UC Federal Documents Archive: Report and Recommendations

September 9, 2014

Prepared by the UC Federal Documents Archive Project Team

Elizabeth Dupuis, UCB (Project Lead)
Ivy Anderson, CDL (HathiTrust)
Colleen Carlton, SRLF
Heather Christenson, CDL (SAG3 liaison and Google Books liaison)
James Church, UCB (GILS liaison)
Elizabeth Cowell, UCSC (CoUL liaison and FDLC member)
Renata Ewing, CDL
Erik Mitchell, NRLF
Kelly Smith, UCSD (GILS liaison)
Emily Stambaugh, CDL (Shared Print liaison)
Kathryn Stine, CDL

Table of Contents

Executive Summary

In December 2013 the Council of University Librarians, with endorsement of the UC Strategic Action Group 3, charged a UC Federal Documents Archive Project Team to design and implement a virtual archive of federal government documents which includes both print and digital copies of each document owned by the UCs. The charge requests confirmation of scope and prioritization of works to be included, confirmation of relationships with relevant partner organizations and initiatives, investigation and identification of effective and efficient approaches for each step of the outlined process, and implementation of a limited scope project (Appendix A).

From January-July 2014 the UC Federal Documents Archive Project Team identified a wide array of issues and undertook investigations related to:

- analysis of collections and metadata focused on RLF and UC campus holdings
- approaches for building the print archive
- implications of shared print model and disclosures
- identification and handling of duplicates and disposition
- approaches for building the digital archive
- comparison and mergers of records and holdings from UCs and HathiTrust
- identification of scanning options
- possibilities for assessing quality of digital scans, and
- assessment of the use and impact of the archive, and of the effectiveness and sustainability of the approach to building the archive.

The Team defined a clearer statement of the goal:

> The UC Federal Documents Archive is designed as a persistent archive that will consist of one print and one digital copy of all US federal government documents owned by the UC Libraries. Print copies may be shelved at a UC Regional Library Facility or a UC campus library; all print copies will be available to library patrons as "library use only." Digital copies will be preserved in the HathiTrust repository.

Based on typical user behaviors, the Team believes the digital copy will be the primary access point for most people. Following the UC shared print philosophy and FDLP principles, the print copies will also be accessible to people. With concern that high circulation might damage the one print archive copy and for better tracking of the items, the Team decided all UC Federal Documents Archive printed items will be available to library patrons for "library use only" at their borrowing library. This approach supports interlibrary lending as allowed by GPO guidelines. All UC Libraries should ensure their circulation policies and procedures facilitate patron requests to meet this guideline. Individual campuses are encouraged to keep copies of items they feel are of high interest and use as well. These general points were discussed with and supported by the UC/Stanford Government Information Librarians.

To guide our thinking, the Team developed a statement of core principles:

a) Design policies, standards, agreements, and approaches that can serve as a national model for a shared print and digital government document archive
b) Focus on creating a UC shared print archive with assured persistence
c) Create a coherent, strategic approach to prioritizing the phases of the archive development
d) Build the initial shared print collection at the Regional Library Facilities

e) Identify a simple approach to selecting resources for digitization and print archiving
f) Employ best practices for efficiently and effectively building the archive
g) Disclose holdings accurately and publicly
h) Build the print and digital archive simultaneously
i) Aim for digital copies to be openly available
j) Leverage the HathiTrust and Google digitization as much as feasible
k) Accept current Google sheet-fed digitization quality
l) Allow campuses to retain autonomy in making local collection decisions for all copies of federal documents beyond those needed for this archive

Proposed Approach to Building the Archive

The Project Team recommends creating the archive in four phases that allow us to leverage the wealth of materials already housed at the RLFs and to work methodically and efficiently with the collections owned by each of the UC Libraries. In lieu of a limited scope prototype project, the Team recommends beginning immediately with Phase One. Phase One allows the UCs to apply, test, resolve, and confirm specifics of the workflows and staffing needed – simultaneously realizing a significant achievement in the near future, and laying the foundation for the long-term.

Phase One:
The first phase prioritizes the substantial print holdings shelved at the UC Regional Library Facilities – approximately 218,600 federal government document titles identified as currently shelved at NRLF and SRLF. From those items, the goal is to formally designate one print copy of each title and volume as the foundation of the UC Federal Documents Archive. This phase will allow us to resolve issues and define workflows more clearly through the actual implementation. The foundation for all future phases, including the assessment metrics and projections for staffing and other costs, will be confirmed. Timeline: October 2014-October 2016 (2 years)

Phase Two:
The second phase focuses on ensuring that the UC Federal Documents Archive offers a complimentary digital copy of all items designated as part of the archive. This work should begin once the complete list of items from Phase One is identified. The initial part of this phase will be focused on the metadata algorithm for matching UC records and HathiTrust records, signaling which items have not yet been digitized by any institution. Initial collection analysis results reveal approximately 86,000 of the titles at the RLFs are already available in HathiTrust, and approximately 31,000 titles are shelved at both RLFs. Ideally the UCs would be prepared to send a steady stream of items for digitization by mid-2015. Timeline: March 2015-March 2018 (3 years, dependent on Google bandwidth)

Phase Three:
The third phase identifies print holdings shelved on campuses across the UC Libraries to formally adopt into the UC Federal Documents Archive, and ensures a digital copy is also made available. This phase begins at the completion of Phase One and will proceed campus-by-campus and/or agency-by-agency, as is deemed most practical and adherent to the core principles. At this start of

this phase, issues about shared print in place, desired participation by each campus, and staffing models will need to have been resolved. Timeline: April 2016 – indefinite (dependent on findings from Phase One)

Phase Four:
Acknowledging that the UC Libraries will continue to collect new federal government information in print format, and that additional digital copies will be made available over time, this phase ensures that the project cycle continues to pick up new materials after all agencies have been addressed once. Additionally this phase addresses non-print formats and the relationship to born digital publications. Timeline: Begin upon completion of Phase Three

While not all issues about the procedures have been resolved, a more detailed overview of each phase and the associated questions continue to be gathered (Appendix B).

Recommendations

1. Approve implementation of Phase One and Phase Two based on the current resources of the RLFs, CDL, and UC Berkeley, and endorse this work by designating the UC Federal Documents Archive as a high-priority strategic project.
2. Confirm preference for a Selective Housing Agreement (SHA) or Memorandum of Understanding (MOU) between RLFs and UC Libraries.
3. Pursue agreement with the U.S. Government Printing Office and California State Library on a modified process for withdrawal of unneeded duplicates of depository titles from UC Libraries that suits the characteristics of a collaborative, large-scale, collection review project.
4. Identify and implement solutions in discovery and fulfillment services to ensure comprehensive access to the records and materials in the UC Federal Documents Archive.
5. Approve a lightweight disclosure approach, attaching a new holding symbol for UC shared print to indicate the item is part of the UC Federal Documents Archive.
6. Approve reliance on the substantial base of digital copies already available in HathiTrust, creating a process by which users of our digital archive can signal if a particular item is of poor quality.
7. Provide feedback about which assessment questions are most critical to pursue.

Assuming this report fulfills the majority of the original charge, the Team could be reformulated with a smaller number of members who confirm their willingness to serve as an action-oriented project team.

Background

The University of California Libraries have built extensive print collections at each of the campuses, with an emphasis on meeting the specific research and teaching needs of the faculty and students at those campuses. Government documents have been a critical component of these collections, with each campus being an active participant in the Federal Depository Library Program (FDLP) and selecting the publications of federal government agencies of relevance to their programs. With a keen understanding of the benefits of shared collections, the UC Libraries have invested deeply in the design and support of the Northern Regional Library Facility (NRLF) and Southern Regional Library Facility (SRLF). In recent years, as the Regional Library Facilities (RLFs) near capacity and as many campuses are pressed to reduce the number of print collections shelved on site, the UC Libraries have investigated strategic approaches to building shared print collections that continue to meet UC Libraries' user needs while reducing unnecessary duplication.

In addition to remarkable print collections, the UC Libraries continue to build similarly renowned digital collections. As a leading partner with the Google Books Project, Internet Archive, and HathiTrust, the UC Libraries have contributed millions of volumes from our print collection to the massive digital corpus shared with other leading research institutions. Whether for enhanced discovery or full text viewing, through digitization and centralized access in a shared repository, UC Libraries' users have unprecedented access to our materials.

To date, however, our approach to digitization has been opportunistic and primarily location-driven rather than collection-based. The items that have been digitized were drawn from the materials that had been shelved at the RLFs by each of the campuses, or were shelved at a particular campus location, without a particular focus on choosing complementary collections amongst campuses. With a comprehensive and intentional collection-focused approach to developing a shared archive of government documents that ensures the preservation of a print copy in tandem with the assured availability of another digitized copy, the UC Libraries will build a critical resource for the benefit of UC Libraries' users. Drawing first from copies already stored at the RLFs and then reaching out for copies shelved on campus as needed, the UC Libraries can make the most of our limited shelving space and use our staff time efficiently across the campuses. Securing a print copy within UC along with a digital copy in HathiTrust will allow campuses to withdraw unneeded duplicates of print titles once included in the archive. Additional copies of government documents included in the archive would be managed at the discretion of the owning campuses – no campus will be required to de-accession its duplicate holdings.

When envisioning a large-scale project, it is critical for the UC Libraries to design an efficient and sustainable model for withdrawal of unneeded duplicates of titles included in the archive. Established procedures exist today for selective federal depository libraries to withdraw unneeded items from their collections, however those processes are not designed for a project of any sizeable scope and scale. The UC Libraries are in a unique position to offer ourselves as a consortium of several non-regional FDLP libraries who are part of a System and who could model the development of a shared print and digital archive for other US libraries. To facilitate this process, UC will seek to reach an agreement with the U.S. Government Printing Office (GPO) and the California State Library (our regional depository) about a mutually agreeable process that is effective for all our purposes.

This project also allows UC to leverage its relationship with other HathiTrust partners engaged in federal documents digitization. The libraries of the Committee on Institutional Cooperation (CIC)

have an active and well-organized sheet-fed digitization project in place with Google for federal documents in their collections, and HathiTrust itself is currently designing and building a Government Documents Registry to support an expanded documents digitization initiative that is currently in the planning stages. Exploring mechanisms to coordinate UC's activities with these other projects will avoid unnecessary duplication of effort.

The UC Strategic Action Group 3 acknowledges the limitations of the UC Libraries' current library catalogs to adequately represent the full, combined view of local holdings and shared print materials. The issues become particularly acute when considering serials with holdings split between NRLF, SRLF and local campuses, though the problem also occurs in relation to titles withdrawn in lieu of a shared print copy. A suitable resolution will be required for the successful implementation of the UC Federal Documents Archive and related collaborative collection projects within the UC Libraries.

Investigation and Analysis

Over a six-month period the Team identified many critical questions and collaborated in small teams to explore specific topics. Each group presented their findings and recommendations for discussion. The following report briefly summarizes each element. In many cases, while all questions were not answered, the Team agreed on a general direction for next steps with further refinements and resolution expected during implementation.

Building the Archive in Phases

As charged, the Team considered possible parameters for a small prototype project which would test the process and allow us to reasonably scale to the larger shared print and digital archive. Conversations covered several approaches and their implications such as whether to create the print archive or digital archive first, and whether to focus agency-by-agency or focus on a given library location. Testing all aspects of the process was important, as was making significant progress with both components of the archive as soon as possible. In January 2014 UC campus libraries were asked to suspend withdrawals of federal government documents in order to ensure sufficient good copies were available for the UC Federal Documents Archive; the Team recognizes that campuses are anxious to continue with collection management decisions for these materials. There are similar synergies for the digital archive with technology partners, making the analysis about that component equally time sensitive.

Serious concerns were raised about conducting a pilot project and coming to a full stop while awaiting approval to proceed. Identifying the policies, staffing, workflows, and partner commitments for a limited scope project would be similar to that needed for the project overall. In May 2014 the Council of University Librarians confirmed their support for re-envisioning this as a continuous project with the structure allowing the project to adjust as needed during implementation. With this new vision, the Team developed an approach that will build on historic contributions, make the most of the RLFs which acknowledged willingness to serve as Archive Holders, and create avenues for UC Libraries to join into the project in ways that would be inclusive, sustainable, and respectful of individual campus' research needs.

The Team recommends the launch of the UC Federal Documents Archive in October 2014 with the start of Phase One. Once the complete list of federal government documents currently shelved at the RLFs is confirmed, Phase Two would begin concurrently. Both Phase One and Phase Two would utilize the current resources of the RLFs, CDL, and UC Berkeley. The completion of Phase One will include an assessment of the project and a recommendation for staffing and costs associated with Phase Three for review and approval before continuing. More description of the steps and lingering questions for each phase is collected for further refinement (Appendix B).

Collection Analysis

One subgroup focused on aspects of the collection analysis and approaches for comparison of records. Their goal was to holistically compare the current federal documents holdings in the RLFs. Using Voyager and Millennium, they determined metadata elements to analyze and extract, title comparison techniques, item and holdings analysis techniques, and elements for candidate lists and output options (Appendix C). Identification of federal government documents records is challenging due to different cataloging practices including some campuses use of LC classification instead of SuDoc classification. Based on the work of this group it seemed wise to continue to support the collection analysis (comparison of titles and holdings; and identification of potential copies for inclusion in the archive, filling gaps, and digitizing) and metadata work in a centralized way rather than offloading portions to each campus.

Snapshot of the initial findings from the NRLF/SRLF holdings analysis:

- RLF total holdings of federal government documents: 218,629
- Titles duplicated at the RLFs (in both SRLF and NRLF): 31, 216
- RLF government document titles currently available in HathiTrust: 86,194
- SRLF: 158,000 bib records: ~280,000 items, ~167,000 monos, ~112,000 serials
- NRLF: 89,000 bib records (from OO8 MARC field) from first pass
- RLF data is organized by location and by format
- Future comparative work will need manual verification

Snapshot of subset of US Department of Agriculture materials:

- All USDA holdings at RLFs: 14,733
    - Monograph titles: 13,408
    - Serials titles: 1,194
    - Other format titles: 131
- Duplication of USDA at RLFs: 2,463
    - Monographs: 2,289
    - Serials: 140
- Duplication of USDA with HathiTrust: 6,477
    - Titles in both RLFs and in HathiTrust: 1,149

While these figures should be considered rough estimates, overall they provide a fuller sense of the scope of work for each phase and will be useful in guiding subsets of work. The Team noted that SRLF's significant government document holdings were largely contributed by UCLA, so we proactively reached out to UCLA and confirmed their support of this proposal.

Another subgroup focused on aspects of comparing the UC print collection records with those of the digital copies available in HathiTrust via Zephyr. Google is developing a more precise algorithm; the Team recommended some members of the UC/Stanford Government Information Librarians

group to work with Kurt Groetsch at Google. The UC Libraries have worked with Google on developing digitization candidate lists and we hope the new algorithm will further refine the output. The idea for querying Merritt to determine how many federal government materials are already web archived and how those overlap with actual printed documents was suggested. Since the current scope of our federal archiving is limited and the focus on born digital documents falls outside the charge, this was determined as a potential future issue for Phase Four. Since these born-digital items will not have a print counterpart, the UC/Stanford Government Information Librarians could be charged to begin investigation of this component in a separate but complementary project, running simultaneous with the early phases of the UC Federal Documents Archive project.

In Phase One significant work still needs to be done to determine the most effective approaches for comparing print holdings and available digital holdings, including determining which partners to work with and how to match holdings with substantial metadata variations.

In Phase Three significant work is needed to analyze the implications of various approaches for filling gaps in partial holdings and contributing new titles, including decisions about non-print formats. Communications with the AULs for Collections and UC/Stanford Government Information Librarians will be helpful. The Team will need to determine costs and efficiencies for working campus-by-campus or distributing a call to all campuses simultaneously as is done with the UC Shared Print journal archiving campaigns. At a minimum the Team should pilot the collection of MARC bib records to determine API bandwidth, OCLC accuracy, record comparison, individual campus engagement, and so on. Special issues such as "bound withs" and sets with odd numbering will necessitate a higher level of checking and metadata review.

Print Archive

Beginning with Phase One, the goal will be to complete the identification, ingestion, marking, and disclosure of one copy of each item currently shelved at the RLFs as part of the UC Federal Documents Archive. In cases where only one copy is at an RLF, it is planned to remain at that RLF. In cases where there are duplicate copies, a decision tree will be developed to determine which RLF becomes the archive holder. Later stages will address filling gaps and introducing additional titles to the print archive. As an initial test, with input from the UC/Stanford Government Information Librarians group, the Team selected the Department of Agriculture to pilot the process at the RLFs. The Department of Agriculture produces documents of great relevance across all regions of California, offers an interesting mix of types of publications to address, and is not too large a subset of materials.

The Team determined that a manual check of all printed federal government documents shelved at the RLFs would be unmanageable. However the idea of sampling the total collection in Phase One, for a confirmation of the availability and condition of the items, would provide the Team with quantitative data that would be valuable for the project assessment. Based on those findings, the Team could suggest a more intensive review process for confirmation before items are disclosed as part of the Archive.

During Phase One all duplicate copies of federal government documents will have a marker added to their record that would allow for generation of lists for withdrawal. The RLF Directors and Operations Managers have expressed concern that a de-duplication project will be time-intensive and staff-intensive for modest shelf space recovery. The markers will allow us to easily identify

these items at some point in the future if such a project is determined to have merit. Some campuses have expressed interest in getting items withdrawn from the RLFs returned to them; most campuses prefer a centralized implementation that does not involve sending items back to the owning campus.

In Phase Two the Team will identify the titles for which we have satisfied the need for both the print and digital copies. The project will produce a regular report of items (titles and holdings) that are fulfilled, enabling each campus to determine if they wish to then withdraw any duplicate copies from their collections. The specific process for withdrawal would require deeper investigation related to RLF Persistence Policy, FDLP guidelines and approved process with the California State Library, interest from UC Libraries, feasibility with RLFs, and the staffing and business model to be developed with Phase Three.

Based on the figures provided in the collection analysis, the RLFs cannot hold all federal government documents intended for the print component of the UC Federal Documents Archive. Thus, the UC Libraries must consider a shared print in place component with the launch of Phase Three. Shared print in place raises additional questions about future changes in individual campus practices, retention decisions, and security of copies which will be discussed with stakeholders across the UC Libraries to reach agreement about the details before undertaking Phase Three.

Persistence Agreements

Print items that are designated as part of the UC Federal Documents Archive will be shelved, as they are now, with some items at the RLFs and some items at UC campus libraries. The Team considered if the creation of our Archive would suggest a change to our policies or practices.

The Federal Depository Library Program (FDLP) allows a depository library to house some of its federal documents at a separate library or institution, including storage facilities. When the other institution is outside the depository library director's immediate authority, a formal Selective Housing Agreement (SHA) is required to ensure that the housing partner complies with all FDLP regulations related to public access, public service, and maintenance of the collection as outlined in *Legal Requirements & Program Regulations of the Federal Depository Library Program*. A template for the SHA is available on the FDLP website. The SHA must be submitted to the Government Printing Office (GPO) and to the California State Library, which serves as our regional depository library and has administrative authority over selective depository libraries in California. A copy must also be kept on file by the depository library and the selective housing facility. Examples of common situations requiring SHA arrangements on the Federal Depository Library Program (FDLP) website include, "Depository library houses materials in offsite storage managed by someone other than library director."

However, earlier communication from the Government Printing Office (GPO) approached storage facilities in a slightly different way, indicating that an SHA is not required if the depository library maintains "administrative purview" of the material it places in the RLF. While the U.S. Federal Government officially owns these materials, from a collection management perspective the UCs treat materials shelved at the RLFs as still being "owned" by the originating institution that selected them, and the RLFs are guided by the Council of University Librarians consisting of the directors of each of the UC Libraries. For these reasons, it seems an SHA is not required for our situation. In this case, GPO strongly encourages an MOU between the depository libraries and RLFs to ensure proper management and public access to FDLP material.

The primary issue for GPO is confirmation of retention of administrative oversight. The Team confirms our understanding that each of the UC Libraries would continue to retain administrative oversight of the materials they collect, whether those materials are shelved on their campus or at an RLF, and whether those materials are designated as part of the UC Federal Documents Archive or not. Following historical policy and practice, by shelving an item at an RLF the UC Libraries are committing automatically to adhering to persistence polices which align well with the terms of the FDLP and the UC Federal Documents Archive. It is understood that the University Librarians of the University of California, by the power invested in them by the UC Regents, continue to be responsible for managing the collections.

Given that completing an SHA or MOU is a fairly simple process and provides a level of shared understanding and safeguarding of federal materials, the Team recommends completing the document that the Council of University Librarians believes best represents our situation.

Shared Print Disclosures

The Team acknowledges the benefits of clearly identifying shared print copies. Disclosure supports:

- Discovery and display of the print archive for library staff and library users
- Resource-sharing among UC members
- Collection analysis informing local and system-level collection decisions

For the UC Federal Documents Archive, UC Archive Holders – NRLF, SRLF, and each UC campus that participates in Shared Print in Place approach – agree to record information about the items included in the archive in OCLC WorldCat, Melvyl, and UC campus catalogs. Our approach to disclosures must be confirmed before Phase One.

The draft UC Federal Documents Archive Disclosure Policy describes the policy, instructions, and metadata standards recommended by the Shared Print Manager for disclosing UC Federal Documents Archive shared print materials to UC libraries, their users, and to the broader library and user community (Appendix D). This policy follows accepted guidelines by UC Libraries for other print archives such as WEST, as described on the UC Shared Print website. The Team expressed concerns about the removal of the original Institutional Symbol, extent of metadata work, potential of increased responsibilities for RLFs and campus libraries, and implied commitment to policies and tools as yet unconfirmed and untested. With the decision to address a very large number of materials in Phase One, the recommendation is to rely on batch processing for formulating retention commitments in OCLC (as much as possible), make an initial 583 "committed to retain" entry on RLF holdings (but not additional secondary and tertiary entries addressing completeness and condition), and to rely on RLF inventory control and sampling (rather than item-by-item archive validation). In short, we are pressed to balance the ideal of keeping the policies and practices of all UC shared print programs very similar, with the realities of limited resources and special requirements of federal government documents.

To move forward, the Team recommends adding the holdings symbols for UC shared print (e.g. ZAPSP, ZASSP) with the subfield to indicate the item is part of the UC Federal Documents Archive. This element would allow us to gather project data during the early phases without extensive or expensive metadata and staffing at this stage.

Digital Archive

Collaboration with key partners to build a collective digital corpus is critical for both efficiency and effectiveness. The UC Federal Documents Archive will initially look for existing digital scans within the HathiTrust corpus, but other digitization projects (such as AgNIC, TRAIL, LLMC, etc.) may also be considered appropriate targets if these are embraced as part of the HathiTrust US Federal Documents Initiative < http://www.hathitrust.org/usgovdocs>. As the UC Federal Documents Archive project moves forward, coordination of digitization plans with the HathiTrust Federal Documents Initiative for materials not yet digitized will be a key goal.

Over the past decade the UC Libraries have partnered with a number of different entities in the digitization of content. The largest quantities of print items from the UC Libraries have been scanned in partnership with Google and Internet Archive. Individual campuses also have the ability to scan items in house. In Phase Two, the conditions for determining the best path for digitization would be developed. UC has a current agreement with Google that could be used for this project; UC needs to provide a clearer sense of the timeline and scale of our planned contributions for ongoing digitization associated with this project. Through CDL, UC Libraries have an ingest process in place for items scanned by Google to be contributed to HathiTrust.

Digital copies identified for the UC Federal Documents Archive will be drawn from HathiTrust as the primary preservation repository. The Team identified several issues of concern such as incomplete and inconsistent identification of federal government documents as part of the public domain, the inability of users to download full documents rather than page-by-page, and remedies for flawed digital copies flagged in the HathiTrust archive. The UC Libraries have representatives on several HathiTrust committees; we recommend that they advocate for and help address these issues as part of future improvements to HathiTrust.

In Phase Two further consideration can be given to partnerships with other reliable groups for identification of digital copies as well as for sources to enhance discovery for library patrons. Accessibility for all users, including users with print disabilities, continues to be a priority for the UC Libraries and will be taken into account in the Phase Two planning and implementation.

Digital Copies and Scanning

The Team considered the issues related to the quality of digital copies and how to manage their inclusion in the Archive. Good metadata is an important element. Quality assurance processes and image quality rating metadata, intended to note the nature and severity of the defects, occur at the time of ingest into a repository. Since the digital copies that form HathiTrust have come from many sources, it would be tedious if not impossible to determine retroactively how to batch the files for testing. Most documentation on this issue indicates that there is no proven automated process for doing this work; it requires human review especially for visual quality.

Internet Archive's standard practice involves conducting quality control during scanning with each page reviewed before moving on to the next. Approaching the issue a bit differently to accommodate a larger scale, Google iteratively runs new batch processes to address quality issues; that approach may address some current problems, overlook some types of errors, and perhaps introduce others. A manual audit conducted by Google and CDL of a small portion of Google's scans found a small percentage of critical problems (~2%) and a significant number of cosmetic problems

(~50% by Google, and ~25% by CDL). Due to the scale of the project we are facing, a one-by-one manual approach based in the UC Libraries is not feasible.

Wherever possible the UC Libraries should take advantage of and advocate for quality certification processes established or implemented by digitization agents or aggregate preservation repositories. In the future, optical character recognition (OCR) confidence scores may be included with Google scanned content and thus included in HathiTrust digital copies. This could be a helpful indicator of the quality of the informational, as opposed to visual, content.

The Team recommends relying on the substantial base of digital copies already available in HathiTrust, and suggests creating a process by which users of our digital archive can signal if a particular item is of poor quality and that can be flagged for review and re-scanning if needed. HathiTrust already solicits feedback on quality issues using an online form; we could use that form or model a separate one for our Archive upon it.

Discovery and Fulfillment

A key element for the successful implementation of the UC Federal Documents Archive and related collaborative collection projects within the UC Libraries is the development of a discovery and fulfillment model that adequately represents the full, combined view of local holdings and shared print materials. The UC Libraries' current library catalogs present numerous limitations, especially in the area of viewing serials holdings from various locations in a coherent way and handling patron requests correctly for multi-volume sets contributed to RLFs from various sources.

Potential solutions include: a) leveraging Melvyl as a discovery layer, b) pursuing a metadata record sharing program, c) implementing a new UC-wide or project-focused discovery system and d) adjusting policies and systems to ensure that requests for materials are not unintentionally cancelled or unfulfilled in distributed collection situations (e.g. a patron request volume one of a set because their library only holds volume two). Several UC task forces are currently exploring the possibilities of related solutions that the UC Federal Documents Archive will look to for direction during Phase One and Phase Two. The UC Federal Documents Archive Project Team will focus on exploring approaches for the collection and packaging of shared metadata records for government documents, which will be an important foundation for supporting discovery locally.

Assessment

The basic question is: What measures and metrics are indicative of a successful process and outcome? The objective is to gather quantitative and qualitative evidence to document findings about: a) collection identification and management, b) impact on individual UC libraries and RLFs, c) effectiveness of the archive for researchers' needs, and d) effectiveness and sustainability of the project model for building and maintaining the print and digital archive.

Brainstormed areas for focus follow, though more thought needs to be given to which of these potential questions will yield the most valuable information and warrant the effort needed to collect the data. Once specific questions are chosen, measures and targets for each will need to be defined. Additionally, some questions may suggest initial baseline data be gathered for future comparison.

A) Collection Identification and Management
   - Accuracy in identifying titles and holdings of federal government documents
   - Efficiency in identifying and managing item-level holdings within UCs
   - Appropriate archive locations are identified
   - Availability of a print copy to create an original or replacement digital scan
   - Pace of adoption of items into the archive and disclosures made
   - Time period and human hours required to complete scope

B) Impact on Individual UC Libraries and RLFs
   - Quantity of shelf space reclaimed (holdings/linear feet per campus and RLF)
   - Feasibility of RLF de-dup process (benefits, costs, outcomes)

C) Effectiveness of the Archive for Researchers' Needs
   - Quantity of use of the print and digital archive
   - Extent digital copy is a satisfactory substitute for print copy
   - Satisfaction with service of the print collection
   - Number of times a digital copy is identified as defective (and corrected)

D) Effectiveness and Sustainability of the Project Model
   - Extent to which the project enables campuses to make strategic de-selection decisions of their own choosing
   - Extent of acceptance of the UC Federal Documents Archive (UC and others)
   - Collection analysis costs
   - Physical consolidation costs (RLFs and campuses)
   - Digitization costs
   - Disclosure and metadata costs
   - Leveraging of complementary initiatives (e.g. HathiTrust)

Recognizing that the project would be able to address only a subset of these questions in any meaningful way, the Team welcomes input about which issues would be most critical.

Staffing and Business Model

The Team outlined many questions about how this project would be staffed and funded. Since this is a new undertaking, the experience gained in Phase One will clarify which elements can be handled programmatically, which elements need human review and decision making, what level of staffing is appropriate for each function, as well as specific responsibilities, required knowledge and skills, number of hours, work locations, systems authorizations, reporting lines, training, and workflows.

The Team recommends conducting Phase One and Phase Two using the current resources of the RLFs, CDL, and UC Berkeley. During these phases the work will focus on materials and functions that will be best informed by staff already familiar with the collections, systems, partnerships, and programs under discussion. It is anticipated that RLF contributions will match current work (e.g.

pull titles for scanning, review of titles and records, add holdings symbols automatically). The Council of University Librarians would aid this work by designating the UC Federal Documents Archive as a high-priority strategic project.

During the first years of this project, the UCs will learn from our other shared print experiments, see more clearly the reality of filled RLFs, and have practical data for this project on which to build a staffing and business model for Phase Three. That proposal would be shared with SAG3 and the Council of University Librarians for confirmation prior to the start of Phase Three. A business model would include definitions of and proposals related shared costs and costs absorbed by each UC campus library. It is anticipated that the plan for new deposits to an RLF would leverage existing campus allocations. Typical shared costs include program management, infrastructure and staff for collections analysis and collection decision-support, support for active archive creation (when appropriate), and coordination with digitization agents.

Partnership with Federal and State Agencies

While the UC Federal Documents Archive project is designed to work with the corpus of federal government documents collected by each of the UC Libraries over many years in their roles as selective federal depository libraries, ideally our work could serve as a useful model for other consortiums and systems. The Team discussed possible partnership opportunities with the Government Printing Office through conversations with Mary Alice Baish both at Federal Depository Library Council meetings and her recent visit to the San Francisco area. As a member of the Federal Depository Library Council, Elizabeth Cowell has a path for opening future conversations as our project develops. Equally important is our collaboration with Tammy Fishman at the California State Library, which is a regional depository that receives all publications distributed to depository libraries by the United States Superintendent of Documents. With the California State Library's extensive connections to and support of California public libraries, our two institutions have a common interest in ensuring and extending access to information, such as that which will be easily available across the state through the digital copies in the UC Federal Documents Archive.

Recommendations

1. Approve implementation of Phase One and Phase Two based on the current resources of the RLFs, CDL, and UC Berkeley, and endorse this work by designating the UC Federal Documents Archive as a high-priority strategic project.
2. Confirm preference for a Selective Housing Agreement (SHA) or Memorandum of Understanding (MOU) between RLFs and UC Libraries.
3. Pursue agreement with the U.S. Government Printing Office and California State Library on a modified process for withdrawal of unneeded duplicates of depository titles from UC Libraries that suits the characteristics of a collaborative, large-scale, collection review project.
4. Identify and implement solutions in discovery and fulfillment services to ensure comprehensive access to the records and materials in the UC Federal Documents Archive.

5. Approve a lightweight disclosure approach, using the OCLC metadata guidelines for print archives, which includes the use of shared print symbols, LHRs, 583, and 561 subfields to record retention decisions; any outcomes of validation (not recommended at this time); and custodial history (original ownership). A subfield is used to indicate the item is part of the UC Federal Documents Archive (583 $f).
6. Approve reliance on the substantial base of digital copies already available in HathiTrust, creating a process by which users of our digital archive can signal if a particular item is of poor quality.
7. Provide feedback about which assessment questions are most critical to pursue.

Assuming this report fulfills the majority of the original charge, the Team could be reformulated with a smaller number of members willing to serve as an action-oriented project team, dedicating a significant portion of time to complete Phase One and Phase Two. Continuing members could include: Elizabeth Dupuis, Colleen Carlton, Heather Christenson, Erik Mitchell, and Emily Stambaugh. A recommended new member is Jesse Silva, the federal documents librarian from UC Berkeley and member of the UC/Stanford Government Information Librarians. All other original members as well as previously identified stakeholders would be called upon as needed for specific aspects of the project design and implementation.

Appendices

Appendix A. Charge

UC Federal Documents Archive Project
December 6, 2013 | Approved by CoUL November 15, 2013

Charge
At the direction of the Council of University Libraries and reporting to SAG3, the UC Federal Documents Archive Project Team (FedArc) is charged to design and implement a virtual archive of federal government documents which includes both print and digital copies of each document. Digitization (or identification of available digital copies) and print retention will be managed in tandem, by identifying single shared print copies of federal documents already at the RLFs and simultaneously scheduling a second copy within the system for digitization. Sheet-fed digitization via Google will be the preferred approach. Where digital copies already exist within HathiTrust, a satisfactory HathiTrust copy would be deemed sufficient and digitization would not be necessary.

The Project Team's tasks include elements such as:

- Confirm a process for identifying materials in scope, such as organized around corporate authors
- Develop principles and shared agreements with all UC Libraries and RLFs to make identification and confirmation of contribution of printed items efficient
- Secure agreement with the GPO for sheet-fed digitization when needed and efficient mass de-accessioning of unneeded duplicate copies
- Investigate relationship to HathiTrust Government Documents Registry and an appropriate process for coordinating with CIC and other HathiTrust partners engaged in government documents digitization
- Determine possibilities for matching records for print and electronic records from UC Libraries (including RLFs) and HathiTrust
- Determine specifications for the shared digital archive, including efficient contributions to HathiTrust for unique items and metadata upgrades needed to identify works as federal documents (e.g. inclusion of SuDoc numbers)
- Articulate specifications for digitization, quality standards, and quality assurance to yield the best quality copy feasible
- Determine processes for deposit and validation of suitably clean print copies for the contribution to the archive at the RLFs, with any associated arrangements for loan and preservation. For example, UC could organize print archiving campaigns to the RLFs based on corporate authors or other selection criteria and organize the processes and tools necessary to analyze and validate holdings, coordinate deposits of gaps in stored holdings, record retention commitments and digitize duplicates where necessary.
- Pursue partnership with Google for sheet-fed digitization, and investigate alternatives for scanning and funding of scanning if Google is not a viable partner or for items which cannot be digitized via Google
- Develop a limited scope project as a proof-of-concept

Project Team (proposed)
- Ivy Anderson (HathiTrust liaison, CDL)
- Heather Christensen (SAG3 and Google Books liaison, CDL)
- James Church (GILS liaison, UCB)
- Elizabeth Cowell (CoUL Liaison and FDLC member, UCSC)
- Elizabeth Dupuis (Project Lead, SAG3, UCB)
- Erik Mitchell (NRLF Director)
- Colleen Carlton SRLF
- Kelly Smith (UCSD)

- Emily Stambaugh (CDL Shared Print Manager)
- Kathryn Stine (CDL) with Renata Ewing acting as Interim

The FedArc project team will coordinate with the Shared Print Strategy Team recently endorsed by CoUL, ensuring that documents archiving at the RLFs meshes well with broader print archiving campaigns and space plans, as well as with Google mass digitization efforts currently underway.

The FedArc Project Team will develop initial plans for a proof-of-concept project to present to SAG3, and will provide regular progress reports to SAG3 on its progress. An initial project plan should be ready for review in six months from the date the group is confirmed, with an estimated timeline of July 2014.

Appendix B. Details and Lingering Questions for Implementation

UC Federal Documents Archive

Phase One – General Steps

The first phase prioritizes the substantial print holdings shelved at the UC Regional Library Facilities – approximately 218,600 federal government document titles identified currently at NRLF and SRLF. From those items, the goal is to formally designate one copy of each title and volume as the foundation of the UC Federal Documents Archive. This phase will allow us to resolve issues and define workflows more clearly through the actual implementation. The foundation for all future phases, including the assessment metrics and projections for staffing and other costs, will be confirmed. Timeline: October 2014-April 2016 (1.5 years)

a) Develop a satisfactory approach for identifying federal government publications in Melvyl
b) Identify all federal documents already shelved at an RLF
(Serials handled separately for attention to volumes/holdings)
c) Claim one copy of each item for the FedDocArc print archive
d) Disclose those items in OCLC according to the OCLC metadata guidelines for shared print
e) Generate a list of all items now adopted into the FedDocArc print archive
f) Generate a list of all items with a second print copy at the other RLF
g) Mark the records of the second print copy for possible future withdrawal

<u>Phase One - Questions</u>

b) What variation do we see in metadata and how do we reconcile versions?
b) Work with title level first then holdings?
b) Need to develop process for comparing volumes/serials
c) Which RLF? SRLF then NRLF? For serials, how address records issues (consolidate vols or wait)?
c) Sample from all holdings for shelf checks to confirm the item is there, complete, good condition, etc; collect information for assessment and possibly different process if many problems are found; some surmise older docs and older agencies may need more checking (consider if Google evaluation guide is a useful tool)
c) Issues of metadata remediation, deposit (if to an RLF), disclosure prep
d) Determine disclosure codes, symbols, and notes
d) Articulate guidelines about what disclosures entails, implies, etc
d) List generation all by CDL/FedDoc project – not asked of each campus
e) What fields/order required to meet all purposes of various lists? (shelf check, campus holdings, digital copy check, publication match, volume holdings, etc)
e) Create public version of what is in the archive?
f) What steps needed to find, flag, mark items and records? (offer back to campuses or handle for them?)

Phase Two – General Steps

The second phase focuses on ensuring that the UC Federal Documents Archive offers a complimentary digital copy of all items designated as part of the archive. This work should begin once the complete list of items from Phase One is identified. The initial part of this phase will be focused the metadata algorithm for matching UC records and HathiTrust records, signaling which items have not yet been digitized by any institution. Initial collection analysis results reveal approximately 86,000 of the titles at the RLFs are already available in HathiTrust, and approximately 31,000 titles are shelved at both RLFs. This phase brings together

these three streams of information to identify specific titles. Ideally the UCs would be prepared to send a steady stream of items for digitization by mid-2015. Timeline: March 2015-March 2018 (3 years, dependent on Google bandwidth)

a) Develop satisfactory algorithm for identifying federal documents with Google
b) Compare the list of all items in the FedDocArc print archive to available digital copies
c) Generate a list of all items that need to be scanned
d) Identify availability of a second copy (RLF or UC Libraries) for scanning
e) Partner with Google for all items they scan through the sheet-fed scanners
f) Partner with others for all remaining items to scan (Internet Archive, UC campuses, etc)
g) Determine how to link the digital copies with FedDocArc print archive records
h) Generate a list of all items with both print and digital copies archived
i) Generate a list of all items with a print copy archive, but lacking a digital copy
j) Create form for reporting problems with digital copies so can be reviewed/addressed

Phase Two - Questions

For a faster start, could we pre-seed the lists with Google so they tell us what needs to be scanned, then compare with what is at RLFs as a start?

Use workflow chart with four cases? 1) volume unique to RLF and in HT, 2) volume unique to RLF not in HT, 3) volume duplicated at RLFs and in HT, and 4) volume duplicated in RLFs and not in HT – each has different path for disclosures, record notes, digitization

d) How prioritize which to call upon? Who does this work?

e) How is Google sheetfed scanning funded (shipping etc)?

j) Coordinate with HathiTrust about reporting bad scans and replacing the files

Phase Three – General Steps

The third phase identifies print holdings shelved on campuses across the UC Libraries to formally adopt into the UC Federal Documents Archive, and ensures a digital copy is also made available. This phase begins at the completion of Phase One and will proceed campus-by-campus and/or agency-by-agency, as is deemed most practical and adherent to the core principles. Timeline: April 2016 – indefinite (dependent on findings from Phase One)

a) Develop a satisfactory approach to identifying all federal government print publications of a given agency, which are not already identified as part of the print archive, and which are held by one or more UC Libraries.
b) Generate a list of items needed for the print archive
c) Make arrangements with UC Libraries with holdings about contributing to the print archive (RLF contributions through their annual allocation or Shared Print in Place)
d) Claim the copy of each item for the FedDocArc print archive
e) Disclose those items in OCLC and add a marker to the records
f) Generate a list of all items now adopted into the FedDocArc print archive
g) Compare the list of all items in the FedDocArc print archive to available digital copies
h) Generate a list of all items that need to be scanned
i) Identify availability of a second copy (RLF or UC Libraries) for scanning
j) Partner with Google for all items they scan through the sheet-fed scanners
k) Partner with others for all remaining items to scan (Internet Archive, UC campuses, etc)

l) Determine how to link the digital copies with FedDocArc print archive records
m) Generate a list of all items with both print and digital copies archived
n) Generate a list of all items with a print copy archive, but lacking a digital copy

Phase Three - Questions

a) Ask campuses if have certain agencies they want to have as shared print in place?
c) Ask campuses about preferred models for contributing, such as using normal allocation or something else)
l) Steps for processes to pull copy, send to one location to send to source of digitization, tracking through to completed scan, return of printed copy, deposit to digital archive, disclosure of digital version
m) Campuses could then de-accession if wished; need to confirm streamlined approach

Phase Four – General Description

Acknowledging that the UC Libraries will continue to collect new federal government information in print format, and that additional digital copies will be made available over time, this phase ensures that the project cycle continues to pick up new materials after all agencies have been addressed once. Timeline: Begin upon completion of Phase Three

Appendix C. NRLF/SRLF Federal Government Documents Holdings Analysis

Summary report

Data was extracted from III and Voyager using the same criteria (e.g. 008 data and location code information). While it is difficult to know for sure that this retrieved all government documents, it did result in a database of 218,629 titles (located in the table all_retrieved_govdocrecords) from both RLFs. Given the fact that this project seeks to identify candidate titles to work from, a comprehensive list is not required to begin work.

Data was normalized using an iterative process (see rough notes at end of this summary). Unique comparison keys were generated to expand title matching (Title Key, Publisher Key, Author Key). In general, normalized OCLC Number exact match comparison resulted in approximately 29,180 matches at the title level. Title/Publisher key, using keycollision normalization techniques resulted in an additional approximately 2036 matches for a total of 31,216 duplicated titles (monographs, serials and other formats). Title/publisher key string length is likely not always sufficient for exact matching and short keys should be reviewed for accuracy.

In addition to analyzing RLF holdings HathiTrust data was acquired to could get an early understanding of HathiTrust overlap. In total, 86,194 titles in the RLF dataset are also available in HathiTrust as a scanned digital object. All but 105 of these titles are available as public domain. Although additional analysis is required, a rough item count of these matched titles represent approximately 253,000 items according to HathiTrust data. In comparison the full harvest of Gov Docs from the HathiTrust dataset revealed 524,197 items (227,366 titles), of which 533,823 items are available as public domain.

This data comparison technique found a number of issues and opportunities that could be addressed in next iterations:

Issues:

1. Database tables contain a limited view of the metadata available in a MARC record. This is notable particularly in Voyager's tendency to extract the first 035 from MARC and return it as network_number. There are cases where this number (http://catalog.library.ucla.edu/vwebv/staffView?bibId=4473861) is a GPO number and as such an OCLC number match fails when indeed the records are duplicated. In this cited example, title_key was successful in identifying a match but with a need to add publisher and other distinguishing keys in for more reliability. In addition, the reliance on title from the index tables makes it very difficult to compare items with brief titles (e.g. annual report, proceedings).

2. A large amount of data normalization is required for this analysis. Voyager and III for example handle OCLC numbers differently and leading zeros needed to be stripped from numbers. In most cases, text comparison was required given the inconsistent use of alpha-numerics in fields (record id in voyager is numeric only, in III is alphanumeric). Other normalization included

3. While title-level comparison is feasible, issue-level comparison has yet to be tackled except beyond simple numeric comparison of items held.

4. Pulling item data from NRLF holdings requires more complex data querying and processing techniques. Additional work is required.

Opportunities:

1. The techniques piloted in this work reveal a path forward for a full title-level comparison between the RLFs. Such a comparison could result in a strong de-duplication checklist that could be prioritized based on item format or item count.

2. Extraction and analysis of MARC records presents an opportunity to fully automate this process and use parallel processing - such a tool would be portable to multiple organizations.

3. While it seems clear that this data may be completely retrievable from OCLC some process would still be required to extract and pre-process item-level holdings. More work is needed to understand if title-key level matches are co-located onto records in OCLC.

Data Extraction Specification

Objective

To create a specification that allows repeated and automated extraction, synchronization and analysis of records from the NRLF and SRLF facilities. The focus of this specification is on Government Document-related resources.

Scope

Bibliographic, Holdings and Item information from NRLF and SRLF databases limited to Government Documents records as defined by authorship or publisher in the first set.

Outcomes

A set of records that have been unified as well as a methodology to extract and analyze them.

Specification

1. Pull gov docs
    a. For UCB Extract all records from III with subfield a 993 $bGov Doc
        i. Found that extracting 993 Gov Doc was unreliable (17k records at NRLF with this key, 600k overall)
        ii. Found that 008 Pos 28 F, location n* returned 168k records
        iii. 008 28 f, location n*, CountryCode ends in u (need to use subsequent processing to strip out 2 character country codes: 89,856k records (
    b. For SRLF: Marc 008 - position 28, f for fed gov docs, 17th, value u for united states
2. Extract all MARC records where a 710 field contains the keywords “department of agriculture” (e.g. 710 1 |a United States.|b Department of Agriculture.)
    a. Note - future efforts to extract could include other records
3. Extract 245 title, 245 h (GMD), 008
4. 035 oclc, unique ids for catalogs (001 for III, 035 for Voyager)
5. ISSN, ISBN
6. Monograph or serial as defined by 008
7. 260 publisher information
8. author (110)
9. All holding info as avail (866 as a broad holding statement - Voyager has multiple, III may not have any - we can look at LHR)
10. All item info as avail (item enum and chron in voyager) (item record in III)

11. return item barcode
12. return owning unit (attached to item record in voyager, use location code in III)
13. lets return item type
14. unique identifiers
15. let's scope to RLF
16. return 300 field - all subfields

Open question about how to bring the two record sets.
how to do the analysis
unify data sets
add column for matched record (item level) (perhaps with ISSN, ISBN, SuDoc normalized)
1. monograph, oclc number match - update
2. monograph title/author hash match - update

Using fingerprint key-collision method we found reasonable match rate success that may include include oclc num and issn/isbn matches.

3. serial oclc number match volume and date match - update
4. serial title/author hash match volume and date match - update
5. xoclc service, xisbn, xissn services - optional

Notes from ETM import and data comparison

1. Imported SRLF into MSAccess db
    a. Imported All Gov Docs
    b. Indexed everything -
2. Queried NRLF data using 008 SRLF criteria, location n*
3. Created a bib and item export from III to preserve data integrity
    a. Imported Bib into Excel to resolve Data import errors
    b. During MS Access import, changed Date 1, Date 2 to short text to avoid data truncation
4. Imported NRLF data into MSAccess
5. Created query to unify data sets
    a. Migrated only unique titles, not items (158,434 titles for SRLF, 89k titles for NRLF)
    b. Added column for NRLF, SRLF
    c.
6. Needed - Sanity Check
7. TBD: Export Data from Access
8. TBD: Create comparison Key
9. TBD: Import comparison Key into Access

Findings:
Strict title comparison, no normalization: 20866
OCLC Number comparison: post cleanup: 15347 (improper matches possible)
titlekey comparison: 27467 (may have improper matches for short strings)

FedDoc metadata analysis

Description of MS Access tables and queries:

1. All_retrieved_GovDocRecords: A unified table showing all SRLF/NRLF holdings and where titles are duplicated
2. HathiGovDocs: A rough list of HathiTrust Government Documents extracted from the HathiTrust datset from April 1, 2014.
3. HathiTrust_URLs: Access URLs and OCLCNumbers for individual items in the HT recordset.
4. nrlf_bibid_volume: A source table of all volume data associated with NRLF titles. Note, this table contains volume information for items not held at NRLF
5. srlf_allfedgovdocs: The source table from SRLF with all title/item data
6. srlf_bibid_volume_barcode: A derivative table from the source SRLF table showing volume information
7. Duplicated_GovDocRecords: A query showing all duplicated gov doc records
8. Duplicated_GovDocRecords_AGRI: A query with results from duplicated gov docs limited to *agri*
9. Duplicated_Monographs: A query showing duplicated gov doc monographs
10. Duplicated_Monographs_AGRI: Same query - limited to agri associated resources
11. Duplicated_non_AM_non_AS: Duplicated alternative formats
12. Duplicated_Serials: All duplicated serials
13. Duplicated_Serials_AGRI: Duplicated serials limited to AGRI titles
14. RLF_GovDocs_With_HathiTrust_DigitalDocuments: A rough query that matches HathiTrust data sources with the RLF data set to show what RLF titles have digital copies in HathiTrust.
15. Duplicated_Titles_and_item_Holding_Info: A report that brings together title duplicates and detailed volume information as well. At this point this report is for illustrative purposes only given the data issues in the NRLF item table.

Appendix D. UC Federal Documents Print Archive Disclosure Policy

UC Federal Documents Archive
Shared Print Disclosure Policy

1 Overview of Disclosure Policy

UC Archive Holders for the UC Federal Documents Archive agree to record information about the items included the archive in union catalogs and other applicable system(s) as established by this policy.

This document describes the policy, instructions, and metadata standards for disclosing shared print archives of federal documents to UC libraries, their users, and to the broader library and user community.

The policy supports the expression of retention commitments for all holdings at the RLFs, and the disclosure of stored holdings as volume-verified (based on volume level inventory control). The policy can be revised in a later phase, if disclosure of validation for completeness and condition are undertaken for campus in place collections.

This policy is an abbreviated one to launch the UC federal documents archive and will be superseded by a fuller UC shared print disclosure policy, currently under development by the shared print teams.

2 Goals of UC Federal Documents disclosure

Disclosure is intended to support three primary goals:

- Resource-sharing among UC members (to be defined in the UC Shared Print Access Guidelines)
- Collection analysis to support local and system-level collection management decisions
- Discovery of shared materials

3 Archive Holder definition

Archive Holders are defined as the location where the physical items are ultimately retained. If materials are held at an RLF, the RLF is the Archive Holder; if at a campus library, the library is the Archive Holder. It is important to identify and associate the location and archival status to support the above goals for discovery, resource sharing and collections analysis.

4 Outline of Disclosure procedures

Archive Holders record information about shared print materials in

1. OCLC WorldCat to support resource-sharing and global discovery

2. Their local catalog and Melvyl for local collection management and duplicate screening support

The specific workflow and sequence of these disclosure actions may vary among libraries.

The disclosure instructions for recording shared print materials in WorldCat follow the OCLC shared print metadata guidelines [https://oclc.org/services/projects/shared-print-management.en.html].

a. Use Shared Print Institution Symbols. Each RLF and each campus has one OCLC Shared Print Institution Symbol to identify the storage facility's Shared Print collections and the library's shared print in place collections. These shared print symbols indicate the location/status of the material and cover materials under any program. Each UC Library and RLF uses its symbol for any current and future shared print collections it manages on behalf of a broader group. (The specific shared print program is identified elsewhere in the records such that the same materials can be contributed to multiple archiving programs and the symbol is used to federate them all at a location.) See Attachment 2 for a list of OCLC Institution Symbols for shared print collections at UC libraries and Attachment 3 for the print archiving programs.

b. Local Holdings Records (LHRs). For each title, create a new Local Holdings Record (LHR) to define the shared print holdings, the shared print Institution Symbol, the print archiving program(s), retention commitment and original owner. See Attachment 1 for detailed entries and Attachment 3 for the archive program names to be used.

   Most of the shared print information is recorded in the 583 Action Note. Each LHR will include one 583 Action Note (first 583 action note only):

   - 583 **‡ a Action**="committed to retain".

c. Update campus holdings records

   - Following OCLC guidelines, and consistent with UC's adopted policy for the substitutive approach to recording holdings, libraries should remove the archived holdings from the original LHR or holdings data, so they will no longer be reflected under the original Institution Symbol. (The original institution symbol is captured in a subfield on the new LHR (561) so the custodial history is kept.) Holdings contributed to an RLF should be removed from the campus library holdings and recorded on the RLF shared print symbol and LHR.

   - The disclosure information in the Archive Holder's local catalog system is the source for future batch record updates in OCLC.

d. Simplified disclosure and display. Full runs recommended. To simplify holdings display, cataloging, and other downstream activities, Archive Holders are encouraged to commit all holdings for a particular publication held at the location (not partial runs).

e. Batch creation of LHRs. UC Archive Holders are encouraged to create the necessary LHRs through batch processing to the extent possible. Libraries should not individually process the volumes and metadata for those titles, but rather should use the least intensive means possible to identify and record the first 583 action note.

   LHRs are transferred using the MARC Format for Holdings Data (MFHD) standard. Details of how libraries may generate and export LHRs will vary depending on the library's local system and available expertise.

f. Online creation of LHRs. For libraries entering less than 100 records, OCLC recommends using Connexion Browser for entry.

g. Discovery in WorldCat. UC materials added to the WorldCat database become discoverable through OCLC interfaces that search and display WorldCat database records.

   - Search and display in WorldCat.org and FirstSearch. Holdings added under the new shared

print Institution Symbol will automatically appear in WorldCat.org and FirstSearch (if subscribed) associated with the name of the new symbol. Shared print holdings can be searched using the "l8" command.

- WorldCat Local. As users of WorldCat Local, UC Libraries may configure their WorldCat Local catalog to display shared print holdings under the shared print symbol.

Attachment 1

LHR Fields and Subfields Required for Disclosure

Required Fields Summary

The following fields are required to identify shared print materials in WorldCat.

- **OCLC control number** of the corresponding WorldCat bibliographic record. This can be the **004**, **014** or **035** field but it must consistently be in the same location in all records. Required for WorldCat.
- **Leader and Directory**
- **001** - Local System Control Number
- **007** - Physical Description Fixed Field
- **008** - Fixed-Length Data Elements
- **022** – ISSN
- **561** - Ownership and Custodial History
- **583** - Action Note(s)
- **852** – Location
- **85x/86x** Coded holdings (formatted holdings pairs) (if available)
- **866/867/868** Summary holdings (text) (if no 85x/86x formatted holdings pairs)

Details for selected fields

Tag and subfield(s)	***Name***	***Description***	***Example***
852 Location: An LHR identifies the holdings for a given title at a given location (Institution Symbol).			
852 ‡a	Location	UC shared print Institution Symbol for the Archive Holder. (See Attachment 2 for a list of UC Institution Symbols).	‡a ZASSP [UC SRLF example]
852 ‡b	Sublocation	Holdings Location Code (HLC) where the archived volumes are physically located at the Archive Holder library or storage facility	

85x/86x or 866 Holdings: The holdings committed to the shared print archive for this title, i.e. the holdings covered by this LHR. Include supplements and indexes that may have been published for the title. Enter these holdings as coded (formatted) detailed holdings if possible, otherwise enter a summary holdings statement.			
85x/86x	Coded holdings (formatted holdings pairs) for basic bibliographic units.		
866/867/868	Summary holdings (text) if no 85x/86x formatted holdings pairs		
022 International Standard Serial Number (ISSN) The ISSN is a very important match point for collection analysis. Include the ISSN in the LHR.			
022 ‡a	ISSN	ISSN for the title record. If there is more than one ISSN in the bibliographic record, use the first one.	
561 Ownership and Custodial History: Identify the original owner(s) of the materials			
561 ‡a History	Institution Symbol	Institution Symbol of library that provided materials to the Archive Holder identified in 852 ‡a Location. If the Archive Holder is the original owner, this would be the library's original or primary Institution Symbol, where the 852 ‡a would contain the library's shared print Institution Symbol.	
561 ‡3	Materials specified	Optional. If used, identifies the holdings originally owned and contributed by the institution identified in ‡a History.	
561 ‡5	Institution	If applicable, the MARC organization code for the original owner.	
583 Action Note: Most of the print archiving information is recorded in the 583 Action Note. Each LHR will include one 583 Action Note as described below: 1) a 583 note describing Retention commitment			
1. 583 Retention note			
583 ‡3	Materials specified	Include if this 583 Action Note describes a different set of holdings than were specified in the LHR holdings fields (85x/86x/87x or 866).	583 $3 v.1-3 INDEX: v.1 SUPPL: v.3 (example of INDEX and SUPPL labels)

		Enter the range of holdings covered and indicate gaps if known.	
583 ‡5	Institution	If applicable, the MARC organization code for the Archive Holder.	
583 ‡a	Action	Type of preservation action. For all FEDARC titles this 583 Retention Note contains "committed to retain". This is not machine controlled vocabulary, but it is controlled vocabulary for print archives. Do not vary.	"committed to retain"
583 ‡c	Time/Date of Action	Date this title was committed to UC Shared Print (YYYYMMDD)	
583 ‡d	Action interval	Use a standard retention end date, regardless of when the holdings were included in the print archive. This date can support future retention reviews and reaffirmation of the commitment to retain.	"December 31, 2040"
583 ‡f	Authorization	Repeatable field containing the name(s) of the archiving program(s). For the UC Federal Documents Archive, enter two Authorization subfields: one for UCL Shared Print and one for the Federal Documents Archive. This field is indexed and supports searching.	‡f UCL Shared Print ‡f Federal Documents Archive
‡u Uniform Resource Identifier	Link to program documentation for print archiving program identified in **‡f)**	URL where documentation for the program is publicly posted and maintained	‡u http://www.cdlib.org/services/collections/sharedprint/agreements_combined.html

Attachment 2:

UC Institution Symbols and Holdings Location Codes

Shared Print Institution Symbol	Institution/Meaning	ILL Supplier	Holdings Location Codes	Uses
University of California Libraries and Storage Facilities				
CUYSP	UCB Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CUVSP	UCD Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CUISP	UCI Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CLUSP	UCLA Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
MERSP	UCM Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CRUSP	UCR Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CUSSP	UCSD Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CUNSP	UCSF Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CUTSP	UCSB Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
CUZSP	UCSC Shared Print in Place	Supplier	Local code	Used for UC Shared Print in Place and UC Bronze Archives
ZAPSP	NRLF Shared Print in Storage	Supplier	Local code	Used for UC Shared Print and for UC WEST Bronze, Silver, Gold Archives
ZASSP	SRLF Shared Print in Storage	Supplier	Local code	Used for UC Shared Print and for UC WEST Bronze, Silver, Gold Archives

Attachment 3

Print Archive Program Subfield $f Authorization

The name of the print archive program is captured in the LHR $f authorization. It is used, in combination with the OCLC shared print institution symbol, to support search and resource sharing of shared collections.

The program name is indexed and it is a repeatable field such that the same materials can be contributed to multiple programs.

For UC shared print projects, enter two Authorization subfields:

- one for UCL Shared Print
- one for a specific project/collection

Table 1: Values for $f Authorization

Subfield $ Authorization Value	Application
UCL Shared Print	Applied to all UC shared print collections (required)
Federal Documents Archive	Applied to project/collection