

The Impact of Digitization on Special Collections in Libraries

Peter B. Hirtle

Digitization has brought significant benefit to the users of special collections. It will also challenge the relative value given to paper originals of rare materials as digital holdings increase. Special collections will need to justify their unique value through a deeper understanding of their holdings and an extension of their scope.

I spend a good part of each day working to make sure that collections of special materials of interest to faculty and staff at Cornell University are available in digital form. But many years ago, before I started a string of positions in which I have been intimately involved with the digitization of special collections material, I taught in the Program in Values, Technology, Science, and Society at Stanford University. Science, technology, and society programs developed from a recognition of the tremendous impact on society of technological innovation. Many of the consequences of technological innovation are unintentional, and not all are necessarily beneficial.

One of the core examples we would use in our teaching was the story of the Manhattan Project. We would challenge the students to think about the scientists who worked on the remarkable scientific and technical achievement of building the first atomic bomb. Were they, we wondered, aware of the moral implications of their work? Should they have been? We would normally finish our discussions with the story of J. Robert Oppenheimer, the director of the Manhattan Project, father of the atomic bomb, and scholar of Hindu philosophy, who, upon the first successful explosion of the bomb, is reputed to have cited a quote from Shiva in the Bhagavad Gita: "I have become Death, destroyer of worlds."

This conference has allowed me an opportunity to reflect on the impact, both intentional and unintentional, the new technologies of digitization will have on special collections. The digitization efforts of the past decade may not have required the same scientific and technological advancement needed to construct an atomic bomb.

Nor are our digitization efforts the moral or ethical equivalent of constructing nuclear weapons, and I am no Oppenheimer. Nevertheless, the accomplishments of the past decade in digitization represent a true technological advancement, one with the potential to alter forever the world of special collections as it now exists.

Ten years ago Terry Belanger in his Malkin lecture tried to predict among other things how changes in technology would alter the nature of special collections work.¹ It is now an appropriate time to acknowledge how much of the future Belanger got right in his predictions, to identify what further developments we can see on the horizon, and to determine whether we can direct or control developments before the fallout from our digitization efforts hurts us all.

The Benefits of Digitization

Before we examine the future, possibly negative, impact digitization will have on special collections, it would be desirable to consider the conclusions that can be drawn from the pioneering digitization projects of the past decade. Many of these projects have led to real progress. Three benefits in particular stand out.

Increased Use

The biggest single benefit that has arisen from our pioneering digitization efforts has been a tremendous increase in the use of digitized material. If you make special collections materials available via the Web with appropriate metadata and software, preferably for free, they will be used.

A good example of this is the experience Cornell and the University of Michigan have had with the Making of America, or MOA, collections. MOA is a collection of nineteenth-century American serials and monographs. In hard copy, prior to digitization, at best a few hundred volumes might circulate each year. At Cornell we initially mounted page images of the volumes and with little advertisement quickly were averaging 4,000 page views a month. This past year, we added searchable text behind the page images and are now at 5,000 views per day. Michigan, which has been following this approach longer, also has over 5,000 page views per day on its site.² Just as impressive to me is that each day someone requests something from one of the twenty-two titles Cornell has available. In hard copy the material may have seemed obscure; when digitized it becomes a core resource.

We can conclude from the MOA experience that people will use material that is digitized, indexed, and freely available on the Net.

New Types of Research

A second major advantage to the digitization of research materials is that it makes possible research that could not otherwise easily be done. An example is at Kentucky, where scientists have been scanning extant manuscripts of *Beowulf* and then using imaging techniques to highlight elements of the text that could not otherwise be seen.³ At Johns Hopkins University, library staff have scanned three variant manuscript versions of the *Roman de la Rose* found at the Walters Art Gallery and Bodleian and Morgan libraries.⁴ They have linked the images to transcriptions of the texts, making it possible to track and see variants online. At Tufts, researchers are combining eighteenth-century texts, images, and geographic information systems to build an interactive atlas of eighteenth-century London.⁵ Digitization can be more than just a substitute for page viewing—it can also generate new and exciting research opportunities.

New Users, New Uses

The first two major contributions of digitization—an increase in traditional use of the material and the development of new avenues of research—are based on the assumption that it is the traditional special collections user who will be using the material. The third major change with digitization is the appearance of new types of researchers using rare books and manuscripts. Again, our experience shows that increasing the availability of special collections material can change the audience using the material. Lexicographers working for the *Oxford English Dictionary*, for example, have been one of the major users of the Making of America collection at Michigan and Cornell, even though the collection was developed with the interests of specialists in American studies in mind. There is some indication that even the elusive K–12 education group might be able to use materials normally reserved for scholars if properly presented and interpreted.⁶

The preliminary pilot projects that Cornell, Michigan, the Library of Congress, and many others have conducted have shown that making special collections material available on the Web can increase use of the material, contribute to new forms of research, and attract new users to the material. It is easy to see why libraries across the country have rushed to develop digitization projects. Our “Manhattan Project” to develop a global digital library is succeeding. But what will be the fallout from converting more and more special collections materials to electronic form?

Changes to Special Collections with Digitization

There are, I would argue, at least five inevitable developments that will occur as the pace of digitization increases. Each of the developments is occurring to some degree now. While it is dangerous to project from current behavior, there is little reason to believe that the trends will soon change. Each development has serious implications for the future of special collections. They are:

Electronic access will replace most uses of printed, paper copies. Reading a book cleanly printed on good paper can be a pleasurable experience. Even reading a poorly printed or abused book found in a special collection can be immensely satisfying because of the physical, tangible connection it has with the past. The physical manifestation of a book or manuscript can also carry information about its date, process of creation, and previous use, none of which can be determined from a digital scan.⁷ For most research purposes, however, electronic access will replace the use of physical artifacts found in special collections.

This assertion is based on two observations. First, while most users in special collections reading rooms enjoy the experience of working with the physical volume, their primary purpose when making a visit is to read the text at hand. They end up in the special collections reading room because that is where the book they want to read is housed. When presented with the opportunity either to use a printed facsimile of an early volume, one that can be checked out of the library and used at one's leisure, or to consult the original volume from which the facsimile was made in a reading room with limited hours and special rules for access, most library users will select the facsimile. Reading, and not the analysis of the physical nature of the artifact, is the primary motivation for using material in special collections.

Second, digital facsimiles have found surprisingly wide acceptance, even at a time when most analysts agree that the ergonomics of the hardware and software for reading a book online are very poor. For example, an unpublished analysis conducted at the University of Texas shows that when both print and electronic versions of recent books are available, students will overwhelmingly select the electronic copy. A study done at the University of Toronto of users of online early Canadiana showed that three quarters of the people who used copies of books online found them more useful than the print originals or microfiche copies, even though most of the users said they preferred the print copies.⁸ Recently, the University of Virginia made twelve hundred titles from its electronic text collection available in Microsoft Reader electronic book format. In less than two months, more than 600,000 copies of the books were downloaded.

The university has since added four hundred more titles and has also made the books available in a format appropriate for the Palm Pilot. In the eight months since the program started, almost 2.5 million books have been downloaded.⁹

The use of digitized books, either as page images or as e-books, is happening in spite of the poor ergonomics associated with most electronic displays and e-book readers. There is almost universal agreement that reading a book on screen or in electronic form is not as pleasing as reading hard copy. Yet in spite of the current poor state of electronic books, people will, when given the opportunity, consult an electronic text over the print original.¹⁰ As the reading environment improves and the availability of electronic materials increases, the preference for electronic surrogates of original material will only increase.

The use of paper originals will decrease. A common assertion is that interest in and use of originals goes up with digitization. Is this assertion true? I know of only anecdotal evidence that addresses this issue, but it would suggest the opposite. At Michigan, the print copies of the Making of America titles are shelved separately from the other print collections; people must ask to have access to the print volumes, making it possible to track their use. Since the Making of America went online there has been only a handful of requests for the print volumes, and half of those were withdrawn once the patrons learned that they could get access to the volumes online.¹¹ Our experience at Cornell has been similar. Some museums have attributed their rising attendance in part to the availability of images online, but this is hard to quantify.

We need published studies with hard evidence that would examine whether digitization increases use of paper originals. For now, though, I think we should assume that the availability of digitized primary source resources on the Internet will lead to a decrease in the use of paper originals.

Given the already-noted preference for digital surrogates, it would be very surprising if use of the originals did not decrease. Most users in all but the finest of research rooms are there simply to read texts that otherwise are unavailable to them because of presumed rarity, fragility, or age. For those researchers who are primarily interested in what a document says (the majority in most research rooms), a digital copy will be sufficient. As a consequence, they will make fewer visits to the research room.

The number of books available as digital facsimiles will increase. The number of rare books available as electronic surrogates is increasing. The number of items already available in digitized form after

only ten years of preliminary work is striking. Most people know that the Library of Congress has digitized over 6 million pages of books, manuscripts, photographs, and films. How many here know, however, that the Bibliothèque Nationale de France will soon have digitized 55,000 titles and over 15 million pages? Smaller institutions such as the University of Göttingen in Germany have digitized over 2,500 titles of eighteenth-century North American travel literature, and I have already mentioned the Early Canadiana Online project. The UMI Digital Vault project plans to digitize the UMI microfilm collection, which represents hundreds of thousands of books and over 5.5 *billion* pages. Already most of the titles in the English *Short-Title Catalogue* (Pollard & Redgrave and Wing) as well as the *Thomason Tracts* are available online. In addition, the Holy Grail—the digitization of the entire Library of Congress—is in sight. A recent Microsoft Reader ad had a timeline for the development of electronic books. The timeline, according to the ad, represents “the best estimates of MS researchers and developers.” The ad predicts that by 2015 industry will unite to convert the entire Library of Congress to e-book form. Even if we double Microsoft’s estimate, it still means that in thirty years most of the volumes in library print special collections will be available online as digital surrogates.

Special collections print holdings will become less special. Today a library might be happy to have in its special collections a few thousand seventeenth-century English books. Given the relative scarcity of the books, the only way the library can provide access to those titles is by having physical possession of the volumes. But is it necessary to maintain a middling collection of rare books when access is no longer tied to physical possession, let alone ownership? How likely is it, for example, that a researcher will visit the Humanities Research Center to look at a volume of early Canadiana if he or she knows that it is one of the 8,000 titles available now or soon to be available through the Early Canadiana Online project and is available for use wherever he or she might be, twenty-four hours a day? How important is it for Cornell to acquire a second mid-seventeenth-century edition of Leonard Mascall’s *Booke of the arte and maner, howe to plant and graffe all sorts of trees*, or even to keep the one edition that it does own, when thirty-four different editions are available in digital facsimile form via Early English Books Online?

Truly unique items will continue to distinguish special collections. Most of the printed items found in special collections, however, are found elsewhere. As digital copies become available for reading, the need to consult the local print holdings will decline. The print copies may still sell for large amounts of money to private collectors,

but few libraries can or should participate in the market if most research needs can be met with digital copies.

Special collections librarianship will change. The fifth inevitable change in special collections brought about by increased digitization will affect the duties and roles of special collections librarians. In most libraries today the question of which staff member will answer a reference question and how it is answered is largely determined by where material is physically housed. A faculty member interested in the Civil War, for example, might learn about some of the library resources from a generic reference librarian while learning about other materials, such as a Lincoln letter, from a special collections librarian. What role will each professional play when the collections are virtual, not physical?

Some libraries have already faced this problem with microfilm. Are microfilm copies of incunables placed in special collections, where the experts who know the source material are located, or do they join general collections, where access is easier? Virtual digital collections compound the problem. Again, our experience with the Making of America collection is suggestive of what changes will take place. Most of the nineteenth-century material found in MOA is housed in the general collection of the library, and so general reference staff have taken the responsibility for answering questions about the material. Soon we will make available, through a grant from the Save America's Treasures program, digital surrogates of a premier collection of abolitionist pamphlets. The pamphlets are housed in the Department of Rare and Manuscript Collections at Cornell, and staff members from that department answer reference questions about them. Yet the MOA collection includes journals with an abolitionist bent, and at some point the library may elect to make the two collections searchable as one. Who will answer reference questions about the combined collection? Will the special collections librarians decide that their mandate extends to general historical topics and assume responsibility for answering all historical questions, even if some of them are about resources not present within the special collections area? Or will general reference staff begin to answer almost all questions about a subject matter and only refer questions relating to the original physical artifact to special collections? The latter approach exploits the specialized knowledge, skills, and abilities of rare book librarians, but how many libraries can afford full-time experts on physical artifacts?

The Future of Special Collections

Perhaps you are starting to understand now why I sometimes wonder if, in encouraging the digitization of special collections, I am not

also leading them on a path to their own destruction. Imagine an environment in which there are digital surrogates available online for most of the printed items in print special collections. And imagine that most users of the special collections prefer to use the online versions than to consult the print originals. If the Library of Congress is available online, why have a special collection at all? Can we really justify the continued existence of special collections to library administrators when use of the collections drops precipitously? As Belanger noted a decade ago, “many educational institutions are going to become increasingly dubious about the appropriateness of maintaining museums of the book on their campuses.”¹²

In short, some of the changes that will take place are dramatic enough that they may call into question the very existence of many special collections. Yet special collections are special, and it behooves us to find ways to preserve them. I do not think we can change the underlying trends previously outlined. There are going to be digital surrogates for more and more material, and more and more people will prefer to work with those surrogates. Rather than bemoaning the fact that people are happy using surrogates, we should accept it and realize that the future of special collections lies elsewhere. I can identify three areas where I think special collections may still have something to contribute, even if Microsoft and its partners control online access to almost all the printed literature in the world.

First, collections must emphasize those elements in their holdings that are truly unique. When it is an issue purely of reading, the printed holdings in most special collections are highly redundant; comparatively few titles exist in a single copy. Manuscript holdings, however, are almost by definition unique. Furthermore, it is unlikely that we will ever be able to digitize even a small fraction of the material found in manuscript collections and archives. Special collections units will need to focus their energy on manuscript and archival holdings. It will not be enough merely to have them, either. Manuscript librarians will also need to develop ways of presenting representations of the holdings online to users in such a way that the users will be motivated to visit the repository. We need to know what kind of people might like to use a collection, what parts in a collection they would find interesting, and how we can present the collection to them in such a way that they will understand that it holds material that they would find valuable, even if the material isn't available online. In short, we need to redouble our ongoing efforts to improve the tools that provide access to unique materials.

Second, we can reinvigorate the idea of special collections as museums. I assume that most people in this room share my belief

that a bunch of TIFF images is no replacement for a leather-bound book. Even Octavo's beautiful PDF versions of some of the great books of the Western world are not satisfactory replacements for the copies at the host institutions.¹³ The reason of course is that the original works have inherent value as artifacts. In special collections in the future, we should stress the artifactual value of the works more, both in collecting material and in presenting it to users. Online surrogates will be able to deliver the content of books; we need to identify and stress the value of books as objects in the special collections museum of the future. It also means that in collection development the emphasis should be on acquiring books that are meaningful as artifacts and not just rare or expensive.

Third, we can become active participants in the conversion of special collections materials. In my discussion above, I presented the imminent conversion of the bulk of the printed literature to digital form as a possible threat to the traditional practices in special collections librarianship. It can also be viewed as an opportunity, but only if the community steps in to manage it. One possibility would be not to leave the conversion of the bulk of the printed literature to UMI and Microsoft. We could instead choose to solicit and undertake the project ourselves. Ten billion dollars was budgeted by the government for the human genome project; how much would it take to convert the unique holdings of the major research libraries in the United States? Would not the ability to be able to offer the equivalent of the Library of Congress to every man, woman, and child in the United States be worth it?

Last, we can look for new collecting areas. I have stressed in this talk the digitization of existing print and manuscript materials, the kind of items that form the bulk of our existing special collections. We may want as well to expand the scope of material that falls under the purview of special collections, especially if collecting, storing, and providing access to the new material is cheaper than traditional special collections material. Which special collections, for example, are actively collecting electronic books? Who has Stephen King's online books or the smaller online literary journals? And who has the Voyager books published on CDs that have been so important in shaping our notion of what a hypertext book is? This kind of material will be collected somewhere. It could be in new units that will develop in libraries to manage twentieth-century material, or it could be in traditional special collections.

It will not be enough just to store and make accessible these electronic files. Special collections librarians can apply to them the same sort of analysis that they use on printed literature. Someone could, for example, investigate the artifactual nature of electronic files: the

ways files are written on disks, the file formats used, and other information that can help date and authenticate electronic information. This is the sort of analysis that special collections librarians do all the time for printed matter; why not apply those same skills to electronic resources? John Gage, chief technology officer at Sun, has recently suggested that establishing the authenticity of various versions of digital files will be the next great technical challenge. This is something special collections librarians know something about, and it could become an area of new expertise.

Conclusion

During the next two to three decades changes in technology will radically alter the way special collections librarians work. The most obvious change will be a dramatic increase in the number of full-text resources available online. Improvements in processing and display technologies that will make reading on a screen (or on electronic paper) comparable to reading a printed book are also likely. Special collections can do nothing to stop these trends. How they adapt to them will determine whether they remain vibrant parts of the library landscape or fall victim to Shiva, the destroyer of worlds.

Notes

A draft of this paper was given at the Fleur Cowles *Flair* Symposium 2000, "The Infinite Library," at the Harry Ransom Humanities Research Center at the University of Texas at Austin on 4 November 2000. This version was presented at the Rare Books and Manuscripts Preconference in San Francisco on 15 June 2001.

1. Terry Belanger, "The Future of Rare Book Libraries," text available from the 16 December 1991 archive of the ExLibris listserv stored at the Conservation Online website <http://palimpsest.stanford.edu>, message from terry@cunixa.cc.columbia.edu, subject: Malkin Lecture.

2. Current figures for Cornell are found at <http://cdl.library.cornell.edu/stat.html>. Figures for Michigan for October 2000 were derived from <http://stats.umdl.umich.edu>.

3. See, for example, W. Brent Seales, James Griffioen, and Kevin Kiernan, "The Digital Atheneum—Restoring Damaged Manuscripts," *RLG DigiNews* 3:6 (15 December 1999), at <http://www.rlg.org/preserv/diginews/diginews3-6.html#technical1>.

4. <http://rose.mse.jhu.edu/>.

5. Gregory Crane, "Designing Documents to Enhance the Performance of Digital Libraries: Time, Space, People and a Digital Library in London," *D-Lib Magazine* 6:7–8 (July–August 2000), at <http://www.dlib.org/dlib/july00/crane/07crane.html>.

6. Anne J. Gilliland-Swetland, "An Exploration of K–12 User Needs for Digital Primary Source Materials," *American Archivist* 61.1 (Winter–Spring 1998): 136–57.

7. The draft report of the Task Force on the Artifact in Library Collections, sponsored by the Council on Library Resources, is just the latest study emphasizing the importance of books as artifacts. See *The Evidence in Hand: The Report of*

the Task Force on the Artifact in Library Collections, available at <http://www.clir.org/activities/details/artifact-docs.html>.

8. Joan Cherry and Wendy Duff, "Improving Access to Early Canadiana: Use of a Digital Collection with Comparisons to Collections of Original Materials and Microfiche. Final Report. December 9, 1999," available at <http://www.fis.utoronto.ca/research/programs/digital/frame.htm>.

9. <http://etext.lib.virginia.edu/ebooks/>.

10. Terry Belanger anticipated this trend in his Malkin lecture in 1991 on "The Future of Rare Book Libraries." He wrote: "In the future, readers are increasingly going to have direct online access to electronic text and data files containing the materials they require; and increasingly, they will perceive that they do not ever need and do not ever want access in printed form to the bulk of this material."

11. Personal communication from Joy Paulson, formerly of the Preservation Department at the University of Michigan.

12. Belanger, "Future of Rare Book Libraries."

13. <http://www.octavo.com>.