

Special Collections at the Cusp of the Digital Age: A Credo

Clifford A. Lynch, Executive Director, Coalition for Networked Information

This essay is an expanded and annotated version of my remarks at the opening of the October 15–16, 2009, ARL-CNI Fall Forum, “An Age of Discovery: Distinctive Collections in the Digital Age.”

Each great research library has its own unique character; special and distinctive collections have always been integral to shaping this character. When ARL came to recognize its semi-sesquicentennial anniversary in 2007, it did so with a magnificent volume titled *Celebrating Research: Rare and Special Collections from the Membership of the Association of Research Libraries*. Such collections link research libraries directly to the core missions of the academy: research, teaching and learning, and public engagement; simultaneously, they represent unique responsibilities for research libraries as stewardship institutions for cultural memory within our society broadly. Leading scholars throughout the centuries have attested to the importance of these collections to research; while humanists are most prominent here, speaking sometimes of great research libraries as the “laboratories of the humanities,”¹ these collections are in fact vital resources across all disciplines, including the sciences.

Such rare and distinctive collections are not, of course, the sole province of research libraries; numerous other academic, public, and special libraries hold important collections of unique materials. Indeed, responsibility for such collections is not limited to libraries: archives, historical societies, and museums have long served as faithful stewards of such collections, and, particularly as we move into the digital age, there is a growing convergence of vision and of opportunities to advance scholarship emerging across the spectrum of such cultural memory organizations. Private collectors have also played a crucial role in the chain of stewardship.

Today, at this forum, we explore the present and the future of such collections. It should be absolutely obvious why the Association of Research Libraries is

sponsoring this forum. But why are the present and future of these collections of such intense and compelling interest to the organization I lead, the Coalition for Networked Information? Our mission is to advance scholarship through the creative use of digital content and advanced information technology. Put simply, special collections are a nexus where technology and content are meeting to advance scholarship in extraordinary new ways. We can see existing special collections are being supplemented and expanded by digital representations of the physical materials; tomorrow's special collections will include a growing proportion of material that has always and only been digital. Information technology is reshaping both stewardship and use of these collections. This essay is a brief, high-level summary of many of the ways in which this is happening; it emphasizes examples rather than comprehensive surveys of developments. Many of the points outlined here are explored in much more depth in sessions at the forum.

First, and foremost, there is the responsibility of stewardship. For our existing special collections, the creation of digital representations² of physical materials offers new pathways to help ensure the survival of the materials in these collections. The digital representations are not substitutes for all purposes, but they can be duplicated and replicated in sites around the world with perfect fidelity and at relatively low marginal cost. The digital representations are both robust and fragile in the way that digital things are, and these strengths and weaknesses are very different from those of the physical collections; given both the physical material and its digital representations, chances are much better that *something* will survive. With the born-digital materials that will comprise tomorrow's special collections, we face new and different challenges in ensuring the long-term integrity and survival of these materials. For cultural memory organizations, these stewardship obligations are paramount—and make no mistake: now that the technology is available and increasingly affordable and well understood, the creation and geographically distributed replication of digital representations of unique treasures is fast becoming an *obligation* of good and responsible stewardship.

Technology is transfiguring our existing physical collections in every dimension: our understanding of the materials, the potential uses and users of the materials, the relationships between the local special collections and the collective worldwide archives of cultural memory. Digital representations are in most regards and for most purposes at least as good as the physical originals (though, as Walter Benjamin has famously observed, they lack the majesty, the aura, of the artifact³). Indeed, practically speaking, digital representations often offer a better engagement

opportunity than the original in museums, archives, or special collections, even in person, and extend that opportunity worldwide through the Internet.⁴

Technological mediation is fundamentally changing scholarship and scholarly practice, from image processing and enhancement, to text mining and information retrieval on large historical corpora. The *Archimedes Palimpsest*, discussed later in this forum, provides a spectacular example of what image processing and enhancement can offer.⁵ These materials can be re-examined and re-integrated through the lenses of modern (digital) technology.

Obviously, we can make fragile materials fully accessible, worldwide, through their digital representations. They can be made available to massive numbers of students, including K–12 and undergraduates engaging in research, as well as to the interested general public. Several talks at the forum will look at the opportunities here. We can also make these materials not only accessible, but re-usable. They can be annotated. Difficult to use, primary scholarly source materials can be transcribed and translated through collective multi-year efforts, perhaps structured to include cadres of students taking a specific course year after year, ultimately producing new critical scholarly editions of these primary sources.⁶ Indeed, providing key source materials through the World Wide Web—particularly content that is complex and impossible to comprehensively describe, such as large collections of historical images—has given rise to vitally important but difficult new curatorial challenges for managers of special collections. These materials evoke and attract a global stream of annotation and commentary, much of it greatly enriching the primary special collections content (for example, by identifying people, places, artifacts, or events depicted in photographs); the

We can re-structure and re-create special collections along logical intellectual lines, and indeed create new “virtual” special collections that facilitate new kinds of scholarly investigation.

volume of this commentary may be too large for the stewards of the collection to effectively even review, and assessing and adjudicating its accuracy may be entirely beyond the capabilities of the hosting institutions. Such widespread attention may in fact give rise to offers of contributions of

related or supplementary materials currently in private hands, or identify linkages to materials held by other cultural memory organizations.⁷

We can re-structure and re-create special collections along logical intellectual lines, and indeed create new “virtual” special collections that facilitate new kinds of scholarly investigation, such as the *Romance of the Rose* collection hosted at Johns Hopkins, discussed later in the forum. This project is trying to image as

many manuscripts of the *Rose* as possible (it expects to have about 150 by the end of 2010); this corpus will allow scholars to trace the evolution and transmission of this key medieval text in new ways, although, interestingly, it will require the development of new tools to permit the parallel examination and analysis of large numbers of variant editions. The re-patriation and re-unification of geographically dispersed special collections is not only possible but increasingly straightforward, and combines cultural diplomacy with new scholarship.⁸ A stunning recent example of the possibilities here is the *Codex Sinaiticus*, which includes the oldest complete New Testament; pages from this work had been scattered across the British Library, the National Library of Russia, St. Catherine's Monastery, and the Leipzig University Library. The pages of the codex, now reunited from all these sources, became available online in 2009.⁹

There are other opportunities to combine stewardship and cultural diplomacy. The British Library, with funding from the Arcadia Foundation, provides an excellent example with the Endangered Archives program. Under this program, the British Library captures digital representations of endangered collections around the world; the library acquires a copy of the representations into its own special collections, while returning another copy to the institution that has responsibility for the endangered (physical) collection.¹⁰ We are seeing efforts to re-create the holdings of national libraries that have been largely destroyed in nations such as Afghanistan.¹¹ We have the potential to redefine relationships between private collectors, scholars, and public collections by digitizing these often-invisible treasures under a wide range of circumstances, either by private agreement or legal mandate (imagine extensions or variations of laws already in place in some nations, notably in Europe, to facilitate the ability of national cultural heritage organizations to retain cultural patrimony being offered for sale by private collectors).

Newly acquired special collections will include more and more digital materials (one prominent recent example is the "papers" of Salman Rushdie, acquired by Emory University, which includes a vast trove of electronic mail).¹² At least at first, the typical case will be digital materials on various portable storage media (floppy disks, tape, hard drives) or even entire personal computer systems, intermixed with printed or other physical materials. The Digital Lives program at the British Library offers a look at the broader range of future complications as, for example, major parts of one's digital life-records move from local storage into cloud-hosted applications or social networking systems.¹³ But while these are an extension of the traditional humanistically focused

collections, documenting the lives and works of important cultural, political, intellectual, or creative organizations or individuals, research libraries will also face a possible convergence or competition with national history museums, disciplinary data archives, and other scholarly memory organizations over massive scholarly and scientific data sets coming from e-research and e-science initiatives. Sayeed Choudhury has argued eloquently, and (in my view) correctly, that these will be an important part of the special collections of the future, though different libraries may choose to place very different levels of commitment on these materials.¹⁴

Let me close by returning to the responsibilities of stewardship. Digital content—whether it be digital representations created from collections of physical materials, or collections of born-digital objects—is both fragile and robust in ways that are very different from purely physical collections. The long-term challenges of preserving digital objects so that they can be meaningfully used in the future are now documented through an extensive literature, and engaged by vibrant worldwide research and development and practitioner communities; steady progress is being made on these very difficult challenges at both technological and operational levels. While the capability of making and distributing perfect digital copies at very low marginal cost offers considerable protection against the natural disasters that have again and again destroyed great physical collections of rare and distinctive materials, human error continues to be a constant and very significant threat to both digital and physical collections.

Less widely recognized are the legal and social challenges within a society that awards little respect to the preservation of cultural memory, or the ways in which the networked information amplifies these challenges; allowing search engines to index a collection on the global Internet attracts legal attacks. Copyright is only one basis for such challenges; others involve libel, privacy, rights to likeness, national security, and even trademarks and patents. And beyond the purely legal, there are cultural conflicts, where some group somewhere demands that material be suppressed, arguing that it is culturally insulting, or perhaps that it represents a part of a body of sacred knowledge.

The battles aren't always legal. As discussed later in this forum, particularly in the haunting presentation by Fred Heath of the University of Texas at Austin, digital collections in areas such as the documentation of human rights violations actually attract sophisticated cyber attacks, the sources of which (state and non-state actors) remain obscure. Special collections hold many types of evidence

that some may want to see destroyed. In the physical world, cultural memory and cultural heritage institutions have all too often been targets in wars between

Effective stewardship of special collections in the digital age will include not just expertise in the curatorial arts and in digital preservation, but also in information security and information warfare, national and international law, diplomacy and public policy.

nations, or in efforts to suppress or control specific populations within a nation. In the digital world these cultural memory institutions can be attacked without crossing the firebreaks into open warfare. Effective stewardship of special collections in the digital age will include not just expertise in the curatorial arts and in digital preservation, but also

in information security and information warfare, national and international law, diplomacy and public policy.

¹ The characterization of the library as laboratory is not new: Christopher Columbus Langdell, appointed Dean of the Harvard Law School in 1870, used it in his "Harvard Celebration Speech," *Law Quarterly Review* 3 (1887): 123. I am indebted to Professor Roy Mersky of the University of Texas at Austin School of Law for educating me in the history of this.

² Until fairly recently, it has been near-universal practice to refer to these digital representations of physical objects as "digital surrogates," a faintly pejorative, sneering phrase that suggests their systematic and intrinsic inferiority to the source physical objects; this is often accompanied by rhetoric implying that *real* scholars *always* need to work with the originals. As I will argue, this is no longer true, at least in a universal and straightforward way, and I've preferred the more neutral term "digital representation" here. I'm grateful to Greg Crane of the Perseus Project at Tufts University for reminding me of the importance of getting the terminology on this right.

³ See "The Work of Art in the Age of Its Technological Reproducibility" (third version, 1939), *Walter Benjamin: Selected Writings Volume 4, 1938–1940*, (Cambridge MA: Belknap Press of Harvard University Press, 2003): 251–283. Note that other translations of versions of this article have used the perhaps-more-familiar title "The Work of Art in the Age of Mechanical Reproduction."

⁴ This is not simply a result of our current ability to take very high-resolution images of manuscripts that are too fragile to handle, though one can readily find endless examples of these today. Very large objects such as sculptures or even buildings can be scanned by lasers to produce extraordinarily high-quality representations. For example, about a decade ago Marc Levoy and colleagues at Stanford University took highly detailed laser measurements of Michelangelo's *David*; the quality was such that the Italian government would not permit the release of the full data set on the Internet; however, the Stanford researchers built a system that allowed viewing of details of specific parts of the statue, including parts that would be inaccessible to a normal museum visitor. See David Koller and Marc Levoy, "Protecting 3D Graphics Content," *Communications of the ACM* 48, no. 6 (June 2005): 74–80; for the general Michelangelo imaging project, see <http://graphics.stanford.edu/data/mich/>. A more recent example, also by coincidence involving Michelangelo, is the *Young Archer* statue from the French Embassy's New York Office for Cultural Services. There's a debate about whether this marble is the work of Michelangelo, and it has gone on 10-year loan for display at the Metropolitan Museum of Art, which has given the embassy a very high-quality three-dimensional copy as a placeholder. See James Barron, "A Statue for a Statue...Sort Of," *New York Times City Room Blog*, October 13, 2009, <http://cityroom.blogs.nytimes.com/2009/10/13/a-statue-for-a-statue-sort-of/>; and Ken Johnson, "Met Asks if Statue Is Work of Genius," *New York Times*, November 6, 2009, <http://www.nytimes.com/2009/11/06/arts/design/06archer.html>.

⁵ See also Reviel Netz and William Noel, *The Archimedes Palimpsest: How a Medieval Prayer Book Is Revealing the True Genius of Antiquity's Greatest Scientist* (Philadelphia: Da Capo Press, 2007) and the Web site <http://www.archimedespalimpsest.org/>. While perhaps the most extensive work has been done in restoring damaged manuscripts and in the study of paintings, the range of opportunities for the creative application of image processing are enormous. For example, by digitizing photographic negatives, it is possible to manipulate the dynamic range of the image to see details that are invisible in the historical prints that accompanied the negatives (a frequently cited project in this area is the work with the glass negatives of the Solomon D. Butcher collection by the Nebraska State Historical Society as part of the Library of Congress American Memory Program). We are beginning to understand that while photographs can be treated as images, photographic negatives might best be thought of as data sets—much like the data sets produced by today's digital cameras in RAW format—that are intrinsically technologically mediated in their use; through this mediation, a digitized negative can produce many different images.

- ⁶ See for example the wonderful work of Christopher Blackwell at Furman University in this area; this is discussed in Christopher Blackwell and Thomas R. Martin, "Technology, Collaboration, and Undergraduate Research," *Digital Humanities Quarterly* 3, no. 1 (Winter 2009), <http://www.digitalhumanities.org/dhq/vol/3/1/000024.html>. The interested reader will also find several other articles of interest dealing with collaborative research and documentation of imaged classics manuscript materials in this issue. See also Blackwell's presentation, "Renewing Scholarship: A QEP-Funded Workshop on Undergraduate Research," http://www.class.uh.edu/mcl/classics/UH_QEP/presentation.html. I'm indebted to Amy Friedlander of CLIR for introducing me to Blackwell's work.
- ⁷ For a well-documented recent example of this, see the report on the experience of the US Library of Congress in mounting image collections on Flickr Commons, available online at http://www.loc.gov/rr/print/flickr_pilot.html.
- ⁸ Clifford A. Lynch, "Repatriation, Reconstruction, and Cultural Diplomacy in the Digital World," *EDUCAUSE Review* 43, no. 1 (January/February 2008): 70–71, <http://www.educause.edu/library/erm08110>.
- ⁹ See <http://www.codexsinaiticus.org/>, and also Helen Shenton, "Virtual Reunification, Virtual Preservation, and Enhanced Conservation," submitted to the Rare Books and Manuscripts, Preservation, Conservation, and Library History division of the 75th IFLA General Conference and Council, Milan, Italy, August 23–27, 2009, <http://www.ifla.org/files/hq/papers/ifla75/163-shenton-en.pdf>. More broadly, see Anne Marie Austenfeld, "Virtual Reunification as the Future of 'codices dispersi': Practices and Standards Developed by e-codices Virtual Library of Switzerland," submitted to the same conference, <http://www.ifla.org/files/hq/papers/ifla75/163-austenfeld-en.pdf>.
- ¹⁰ See the Endangered Archives Web site, <http://www.bl.uk/about/policies/endangeredarch/homepage.html>. See also, for example, the programs to digitize endangered archives in Timbuktu, e.g., John Noble Wilford, "Project Digitizes Works from the Golden Age of Timbuktu," *New York Times*, May 20, 2008, http://www.nytimes.com/2008/05/20/science/20timb.html?_r=1&emc=eta1.
- ¹¹ See the Afghanistan Digital Library, <http://afghanistandl.nyu.edu/>.
- ¹² For a discussion of Rushdie's archives at Emory, see Mary J. Loftus, "Rushdie Hour," *Emory Magazine*, Spring 2008, http://www.emory.edu/EMORY_MAGAZINE/2008/spring/rushdie.html. More generally, see the 2008 white paper by Matthew Kirschenbaum, "Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use," <http://www.neh.gov/ODH/Default.aspx?tabid=111&id=37>.
- ¹³ See the Digital Lives Web site, <http://www.bl.uk/digital-lives/>, in particular the material from the February 2009 First Digital Lives Research Conference, and especially Simson Garfinkle and David Cox, "Finding and Archiving the Internet Footprint," <http://www.simson.net/webprint.pdf>. See also Neil Beagrie, "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine*, June 2005, <http://www.dlib.org/dlib/june05/beagrie/06beagrie.html>. Also see Kieron O'Hara, Richard Morris, Nigel Shadbolt, Graham J. Hitch, Wendy Hall, and Neil Beagrie, "Memories for Life: A Review of the Science and Technology," *Journal of the Royal Society Interface* 3, no. 8 (June 2006): 351–365. And see Gordon Bell and Jim Gemmell, *Total Recall: How the E-Memory Revolution Will Change Everything* (New York: Dutton, 2009).
- ¹⁴ See Sayeed Choudhury and Timothy L. Stinson, "The Virtual Observatory and the Roman de la Rose: Unexpected Relationships and the Collaborative Imperative," *Academic Commons*, December 16, 2007, <http://www.academiccommons.org/commons/essay/VO-and-roman-de-la-rose-collaborative-imperative>; see also Clifford A. Lynch, "The Impact of Digital Scholarship on Research Libraries," *Journal of Library Administration* 49, no. 3 (April 2009): 227–244.

© 2009 Clifford A. Lynch



This article is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/us/>.

To cite this article: Clifford A. Lynch. "Special Collections at the Cusp of the Digital Age: A Credo." *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 267 (Dec. 2009): 3–9. <http://www.arl.org/resources/pubs/rli/archive/rli267.shtml>.