

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>.