, Council on Library and Information Resources (September, 2007), <http://www.clir.org/activities/details/lstdi.pdf> (accessed September 1, 2007).
- ¹⁴ Austin J McLean. "Early British Printing Meets the Electronic Age: A Large-scale Digitization Case Study," *Microform & Imaging Review* 30, no. 4 (2001): 134.
- ¹⁵ Diana Kichuk, "Metamorphosis: Remediation in *Early English Books Online (EEBO)* *Literary and Linguistic Computing* 22 (September 1, 2007): 291–303.
- ¹⁶ See Review "Google Books: What's not to Like" (April 30, 2007), <http://blog.historians.org/articles/204/google-books-whats-not-to-like> (accessed August 17, 2007).
- ¹⁷ Richard Ovendon, "Being Digital in Bodley: Early Books and Research Libraries in the Google Era," (de)materializing the Early Modern Text: Early English Books Online in Teaching and Research Conference, Bath Spa University, 8–9 September, 2005.
- ¹⁸ See further examples in Shawn Martin, "Digital Scholarship and Cyberinfrastructure in the Humanities: Lessons from the Text Creation Partnership," *Journal of Electronic Publishing* 10 (Winter 2007), <http://hdl.handle.net/2027/spo.3336451.0010.105> (accessed September 1, 2007).
- ¹⁹ For an interesting take on this argument, see Kornelia Tancheva, "Recasting the Debate: The Sign of the Library in Popular Culture," *Libraries & Culture* 40, no. 4 (2005) 530–546.