



Increasing Access to Special Collections

Ricky Erway

OCLC Research,
erwayr@oclc.org

Abstract

In an environment where we increasingly have access to a collective collection of digitized books, special collections will become increasingly invisible if they are not accessible online. In an era of increasing expectations and decreasing budgets, finding ways to streamline some of our processes is the best way to enable us to do more with less. This report details a number of investigations into how access to special collections can be increased. It includes guidance running the gamut from digitization and rights management to policies and procedures.

Key Words: special collections; archives; digitization; access; OCLC Research

Given large backlogs and the many other challenges associated with special collections, it sometimes seems like we are heading into a tunnel and can barely see the light at the other end. But we are not in there alone. Collaborative efforts, where those who have made progress join with those just setting out, are a great way to advance toward the light. Only then will we realize our full potential in research, education, and society.

OCLC Research: Mobilizing Unique Materials

OCLC Research works with staff from partner institutions on efforts to ensure that special collections emerge from the dark into the light of day. This paper shares some highlights from a number of investigations in the hope that others can learn how to streamline some of their processes, making their special collections a little less special and a lot more accessible.

***Shifting Gears: Gearing Up to Get into the Flow* (Ricky Erway and Jennifer Schaffner, 2007)**

Shifting Gears argues for scaling up the digitization of special collections, with a particular focus on adopting a 'just do it' attitude towards digitization to make special collections more radically accessible.

Its recommendations are intended to help special collections keep pace with mass digitization of books. As we increasingly share a collective collection of books, it is special collections that distinguish cultural repositories. We can hide those treasures in backlogs or behind custom portals — or we can push them out into the light of day.

These are some of the recommendations:

- Due to the special nature of these often unique materials, we will always preserve the originals to the best of our ability. The chief purpose of digital versions should therefore be to improve access.
- Vast quantities of digitized primary materials trump a few superbly-crafted, selective showcases from our special collections. Minimal description and faster scanning allows us to expand use, while limiting access to those who can visit in person restricts use.
- Don't be selective; carefully appraised and acquired collections are worthy of digitization, so either scan as materials are accessioned or scan materials that are often requested.
- Collection-level description with some representative scans may be enough to determine where to apply further effort. Iterate once you've identified high-interest materials.
- Early grant-funded digitization efforts often had 'special project' status, implying that providing access is not mission-essential. Grant-giving agencies should be encouraged to support projects that put permanent processes in place for ongoing operations. We need dedicated budget, staffing, and infrastructure for on-going digitization.
- Work with private partners to develop scanning approaches and industrial-strength workflows to suit particular non-book formats.
- We need to stop obsessing about describing at the item level. Everything that is digitized does not need to be painstakingly described. Start at the top, at the collection level, then think how to group materials and where further description is needed. Researchers

do not mind looking through a number of images to find the right ones.

- Too often we create expensive but little-used websites. Usage will increase when we expose our content to search engines and aggregators, which are more successful at reaching broad audiences.

One area where people are still feeling conflicted is in selection for digitization. While we may wish we could digitize everything in our collections, we should consider digitizing whole collections or significant portions of high-demand collections whenever we can. One way to gauge demand is to digitize some samples and see what the interest is before scanning more. Sometimes we have to digitize selections from many collections for a collaboration or for a grant-funded project — or to show off our stuff (internally, to funders, or to potential donors.) Unfortunately, later, when we have time to digitize complete collections, it can be more efficient to rescan than to sort out what has already been done. We need to ask ourselves if this approach moves us in the direction we want to go — or if it is an unhelpful diversion. Another approach to selection is to look beyond our own institution to see what has been digitized and where we can make a unique contribution, either in a subject area where our institution excels or by digitizing things unique to our institution. And we can get the best of all worlds: if we digitize as much as we can, it is then easier for curators to put together an exhibit that highlights our collections, to contribute to a topical aggregation, and to meet researchers' needs.

***Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives* (Jackie M. Dooley and Katherine Luce, 2010)**

A 2009 survey of 169 special collections and archives in research libraries in the US and Canada shows that digitization of special collections and increasing user access to those collections are of critical importance to research libraries. The survey report reveals that a lot of rare and unique material remains hidden from users and the backlogs continue to grow. Half of archival collections have no online presence at all. The size of collections has grown dramatically, as much as 300% for some formats. And monetary resources are shrinking (75% of general library budgets have been reduced) at the same time that user demand is growing (use of all types of material has increased across the board and user demand for digitized collections remains insatiable). Staff are struggling to keep up and the current tough economy renders

'business as usual' impossible. The survey also shows that management of born-digital archival materials is still in its infancy; it was rated among the most challenging issues in managing special collections (along with space and digitization).

The recommendations in the report focused on issues that warrant shared action. These are just a few of the recommendations:

- Develop and liberally implement exemplary policies to facilitate rather than inhibit access to and interlibrary loan of rare and unique materials.
- Define the characteristics of born-digital materials that warrant their management as 'special collections' and define a reasonable set of basic steps for initiating an institutional program for responsibly managing born-digital archival materials.
- Determine the scope of the existing corpus of digitized rare books and develop models for large-scale digitization of special collections.

Well-intentioned Practice for Putting Digitized Collections of Unpublished Materials Online (2010)

One aspect of modeling for large-scale digitization is reconsideration of our approach to rights management. An OCLC Research event named 'Undue Diligence' led to a community of practice for risk analysis regarding rights associated with digitization of unpublished collections. Digitization of special collections can be inhibited by excessive concerns about intellectual property or privacy rights. *Well-Intentioned Practice* establishes a community of practice based in risk analysis and adoption of the fair use guidelines in the US copyright law.

The primary responsibilities of cultural materials repositories, the stewardship and support for research and learning, require us to provide access to materials entrusted to our care. Establishing a reasonable community of practice allows us to place collections of unpublished materials online for the purpose of furthering research and learning. It promotes a well-intentioned, practical approach to identifying and resolving rights issues that is in line with professional and ethical standards. While the document was developed with US law in mind, it is hoped that the spirit of the document will resonate in non-US contexts.

The document recommends that if your institution has legal counsel, you should involve them in adopting this approach; post-adoption, only seek their advice on specific questions. Other points are to:

- Select collections wisely, starting with those that are of high research interest, but avoid those that warrant exceptional caution (such as contemporary literary papers or sensitive medical records). In the US, we would assess the advantages and risks of relying on the 'fair use' provision in our copyright law.
- Use archival approaches to make decisions, such as checking donor files and accession records for permissions, rights, or restrictions; assessing rights and privacy issues at the appropriate level (most often at the collection- or series-level); and if there is an identifiable rights-holder at that level, attempt to contact them and get permission.
- Document what you learn about the rights status in the description of the collection. Document your actions, processes, findings, and decisions and share them with your professional community.
- Adopt a liberal take-down policy and an appropriate disclaimer and provide them to users of the online collections.
- Prospectively, work with donors to identify possible intellectual property issues and get relevant contact information. Ask donors to identify sensitive materials that may be in the collection. Suggest that donors transfer copyright to the institution or license their works under a Creative Commons license.
- Above all, ensure that no restrictions are placed on content that is already in the public domain, get permission to digitize the materials for unrestricted access, and guard against limitations or restrictions on fair use rights.

The principles laid out in the *Well-intentioned Practice* document have been endorsed by many organizations and prominent individuals. One of the adopters said that we need to 'leverage the full range of rights management strategies and the strength of the fair use argument in order to realize [our] missions as teaching and research institutions.' ... 'The risk of conflict over intellectual property rights is small because the challenges will be few, and can be addressed and rectified without litigation. The benefits to education and research are enormous and outweigh the minimal risks.'

***Good Terms: Improving Commercial-Noncommercial Partnerships for Mass Digitization* (Peter Kaufman and Jeff Ubois, 2007)**

Sometimes making progress requires combining forces to obtain needed funding and technological capabilities. Special collections may be attractive to third-party, private-sector partners and a partnership can facilitate digitization that the institution could not otherwise afford. *Good Terms* offers guidance for those engaging in commercial partnerships to ensure broad access to collections.

The recommendations include:

- Be sensitive to private partner needs to protect business and technology secrets, but insist on your own right to discuss aspects pertaining to your broader community. Deals, such as the Google library partnerships, involve some of the most complex decisions libraries will face; they can be improved through consultations with peers.
- You must have input into the specifications of quality and formats and be clear about exactly what you will receive and that you will own those deliverables.
- Avoid contract terms that make it difficult or impossible to offer scholars the kinds of functionality — including automated or bulk access to collections — that can support innovative research and that allow the development of new functionality.
- Preserve the right to combine parts or all of your digitized content with collections at other institutions or to include it in aggregations.

If these terms cannot be secured, then the consequences of compromise should be fully understood. Above all, if there must be restrictions on ownership, access, and distribution, it should be time-bounded and should not survive termination of the agreement.

***Rapid Capture: Faster Throughput in Digitization of Special Collections* (Ricky Erway, 2011)**

As institutions begin digitizing collections programmatically rather than on a special projects basis, some parts of the process can be routinized in ways that speed throughput and contribute to the overall success of the project.

Rapid Capture is a sampling of innovative approaches in the capture of non-book formats that others may want to consider.

Rapid Capture looks at the moment of actual digitization of a variety of formats and highlights approaches that achieve substantial production throughput. There are nine vignettes that describe quite succinctly what they did, how they did it, and the result.

Here are a few highlights.

- The Bancroft Library at the University of California, Berkeley, digitized correspondence of the 19th-century environmentalist John Muir by outsourcing the scanning from microfilm. From 22 reels of microfilm, they captured almost 25,000 image files, with a daily throughput of 4800 images per day. Microfilm reels are not 'original' materials, so they can be shipped off-site and scanned for a fraction of the cost of doing the work in-house from the originals. The Bancroft has successfully reduced costs by more than 80% using this methodology.
- The Archives of Traditional Music at Indiana University are reformatting audio cassette tapes for preservation and access. They have developed safe practices for parallel transfer, which is the digitization of multiple recordings simultaneously. An audio engineer monitors the audio from three cassette decks at the same time via software that automatically brings each of the sources into primary focus, by increasing the audio volume. The technician can always hear all three recordings, so has continuity, but focuses on one at a time. This approach allows comparison from one source to another in batches of similar materials, so the technician can listen for subtle shifts in fidelity.
- The University Archives at the University of Minnesota are digitizing legacy university publications, averaging 500 pages per hour. This is now a routine function for high-use content and items held in duplicate. In the most recent year, they have scanned almost 220,000 pages on a single scanner. The work was incorporated into the duties of student workers, totaling only 15 student hours per week. The UM Archives are unusual in that they are, for the most part, scanning and discarding the original material. The content has informational value, but the physical items have little, if any, artifactual value. The quarter

of a million pages scanned in a year have freed up hundreds of feet of storage space — and have significantly improved access.

Sharing Special Collections

In order to promote ways to get materials more efficiently and effectively into the hands of users, a Sharing Special Collections Advisory Group was formed. It is made up of both special collections and interlibrary lending (ILL) staff and has studied numerous ILL policies and workflows.

Two main tasks were identified during discussions: streamlining work flows when handling interlibrary loan requests for rare and unique materials and exploring how best to go about building trust between institutions sufficient to allow the physical lending of special collections materials. The working group has developed an analysis of more efficient and flexible collaborations between ILL and Special Collections departments. The group has also compiled a model policy, synthesized from special collections loan policies from institutions with much experience in this area.

Capture and Release: Digital Cameras in the Reading Room (Lisa Miller, Steven K. Galbraith *et al.*, 2010)

There are other ways we can make special collections more accessible. *Capture and Release* argues for allowing, if not encouraging, the use of personal digital cameras in reading rooms. Use of personal cameras means that when a researcher finally finds the document he has been seeking, he can capture an image, return the document to the collection, and study it later.

The report identifies benefits of cameras in the reading room:

- Digital cameras are gentler on collection materials than are photocopiers. Upending materials to position them on a machine risks more damage to materials than photographing them while they are face-up and appropriately supported. The materials are not subjected to the intense light of a photocopier, but rather are photographed with ambient lighting.
- Digital cameras maximize researchers' precious time in the reading room and end the waiting, fees, and paperwork for photocopies.

Camera use also enables getting copies of oversize materials and bound volumes that are often excluded from photocopying policies – and cameras capture color images.

- Allowing personal digital cameras outsources duplication tasks to the user, freeing staff to perform other work and reducing photocopier maintenance and supplies.
- Digital cameras lessen the repository's risk for copyright infringement. When a repository makes copies of copyrighted documents for users, it runs the risk of engaging in direct or indirect copyright infringement.
- As much as we would like to deliver all our collection materials online, it is still beyond our grasp. Digital cameras are research tools that reach across this online/offline divide, one researcher at a time.
- Most importantly, digital cameras facilitate use. Researchers with limited time can cover more collection materials during their visit by photographing relevant materials for in-depth study later.

The report goes on to list suggested practices and lays out a grid with a sliding scale of varying levels of accommodation on fifteen facets.

***Scan and Deliver: Managing User-Initiated Digitization in Special Collections and Archives* (Jennifer Schaffner, Francine Snyder, and Shannon Supple, 2011)**

Scan and Deliver investigates policy issues and presents a spectrum of practices related to patron-initiated digitization of materials. The report identifies a range of different workflow tracks for review, decision-making, scanning, and delivery.

Review considers the means for making requests (verbal request or detailed form) and how requests are approved. It also touches on copyright, privacy, and other legal issues — and where the responsibility lies. *Decision-making* considers whether or not the institution will keep the images — which makes an enormous difference in the workflow and has ramifications for metadata, storage, and quality — and whether the whole unit (folder or volume or series or collection) is scanned. *Scanning* considers who will do the digitizing and description and whether you perform quality control (often a bottleneck in digitization workflows). Finally, *delivery* considers how you will deliver digital copies to the user (online, on disk, via email, on a thumb drive, thru

Flickr...). An institution considering scan-on-demand services can decide which part of the spectrum for these three areas they would like to inhabit.

***The Quest for the Holy Grail: Single Search across an Institution's Collections* (Leah Prescott and Ricky Erway, 2011)**

Single Search considers bringing together an institution's typically disjointed access systems. Let us assume we have followed all the advice and we have succeeded in getting a lot of special materials ready to be accessed. What do we generally do next? We build separate portals. Maybe not one for every collection, but chances are we are still building a number of silos across the libraries, archives and museums across an institution. It would be far preferable to offer a single search interface so that users would not have to guess which portals to investigate and then figure out how to navigate each system.

Representatives from nine implementers of single-search systems shared their experiences to help others who are considering single search. In the report they cover institutional considerations such as motivation, user needs, collection management practices, institutional priorities, organizational structure, institutional culture, and funding. They provide an overview of technological considerations such as whether to retain multiple catalog systems or combine all their data into one system; whether to harvest to a central repository, employ federated search, or use central indexing; use of digital asset management systems; and benefits of open source and commercial approaches. They share their ideas about managing metadata, including the use of standards, vocabularies, mapping, and crosswalks. And they look at access considerations, including how to provide access to digital objects, issues associated with rights management, and getting user feedback.

But, this begs the question, is not each single search implementation still a silo? In order to get the content into users' flow, we need to get it to aggregators and make it crawlable by search engines.

Born-Digital Special Collections

The born-digital project is just getting started. It will offer tips for those managing materials that never were in analog form, focusing on effective

management of born-digital materials as they intersect with special collections and archives in research libraries.

As a first step we provided a simple definition of born digital: items created and managed in digital form. We recognized that different people may have entirely different things in mind when they use the term 'born-digital,' so we identified nine different types of born-digital material (there are, no doubt, others):

- Digital photographs
- Digital documents
- Harvested Web content
- Digital manuscripts
- Electronic records
- Static data sets
- Dynamic data
- Digital art
- Digital media publications.

We will be specific about which types of content are included in our investigation and which are not. Working with a number of advisors, we intend to identify the skills and practices in the archival tradition that will be of value in the preservation of, and access to, materials that were born digital. We will also assemble a set of minimal steps that can be taken to begin the process of managing these materials, while ensuring that no irreversible harm is done.

***The Metadata is the Interface: Better Description for Better Discovery of Archives and Special Collections* (Jennifer Schaffner, 2009)**

The Metadata is the Interface looks at how well the metadata we create serve discovery. In order to learn how people do research in archives and special collections, we conducted an extensive literature review of user studies and usability studies, creating a synthesis of over 80 reports and articles about discovery of archives and special collections. There have been many good efforts to learn what our users really want.

The research was distilled in order to make recommendations about how to make improvements in descriptive practices that will improve discovery. Researchers are autonomous when they work online, so archivists are

distanced from the discovery process. The best help we can offer is through the descriptions of our collections. Yet, while we manage and describe our collections by provenance and present them in that context, our users typically want to find primary resources by keyword or topic — and they want the search results ranked by relevance.

Our current descriptive structures and standards are not giving them what they want. We need to find better ways to describe our collections so that users will find them. The author says, 'People expect to find archives and special collections on the open Web using the same techniques they use to find other things, and they expect comprehensive results. Invisibility of archives, manuscripts and special collections may well have more to do with the metadata we create than with the interfaces we build. Now that we no longer control discovery, the metadata that we contribute is critical. In so many ways, the metadata is the interface.'

There are several ways we can respond to the findings. We can reduce the amount of effort we put into metadata, without negatively impacting discovery — and hopefully improving it. We have to do some description for the material to be found, but then there are options. For collections (or parts of collections) of high interest, we can provide more description. We should remember that researchers often search on subject words, so assigning subjects should become a priority. We can let online users add tags to the descriptions to help others find the material. When an in-person researcher is going through part of a collection, we can ask them to make notes of the contents (correspondents, date range, maybe even transcriptions). They may be happy to give back to the institution — and they may very well know more about the contents than any staff member might. And again, not all researchers know to come to a particular archive for what they seek. We need to participate in aggregations like Europeana and WorldCat — and, most importantly, make sure that search engines index our descriptions.

There is Light at the End of the Tunnel ...

Numerous projects have been described, all of them ultimately focused on increasing access to, and use of, special collections. Especially now, when there is increased institutional and user interest in special collections — coupled with flat or decreasing budgets — finding ways to streamline

processes is a good way to enable us to do more with less, and to bring special collections into the light.

The European cultural heritage community would find it illuminating to undertake a survey like *Taking Our Pulse*. It would likely suggest next steps that could be addressed collaboratively in the European context. At a minimum, it would identify areas in special collections management that are fairly well under control and areas where institutions want assistance, thus helping to focus consortial efforts where they are most needed. A lot has been accomplished to help libraries increase access to their special collections, but we know there is more yet to be done.

Documents Mentioned

Shifting Gears: Gearing Up to Get into the Flow. <http://www.oclc.org/research/publications/library/2007/2007-02.pdf>

Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives. <http://www.oclc.org/research/publications/library/2010/2010-11.pdf>

Well-intentioned practice for putting digitized collections of unpublished materials online. <http://www.oclc.org/research/activities/rights/practice.pdf>

Undue Diligence: Seeking Low-risk Strategies for Making Collections of Unpublished Materials More Accessible, <http://www.oclc.org/research/events/2010-03-11.htm>

Good Terms: Improving Commercial-Noncommercial Partnerships for Mass Digitization. <http://dlib.org/dlib/november07/kaufman/11kaufman.html>

Rapid Capture: Faster Throughput in Digitization of Special Collections. <http://www.oclc.org/research/publications/library/2011/2011-04.pdf>

Sharing Special Collections <http://www.oclc.org/research/activities/sharing>

Capture and Release: Digital Cameras in the Reading Room. <http://www.oclc.org/research/publications/library/2010/2010-05.pdf>

Scan and Deliver: Managing User-Initiated Digitization in Special Collections and Archives. <http://www.oclc.org/research/publications/library/2011/2011-05.pdf>

The Quest for the Holy Grail: Single Search across an Institution's Collections. <http://www.oclc.org/research/publications/library/2011/2011-17.pdf>

Born Digital Special Collections <http://www.oclc.org/research/activities/borndigital>

The Metadata *is* the Interface: Better Description for Better Discovery of Archives and Special Collections. <http://www.oclc.org/research/publications/library/2009/2009-06.pdf>