

**Developing a North-American Strategy
To Preserve & Manage Print Collections of Monographs**

October 27 – 28, 2010

Summary of a Planning Meeting funded by the Institute for Museum and Library Services (IMLS)

Approximately 30 library leaders from across the United States gathered in Chicago, Illinois on October 27 and 28, 2010 to define characteristics of a collaborative monograph archiving model. (See Attachment 1 for a complete list of attendees.) Through a combination of plenary and breakout sessions, attendees identified themes, concerns, possible implementation scenarios, and issues that require testing or research to confirm a framework for future action.

DAY 1: OCTOBER 27, 2010

Plenary session 1

The opening plenary session served as an open-ended discussion to define the context for further consideration. Several major themes and issues emerged.

1. There is significant overlap between Hathi Trust materials, library storage facilities, and campus collections (based on analysis by OCLC Research), which could support digital delivery backed up by archived print copies. One major obstacle is the degree to which access to the digital copy is limited by copyright restrictions: only about 25% of Hathi materials are currently in the public domain. However, a key finding of OCLC's research in the Cloud Library project is that a bilateral agreement between a given library and storage facility would not provide sufficient coverage: a network of shared print repositories with explicit agreements would be necessary to provide access to a large enough shared print collection to enable collection management decisions at individual libraries.
2. "Bibliographic indeterminacy" is a significant problem. It is difficult to compare title holdings where OCLC numbers and ISBNs may be missing, where OCLC numbers may be obsolete, erroneously unique, or duplicated.
3. Monograph archiving may be more politically difficult and more costly than journal archiving. For serials, there is an enormous level of duplication across a small number of titles. For monographs, the pattern is the opposite: a relatively small amount of duplication across a very large number of titles. Questions include how to achieve actual cost and space savings while supporting pride in campus collections and providing adequate delivery. One possibility is to emphasize preserving the unique rather than eliminating the duplicative.

4. It will be important to study actual user behaviors regarding browsing and use of digital and print copies. It may be necessary to provide digital surrogates for browsing and skimming (such as “Inside the Book”). Faculty attitudes are evolving based on experience with electronic journals.

Decisions about print monograph collections disproportionately affect faculty and students in the humanities. It will be important to build the infrastructure for the next generation of scholars. It is also important to note that use of print monographs has been declining even before e-books were available (based on ARL circulation statistics).

Breakout session 1: Affinity Groups

During the first set of breakout sessions, attendees divided into groups according to their primary roles. The main points identified in those groups are outlined below.

1. University Librarians

The group defined this goal for the overall effort: develop a framework to collectively manage print collections with less cost in the context of digital collections

Some concerns were expressed:

- Over promising what can be accomplished
- Copyright restrictions (how accessible can we really make materials?)
- Cost savings (the assumption is that significant costs will be eliminated but new costs will appear).
- Subject expertise: How will subject expertise be deployed? How can expertise be made available at the network level when it is currently affiliated with the individual institution?

The group identified some steps or information needed in order to move to a new model.

- Develop better knowledge of what is already held in the collective collection. What are the gaps? What is the overlap? Are items listed in the catalog actually on the shelves?
- Explore copyright restrictions, including pushing the envelope on lending of in-copyright digital material.
- Analyze collections in a more granular way, to identify and focus on scholarly monographs.
- Engage faculty and scholarly societies about how to build shared collections and what should be kept. Involve experts outside the libraries. Develop a positive vision for shared collections and services.

2. Collection officers

This group identified several areas of focus:

- Develop a positive vision. There is a tension between downsizing print collections with access to archived print materials vs preserving the scholarly record, which has great importance to the

scholarly community. One approach may be to prioritize scantily-held works in digitization efforts. We need to ensure that current economic pressures do not decimate the national collection. To be realistic, we should consider what constitutes acceptable losses. Another approach may be to concentrate on widely held copies to maximize the number of items that could be held in few instances. Another aspect of the vision centers around concern about which items require preservation as artifacts.

- Develop better data. It is important to understand – and to improve --the accuracy of information about holdings. Disclosure of archived holdings will be very important, through systems that can provide data effectively both at the local and network levels. Automated holdings disclosure and collection analysis will be important to facilitate widespread use.
- Determine appropriate kinds and levels of access. Users should be able to access these materials in the format that they choose. Are digital surrogates reliable and accepted?
- Press on copyright and licensing issues
- We need to take on the notion of what is an acceptable loss.

3. Consortium leaders

This group considered the following questions.

- How much duplication of holdings is there? There is a need for more visibility of the holdings in storage facilities in order to understand what is already contained in access-controlled and environmentally-controlled environments. Better information is needed about duplication and uniqueness in collections: some analyses of WorldCat indicate a certain level of uniqueness but closer examination shows false uniqueness. There is more overlap than we think. Even though there are significant issues with data and matching, we should not obsess about perfect data, but begin with what is known in order to foster cooperation.
- What is the role of consortia? Existing consortia represent trust networks but they may not scale. Are they equipped to support shared print programs for monographs? Most consortia don't have operational leadership or capabilities around this issue. Where do the conversations need to continue over time? Who should manage and lead them? Are the right people in the right roles?
- Who can provide the infrastructure for collection management at network scale? Every institution will fall along a different place along the spectrum of preserving the scholarly record or saving space. Every institution will have a different perspective. Groups with a national constituency (e.g. CRL, OCLC, Hathi Trust) may provide infrastructure while regional consortia may provide relationships and operational support. Journal archiving efforts grew organically from regional consortia, can monograph archiving efforts follow the same pattern?

During the plenary discussion session which followed the breakout groups, attendees outlined the following topics for further research or exploration:

- What is the overlap between campus collections and existing storage facilities? OCLC Research has done some work in this area, can this be extended to additional institutions?
- Do we know what we actually have in library collections? How accurate is catalog data for identifying holdings, and are those items actually still in the collection? Materials may be withdrawn, lost or misshelved. Conduct a sample-based inventory of catalog holdings and items actually on shelves. Holdings data for storage facilities is considered to be more accurate, but is it?
- Do we know how users actually use monographs? There are a lot of assumptions about what users want, how they behave and how their behavior is evolving. How do they think of the monograph? How do they use it? How much are researchers buying individually (digital or POD), and at what price point?
- What would it take for libraries and scholars to have confidence in digital access? What provisions would make institutions comfortable? How does the quality of digital books affect access?
- Which materials are the most likely candidates for collaborative retention? Can we identify scholarly monographs at a more granular level? Scholars can be intensely interested in a body of material. Can that interest be operationalized at the network level? Where are there real benefits that can be gained from the band of scholars?

Breakout session 2: Discussion Topics

In the second set of breakout sessions, attendees divided into groups to discuss several different topics. Summaries of these discussions are outlined below.

1. Digital Surrogates: Describe the relationship between and issues related to digital surrogates and print archive copies.

A digital surrogate is a digitized or digital copy of a manifestation of a work. Some sources of digital surrogates are HathiTrust, Google Books, Internet Archive, licensed ebooks (purchased or leased), other local or network projects.

Digital surrogates may serve these purposes: preservation, print replacement, artifactual representation, computation or data mining, expanding access, portability, ability to manage/place collections more remotely.

What do we need in a digital repository or service in order to consolidate or collectively manage print collections?

- Open standards-based formats
- Accurate information about quality/completion
- Usability for a given purpose (most need is not artifactual)
- Discoverability (increased discoverability of in-copyright materials may lead to increased use; not necessarily so for public domain items)

- Reliability/permanence/trust

The group summarized the following points for further exploration:

- Who will retain print copies, what are the incentives to store vs not store?
- Ebooks that correlate to highly-duplicated materials are the first priorities
- Need research projects to study relationship of online discoverability and print use
- How do we balance mass digitization (Google, Hathi, Internet Archive) and publisher digitization [as materials to rely on?]

2. Bibliographic information: Outline the requirements for a bibliographic information/disclosure system (or alternative systems) that would enable large-scale collaboration on monographs among libraries.

The group discussed issues related to systems and data to support print archiving. What characteristics are specific to monographs? How does this effort fit with journal archiving initiatives? CRL is talking with CDL and Ithaka to develop a print archives registry and collection analysis system for serials archiving. Discussions are underway with OCLC to design methods to disclose print archiving commitments through WorldCat.

Participants identified the following issues and approaches:

- Build on the plans being developed for journal archiving: WEST recommending use of existing OCLC WorldCat features for disclosure and resource sharing, and an archives registry and collection analysis service to be developed by CRL. Don't develop separate architectures for journal and monograph archiving.
- What is the impact of separate records for print & digital, how do we refer users to the archived copy? There needs to be a link between print copies, digital copies, and copies in library storage. Should we be concerned about exactly which print volume was the source used for the digital copy?
- It will be important to keep representatives of all editions; how to define edition?
- Need to agree that cataloging is a requirement for print archiving, the same as retention and access agreements. Need community standards for using Institution Symbols, 583 Preservation Action Notes.

3. Service models and business models: Outline the characteristics of service models and business models to sustain retention commitments and provide new modes of discovery and access to retained copies.

- There are declining incentives (and possibly a declining time horizon) to keep print.
- How can consortia facilitate commitments? Should there be different types of members (roles, e.g. libraries which retain vs those which use? Examples of partner categories: Sustaining vs Retaining (as in Hathi Trust).

- What services are required? Storage, retention, discovery, delivery, data mining, digitization, legal services, subject expertise
- Agreements are necessary to define expectations and commitments
- What kinds of organization(s) are necessary to manage this?

Plenary Session 2: First day wrap-up

The end-of-day discussion resulted in three well-defined statements:

- Need sharply-defined scenarios describing possible future approaches
- Need to define the pain, i.e. the problem to be solved
- Frame the goal as “save the long-form argument”.

DAY 2: OCTOBER 28, 2010

Plenary Session 3

Building on a suggestion from the group, the agenda for the second day was revised to focus on six scenarios (outlined by the Steering Committee). Some working assumptions are: 1) archiving would be based on a distributed model; 2) some kind of retention and access agreements would be developed; 3) a system infrastructure for disclosure and resource sharing would be defined. All of these characteristics would need to be defined for any of the approaches below, and are not addressed separately.

Each of the following scenarios emphasizes a focus on collaborative preservation of monographs selected or identified in one of the following ways.

1. **Already in storage.** Under this scenario, libraries would identify and disclose low use monographs already housed in library storage facilities. These materials are already shelved in a protected environment and they may be costly to deaccession (and thus are likely to be retained).
2. **In Hathi Trust and in the public domain.** This scenario would use the approximately 1 million public domain titles currently in the Hathi Trust as the basis for identifying corresponding print holdings for archiving.
3. **By class range, subject, or discipline.** Under this scenario, participants would identify collectively a set of domains or class ranges as a proactive way to focus print archiving efforts.
4. **Library volunteers titles or subjects to preserve.** Under this scenario, participating libraries would volunteer commitments to certain titles or subject areas as opposed to having a community defined direction. The goal is to create the lowest barrier to entry for participation, but it may result in the most complex or diffuse organizational model and the archived contents may be difficult to explain.

5. **Branch Library Closings:** This scenario drives archiving and deselections decisions based on the fact that institutions are closing branches and need to decide what to do with the collections. The advantage of this approach is that those collections need to be processed anyway. However, most branch closings involve science libraries, which do not include many monographs.
6. **In Hathi Trust and published through 1963 or 1976:** This scenario is similar to Scenario #2, but covers all titles published through 1963 (which would include those which required explicit copyright renewal) or 1976 (those published before copyright term was changed to author's life plus 50 years). This approach would significantly increase the pool of materials available to be archived.

A straw vote was conducted to identify the most promising scenarios. The votes indicated interest in the scenarios based on 1) already in storage, 2) in Hathi Trust, both in the public domain and published up to 1976 (i.e. a combination of scenarios 2 and 6), and 3) by class range, subject, or domain.

Breakout session 3: Discussion of scenarios

Attendees divided into groups to discuss and report on these scenarios further.

Scenario 1: Already in storage

- Advantages: volumes already in a protected environment, facilities already provide some degree of access/delivery, relatively quick to initiate this approach after identifying these items, opportunity to experiment with delivery models, understanding the capacity and contents of these storage facilities, and extending their utility, would carry weight with university administrations, relatively high confidence that these copies actually exist because each volume has been ingested individually into storage, then maintained in access-controlled facilities.
- Disadvantages: Relatively random selection of materials, difficult to predict which items will be covered, may or may not have digital equivalent in Hathi or other resource. Instead of random selection, consider identifying the subset of stored materials that have a broad overlap with partner collections to create a more focused value proposition.
- Assume an opt-in approach among the 40-50 high-density facilities (need to define), pilot with a small number.
- Value proposition for large libraries (more likely to have storage facilities): contribution to greater good, less effort on selecting what to retain, direct delivery from storage relieves demand on ILL, collections already viewed as shared in some cases, there is an existing service structure, perhaps financial compensation for retention commitment
- Value proposition for medium/small libraries: Justify investment in facilities, can deselect based on others' retention, right to contribute unique items (access to storage), enhanced delivery service above and beyond ILL

- Suggested timeline: in 2011, disclose stored holdings and retention agreements, develop collection analysis system (planned by CRL)

Scenario 2+6: In Hathi Trust and also in public domain or published through 1963 or 1976

- Advantages: Clear link between digital availability and print preservation, advances the transition to digital delivery, supports testing use of digital copies, organizational structure already in place to lead the development (Hathi)
- Disadvantages: Hathi monographs are generally held by fewer libraries and disproportionately held by research libraries (perhaps less value in this approach for other libraries), may be relatively fragile and somewhat rare (may limit access), requires willingness to provide digital copies where copyright status is unknown (perhaps with take-down policy if protested)
- Consider additional ways to define the pool of materials:
 - U.S. titles published pre-1976 (over 820,000 in Hathi)
 - Identify widely held titles (e.g. 24% held by >100 libraries)
 - Match to titles already in storage

Scenario 3: By class range, subject, discipline

- Advantages: Engages scholars to address resource questions, elevates conversation from inventory management to scholarly communications, aligns libraries to scholarly enterprise, feeds digitization efforts, libraries can reclaim space efficiently by deselection in a defined shelving area, institutional preservation of a certain domain, some disciplines (eg Z's) have high duplication but a smaller constituency
- Disadvantages: May exacerbate concerns of humanities scholars that their materials will be removed
- Recommendation: focus on history and literature, work with scholarly societies
- Value proposition: a way to engage scholars at a national level and thereby help with local collection strategies; better understanding of scholarly processes and aligning those with libraries; testing use of data about the collective collection as part of the argument about the need for large-scale collaboration; potential for feeding titles into the digitization process.
- Timeframe: In 2011, conduct preliminary research to identify materials, determine how to conduct the discussion with scholars and bring to local campuses, perhaps through a planning grant

Preference among three scenarios

Subsequent discussion revealed significant interest in all three scenarios, with one likely approach being to combine all three into a single follow-on project: Pre-1976 humanities titles represented in Hathi Trust and in storage facilities.

Final plenary session

The meeting discussions particularly on the second day resulted in identifying the following research and demonstration topics as likely projects for future action.

Areas for further research or analysis:

1. Optimal copies research. How many copies are needed to support collaborative print monograph preservation? What does “optimal” mean for monographs?
2. Library plans, goals, and constraints. Survey broad set of libraries (including small and mid-size) about plans to divest monographs and whether/what kind of collaboration they might support.
3. Overlap of materials in print collections and Hathi Trust that are in the public domain and already-identified storage facilities.
4. Circulation patterns. Study interlibrary lending/borrowing for monographs and local circulation if possible.
5. User behavior. How much are they spending outside libraries to get content?
6. Demand for print. How much demand for print in light of digital availability? Does presence of digital version increase or decrease use of print? Can libraries increase the use of print if positioned differently in Google and the catalog?
7. Cataloged holdings vs actual inventory. Sample-based study of holdings in catalogs compared to items in library.
8. Costs to deduplicate storage facilities (e.g. OhioLink) and library collections.
9. Leadership and ongoing coordination. What entity(ies) are in a position to lead and support long-term coordination on this issue?

Attendees also identified the following potential demonstration projects

1. Planning grant for subject-oriented project including discussion with society executives, develop plan to engage professional associations
2. Project to test actual and proclaimed user behavior, including browsing behavior and dependence on a shared collection (in storage, possibly in libraries, print only and print-and-electronic).
3. Project to expose Hathi orphan works digital copies for use.

DAY 3: OCTOBER 29, 2010

Members of the Steering Committee (representing Lyrasis, California Digital Library, CIC, and CRL) met on the day following the group meeting to identify next steps. They agreed to continue serving as an ad hoc coordinating body to promote further discussions and definition of specific grant projects over the next few months.

OUTCOMES AND FOLLOW UP

Throughout the workshop and subsequent Steering Committee meeting two ideas garnered particular attention. The first is the notion that identifiable bodies of material exist through the intersection of monographs that are:

- a) already in storage facilities, and/or;
- b) in Hahti, and/or;
- c) in a particular domain or domains.

These bodies of material could serve as the test bed for further development of the issues and concepts discussed, and a near-term project idea emerged. This project would focus on particular domain(s), particularly those well-represented in Hahti and, if identifiable, already in storage facilities, and would include planning and engagement with professional associations. Bob Kieft (Occidental College) is initiating discussion with Modern Language Association executives to take advantage of opportunity to meet at the MLA Annual Convention in early January 2011. Working with Bob will be Bernie Reilly (CRL), Martha Brogan (University of Pennsylvania), Michael Stoller (New York University), Roger Schonfeld (Ithaka S+R) and a representative from the Hahti Trust, most likely CDL's Ivy Anderson.

A second idea to be pursued in the near-term is gathering data from a broad set of libraries about plans to divest monographs and attitudes toward collaboration. There is consensus that ARL libraries will be moving forward in any case with actions both to retain and divest of monographs. The likelihood of success in collaborative efforts may increase through involvement of small and mid-size academic libraries. As an initial step to understand attitudes toward collaboration, Roger Schonfeld has included questions related to this workshop's topics in the 2010 Ithaka S+R Library Survey on collection management strategies (currently underway) and will share results.

Subsequent projects will be included in the final report of this workshop to the Institute for Museum and Library Services (IMLS) by the end of January 2011.

Workshop Attendees		
Name		Institution
Ivy	Anderson	California Digital Library
Kim	Armstrong	Committee on Institutional Cooperation (CIC)
Joni	Blake	Greater Western Library Alliance (GWLA)
Steve	Bosch	University of Arizona
Martha	Brogan	University of Pennsylvania
Timothy	Cherubini	LYRASIS
Paul	Courant	University of Michigan
Sharon	Farb	University of California, Los Angeles
Mike	Handy	Library of Congress
Aisha	Harvey	Association of Southeastern Research Libraries (ASERL)
Martha	Hruska	University of California, San Diego
Peggy	Johnson	University of Minnesota
Paula	Kaufman	University of Illinois
Anne	Kenney	Cornell University
Bob	Kieft	Occidental College
Tom	Leonard	University of California, Berkeley
Michael	Levine-Clark	University of Denver
Constance	Malpas	OCLC
Lizanne	Payne	Center for Research Libraries
Bernie	Reilly	Center for Research Libraries
Deb	Rollins	University of Maine at Orono
Judy	Russell	University of Florida
Mark	Sandler	Committee on Institutional Cooperation (CIC)
Jay	Schafer	University of Massachusetts, Amherst
Karen	Schmidt	Illinois Wesleyan University
Roger	Schonfeld	Ithaca S+R
Peggy	Seiden	Swarthmore College
Emily	Stambaugh	California Digital Library
Michael	Stoller	New York University
Don	Waters	Andrew W. Mellon Foundation
Mark	Watson	University of Oregon
John	Wilkin	University of Michigan
Bob	Wolven	Columbia University
Amy	Wood	Center for Research Libraries