



Research Library Issues

RLI 267 December 2009

A Bimonthly Report from ARL, CNI, and SPARC

A Special Issue on Distinctive Collections

Introduction 1–2

Special Collections at the Cusp of the Digital Age: A Credo 3–9

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

Moving Special Collections Forward in an Age of Discovery:

Themes from the ARL-CNI Forum 10–19

Lisa R. Carter, Head, Special Collections Research Center, North Carolina State University Libraries
and Visiting Program Officer for the ARL Special Collections Working Group

The Collaborative Imperative:

Special Collections in the Digital Age 20–29

Anne R. Kenney, Carl A. Kroch University Librarian, Cornell University

The Changing Role of Special Collections

in Scholarly Communications 30–42

Donald J. Waters, Senior Program Officer, Scholarly Communications, Andrew W. Mellon Foundation

A Special Issue on Distinctive Collections

Introduction

In spring 2009, the ARL Working Group on Special Collections released a discussion report that identified key issues in the management and exposure of special collections material in the 21st century. The report adopted a broad definition of “special collections,” encompassing distinctive material in all media and attendant library services. The main focus was on 19th- and 20th-century materials, including emerging digital materials and media, but most of the report applies with equal force to collecting and caring for materials from previous centuries. While the report focused on special collections in North American research libraries, it has potential application more broadly.

The report provides overviews of and makes recommendations in three areas:

- Collecting Carefully, with Regard to Costs, and Ethical and Legal Concerns
- Ensuring Discovery and Access
- The Challenge of Born-Digital Collections

Overall the report highlights the need for research library leadership to support actions that will increase the visibility and use of special collections and promote both existing and developing best practices in the stewardship of special collections.

In releasing the report, the working group invited discussion among the many professionals who are charged with the perplexing challenges of handling rare, unique, or unusual material about the extraordinary challenges they face as collectors and stewards of special collections in libraries and archives in the 21st century. To stimulate these discussions, ARL hosted a Web conference, a blog, and—in partnership with the Coalition for Networked Information—held a forum in which some 200 people participated.¹

To sustain the growing momentum of community interest in rethinking the roles of distinctive collections, this issue of *Research Library Issues* is devoted to the topic. Included are a few of the outstanding papers delivered at the forum along with pointers to the forum proceedings available on the ARL Web site and other

resources identified by speakers. Also presented is an excellent synthesis of the major themes that emerged at the forum in presentations and audience questions that was written for ARL by North Carolina State University's Lisa Carter.

CNI's Clifford Lynch writes in his essay, "special collections are a nexus where technology and content are meeting to advance scholarship in extraordinary new ways." The papers included here, as well as those referenced, describe some of the "extraordinary new ways" that librarians, archivists, and scholars are coming together and using technology to identify and steward distinctive collections while simultaneously expanding their visibility for use by a worldwide audience. In publishing this special issue of *RLI*, ARL hopes that it will serve to inform and inspire yet more innovative initiatives, and underscore the working group report's key message that this is truly a time of great opportunity to grasp the challenges before us and help shape a glorious future for the extraordinary resources found in special collections in North America.

¹ The working group report, the archive of the Web conference, and the blog are available online at <http://www.arl.org/rtl/speccoll/SCwebcastjuly07.shtml>. The proceedings from the ARL-CNI forum are available at <http://www.arl.org/resources/pubs/fallforumproceedings/forum09proceedings.shtml>.

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>.

Moving Special Collections Forward in an Age of Discovery: Themes from the ARL-CNI Forum

Lisa R. Carter, Head, Special Collections Research Center,
North Carolina State University Libraries and Visiting Program Officer
for the ARL Special Collections Working Group

This year's ARL-CNI Fall Forum, "An Age of Discovery: Distinctive Collections in the Digital Age," highlighted the opportunities special collections provide to engage users and realize the teaching, learning, and research missions of libraries and universities. Two hundred library directors and special collections librarians and archivists gathered at the Mayflower Hotel in Washington DC on October 15–16, 2009, to consider the value proposition of and innovation possibilities inherent in collections of rare books, archives, and other unique materials. The forum was based on the extensive and cumulative work of the ARL Special Collections Working Group and, specifically, its recent discussion report, *Special Collections in ARL Libraries*.¹ Scholars, archivists, librarians and teachers spoke passionately about the value of special collections; how collaborations and integrating special collections into the enterprise can promote their use, garner them increased attention, and provide additional resources; how students develop life-long learning and research skills working with rare materials; and what issues are more prominent now that many of our distinctive collections are available, and even born, digitally. Presentations and recordings from the forum are available from the ARL Web site.²

Several themes recurred throughout the forum, suggesting areas for future exploration, effort, and emphasis, including:

1. Use Drives Special Collections Activity
2. Special Collections Are Central to the Academic Enterprise
3. Digital and Collaboration are Necessary

These themes point to next steps the research library community must take to strategically address the needs of 21st-century researchers—students, faculty, and life-long learners—and connect them to our most unique collections.

Use Drives Special Collections Activity

Use as the driving force in the value proposition of collecting, maintaining, and providing access to special collections surfaced early in the forum and was reinforced repeatedly throughout. From that positioning, speakers articulated the impact that unique, rare, and primary resources are having on learners of every scholarly level by sharing innovative projects and examples from collections. Using engrossing examples as diverse as brilliant illuminated manuscripts, heart-wrenching human rights Web sites, fragmented Buddhist scrolls, and newly revealed Archimedes drawings, presenters addressed the critical matter of getting materials into the hands of users more quickly and in ways that promote dynamic and meaningful advancement of knowledge.

Speakers advocated rethinking strategies for resource allocation, processing and digitization workflows, and promoting special collections in the context of use. Mark Greene urged against “protective thinking” that leads to inefficient processing, highly selective digitization, and delays in expeditious discoverability by the widest audience, including K–12 and undergraduate students alongside “qualified” researchers. G. Wayne Clough shared the Smithsonian’s work to support a learning journey that starts before a visit to a collection, creates tangible memories during, and continues long after, suggesting that collections want to be “petted.” Don Waters promoted framing the investment in special collections and archives in the context of scholarly objectives and improving the efficiency of research.

Several speakers provided evidence of how use is changing with the advancement of digital technologies. Now that digital delivery is an expectation, metadata must facilitate deep discovery and user contribution should be harnessed to enrich future research. In a Web 2.0 environment, special collections need to be findable at the surface of the Web, open for creative reuse and placed well within users’ fluid virtual work spaces. Clough encouraged libraries to expose collections for the challenging, organic aggregations of knowledge that they are.

Increasingly, use can be leveraged to increase future research. Jacqueline Goldsby discussed her success with graduate student scholars working on the

University of Chicago’s Mapping the Stacks project to survey, identify, and process African American collections, lending their subject expertise to under-described collections. Sarah Shreeves described how students participating in



Listen to Jacqueline Goldsby on the value of a collective model of humanist inquiry that breaks down silos and enriches our understanding of mid-20th-century African American literary and cultural studies. [4 min.]

<http://www.arl.org/bm~doc/rli-267-goldsby.mp3>

Photo: David Christopher

the Ethnography of the University project at the University of Illinois at Urbana-Champaign (UIUC) seed future study by depositing their output into the institutional repository. These projects show the added value of engaging students with special collections in a curricular context and illustrated Waters’ observation that bringing users efficiently into processing streams requires solid infrastructure and scholar-friendly data-entry tools.

Summarizing various speakers’ reflections on use, Alice Prochaska noted that special collections are not distinctive just because they are unique but also because of what their stewards do with them to promote use. Placing students and scholars at the center of the value proposition brings strategically built collections into alignment with the academic

mission. Effective assessment methodologies can then drive advocacy and ensure allocation of resources in the broader university environment.

Special Collections Are Central to the Academic Enterprise

Throughout the forum, this call to align special collections with the core mission and activities of the research enterprise provided a revised perspective for addressing the challenges of engaging scholars and advocating for resources.

Several speakers explored aligning special collections with the teaching and learning mission of research institutions. Beginning with the first panel, Steve Nichols acknowledged that traditionally special collections have been viewed as “eccentric” and marginal to undergraduate education, but suggested that they should instead be seen as intimately aligned with the teaching, learning, and research directive of research universities. This exposure must be more than show and tell and integrated into “the fabric of the curriculum,” as Barbara Rockenbach noted. The Boyer Commission report, *Reinventing Undergraduate Education*,³ was referenced by both Rockenbach and Shreeves, who argued that

engaging students with primary sources supports inquiry-based learning, hands-on exploration of meaning, and inquisitive habits of mind. Rockenbach's experiences with advancing the use of special collections in the classroom highlighted the aggressive outreach to faculty that this requires and the importance of promoting special collections as teaching space.

Shreeves discussed UIUC's living-learning project, *Ethnography of the University*, which requires students to employ archival research in analyzing issues in the academic environment they inhabit. Shreeves noted that using archives and publishing in the institutional repository helps students better understand the implications of their original research and the modern research cycle, and that assessments show that students feel more engaged than they do in other class work. Greene described another successful approach in which innovative teaching grants from the American Heritage Center are given to faculty to build undergraduate courses around primary sources.

Alignment is not limited to fostering curricular engagement. The value of special collections can be greatly enhanced by engaging the learning concerns

of broader communities. Special collections can provide intense personal connections for some users. Ian Wilson and Fred Heath reminded the audience that a broader constituency can play a key role in generating support for an institution's mission, whether that constituency is a group of genealogists or human rights advocates. Josh Greenberg gave the example of the unexpected turnout of hundreds of New Yorkers to the New York Public Library's *Design by the Book* debut video-screening party to learn about book arts. Clough spoke of reaching over 4,000 people through the Smithsonian's offering of free, educational, collection-oriented webinars. If the core mission of the research library is to advance knowledge, special collections play a

critical role in advancing inquiry at a most basic level.

As special collections increasingly move to the center of the research library, activities and resources that increase their accessibility must be mainstreamed. Assets, skills, and talents across the institution can be leveraged to expose distinctive collections alongside other information resources. Speakers noted



Listen to Ian Wilson describe how special collections offer leadership opportunities in the e-world. [4 min.]

<http://www.arl.org/bm~doc/rli-267-wilson.mp3>



that permanent funding, along with requirements for collaboration, efficiency, and policy development must come from the top. Archivists and special collections librarians must constantly demonstrate alignment with the organizational mission, engage institutional colleagues and provide evidence of how less-compelling topics contribute to research, teaching, and learning. Ken Hama specifically noted the slow, iterative nature of incorporating special collections into the enterprise, requiring persistent reframing of special collections in the context of institutional goals. The current financial environment precipitates the need to share models for integrating special collections into the main information-management and discovery workflows in the research library.

Digital and Collaboration Are Necessary

Special collections present opportunities for research libraries to enrich transformations affecting research and research-intensive institutions. Libraries and their special collections have a strong record of embracing the digital information environment and collaboration to support changing modes of research. Yet presenters urged that more attention be given to developing and ensuring sustainability and building solid infrastructure in both areas.

Speakers promoted the digitization of collections as key to connecting special collections with users, but quickly moved beyond the advantages of digital surrogates to the expanded options presented by the digital world. Collections can be transformed online; for example, providing access to details that cannot be seen with the naked eye and reuniting pieces of disparate collections virtually. Will Noel shared astonishing examples of digital transformation and rediscovery of unique items based on his work with the Archimedes Palimpsest and other projects at the Walters Art Gallery. Richard Saloman highlighted the emerging capability to create globally interacting, digital, special collections in his description of the Early Buddhist Manuscripts Project. These and other presentations demonstrated that real transformation occurs when digital collections are turned over to the users for unexpected interpretation and reuse.

In order to facilitate original and creative knowledge building such as non-consumptive research, mash-ups, and “citizen science,” libraries must do more to establish a stable foundation on which users can work. Large-scale digitization; open access; flexible, minimal, automated metadata; user-friendly

tools; and effective policy setting can allow user subject experts to participate in exposing collections. Libraries will need to more effectively use existing tools and infrastructure and innovate new solutions where necessary. Greenberg encouraged libraries to “write for Google” and “write for referral” to surface collections in researchers’ pathways. Tracy Seneca offered a different twist, reporting on a repurposable method of creating digital archives by harvesting openly accessible Web sites. Speakers gave a wide variety of examples demonstrating how repurposing the context of use provides meaning in a way that is at the heart of the research enterprise. Anne Kenny discussed how special collections can be used to build or enrich digital communities and how passionate those communities become about the digital collaborative space. Yet to enable this knowledge building, libraries must move away from one-off, boutique, digital projects to solidify digital programs. As special collections face the challenges of sustaining digital surrogates and born-digital materials, permanent funding and infrastructure must be allocated to ensure that those collections are accessible and authentic for the long run.

Digitization and digital curation are no longer specialized activities; they are a part of the life-cycle management of special collections. The challenging but critical tasks for success are the policy setting, infrastructure building, and

training. Speakers acknowledged that while this work must take place in each library, this work will not be successfully undertaken by libraries acting alone. Throughout the forum, speakers provided concrete examples of successful and innovative collaborations within and across institutions and between institutions and collections users. Kenney advocated for collaborative strategies that bind research libraries together as we work with commercial partners on large-scale digitization of special collections. Greenberg agreed that we must work as a broader, online ecosystem. Waters called for new and reliable ways to link collections across institutional boundaries. Reflecting on the successes of the MetaArchive Cooperative, Katherine Skinner



Listen to Katherine Skinner on the need to engage in strong collaborative networks to preserve and expose the digital cultural record. [4 min.]

<http://www.arl.org/bm~doc/rli-267-skinner.mp3>

noted that we need to create durable associations of autonomous entities collaborating to achieve common or compatible goals while maintaining

flexibility. Competition and protectionist attitudes must give way to institutional humility and stronger collaborative networks. After all, our users do not care which institution owns the original or provides the digital surrogate, they just want unfettered access.

Further, purposeful collaboration takes effort, flexibility, and persistence to achieve the full potential of cooperative activity. Will Noel noted that data management is a major challenge for cooperative efforts, while Skinner added that collaboration demands an organizational structure in order to work. Discussion emphasized that development of infrastructure for collaboration is best kept lightweight, distributed, and virtual, keeping in mind that open, dynamic collaboration is useful for access and exposure, while a closer, constricted organization is necessary to protect and sustain collections. Regardless, consensus confirmed that programmatic digital and collaboration infrastructure are *the* key investments in effectively connecting researchers with distinctive collections.

Conclusion: An Investment in the Knowledge Economy

In closing, Ian Wilson reminded the audience that the future of special collections offers opportunities for leadership at the edge of evolving research practice. Taking advantage of the virtual space is like exploring a new land, where shared risk and calculated investment can result in significant impact. The way scholars learn and process knowledge is changing. Users now commit to search strategies rather than memorization of facts and details. The opportunity to engage the learning process via the raw materials of knowledge, rare objects, and primary sources, is greater than ever before. Moving forward, investment in special collections will require user-centered mission alignment, resource reallocation towards mainstreaming and sustainability, and the commitment and trust-intensive work of collaboration; but such an investment offers a rich, rewarding, and transformative contribution to advancing knowledge.

¹ *Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections* (Washington DC: ARL, 2009), <http://www.arl.org/bm~doc/scwg-report.pdf>.

² An Age of Discovery: Distinctive Collections in the Digital Age, Proceedings from the ARL-CNI Fall Forum, October 14–15, 2009, Washington DC, <http://www.arl.org/resources/pubs/fallforumproceedings/forum09proceedings.shtml>.

³ The Boyer Commission on Educating Undergraduates in the Research University, *Reinventing Undergraduate Education: A Blueprint for America's Research Universities* (Boyer Commission, 1998), <http://naples.cc.sunysb.edu/Pres/boyer.nsf/>.

An Age of Discovery: Distinctive Collections in the Digital Age: Proceedings from the ARL-CNI Fall Forum, October 14–15, 2009

Freely available on the ARL Web site at <http://www.arl.org/resources/pubs/fallforumproceedings/forum09proceedings.shtml>

Welcoming Remarks

Brinley Franklin, Vice Provost, University of Connecticut, and President, ARL
Audio [MP3 2 min.]

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

Clifford Lynch, Executive Director, CNI
PDF & Audio [MP3 5 min.]

Why Are Special Collections So Important? Exploring the Value Proposition of Special Collections

Introductory Remarks

Mark Dimunation, Chief of the Rare Book and Special Collections Division, Library of Congress
Audio [MP3 9 min.]

Why Special Collections Matter Now More than Ever Before

Stephen G. Nichols, Professor of French and Humanities, Medieval French Literature, Johns Hopkins University
Audio [MP3 31 min.]

Existential Archives: Looking for the Value Propositions of Archives and Special Collections

Mark A. Greene, Director, American Heritage Center, University of Wyoming
PDF & Audio [MP3 30 min.]

The Changing Role of Special Collections in Scholarly Communications

Donald J. Waters, Senior Program Officer, Scholarly Communications, The Andrew W. Mellon Foundation
PDF & Audio [MP3 26 min.]

Building on Our Strengths: Opportunities for Special Collections in the Digital Age

Introductory Remarks

Nancy E. Gwinn, Director, Smithsonian Institution Libraries

Keynote Remarks: An Age of Discovery: Distinctive Collections in the Digital Age

G. Wayne Clough, Secretary of the Smithsonian Institution
PDF & Audio [MP3 29 min.]

Reactor Comments:

Ian E. Wilson, Librarian and Archivist of Canada Emeritus, President of the International Council on Archives, Strategic Advisor, University of Waterloo
Audio [MP3 10 min.]

Alice Prochaska, University Librarian, Yale University
Audio [MP3 8 min.]

Collaboration to Build Cross-Institutional Collections

Introductory Remarks

Bill Landis, Head of Arrangement and Description & Metadata Coordinator in Manuscripts and Archives, Yale University Library
Audio [MP3 2 min.]

Collaboration to Build Cross-Institutional Collections: Ancient Scrolls and Stone Tablets

Richard Saloman, Professor, Sanskrit and Prakrit Language and Literature, Epigraphy, Ancient Indian History, and Gandharan Studies, Department of Asian Languages and Literature, University of Washington
Slides [PPS 38.7 MB] & Audio [MP3 25 min.]

MetaArchive Cooperative: A Collaborative Model for Digital Preservation

Katherine Skinner, Executive Director, Educopia Institute and Program Manager, MetaArchive Cooperative
Slides [PPS 3.8 MB] & Audio [MP3 27 min.]

Lessons from Two Inter-Institutional Projects: the Islamic Digital Resource Project at the Walters Art Museum and the Archimedes Palimpsest Project

Will Noel, Curator of Manuscripts and Rare Books, Walters Art Museum
Audio [MP3 21 min.]

Integrating Special Collections into the Enterprise**Introductory Remarks**

Susan K. Nutter, Vice Provost and Director of Libraries, North Carolina State University
Audio [MP3 3 min.]

Integrating Special Collections into the Enterprise: A Case Study of the Yale Center for British Art

Ken Hamma, independent consultant working with the Yale Center for British Art, the Yale Office of Digital Assets and Infrastructure, and the Andrew W. Mellon Foundation; formerly Executive Director for Digital Policy and Initiatives of the J. Paul Getty Trust
PDF & Audio [MP3 18 min.]

Special Collections in the Digital Age: Partnerships and Collaboration

Fred Heath, Vice Provost and Director, University of Texas Libraries
Slides [PPS 23.2 MB] & Audio [MP3 27 min.]

Special Collections as Laboratories for Researchers and Students**Introductory Remarks**

Sarah M. Pritchard, University Librarian, Northwestern University
Audio [MP3 4 min.]

Mapping the Stacks: A Guide to Black Chicago's Hidden Archives

Jacqueline Goldsby, Associate Professor, Department of English & the College, University of Chicago, and Visiting Associate Professor, Department of English, New York University
Audio [MP3 18 min.]

Archives, Undergraduates, and Active Learning: Case Studies from Yale University Library

Barbara Rockenbach, Director of Undergraduate & Library Education, Yale University Library
Slides [PPS 3.7 MB] & Audio [MP3 21 min.]

Student Research on the University, in the Archives, and in the IR

Sarah L. Shreeves, Coordinator, Illinois Digital Environment for Access to Learning and Scholarship (IDEALS), University of Illinois at Urbana-Champaign
Slides [PPS 4.4 MB] & Audio [MP3 18 min.]

What Changes with Digital Content?**Introductory Remarks**

Clifford Lynch, Executive Director, CNI

What Changes with Digital Content?

Josh Greenberg, Director of Digital Strategy and Scholarship, New York Public Library
PDF [13.4 MB]

What Changes with Digital Content? Three Key Issues

Anne R. Kenney, University Librarian, Cornell University
PDF & Slides [PPS 5.2 MB]

What Changes with Digital Content? Web Archiving

Tracy Seneca, Web Archiving Coordinator, California Digital Library
Slides [PPS 2.6 MB]

Concluding Keynote Remarks

Ian E. Wilson, Librarian and Archivist of Canada Emeritus, President of the International Council on Archives, and Strategic Advisor, University of Waterloo
Audio [MP3 26 min.]

Alice Prochaska, University Librarian, Yale University
Audio [MP3 3 min.]

Appendix

Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections, March 2009
PDF

The Unique Role of Special Collections, Special Collections Web Conference, July 7, 2009, Archive

© 2009 Lisa R. Carter



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: Lisa R. Carter. "Moving Special Collections Forward in an Age of Discovery: Themes from the ARL-CNI Forum." *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 267 (Dec. 2009): 10–19.
<http://www.arl.org/resources/pubs/rli/archive/rli267.shtml>

The Collaborative Imperative: Special Collections in the Digital Age

Anne R. Kenney, Carl A. Kroch University Librarian, Cornell University

The recent ARL-CNI Fall Forum, “An Age of Discovery: Distinctive Collections in the Digital Age,” offered a wonderful venue to reconsider the role of special collections—both physical and digital—in an age that places emphasis on ubiquitous access, social networking, and the promise of Web 2.0. The forum included presentations by scholars, librarians, curators, and technologists, and the place was packed. A number of themes emerged during presentations and the ensuing Q&A sessions, including calls for greater collaboration across institutional boundaries as well as with content creators and users. It was not lost on most attendees that the kinds of issues being discussed would not have been seriously considered by many even five years ago. Much has changed, but one thing is clear: as special collections face a new renaissance in the Digital Age, research libraries are challenged to reconsider institutional practice, and especially the collaborative imperative that connects institutions, digital communities, and the users we serve.

Digitizing Special Collections

“...large-scale digitization is an exciting option that will almost certainly become a fact of life for a significant number of special collections librarians and archivists in the near future.”¹

The recent report on special collections in ARL libraries noted that the focus of large-scale digitization increasingly will be on special collections materials as the sweep to digitize general stack collections comes to an end. Certainly special collections have been digitized over the past two decades, but the scope and expense associated with mass digitization is out of reach for most research libraries without external funding or external partners. The collaborative imperative should bind research libraries together as we move into the era of mass digitization of special collections and permeate our relations with

commercial partners. When it comes to rare and unique materials, libraries and archives should be able to negotiate from a position of strength, which will be enhanced if we approach this collectively. Simply put, can we resist the

Principles to Guide Large-Scale Digitization of Special Collections

Principle 1: Distinct collections demand extra vigilance in digitization.

Principle 2: Libraries must respect any donor-imposed restrictions on the digitization and use of materials.

Principle 3: Libraries should seek the broadest possible user access to digitized content. This includes patrons of other libraries and unaffiliated researchers.

Principle 4: Libraries should receive copies of all digital files generated from their collections, with the option for complete local access to the files (to the extent that copyright law allows).

Principle 5: Any enhancements or improvements to the digitized content should be shared on a regular basis with the supplying library.

Principle 6: Restrictions on external access to copies of works digitized from a library’s holding should be of limited duration.

Principle 7: Libraries should refrain from signing nondisclosure agreements (NDAs) as part of digitization negotiations.

Principle 8: Libraries should ensure that the confidentiality of users is protected in the vendor’s products.

Principle 9: Libraries should refrain from charging fees or royalties for access to or non-commercial use of public domain materials held in their collections.

The author wishes to acknowledge the work of Peter B. Hirtle, Senior Policy Advisor, Cornell University Library, in developing this set of principles.

temptation to enter into special deals for special collections digitization that may offer short-term gains but ultimately be of disservice to our institutions and our users?

At the May 2009 ARL Membership Meeting, approximately 100 member directors participated in a real-time survey that involved the use of clickers and a set of questions focusing on multi-institutional collaboration.

Among questions posed was one in which they were asked whether they would be “willing to commit my institution to forego one-on-one arrangements with commercial entities around digitization of special collections materials in favor of collective arrangements involving multiple research libraries.” Their responses were encouraging: 89% of audience members either strongly agreed (56%) or agreed (33%) with this statement. Only 11% disagreed and only 1% strongly disagreed.

What might be considered critical in collective arrangements governing contracts for the mass digitization of special collections? Following the May ARL meeting, a number of us started a

list of principles and came up with nine. I want to acknowledge especially the work of Peter Hirtle (Cornell) who drafted these statements and Madelyn Wessel (University of Virginia) who provided invaluable comments. Thanks

are also due to a number of ARL library directors and other professional colleagues who have reviewed the list.²

Principle 1: Distinct collections demand extra vigilance in digitization.

When digitizing distinct collections, special attention should be paid to the nature of the material being digitized. No blanket digitization standard should be applied to all materials. Instead the inherent characteristics of the items should determine the level of care. Rare, unique, or fragile items should be digitized according to the highest professional standards in terms of handling, security, and scan quality. This may often require on-site conversion with specialized equipment. Digitization should be conducted in a way such that it is not necessary to revisit the process in the future as repeated digitization may lessen the artifactual value of originals.³ Material that is more common or does not contain significant artifactual integrity can be digitized in a fashion that fosters widespread access to the most amount of material.

Principle 2: Libraries must respect any donor-imposed restrictions on the digitization and use of materials.

Special collections material is often acquired from donors with express limitations on its use, even when that material is ostensibly in the public domain. In negotiating with commercial vendors, libraries must ensure that the terms of any applicable donor agreement are respected. In negotiating with donors, librarians and archivists should educate them about the desirability of making materials accessible online.

Principle 3: Libraries should seek the broadest possible user access to digitized content. This includes patrons of other libraries and unaffiliated researchers.

For over a century, libraries have participated in successful resource-sharing cooperatives that have made content widely accessible. The same spirit should govern commercial digitization activities. Libraries should resist arrangements that result in onerous subscription charges for access to resources digitized from their collections. In the best of all possible worlds, there would be some level of free access to all content, with only special value-added services restricted to a subscription model.

Principle 4: Libraries should receive copies of all digital files generated from their collections, with the option for complete local access to the files (to the extent that copyright law allows).

Libraries should insist on the right to provide free local access to digitized materials from their holdings. They should determine on their own what constitutes a fair use of those digital files and make them available accordingly. Nothing in the contract with the commercial entity should limit the library's right to make a fair use determination. Material that is of uncertain copyright status should be excluded from commercial products.

Principle 5: Any enhancements or improvements to the digitized content should be shared on a regular basis with the supplying library.

In addition to making material available to the public, research libraries should seek to provide context to aid in the understanding of that material. This is especially true with special collections materials, which often must be interpreted or analyzed. In order to preserve and provide context for digitized distinct collections, it is important that the contributing library receive on a regular basis copies of enhanced content and metadata about that content. This could include upgraded or replaced image files as well as corrected or improved OCR text.

Principle 6: Restrictions on external access to copies of works digitized from a library's holding should be of limited duration.

In order to allow a commercial partner time to recover its investment in digitization, it may be necessary to grant to that entity exclusivity over the use of the digital files for a period of time. The ultimate goal, however, is to "ensure the results are widely available for scholarship."⁴ The period of exclusivity, therefore, should be limited, preferably to no more than seven years. After that time period the library should be able to distribute freely any file digitized from its holdings. It should also be able to aggregate the content with other resources from its own collections and those of other institutions and to expose the content to data mining and other new ways of exploiting it.

Principle 7: Libraries should refrain from signing nondisclosure agreements (NDAs) as part of digitization negotiations.

At the ARL Membership Meeting, library directors were asked whether they

would be willing to commit their institution “to making public the content of publisher agreements, including pricing, special arrangements, and other privileges.” Thirty-five percent of audience members indicated they would commit to public disclosure “under all circumstances,” and forty-four percent indicated they would “under most circumstances.” The ARL Board also supported a resolution from the Scholarly Communication Steering Committee to “strongly encourage ARL member libraries to refrain from signing agreements with publishers or vendors, either individually or through consortia, that included non-disclosure or confidentiality clauses.” The values of transparency and public disclosure that underlie state open records laws should guide library transactions whether their home institutions are public or private. Libraries should respect that commercial partners may need to protect certain business and technological secrets, but not agree to keep licenses or core financial arrangements confidential. Libraries must “insist on their own right to discuss aspects pertaining to their broader community.”⁵

Principle 8: Libraries should ensure that the confidentiality of users is protected in the vendor’s products.

The confidentiality of usage data is one of the guiding principles of the library profession. In almost every state, library usage data is also protected by law. If a library digitized and made accessible to its patrons resources from its holdings, it should hold in confidence any personally identifiable information associated with the use of that material. The same principle should apply to material digitized by a commercial entity working in partnership with the library. Libraries should insist that personally identifiable information is scrubbed from commercial log files and content management systems. Alternatively, commercial systems must offer library patrons the option of reading and working anonymously.

Principle 9: Libraries should refrain from charging fees or royalties for access to or non-commercial use of public domain materials held in their collections.

The combination of digitization technologies and Internet distribution can radically transform how researchers make use of special collections materials. As the Budapest Open Access Initiative has noted, “removing

access barriers...will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.”⁶ The most creative uses of our shared cultural heritage can only occur, however, if the public has the ability to access and use public domain source materials without onerous permissions processes or the imposition of fees. Therefore, in the spirit of the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, all non-commercial users should have “a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship.”⁷ If fees are to be assessed for the use of digitized public domain works, those fees should only apply to commercial uses. Cornell University Library recently lifted all restrictions on the use of its public domain reproductions.⁸

Building Digital Communities

As we move to digitize special collections on a massive scale, we should not ignore the broader ecosystem of the Internet that incorporates social networking in the use of content, as exemplified by Wikipedia and Flickr Commons. Providing effective digital access to the treasures of research libraries will require us to appreciate—and accommodate—digital communities. Research libraries have the opportunity to build community around content, to build content around community, and to provide a home for digital creators. Several examples illustrate these points.

FamilySearch.org is a service provided by the Church of Jesus Christ of Latter-Day Saints. This Web site provides a gateway to the millions of genealogical records that the church has gathered and made available, increasingly online. In early 2006, FamilySearch.org provided an online tool for volunteers to index digital images of vital records. In April 2009, a major milestone was reached when the 250-millionth record was indexed by one of the over 100,000 volunteers from around the world. Each record is actually indexed by two individuals for accuracy, with discrepancies checked by a third person. Currently, FamilySearch.org reports that volunteers are indexing over a million names per day.⁹ Mass digitization requires mass metadata creation and, by building digital communities around content, the work is being done quickly.

Online content plus a motivated community equals a very powerful collaboration.

Research libraries can also build content around communities. The Institute of Museum and Library Services has recently funded an innovative project at the University of California, Santa Cruz to digitize material from the Grateful Dead Archive and make it available on a Web site, the Virtual Terrapin Station. What makes this project so exciting is that the “Deadhead” community is invited to participate in building the collection. The Grateful Dead were distinctive in allowing fans to photograph and record their concerts. The project will provide tools to the public to encourage their contributions and the curation of a large, socially constructed archive. The group’s musical legacy is worthy of preservation, but so too is the social/cultural phenomenon surrounding them. By building content around this community of fans, a much richer historical record will be preserved and made accessible. This ability to connect content and community is a key theme in the digital domain. Other examples of this phenomenon are seen in the development of “digital memory banks” to document significant events, such as those that have been established to upload stories, photographs, and documentation on 9/11 and hurricanes Katrina and Rita.¹⁰

A third example of building community is represented in the Rose Goldsen Archive of New Media Art.¹¹ The Goldsen Archive has provided a safe harbor for an international community of independent digital media artists for close to a decade. The Web site offers community and private space for artists to work, share, discuss, experiment, vent, adapt, perform, exhibit, and preserve their work. A key to the site’s success is providing a trusted, secure environment for a highly distributed fringe group of creative artists. This international community of over 1,250 artists and theorists working at the edge of contemporary practice connect with each other through an online new media e-mail list, -empyre-. The University of New South Wales in Australia maintains the server for the group and the current moderators are two faculty members at Cornell University. This past January, the group was asked to share New Year’s resolutions on new media art. Many responded, including a digital artist from India who wrote, “I promise I’m not afraid of prolonged power cuts.” Another from Europe vowed that “between 1:00 a.m. and 4:00 a.m., I will not delete files I think I won’t need again.” And a third, somewhat jaded artist from Australia commented, “I’m glad hardly any institutions really bothered to collect Internet art—it will make it so much more valuable in the future.” These and other postings revealed how

vibrant this digital arts community is, but also how vulnerable and suspicious it can be. As research libraries work to document such movements, it will be critical for librarians and archivists to engage sensitively with the community to help preserve and protect its work. We may not have all the answers but, without this engagement, we might not even know what questions to ask.

Serving Users

Joshua Greenberg, Director of Digital Strategy and Scholarship at the New York Public Library (NYPL), recently called NYPL “the library of the unaffiliated,” an epigram that suggests the future of all research libraries in the digital world. Mass digitization of special collections and online access can lead to mass consumption. Users will come in all shapes and sizes, with varying needs and levels of preparation, and they will come from everywhere. Their numbers will extend well beyond the scholarly community. Over the past nine years, Cornell has provided digital access to documentary evidence on the tragic Triangle Factory Fire of 1911 that took the lives of 146 young women and girls.¹² The materials include content from the records of the International Ladies’ Garment Workers’ Union and other unions, photographs, first-hand testimonies from survivors, and documentation on the resulting investigations and reforms. The site was originally created to respond to the steady flow of requests for information on the Triangle Fire that Cornell received from middle- and high-school students. A visitors’ book added in 2001 to the site contains hundreds of postings, including very recent ones, which help document the diversity of audiences served. Users include students, teachers, scholars, political activists, family members of the women who worked in the thread and needle trades, as well as fire marshals from around the country who have made the site mandatory reading in their training programs.

Most users are profoundly grateful for digital access to such resources, but increasingly the “unaffiliated” are beginning to expect special collections Web sites to provide services and support comparable to what they can obtain elsewhere on the Internet. Recently, Cornell upgraded its Making of America (MOA) digital collection of 19th- and early 20th-century materials¹³ and, as is often the case, there were some bugs that needed to be fixed. A number of habitual users of the site (most of whom are not affiliated with Cornell) complained and the staff responded to the satisfaction of most of them. One power user, however, went on the offensive when it looked like a certain feature

would take more time to restore. She wrote to colleagues on several e-mail lists, urging them to contact Cornell and “demand its restoration.” She added: “This is really too important to take sitting down.” I’ll spare you my first reaction to this e-mail, but it did raise the question of how many library resources should be devoted to non-Cornellians. When MOA first launched, it was argued that, if we are going to make material digitally available to the Cornell community anyway, it was a marginal overhead to make it accessible to the world as well. As it turns out, that’s not quite true. The vast majority of users of our digital content have no Cornell relationship and, when they write with their concerns and their questions, we devote IT and reference staff resources to them—on a fairly steady basis. Most express gratitude to Cornell for making this material accessible. What’s interesting about this current situation is the user’s sense of entitlement. I might resent her tone, but she’s got a point. If we are going to offer up our holdings to the world, we have an obligation to meet certain expectations.

Making material freely available does not make it free. Balancing our commitments to open access and responsible stewardship of our institutional resources in these hard economic times requires dedication, flexibility, and a rethinking of business as usual.

Conclusion

Earlier this fall, Katherine Reagan spoke at a Grolier Club symposium on “Books in Hard Times” and characterized two possible fates awaiting special collections:

1. The Special Collection Grave-Yard, where physical items go to reside and are rarely used once they are digitized, and
2. The Special Collections Renaissance, in which digital access leads to ever greater use of the originals.¹⁴

Whatever the future is along this spectrum, research libraries will need to consider the changes that attend digital access on a grand scale. As we transplant special collections to an online environment, we should avoid the temptation to transplant traditional approaches that do not accommodate the profound differences that await us, where institutional borders blur, where digital communities thrive, where the unaffiliated seek to use our materials in ways not fully imagined. Our success may well depend on our ability to seize the collaborative imperative that links institutions to a participatory information environment.

- ¹ *Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections* (Washington DC: ARL, 2009): 22, <http://www.arl.org/bm-doc/scwg-report.pdf>.
- ² Others who reviewed these principles include Paul Courant (University of Michigan), Chuck Henry (Council on Library and Information Resources), Wendy Lougee (University of Minnesota), Cliff Lynch (Coalition for Networked Information), Maura Marx (Open Knowledge Commons), James Neal (Columbia University), Merrilee Proffitt (OCLC), Bernie Reilly (Center for Research Libraries), Brian E. C. Schottlaender (University of California, San Diego), Winston Tabb (Johns Hopkins University), Karin Wittenborg (University of Virginia), and Ann Wolpert (Massachusetts Institute of Technology).
- ³ Task Force on the Artifact in Library Collections, *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections* (Washington DC: Council on Library and Information Resources, 2001), <http://www.clir.org/pubs/reports/pub103/contents.html>.
- ⁴ Merrilee Proffitt and Jennifer Schaffner, *The Impact of Digitizing Special Collections on Teaching and Scholarship: Reflections on a Symposium about Digitization and the Humanities* (Dublin, Ohio: OCLC Online Computer Library Center, 2008), <http://www.oclc.org/research/publications/library/2008/2008-04.pdf>.
- ⁵ Peter B. Kaufman and Jeff Ubois, "Good Terms—Improving Commercial-Noncommercial Partnerships for Mass Digitization: A Report Prepared by Intelligent Television for RLG Programs, OCLC Programs and Research," *D-Lib Magazine* 13, no. 11/12 (November 2007), <http://www.dlib.org/dlib/november07/kaufman/11kaufman.html>.
- ⁶ Budapest Open Access Initiative, February 2002, <http://www.soros.org/openaccess/read.shtml>.
- ⁷ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, October, 2003, <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>.
- ⁸ Peter Hirtle with Tricia Donovan, "Removing All Restrictions: Cornell's New Policy on Use of Public Domain Reproductions," *Research Library Issues*, no. 266 (October 2009): 1–6, <http://www.arl.org/resources/pubs/rli/archive/rli266.shtml>.
- ⁹ See official press releases at http://www.familysearch.org/eng/home/News/frameset_news.asp?PAGE=home_press.asp.
- ¹⁰ See, for example, <http://hurricanearchive.org/>.
- ¹¹ Rose Goldsen Archive of New Media Art, <http://goldsen.library.cornell.edu/>.
- ¹² The Triangle Factory Fire, <http://www.ilr.cornell.edu/trianglefire/>.
- ¹³ The Making of America, <http://digital.library.cornell.edu/m/moa/>.
- ¹⁴ Katherine Reagan, "Books in the Age of Anxiety," presentation at the "Books in Hard Times" Grolier Club Symposium, September 22, 2009.

© 2009 Anne R. Kenney



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: Anne R. Kenney. "The Collaborative Imperative: Special Collections in the Digital Age." *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 267 (Dec. 2009): 20–29.
<http://www.arl.org/resources/pubs/rli/archive/rli267.shtml>.

The Changing Role of Special Collections in Scholarly Communications

Donald J. Waters, Senior Program Officer, Scholarly Communications,
Andrew W. Mellon Foundation

Presented at the ARL-CNI Fall Forum on "An Age of Discovery: Distinctive Collections in the Digital Age," Washington DC, October 14, 2009

In October 1995, I was co-chair of the Task Force on the Archiving of Digital Information. I stood before the assembled membership of ARL and reported dutifully on the progress of the task force.¹ I noted in my talk then that William Safire had recently devoted his wonderful "On Language" column in the *New York Times Magazine* to the topic of kids' slang. He advised that "if you want to stay on the generational offensive, when your offspring use the clichéd 'gimme a break,' you can top that expression of sympathetic disbelief with 'jump back' and the ever-popular riposte 'whatever.'" However, he also noted that some expressions, such as "I'm outta here" or "I'm history," had become very much dated. Quoting from a study of slang, Safire pointed out that "I'm history," is "a parting phrase modeled on an underworld expression referring to death, and it has both inspired and been replaced by the more trendy expression, 'I'm archives.'"²

Today, according to a recent article in the "Sunday Styles" section of the *New York Times*, the trendy have taken their slang to an even higher level of sophistication. They are now studiously avoiding being associated with mundane activities such as "hosting" or "selecting," and are instead opting to engage in the more up-to-the-minute and stylish activity of "curating." The Oxford dictionary defines the standard meaning of "to curate" as "to look after and preserve." However, this sense of the word has been supplemented with a variety of non-traditional uses. The *Times* reported that "The Tipping Point, a store in Houston that calls itself a sneaker lifestyle shop, does not just sell a collection of differently colored rubber soles....No, its Web site declares, the store 'curates' its merchandise." Similarly, "Etsy, the shopping Web site devoted to

handmade and vintage goods, routinely brings in shelter magazine editors, fashion designers and design bloggers to serve as ‘guest curators.’” And “promoters at Piano’s, a nightclub on the Lower East Side [of Manhattan], have recently announced on their Web site that they will ‘curate a night of Curious burlesque.” Now if all of your competitors are “curating” merchandise, you do not want to be known as someone who merely “buys and sells” and, similarly, if all your rival nightclub promoters are “curating” parties, why in the world would you want to be left to be merely “hosting” one?³

In 1995, I was simply astounded at how change in popular jargon was so closely tracking a controversial definitional change in more esoteric circles. You’ll remember that one of the results of the task force was to loosen the definition of archival practice and extend some of its core concepts to define the practice of collecting and preserving digital information.⁴ This definitional extension has now largely been accepted and even superseded, but at the time of its formulation, it was met with howls of protest from purists who felt that the task force was demeaning the value of true archival work by describing work on the ephemera of bits and bytes in the same terms. Find your own word, they said.⁵ And today here we go again as the popular culture is closely tracking a more esoteric extension of the meaning of the term “curation” from museum practice to the definition of how effectively to manage and preserve floods of digital data produced by sensors of various kinds including telescopes, gene sequencers, and book scanners.⁶

What, if anything, do these various semantic extensions say about the value today of special collections, whether in artifactual or digital form? I will return to this specific question at the end of this paper. In the meantime, I want to explore some ideas about how best to construct the value proposition justifying investment in special collections, and about the areas of work that are likely to be most fruitful to advance scholarly communications.

The Definition of Special Collections

“Special collections” is used in various senses for various purposes, sometimes referring simply to rare books and manuscript materials, and sometimes more generally to materials that are used as primary sources of evidence as opposed to secondary sources. In the recent working group report on *Special Collections in ARL Libraries*, “special collections” are defined “ecumenically” to include “any kind of vehicle for information and communication that lacks readily available

and standardized classification schemes, and any that is vulnerable to destruction or disappearance without special treatment.”⁷ In this sense, special collections are those materials containing primary evidence for scholarship that require special treatment in their description or handling.

A value proposition is important because the costs of these special treatments can be quite substantial. At its most simplistic, the value proposition for special collections is that scholarship broadly across fields in the humanities, social sciences, and the sciences just cannot proceed without corollary investment in the acquisitions and carrying costs of the primary-source evidence needed to sustain and advance those scholarly fields. But how can or should a particular institution justify particular investments in particular kinds of collections? Tomes have been written on this more specific question. Institutional missions, areas of special expertise, previous investment in particular areas of scholarship, growth trajectories in new areas, and special opportunities presented by relationships with donors and private collectors are all among the factors that play a role in particular value propositions.⁸ It is undoubtedly the complex nature of the interaction of these factors that accounts for the wide and rich variation among research libraries and archives in the kinds and level of their investment in special collections.

Added to the complex factors we know to be at work, the overall environment for scholarly communications has changed in startling ways and with these changes has emerged a new kind of conventional wisdom about special collections. Over the last 15 years there have been substantial not-for-profit and commercial investments in the electronic availability of back- and front-lists of journals and books that are of interest to scholars. What JSTOR, Project Muse, Elsevier and Wiley (among others) accomplished in the '90s for journals surely has many parallels to what Amazon, Google, and the Internet Archive (among others) have accomplished in the first decade of the new century for books. However, the massive Google books digitization project stands as a buoy marking the sea change that has occurred. As a way of taking account of these changes in the special collections arena, the conventional wisdom is to say that because books and serials are now more commonly available to wide audiences in the form of online networked information, what now makes libraries distinctive is not their book and serials holdings but their special collections.⁹ Building on this conventional wisdom, it seems to follow logically that the value proposition for institutional investment in special

collections is that such investment is worthy because it will enhance the distinctiveness of the institution.

A Critique of the Conventional Wisdom

This conventional wisdom about the distinctiveness of special collections compared to the commonness of book and serial collections certainly provides a useful heuristic and helps focus much needed attention on the requirements for building special collections into more useful scholarly resources. However, there are a variety of dangerous traps in the logic about common and special collections. First, system-wide analyses of research library holdings have suggested that books and serials that are being digitized are not so commonly

...there are a variety of dangerous traps in the logic about common and special collections.

held in libraries as one might have expected.¹⁰

With the lack of overlap, libraries cannot readily assume that their physical copies are represented in the common online collections, are held physically elsewhere, and thus can be readily

discarded. Instead, the digitization process may be accelerating the process of converting books and serials from circulating collections to collections of artifacts that need special treatment.¹¹ The most logical special treatment is not simply moving these artifacts into off-campus shelving but into more deeply rationalized and cost-effective shared shelving. In September 2009, both the Council on Library and Information Resources (CLIR) and the University of California received grants from the Mellon Foundation for separate but complementary, multi-institutional efforts to define the terms of and conditions needed to accelerate research library use of deeply shared storage facilities. This work builds on extensive previous work, especially by the Center for Research Libraries, the University of California, and OCLC's Programs and Research division.¹²

A second concern about the distinction between common and special collections is whether common collections that move online still require careful metadata treatments. Google, Amazon, JSTOR, and others with large aggregations of books and serials now provide access to inverted indexes, frequency analyses, and certain kinds of dynamically computed metadata such as a list of older works cited by a particular work and newer works that cite it. These search and discovery tools are proving to be a boon to scholarship. However, moving book and serial collections to the network has amplified,

rather than dissipated, other quality-control and metadata problems that are difficult to solve algorithmically and do require continuing special treatment. JSTOR has maintained the gold standard for descriptive metadata in its serials collections.¹³ However, Geoffrey Nunberg has recently pointed to a variety of general errors in Google's book collection that are particularly troublesome for scholars who depend in their work on careful description of ordinary features such as series, edition, volume, and publication date.¹⁴ In addition, the Council on Library and Information Resources will soon be releasing reports of extensive Mellon-funded studies by scholars in four different fields— linguistics; Latin American literature; history; and media history and cultural studies—that document vexing and ongoing quality-control problems in the book collections digitized by both Google and the Open Content Alliance.¹⁵ Mellon also made a grant this summer to the University of Michigan for a systematic characterization of quality-control issues in the HathiTrust collections.

A third trap in the logic about the common and special collections lies in the largely unexplored area of what the Proposed Settlement Agreement for the Google book digitization project has called "non-consumptive research."¹⁶ Joseph Esposito, Clifford Lynch, and others have often pointed out that the bulk of reading in the future will not be done by humans but by computers.¹⁷ Non-consumptive research refers to such a kind of reading. Overall, our experience with non-consumptive research on texts is limited, especially in fields of the humanities outside of linguistics, but we have learned a good deal from the NORA, MONK, and SEASR projects at the University of Illinois at Urbana-Champaign. Teams of scholars led by John Unsworth, Martin Mueller, and others have found that computers are powerful readers when working on simple discovery tasks, but for advanced scholarly analysis, the machines are largely illiterate unless they are working on well-prepared and well-marked-up texts.¹⁸ Different kinds of inquiries require different kinds of markup, often overlapping, and only some of the markup can be accomplished by algorithm given current technologies. Moreover, texts created by optical character recognition often need even further correction and preparation for sophisticated reading by machine. I assume that these various kinds of human intervention would be permitted on the texts stored in the non-consumptive research centers that the Google Settlement would establish. If not, much useful work could be done on public-domain materials even though the utility would be limited to special scholarly audiences in specific disciplines. In any case, the special

markup and error correction treatments required to make non-consumptive research, as opposed to simple search and discovery, truly useful to scholars puts the online collections of books and serials into a category that is far from common and more like the incarnation at the network level of the physical special collections that we know and love. Special collection skills and expertise are not unnecessary at the network level, they are simply operating in a different context.

The final trap I would mention lies in the suggestion that special collections are what give libraries and their home institutions their distinctiveness. Surely, special collections can be a source of pride, expertise, and excellence, and these

...taken to an extreme, the argument about institutional distinctiveness can also limit scholarly productivity by provoking the impulse to protect silo-like boundaries around collections....

qualities can motivate deep and useful investments. However, taken to an extreme, the argument about institutional distinctiveness can also limit scholarly productivity by provoking the impulse to protect silo-like boundaries around collections, thereby hindering the natural scholarly impulse to create and explore links

among related special collections across various holding institutions. Many have called for more openness within and connections across special collections,¹⁹ but many barriers remain. I particularly invite library directors to

I particularly invite library directors to take a close look at the rights and permission statements that they have readers sign to use their special collections. Perhaps they will be as surprised as I was at the general, blanket, and highly restrictive claims their institutions make to usage rights over this material.

take a close look at the rights and permission statements that they have readers sign to use their special collections. Perhaps they will be as surprised as I was at the general, blanket, and highly restrictive claims their institutions make to usage rights over this material.²⁰

I conclude from this brief critique of the conventional wisdom about the commonness of book and serial collections and the distinctiveness

of special collections that we need to refine our value proposition. The common versus distinctive opposition is simply too crude to get us very far. What is important about books and serials is that moving digital surrogates and newly produced works to the network level generates aggregations operating at a scale that advances existing lines of inquiry and opens new ones and makes scholars and students more productive, even when using individual works. These same criteria must form the heart of the value proposition for special collections.

Because special collections in the humanities, social sciences, and the sciences are full of primary-source materials, they are the fuel of scholarship in these areas. However, before making investments in them, libraries must answer: How would the investment advance existing lines of inquiry and open new ones? How would it make scholars and students more productive? Let me now offer for your consideration three potentially fruitful areas of activity for enhancing the value of special collections.

Processing Special Collections

First, while there are many well-known, well-described, and heavily used special collections, the overwhelming problem that many research librarians have articulated in multiple conference papers and reports is the mountain of collections that remain unprocessed. Carol Mandel referred to the problem memorably as being like the “unwelcome white elephant” that eats you out of house and home.²¹ CLIR’s Hidden Collections program is one small attempt at a solution. Perhaps more important has been the growing adoption of the “more product, less process” approaches that Mark Greene and Dennis Meissner have so effectively advocated.²² Processing tools like the Archivists’ Toolkit and Archon have emerged and developers of both products are now working together to create a single unified product that consolidates the best features of

Deep knowledge of the collections is simply not sufficient for determining priorities for processing. Priorities must also be assessed against criteria of scholarly value

each and is better designed to operate and interoperate with related open source tools such as OLE, the Open Library Environment, and CollectionSpace, which is a museum-oriented system. We still lack the equivalent of a bibliographic utility for the detailed descriptions of special collections.²³ And because there is such

a mountain of materials to be processed, not as much attention has been focused as it should be on methods for efficiently determining priorities.

With Mellon support, a number of institutions have experimented with assessment tools to determine priorities for processing various types of collections. Although these tools now need to be accumulated, evaluated, and appropriately refined, libraries do need to use them more widely because it is amply clear from early experiments that they help focus library attention on the needs of scholars. Deep knowledge of the collections is simply not sufficient for determining priorities for processing. Priorities must also be assessed against

criteria of scholarly value, and for such assessments deep knowledge of the research and curricular priorities in various disciplines is also needed. In their forthcoming book from the Oxford University Press, Fran Blouin, Head of Michigan's Bentley Library, and Bill Rosenberg, a historian, analyze in detail the causes and consequences of the gulf in understanding that now exists between special collections librarians and scholars.²⁴ I urge you to read it when it is available. We urgently need creative solutions.

One way of gathering the deep scholarly knowledge needed to sort out priorities for special collection activity is to bring scholars and students directly into the special collections processing streams. Professor Jackie Goldsby is on the program of this meeting and will be speaking about pioneering efforts that she and her students have made with librarians and archivists at the University of Chicago and in the archives of various other institutions in and around Chicago. Mellon has funded similar initiatives at Columbia; Johns Hopkins; the University of California, Los Angeles; and the Huntington Library. All the results are not in yet, but what we do know is very promising, with benefits all around for the scholars and students, the library, and the university.

Contribution Mechanisms

These programs illustrate one approach to the second fertile area of development for special collections to which I would draw attention: namely, finding efficient and productive ways to engage scholars and students in the development of special collections as scholarly resources. We have all heard about the Web 2.0 types of activities that try to draw readers in by adding tags or other forms of annotation to library records and surrogates. These are fascinating initiatives, but bringing scholars and students directly into the cataloging process is both more risky, and potentially more rewarding because of the deep engagement it can produce. Let me offer a few other examples to stimulate your thinking about how scholars and students could be productively engaged.

The Medici Granducal Archive in Florence, Italy, has a treasure trove of information about the Italian Renaissance that is almost entirely unprocessed. The Medici Archive Project (MAP), an organization based in New York, regularly provides residential research fellowships for visiting scholars, and hit on the idea in 1999 to develop a scholar-friendly data-entry system and require its fellows to spend a portion of their time cataloging the files they were

researching. One of the outstanding results of this project was the creation of a names identity database—a prosopography—that helps scholars sort out the identities of the formal and personal names that appear throughout the letters and other documents in the archives. After a decade of use, the data-entry system now needs to be upgraded, and MAP is using the occasion also to reconceive its fellowship programs. It will continue to have a small number of residential fellows, but is now planning for them to be of shorter duration so that it can also establish long-distance fellowships for individuals as well as a program for distance learning. For both of these new initiatives, MAP would digitize relevant files for the research or course topic but then still require the fellows and the students under the supervision of the course instructor to catalog at a distance these files as part of the interaction.

Another example is the work of Greg Crane, the classicist at Tufts, who established the Perseus database and has lately been hugely imaginative and productive in thinking about “What to do with a Million Books.” He and his research team have selected a corpus of books from the classical canon, worked with the University of Toronto and other libraries to ensure that these works find their way into the work flow of the mass digitization projects. His team then obtained library assistance and created a fully faceted, master bibliography of

...one solution to the metadata problems that are rampant in Google books might be to distribute the effort to self-organizing scholarly teams that care about specific parts of the corpus and will invest the necessary effort to correct and make it usable for scholarly purposes.

these selected works.²⁵ This initiative demonstrates that one solution to the metadata problems that are rampant in Google books might be to distribute the effort to self-organizing scholarly teams that care about specific parts of the corpus and will invest the necessary effort to correct and make it usable for scholarly purposes.

Crane and his team are also working with information specialists to engage other scholars and their students in developing and implementing the methods for applying linguistic markup to the corpus to facilitate machine analysis. Crane’s efforts seem to me to provide a model that could easily be emulated by other scholarly teams in other fields.

Connecting Collections

This brings me to the third area of development I would ask you to consider: Can we develop new and reliable methods to link related special collections across institutions? We have been exploring this area at the Mellon Foundation

in several venues. Staff members of the archives at Boston University and Woodruff Library in Atlanta are together building a deeply integrated shared catalog of their holdings of the papers of Martin Luther King Jr. Into the project, they have drawn the scholarly editor of King's papers, whose team is contributing the vast knowledge it has accumulated about attribution, dating, provenance, and people. The project is also now considering how to draw in a third archives, the King Center in Atlanta.

The Integrating Digital Papyrology project based at Duke, with University Librarian Deborah Jakubs as one of the principal investigators, has gone a step beyond building a unified catalog by integrating three historically separate databases about essentially the same corpus of papyri: one containing bibliographic information; another containing images of the papyri; and the third containing transcriptions. Project staff are now in the process of adding an editorial overlay so that scholars can efficiently make new peer-reviewed entries into the database.

The Mellon Foundation also recently made a grant to a group of university presses, led by the Indiana University Press, all of which specialize in the publication of ethnomusicology. These presses have chosen to use as part of their publishing platform the database of Indiana's EVIADA project, a digital archives of ethnomusicological field video, so that primary source evidence can be closely linked to newly published monographs.

There are many other examples that I could offer from Mellon-funded programs, including the Roman de la Rose Digital Library led by Stephen Nichols, the Parker Library on the Web at Stanford and Cambridge, Electronic Enlightenment at Oxford, the Founding Fathers' papers at the University of Virginia, the Long Civil Rights Movement project at the University of North Carolina, and the Stalin archives at Yale. However, I hope I have said enough to convince you that a value proposition for special collections that is framed in terms of scholarly objectives is enormously attractive and opens a rich area for innovation and the pursuit of new lines of inquiry in a variety of scholarly fields.

* * * * *

Now, in closing, let me return to a question that I raised at the beginning: What, if anything, do the various slang expressions about archives and curating that I mentioned at the beginning say about the value today of special collections, whether in artifactual or digital form? A flip answer would be to quote George Bernard Shaw, who once wrote that "people exaggerate the value

of things they haven't got....Everybody worships truth, purity, unselfishness for the same reason—because they have no experience of them."²⁶ Following Shaw, we could simply dismiss the slang as the inflated, self-important expressions of the unknowing. But we know Shaw to be wrong and so I commend to you the response of Laura Hotman, a senior curator at the New Museum of Contemporary Art. The author of the *Times* article on curating asked Ms. Hotman what she thought of the slang expression. "It doesn't really bother me," she said. "Actually, I'm hoping its popularity will spawn a reality show—maybe 'Top Curator.'"²⁷ Wouldn't that be fun!

Acknowledgements: I would like to thank my Mellon Foundation colleagues, Helen Cullyer and Susanne Pichler, for their help at various stages in the composition of this paper. Paul Duguid also read a version of the paper and offered a number of helpful comments.

- ¹ Donald J. Waters, "The Implications of the Draft Report of the Task Force on Digital Archiving," in *Proceedings of the 127th Membership Meeting of the Association of Research Libraries* (Washington DC: ARL, 1995), <http://www.arl.org/resources/pubs/mmproceedings/127mmwaters.shtml>.
- ² William Safire, "Kiduage," *New York Times Magazine*, October 8, 1995: 28, 30, <http://www.nytimes.com/1995/10/08/magazine/on-language-kiduage.html>.
- ³ Alex Williams, "On the Tips of Creative Tongues: Curate," *New York Times*, Sunday Styles section, October 4, 2009: 1, 8, <http://www.nytimes.com/2009/10/04/fashion/04curate.html>.
- ⁴ Donald J. Waters and John Garrett, eds., *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (Washington DC and Mountain View CA: The Commission on Preservation and Access and the Research Libraries Group, 1996), <http://www.oclc.org/research/activities/past/rlg/digpresstudy/final-report.pdf>.
- ⁵ *Ibid.*, 46, n. 5.
- ⁶ See for example, Philip Lord et al., "From Data Deluge to Data Curation," in *Proceedings of the UK e-Science All Hands Meeting 2004* (Nottingham UK: UKOLN, 2004): 371–375, <http://www.allhands.org.uk/2004/proceedings/proceedings/proceedings.pdf>; Elizabeth Yakel, "Digital Curation," *OCLC Systems & Services* 23, no. 4 (2007): 335–340.
- ⁷ *Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections* (Washington DC: ARL, 2009), <http://www.arl.org/bm~doc/scwg-report.pdf>.
- ⁸ See for example, Mary Lynn McCree, "Good Sense and Good Judgment: Defining Collections and Collecting" (1975), reprinted in Maygene F. Daniels and Timothy Walch, eds., *A Modern Archives Reader: Basic Readings on Archival Theory and Practice* (Washington DC: National Archives, 1984): 103–113; Linda J. Henry, "Collecting Policies of Special-Subject Repositories," *American Archivist* 43, no. 1 (1980): 57–63; William L. Joyce, "The Evolution of the Concept of Special Collections in American Research Libraries," *Rare Books and Manuscripts Librarianship* 3 (1988): 19–29; Michael T. Ryan, "Developing Special Collections in the '90s: A fin-de-siècle Perspective," *Journal of Academic Librarianship* 17 (1991): 288–293; Susan M. Allen, "Special Collections Outside the Ivory Tower," *Library Trends* 52, no. 1 (2003): 60–68; Frank Boles, *Selecting & Appraising Archives & Manuscripts* (Chicago: Society of American Archivists, 2005).
- ⁹ See for example, Beth M. Whittaker, "'Get It, Catalog It, Promote It': New Challenges to Providing Access to Special Collections," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 7, no. 2 (2006): 121; Rick Anderson, "Future-Proofing the Library: Strategies for Acquisitions, Cataloging, and Collection Development," *Serials Librarian* 55, no. 4 (2008): 563; Paul S. Koda, "The Past Is More than Prologue: Special Collections Assume Central Role in Historical Research and Redefine Research Library Collections," *Library Quarterly* 78, no. 4 (2008): 474; Sarah M. Pritchard, "Special Collections to the Fore," *Portal* 9, no. 2 (2009): 177; cf. Anthony Grafton, "Future Reading: Digitization and Its Discontents," *New Yorker*, November 5, 2007: 50–54.
- ¹⁰ Brian F. Lavoie et al., "Anatomy of Aggregate Collections: The Example of Google Print for Libraries," *D-Lib Magazine* 11, no. 9 (2005), <http://www.dlib.org/dlib/september05/lavoie/09lavoie.html>; Roger C. Schonfeld and Brian F. LaVoie, "Books without Boundaries: A Brief Tour of the System-Wide Print

- Book Collection," *Journal of Electronic Publishing* 9, no. 2 (2006), <http://dx.doi.org/10.3998/3336451.0009.208>; Brian Lavoie and Lorcan Dempsey, "Beyond 1923: Characteristics of Potentially In-Copyright Print Books in Library Collections," *D-Lib Magazine* 15, no. 11/12 (2009), <http://www.dlib.org/dlib/november09/lavoie/11lavoie.html>.
- ¹¹ Stephen G. Nichols and Abby Smith, *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections* (Washington DC: Council on Library and Information Resources, 2001), <http://www.clir.org/pubs/reports/pub103/contents.html>.
- ¹² Bernard F. Reilly Jr., *Developing Print Repositories: Models for Shared Preservation and Access* (Washington DC: Council on Library and Information Resources, 2003), <http://www.clir.org/pubs/reports/pub117/pub117.pdf>; Brian E. C. Schottlaender, "You Say You Want an Evolution...: The Emerging UC Libraries Shared Collection Concept," *Library Collections, Acquisitions, and Technical Services* 28, no. 1 (2004): 13–24; Lizanne Payne, *Library Storage Facilities and the Future of Print Collections in North America* (Dublin OH: OCLC Programs and Research, 2007), <http://www.oclc.org/programs/publications/reports/2007-01.pdf>; Constance Malpas, *Shared Print Policy Review Report* (Dublin OH: OCLC Research, 2009), <http://www.oclc.org/research/publications/library/2009/2009-03.pdf>.
- ¹³ Roger C. Schonfeld, *JSTOR: A History* (Princeton: Princeton University Press, 2003):156–160; Roger C. Schonfeld and Ross Housewright, *What to Withdraw? Print Collections Management in the Wake of Digitization* (New York: Ithaka S+R, 2009): 9–11, <http://www.ithaka.org/ithaka-s-r/research/what-to-withdraw>.
- ¹⁴ Geoffrey Nunberg, "Google's Book Search: A Disaster for Scholars," *Chronicle Review*, August 31, 2009, <http://chronicle.com/article/Googles-Book-Search-A/48245/>. See also Paul Duguid, "Inheritance and Loss? A Brief Survey of Google Books," *First Monday* 12, no. 8 (2007), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847>; Grafton 2007, 52–53; and Robert Darnton, "The Library in the New Age," *New York Review of Books*, June 12, 2008: 72–80.
- ¹⁵ *Ghostlier Demarcations: Large-Scale Digitization Projects and Their Utility for Contemporary Humanities Scholarship* (Washington DC: Council on Library and Information Resources, forthcoming in 2010).
- ¹⁶ Proposed Settlement Agreement, *Authors Guild Inc., Association of American Publishers Inc., et al. v. Google Inc.*, case no. 05 CV 8136-JES, US District Court, Southern District of New York, filed Oct. 28, 2008: 11–12, 79–83, http://www.googlebooksettlement.com/r/view_settlement_agreement.
- ¹⁷ Joseph Esposito, "The Processed Book," *First Monday* 8, no. 3 (March 2003), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1038/959>; Clifford A. Lynch, "Open Computation: Beyond Human-Reader-Centric Views of Scholarly Literatures," in Neil Jacobs, ed., *Open Access: Key Strategic, Technical, and Economic Aspects* (Oxford: Chandos Publishing, 2006): 185–193, <http://www.cni.org/staff/cliffpubs/OpenComputation.pdf>. Also see for example, Gregory Crane, "What Do You Do With A Million Books?" *D-Lib Magazine* 12, no. 3 (2006), <http://www.dlib.org/dlib/march06/crane/03crane.html>.
- ¹⁸ See for example, Tanya Clement et al., "How Not to Read a Million Books," paper presented at the Seminar on the History of the Book, Rutgers University, New Brunswick NJ, March 5, 2009, <http://www3.isrl.illinois.edu/~unsworth/hownot2read.rutgers.html>; Brian L. Pytlik Zillig, "TEI Analytics: Converting Documents into a TEI Format for Cross-Collection Text Analysis," *Literary and Linguistic Computing* 24, no. 2 (2009): 187–192; see also Gregory Crane et al., "Classics in the Million Book Library," *Digital Humanities Quarterly* 3, no. 1 (2009), <http://www.digitalhumanities.org/dhq/vol/3/1/000034.html>.
- ¹⁹ See for example, Daniel Traister, "Is There a Future for Special Collections? And Should There Be?" *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 1 (2000): 70–72; Koda 2008, 479–480; Diane Zorich et al., *Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives, and Museums* (Dublin OH: OCLC Programs and Research, 2008), <http://www.oclc.org/programs/publications/reports/2008-05.pdf>.
- ²⁰ For an exception, see Peter Hirtle, "Removing All Restrictions: Cornell's New Policy on Use of Public Domain Reproductions," *Research Library Issues*, no. 266 (Oct. 2009): 1–6, <http://www.arl.org/resources/pubs/rli/archive/rli266.shtml>.
- ²¹ Carol Mandel, "Hidden Collections: The Elephant in the Closet," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 5 (Fall 2004): 106.
- ²² Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68, no. 2 (2005): 208–263.
- ²³ Whittaker 2006, 107.
- ²⁴ Francis X. Blouin Jr. and William Rosenberg, *Processing the Past: Contesting Authority in History and the Archives* (Oxford: Oxford University Press, forthcoming in 2010).
- ²⁵ Crane 2006, 2009; Christopher Blackwell and Gregory Crane, "Conclusion: Cyberinfrastructure, the Scaife Digital Library and Classics in the Digital Age," *Digital Humanities Quarterly* 3, no. 1 (2009), <http://www.digitalhumanities.org/dhq/vol/003/1/000035.html>.
- ²⁶ George Bernard Shaw, "The Man of Destiny," in *Plays: Pleasant and Unpleasant: The Second Volume Containing the Four Pleasant Plays* (Chicago: Herbert S. Stone and Co., 1905): 311.
- ²⁷ Williams 2009, 8.

© 2009 Donald J. Waters



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: Donald J. Waters. "The Changing Role of Special Collections in Scholarly Communications." *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, no. 267 (Dec. 2009): 30–42.

<http://www.arl.org/resources/pubs/rli/archive/rli267.shtml>.

Research Library Issues

A Bimonthly Report from ARL, CNI, and SPARC

Research Library Issues is a freely available, online-only publication released six times per year by the Association of Research Libraries
21 Dupont Circle NW, Suite 800
Washington DC 20036
voice: 202-296-2296
fax: 202-872-0884
ISSN 1947-4911

Executive Director: Charles B. Lowry

Editor: G. Jaia Barrett

Assistant Editor: Kaylyn Groves

Designer: Kevin Osborn, Research & Design Ltd., Arlington, Virginia

© 2009 Association of Research Libraries

ARL policy is to grant blanket permission to reprint from *Research Library Issues* for educational use as long as full attribution is made. Exceptions to this policy may be noted for certain articles. This is in addition to the rights provided under sections 107 and 108 of the Copyright Act. For commercial use, a reprint request should be sent to ARL Publications pubs@arl.org.

Current and back issues are freely available on the ARL Web site.

The Web site is also where you may sign up for alerts of new releases of *Research Library Issues* and learn about submitting ideas for articles.

Visit <http://www.arl.org/resources/pubs/rli/>.