

# **UC Federal Documents Archive Project: Progress Update and Recommendations for Next Steps**

*September 23, 2017  
(Approved February 13, 2018)*

## **FedDocArc Implementation Oversight Team:**

Elizabeth Dupuis (Project Lead, DOC, UCB)

Paul Fogel (Google Books Liaison, CDL)

Lynne Grigsby (Systems Lead, UCB)

Cathy Martyniak (SRLF)

Erik Mitchell (NRLF)

Vincent Novoa (UCR)

Jesse Silva (Collection Lead, GILS, UCB)

Emily Stambaugh (Shared Print, CDL)

Tin Tran (SRLF)

# UC Federal Documents Archive Project: Progress Update and Recommendations for Next Steps

In 2014 the UC Libraries endorsed the “UC Federal Documents Archive: Report and Recommendations,” approving the development of a shared UC archive of US federal government documents to create a more coherent and coordinated collection, improve access through increased digitization, and regain shelf space from unneeded duplication. FedDocArc is designed as a persistent archive that consists of one print and one digital copy of all US federal government documents owned by the UC Libraries. The project scope was envisioned as four interrelated phases. The Council of University Librarians (CoUL) charged the FedDocArc Implementation Oversight Team, reporting through the UC Libraries’ Direction and Oversight Committee (DOC), to undertake the first two phases utilizing the current resources of the UC Regional Library Facilities (RLFs), California Digital Library, and UC Berkeley. This report provides a progress update of the work accomplished in the past two and half years, and an outline for the planned next steps.

<b>Executive Summary</b>	2
<b>Project Phases and Outcomes</b>	2
● Phase One: RLF Collections for Print Archive	2
● Phase Two: RLF Collections for Digital Archive	3
● Phase Three: UC Campus Collections for Print and Digital Archives	4
<b>Key Issues</b>	4
● Selective Housing Agreement	4
● California State Library	5
● Workflows	5
● ERIC Microfiche	5
● Metadata Comparison of RLF Records	5
● RLF Shelf Check Sample	6
● RLF Monograph Record Comparisons	6
● RLF Serials Comparisons	6
● Offers Process	7
● Disclosures	8
● Digitization	8
● Pilot of Campus Collections Reviews	9
● Relationship with HathiTrust	9
<b>Looking Ahead</b>	10
Appendix A: Goal and Core Principles	12
Appendix B: Selective Housing Agreement	13
Appendix C: Google Project Cost Guidelines and CDL Statement of Work	16
Appendix D: Proposed Schedule for Campus Collections Reviews	18

# Executive Summary

From January 2015 to September 2017 the FedDocArc Project has made significant progress. In that period the UC Libraries have completed the foundation of the print and digital archive from the items shelved at the Regional Library Facilities (RLFs), including articulating processes for retention disclosures, engaging other depository libraries, and coordinating with digitization partners. Although the plan was to work through each of the four phases sequentially – first completing the foundation of the print and digital archive from the items shelved at the RLFs – several campuses were eager to engage sooner, so we experimented with a campus pilot as well. By undertaking elements of multiple phases, we have gained a richer picture of the opportunities and challenges, suggesting ways to adapt and improve our processes as we continue the project.

One of the most challenging aspects has been the advancing the project with current staff from Berkeley, the RLFs, and CDL – all of whom have other primary responsibilities as well. At one point Berkeley simply did not have the programming capacity to proceed so part of the FedDocArc Project was paused and the timeline extended. This spring Berkeley hired a temporary programmer to support this project, bringing welcomed additional human resources and renewed focus; we have now caught up with processing all decisions about holdings at the RLFs as well as developed the specifications for an improved approach to managing future collection comparisons and decisions. If the UC Libraries are satisfied with the general pace of progress and planned approach for next steps, no significant changes are needed in order to continue the work of the FedDocArc Project at this time. As Berkeley deals with budget cuts, changes may be required.

## Project Phases and Outcomes

The FedDocArc Project, following the original goal and core principles as outlined in Appendix A, has completed portions of work related to three of the four phases outlined in the original report, allowing us to leverage the wealth of materials already housed at the RLFs, to synchronize with the digitization partnerships already created, and to work methodically and efficiently with the collections owned by each of the UC Libraries.

### Phase One: RLF Collections for Print Archive

**Goal:** Review the print collections of US federal government document titles currently shelved at the UC Regional Library Facilities (NRLF and SRLF); retain and disclose one print copy of each title and volume; remove duplicate copies for digitization if a digital copy is not available in HathiTrust, to send to a requesting California depository library, or to process for withdrawal.

**Process:** Create a joint database of records from NRLF and SRLF, determine markers for identifying federal government documents in print, determine markers for matching records, generate separate list for single-volume monographs from other items, develop criteria for determining which RLF copy would be retained, mark the action designated for each copy, generate lists to process for the given action.

**Highlights:**

- The quantity of records from the RLFs and complexity of matching exceeded the ability of our original database that relied on developing scripts, producing Excel spreadsheets for review, and ingesting the decisions from the spreadsheets back into the database. To ensure the initial items were handled correctly, we reviewed lists a second time before adding the disclosures.
- In spring 2017 the Berkeley team, with the addition of a programmer hired by Berkeley to work on this project, began prototyping a completely new database designed to better address the lessons learned over the first two years. The new database is anticipated by March 2018.
- Articulation of various indicators for federal document to better compare/match records
- Articulation of various scopes for lists (e.g. multi-volume monographs and serials together)
- Some records require visual review of the physical items especially with different cataloging practices for volumes/parts/years in the holdings
- Identification of non-print formats can be difficult to determine by the records
- UC Government Information Librarians made valuable contributions in reviewing lists

**Outcomes:**

- 54,007 items disclosed at NRLF
- 140,073 items disclosed at SRLF
- 18,124 items offered before withdrawal
- 102 items transferred to other depositories from the offers claimed
- Developed steps for disclosure of monographs and serials as part of FedDocArc
- Developed procedures and forms for efficiently offering titles to other depository libraries

## Phase Two: RLF Collections for Digital Archive

**Goal:** From the duplicate items identified in the print collections of US federal government document titles currently shelved at the UC Regional Library Facilities (NRLF and SRLF), pull and send items for digitization by Google or in house; ensure the record for the print item is pulled in the case of sheet-fed digitization; ensure digital copies are designated as US federal documents and contributed to HathiTrust.

**Process:** Create a list of duplicates flagged to be digitized; sort for sending to Google or handling in-house; coordinate Google shipments with CDL; check out and track items to ensure records are handled appropriately; add costs for shipment preparation to CDL-funded reimbursement for Google digitization activities.

**Highlights:**

- Carefully manage pacing so not to duplicate what we are sending to Google from various UCs but also batching enough for sheet-fed scanning to be sent at one time
- Pulling from NRLF involved developing special pull list steps, identification of items not meeting Google criteria to channel for alternate scanning, and extra coordination of shipment/weight/date

**Outcomes:**

- 6,953 scanned by Google from NRLF

## Phase Three: UC Campus Collections for Print and Digital Archives

**Goal:** Compare the print collections of US federal government document titles currently shelved at each UC campus library with the items previously disclosed as part of the print archive; retain and disclose one print copy of each unique title and volume; remove unwanted duplicate copies for digitization if a digital copy is not available in HathiTrust, to send to a requesting California depository library, or to process for withdrawal.

**Process summary:** Campus sends all exported MARC records to Berkeley to identify items of overlap and unique items for various actions (retention on campus or at an RLF, digitization by Google or in house, or offers/withdrawal; Berkeley sends campus various sets of lists for their review and actions; campus decisions are sent back to Berkeley to update the database for tracking

### Highlights:

- Determined it is more efficient to work with one campus at a time for their full collection rather than with all campuses for an agency-by-agency review
- Pacing and staffing at a campus is influenced by their ability to assign current staff to the project and to develop procedures that work best for them
- Shared print copies can be sent to RLF or kept on campus as shared print in place; currently all print items are held at the RLFs
- Developing new approach to let campuses know all the items already part of the print and digital archive so they can withdraw those titles sooner if they wish
- Berkeley – Experimented sending specific titles to send to Google sheet-fed scanning
- Riverside – Agreed to participate as first campus starting summer 2015; completing monographs and preparing for serials review
- San Diego – Agreed to participate as next campus starting fall 2017; beginning with items possible to withdraw, then monographs to contribute

### Outcomes:

- 25,662 items scanned by Google from UCR
- Set pacing agreements with each campus to keep process moving in a manageable and predictable pace for all involved, ideally each step as six months

## Key Issues

### Selective Housing Agreement

In April 2015 we filed a final version of the Selective Housing Agreement signed by all members of CoUL to the Government Publishing Office and the California State Library, our regional depository (see Appendix B). Confirmation of this document was a critical step toward reassuring all parties of UC Libraries' commitment to create a shared print set first, then provide access to a complimentary digital copy.

## California State Library

The California State Library (CSL) is the designated regional library for our state, and as such, oversees the Federal Depository Library Program (FDLP) libraries in California. UC Libraries have maintained a successful, long-standing working relationship with CSL. Because of this relationship, CSL has been a supportive partner of FedDocArc Project from the beginning and CSL's support can best be seen in the flexibility UC Libraries now have with the FDLP disposal process. FDLP procedures dictate a cumbersome disposal process when libraries wish to weed and discard government documents from their collections. For the purposes of creating the FedDocArc, UC Libraries are allowed to offer directly to California libraries anything published after 1970 (unless it pertains to a small number of subjects about California or California issues), thus eliminating the time consuming step of CSL's own review of the disposal list. UC Libraries can also check current offers against previous offers approved by CSL, further streamlining the offers process for both organizations. UC Libraries have agreed to share our growing list of offers with CSL to help them with checking titles from other libraries (much of CSL's FDLP collection is not cataloged). UC looks forward to continuing our relationship with the CSL.

## Workflows

The Berkeley team has developed many processes and workflows – including designating file conventions and tracking, managing and updating the master database, generating review lists, reviewing and marking decisions on the lists, pulling copies for other scanning, pulling copies for Google scanning, handling post-scanning files, offering unneeded titles, making record changes for batch suppression/withdrawal, making record changes for batch disclosure, checking RLF shelves, recycling unclaimed items, checking and improving HathiTrust files, and more. Based on experience of these first two years, several steps have been refined and we will continue to improve our efficiency with time.

## ERIC Microfiche

CoUL endorsed the transfer of Berkeley's complete set of ERIC microfiche documents to the NRLF to become part of FedDocArc. Other UC Libraries had stored incomplete sets and it was determined to be most efficient to adopt the Berkeley set in one step rather than compare and consolidate holdings. In doing this, all other UC Libraries could withdraw their holdings if they no longer wished to keep them. Berkeley provided the cataloging for all items to allow search by their ED number and to include URLs for e-versions when available. Requests for items from the ERIC microfiche collection come with specific document identifiers that allow NRLF staff to pull the precise document, create a digital scan of the item, and fulfill the request as they do with journal articles. Since the collection records were made available this way there has been a huge increase in use. Unfortunately because these items have newer cataloging dates, they often appear higher in search results and users who are unfamiliar with how to read bibliographic records are often disappointed to learn that the items is an abstract rather than the full text.

## Metadata Comparison of RLF Records

With shared commitment of library staff at UCB/NRLF and UCLA/SRLF we gathered the records for items at the RLFs that met criteria indicating they are US federal documents. After much work to

massage the data in compatible forms, we were able to create a MySQL database from which we generated lists. Through our work the pilot year we learned that:

- we needed to work with all records rather than just those thought to match selected characteristics in order to produce the best quality matches for government documents
- we needed different software to cope with the enormous number of records from NRLF, SRLF, and all of the 10 UC Libraries
- we needed database architecture that allowed matching and decisions to be handled online rather than through generation of spreadsheets

We have developed the new database in MySQL, coded in Python, but is designed to incorporate some of our findings about better matching as well as to track disposition of items to reduce the need for duplicate checking of work to confirm decisions. Now that we have completed the initial work with collections at the RLFs and UCR, we will pull fresh exports of these database holdings and begin to use the new database.

## **RLF Shelf Check Sample**

As recommended by the original report, at the start of our work we sought to check how reliable our metadata records in the ILS are regarding the status of our government documents shelved at the RLFs. A sample set of 1000 items was generated following the ration of US federal documents at each RLF (~63% SRLF / 37% NRLF), SRLF and NRLF staff checked their items over winter break 2014, finding the RLF record status not reflecting the actual status of the item for only 4 items, representing 99.996% accuracy. The shelf check and problem solving work took an average of 1.5 minutes per item. From this work we agreed to make action decisions for items without a second pre-shelf check.

## **RLF Monograph Record Comparisons**

Our initial review focused on print monographs at the RLFs. Through this experience we learned to separate out special collections, non-print formats, and multi-volume sets. We generated lists of records that indicate exact matched copies were held at both NRLF and SRLF; in this case NRLF retained the print copy and SRLF gained the freed shelf space from removal of duplicates. We also generated lists of records for monographs held uniquely by either NRLF or SRLF and included all those items in the print archive. We developed guidelines and a standard set of codes for decisions to enable us to enlist help of other library staff in future reviews while ensuring consistency in decisions.

## **RLF Serials Comparisons**

Serials require human review for accurate comparison of holdings, especially with the complications of numbering of issues/parts for government publications and a wide variety of approaches to binding and cataloging. We developed guidelines for reviewing records and marking action items. Members of the UC Government Information Librarians (GILS) group volunteered to review subsets of individual issues from the list of federal document serials housed at the RLFs with Jesse Silva providing initial training and service as a resource person. Additionally Jesse Silva and Glenn Gillespie piloted a process for physically comparing NRLF and SRLF holdings to clear up the initial set of problem records for which the metadata provided inadequate to make an accurate assessment about duplication.

## Examples of Titles Needing Further Review

In doing reviews of serials and multi-volume monographs, the FedDocArc content review team quickly realized that we would run into titles which require additional in-depth checking of the holdings because of the different practices in the recording of holdings between the northern and southern campuses. Below are three examples illustrative of the more common differences in practice.

Title	NRLF Holdings	SRLF Holdings
Departments of Labor and Health, Education, and Welfare appropriations for 1977 : hearings before a Subcommittee of the Committee on Appropriations, House of Representatives, Ninety-fourth Congress, second session / Subcommittee on the Departments of Labo	J61.A6 94th v.45-48 v.45; J61 .A6 94th v.45-48 v.46;	HV85.U58d 1976 pt.01, pt.02, pt.03, pt.04, pt.05, pt.06, pt.07
Indian education.	E97.U65 no.1-79	No call number Yr.1914-15, yr.1968-69 , yr.1969-70 , yr.1970-71
Composite report.	UA12.U56 1959 1959	UA12.U535 1959c Vol 1, 2

Because of these issues, FedDocArc team recommends a component of human checking of certain classes of records so we do not mistakenly discard items we need to retain.

## Offers Process

For US federal documents there is a well-recognized “needs and offers” process that allows depository libraries to make excess titles available to other depository libraries before withdrawing them. It was important that we find a process effective for geographically dispersed and large-scale collections. Initially the FedDocArc Team explored use of the open source software developed by ASERL, however that software could not be easily modified to support the needs of our project. Because we sought a solution that could be developed quickly, required little funding for development or maintenance, and offered transparent and easy communication across multiple entities, we crafted a lightweight solution based on Google spreadsheets and Google forms. This turned out to be a very flexible and efficient approach for sharing lists of unneeded duplicate titles with other depository libraries in California and allowing them to claim desired titles.

Our first list was monographs from SRLF for which we had a print copy at NRLF and a digital copy already available in HathiTrust. We piloted steps for pulling the designated items from the shelves, tracking the freed shelf space and barcodes for reuse by new deposits, sorting and publishing the



items by their pre-1970 and 1970+ publication dates, fulfilling claims from other depositories, withdrawing the records from the RLF and owning campus catalogs, and properly disposing of the remaining items. This same process was effectively used by UCR during their offer stage too.

## Disclosures

As recommended in the original report, FedDocArc print archive disclosures were designed according to OCLC metadata standards where they existed. We developed a shared print monograph disclosures process ourselves since OCLC developments proceeded slower than our needs. Records of all items at NRLF will have the location nu03b, item messages indicating shared print and building use only, and an internal note designating retention commitments to December 31, 2035 as part of our US Federal Documents project. Similarly, SRLF items will have the location srucl3, item messages indicating shared print and building use only, and a 583 statement designating retention commitments to December 31, 2035. Serials have been disclosed through OCLC; monographs will be disclosed through OCLC once a process has been confirmed that is compatible with their system and ours.

## Digitization

We have been able to leverage the CDL-coordinated UC/Google Books partnership to operationalize the destructive (i.e. sheet-fed) scanning of US federal government documents. CDL has integrated the coordination and logistics involved in FedDocArc with the larger project, and has also managed the reimbursement to campuses for costs incurred in these activities at the same rate as those associated with the general Google Books project work.

In 2015 we piloted a sheet-fed digitization project between Google and Berkeley, sending approximately 500 items/month to begin. Since April 2016, UCR has averaged 1,200 items/month for a total of over 25,000 items. Through this work we gained knowledge of how long it takes to pull and process each batch of items (roughly 1.5 minutes to pull from the shelves and another 1.5 minutes for list prep work and packing onto Google trucks). We also developed a better understanding of factors that make an item 'unscannable' for Google – such as too tight binding, brochures, and oversize foldouts. For these we explored other options, including piloting in-house scanning at select UC campus libraries, ensuring that the digital file is processed and contributed to the HathiTrust corpus. Our initial list of eligible items were derived from Google generated candidate lists; in the future we plan to contribute items from the RLFs and other UC libraries, and we have agreed to work with the HathiTrust committee developing the registry of US federal documents to fill gaps they identify in their digital holdings. We developed criteria for items not recommended for destructive scanning; these items are possible to digitize through non-destructive scanning by Google or in-house.

In 2015 staff from CDL, Berkeley, and NRLF adapted the steps developed from the sheet-fed scanning pilot at Berkeley to fit the special circumstances of the RLFs. The project identified a set of titles for which we had copies at both NRLF and SRLF but no digital copy was available. The SRLF copy was designated for the print archive and the NRLF was designated for scanning. A small group of Berkeley library employees spent two days at NRLF pulling the designated items from the shelves, checking out items to the special Google account, and packing items for shipment. In the pilot, we also developed a process for tracking the freed shelf space and barcodes for reuse, creating shipment manifests and batch of MARC records, shipping the items to Google, and

withdrawing records from the RLF and owning campus catalogs. Per an agreement with the California State Library and Google, once disbound and scanned, the items are disposed at the Google scanning facility. From our experience, within a few months of scanning the items are viewable in HathiTrust.

Our preference is to send federal documents through the Google scanning partnership, but some items do not meet the Google criteria. The category of items includes volumes either too small or too large, items with tight binding, and items requiring delicate care. At Berkeley the Digital Projects team has overseen a project to handle these special items, image all pages, and submit those digitized copies to HathiTrust. All NRLF titles identified in the November 2015 pilot pull project that were not yet scanned but did not meet the Google criteria are completed and now accessible in HathiTrust. Currently the unit is working on USGS journals that have too many foldouts for Google to accept. These will be scanned with anticipated completion by end of 2017.

## **Pilot of Campus Collections Reviews**

In response to interest from individual campuses to review federal documents shelved on their campus, and from the UC Libraries to share as many volumes for high-quality digitization as soon as possible, we began a pilot of Phase 3 a year earlier than planned. UC Riverside accepted the invitation to help work out the processes we could use with other campuses. The general steps include: getting a full export of all the campus' MARC records, comparing records of that campus with the records identified as part of FedDocArc print archive so far, and processing items for retention, scanning, or offers/withdrawal. CDL works with the campus to review the Google Books Project's guidelines regarding funding for campus operations and to formalize a statement of work appropriate for their collection. The project funds provided to the campus supports essential temporary staffing who focus specifically on this project (see Appendix C). UCR anticipates completion of their monograph review by end of 2017 and is ready to begin their serials review. UC San Diego accepted the invitation to become the second campus contributing to the project. Since only one campus can be reviewing certain batches of items at a time, we will begin UCSD's review of monographs as soon as UCR is finished. We will continue to look for efficiencies and develop more structured coordination with due dates for certain phases, ideally each as 3-6 months, in order to allow more libraries to plan ahead. The timing of involvement of all UC Libraries will be confirmed through discussions with CoUL. The FedDocArc team has outlined two options with a recommendation to continue with the general structure shown as option A (see Appendix D).

## **Relationship with HathiTrust**

UC Libraries are committed to providing permanent public access to the content of United States federal government documents; in partnering with HathiTrust, we are beginning to provide this level of access to the rest of the world. Early in our work the UC Libraries noted that, despite federal documents being in the public domain, many documents are not labeled as such. UC Libraries asked HathiTrust for the metadata signals they use to designate items as US federal documents and are ensuring that the records we pass to Google for the items we send for digitization, and send to HathiTrust for ingest via CDL's Zephir, meet these criteria; HathiTrust has been efficient in correcting titles that should be in the public domain thus enabling users to view the full text.

The HathiTrust US Federal Government Documents Registry aims to create comprehensive holdings of this corpus. As they explore options for completing their holdings, the FedDocArc project anticipates being an active contributor. One strategy is to use the already generated Essential Titles List (ETL); the ETL is a list that libraries should consider retaining in print and was developed by GPO in consultation with the FDLP community. HathiTrust is using this list to fill in titles, and UC Libraries have contributed early volumes to complete the holdings of the *Statistical Abstract of the United States*.

Another option is the ingest of digital collections of federal agencies. For example the US Geological Survey (USGS) provides high quality digital scans for many of their publications on their web site. Rather than create new scans of these items – many with multiple, large sized foldouts – it is useful to explore partnerships to ingest these materials into HathiTrust to create even more visibility and access as part of the digital archive we are building. In the meantime, Berkeley is piloting in-house scanning of full volumes including the critical fold-out maps and contributing those to HathiTrust.

Public access to government information is an important factor. Currently the general public is not allowed to download an entire document, which can be frustrating for their work. As such the UC Libraries continue to work to remove these barriers to access.

## Looking Ahead

At the current pace, is anticipated that this project will take six years to complete. Unless otherwise directed by UC Libraries' leadership, by January 2019 the FedDocArc Project will:

- Launch a new master database, capable of handling the scale of records involved, tracking final decisions made for individual items, and gathering statistics
- Run a fresh list of print federal documents in the RLFs that were not captured in the initial lists, adding disclosures for unique titles, scanning, or withdrawing as appropriate
- Continue work with UCR to complete the review of their holdings
- Initiate work with UCSD to complete the review of their holdings
- Confirm a general schedule for all UC Libraries interested in being involved
- Develop approach to generate list of items already in the print and digital archive so all UC Libraries could take local initiative to withdraw those items if desired before their future participation to contribute their unique local holdings
- Collaborate with HathiTrust on intentional development of the federal documents in their holdings with the goal of filling gaps and making their holdings comprehensive
- Advocate with HathiTrust to allow the general public to download entire federal documents, thereby allow UC Libraries and all other institutional partners to fully meet their commitment to permanent public access
- Explore further coordination within the UC Libraries for titles/items being locally scanned that will become part of the digital archive
- Develop document for UC library staff to identify items disclosed as part of FedDocArc

### **Anticipated future stages will involve:**

- Synchronize disclosures with national models when confirmed (coordinate with OCLC and HathiTrust retention disclosure policies and services)
- Document any special issues related to shared print in place items for FedDocArc
- Development of review approaches for non-print formats (microformats, maps, CDs, etc)
- Design of sustainable system for discovery and comparison of FedDocArc items for campus and RLF library staff (functional for contributing items) and users (functional for public access), with consideration for changes to policies for future deposits and digitization

The FedDocArc Implementation Oversight Team has considered questions such as:

- Do we need a different staffing or budget model?  
Until Berkeley, NRLF, or SRLF raise a concern, the current model has minimal overhead requiring no pooling or transfer of funds beyond CDL cost reimbursement. As the FedDocArc project moves to work happening at each campus, the workload will be increasingly shared. The funding currently managed by CDL to support digitization with Google Books has been invaluable, although the reimbursement model may need future modification to incorporate some cost sharing by campuses.
- Is the campus-by-campus approach effective?  
This approach has the benefit of working in a focused way with RLF quotas and Google shipments, allowing individual campuses to choose to keep duplicated titles useful locally. Government documents present challenges in determining what should be included overall, so the current approach has helped us build that list. An alternative approach could be designed like the Journal Archiving Campaigns. Collection analysis, multi-institution decision comparisons, and the creation of a master list are possible future partnerships with the HathiTrust Federal Registry.
- Could we increase our digitization by changing our project model?  
The UC Libraries have benefited greatly from partnering with Google, especially on the sheet-fed digitization that produces a higher-quality file. Additionally since 2014 Berkeley has focused on specific titles to digitize in-house. As the FedDocArc Project engages other campuses, we can draw upon their scanning capabilities also. The limiting factors at this time are the speed we can identify items for which we have a second print copy, the quotas we can send to Google, and our in-house staffing for local scanning.
- How would a UC-wide shared integrated library system influence this project?  
In an ideal world, this type of shared system could be very valuable and potentially much more efficient. No such system has been approved or developed thus far, as such it makes sense to continue to make progress using the current vision.

# Appendix A: Goal and Core Principles

## 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, including the general public, as “library use only.” Digital copies will be preserved in the HathiTrust repository.

## 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



## **Selective Housing Agreement for the University of California's Federal Documents Archive**

Last updated: March 2015

This Agreement for Selective Housing covers U.S. Government publications distributed through the Federal Depository Library Program (FDLP) and deposited into the University of California's Federal Documents Archive (FedDocArc).

FedDocArc is a cooperatively developed shared print archive created and maintained by the ten campuses of the University of California (UC). The archive will contain one copy of every print document received by a UC Library through the FDLP. Items deposited into the archive are viewed as the primary UC copy. The Library that designates an item for the archive continues to maintain control over the item and is referred to as the "controlling library." Items in the archive may be housed at one of the Regional Library Facilities (RLFs— UC's designated library shelving facilities) or on the campus that originally received it. Items deposited into the archive may not be weeded. All items in the archive will be accessible to anyone in the UC Libraries and by members of the general public.

This agreement is made in March 2015 by and between the Libraries of the ten campuses of the University of California (Berkeley, Davis, Irvine, Los Angeles, Merced, Riverside, San Diego, San Francisco, Santa Barbara, and Santa Cruz) and the RLFs (overseen by the Libraries at UC Berkeley and UC Los Angeles). This agreement is entered into for the purpose of developing and maintaining a cooperatively developed shared print archive of U.S. government publications by the UC Libraries. The publications are lent following UC lending rules and procedures, but remain the property of the U.S. Government Publishing Office under the shared control of the UC Libraries.

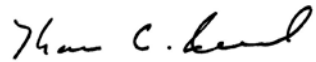
### **In pursuance of this agreement, all member institutions agree to:**

- Share responsibility for carrying out the provisions of this agreement for the U.S. Government publications designated for/deposited into the FedDocArc.
- Continue to make available for free and unrestricted use all Federal depository publications outlined in this agreement to the general public.

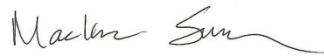
- Return from the RLF to the controlling UC campus library, to be housed at that library, any publication outlined in this agreement, upon request. These publications will continue to remain part of the FedDocArc and governed by this agreement
- Maintain a piece-level accounting of the Federal depository publications selectively housed under this agreement in the FedDocArc.
- Ensure all publications outlined in this agreement are identified as Federal depository property and date of initial receipt is recorded.
- Adhere to the retention/weeding procedures of the UC and follow the FDLP disposition guidelines overseen by the California State Library.
- Abide by all UC borrowing agreements and policies for government publications.
- Process depository material to ensure timely access to publications and for claims.
- Ensure each institution in this agreement is in compliance with Title 44 of the United States Code, Chapter 19, the Legal Requirements and the Program Regulations of the Federal Depository Library Program and terms outlined in this agreement.
- Transfer designated publications to the FedDocArc on an as needed, voluntary basis.
- Keep administrative records indicating the location of publications involved in this agreement.
- Provide collection development tools and information to each member of this agreement for the cooperative development of the FedDocArc.
- Accept responsibility for returning full administrative control back to the controlling library for all publications upon termination of this agreement.

All parties agree to review this agreement every 10 years and update as necessary.

On behalf of the UC, we the undersigned agree to the terms recorded in this Selective Housing Agreement and send this agreement and all future updates to the California State Library and Government Publishing Office.



Thomas C. Leonard, UC Berkeley /  
Northern Regional Library Facility



Mackenzie Smith, UC Davis



Lorelei Tanji, UC Irvine



Ginny Steel, UC Los Angeles /  
Southern Regional Library Facility



Donald Barclay, UC Merced



Steven Mandeville-Gamble, UC Riverside



Brian E. C. Schottlaender, UC San Diego



Karen A. Butter, UC San Francisco



Denise Stephens, UC Santa Barbara



M. Elizabeth Cowell, UC Santa Cruz



# UC Google Books Projects: Guidelines Regarding Funding for UC Campus Operations

## **Funded by CDL (CDL is reimbursed by Google for most costs)**

The following costs will be funded by CDL. CDL will be reimbursed by Google for a majority of the expense.

- Book pulling staff, including % benefits\* relative to hours worked
- Direct operational supervision staff, including % benefits\* relative to hours worked
- UC programmer time, including % benefits\* relative to hours worked to write initial code for generating picklists and metadata extraction. This does not include ongoing operational tasks.
- Temp agency fees directly associated with the Google project
- Minimal site preparation costs

*\* Campuses should indicate their benefits rate on the costs spreadsheet*

## **Campus contributed costs**

The following costs are not funded by CDL or Google.

- Barcode scanners
- Senior management overhead or any management above direct supervision of operations
- Local scanning of rejected items

## **Costs that will be considered for funding by CDL**

CDL does not typically fund these items, but each can be considered on a case-by-case basis. Payment may be dependent on CDL being able to charge back to Google.

- Barcoding
- Item record creation
- Consumables directly related to the project that will not be repurposable at the end of the project
- ILS costs

## **Process**

- Campus submits the Statement of Work form and Costs Spreadsheet for the upcoming fiscal year to CDL in April.
- CDL will carry out a Transfer Of Funds to the campus, based on the Statement of Work
- The process will be repeated annually for the duration of the project.

## **Fed Doc Arc Project**

- CDL will fund costs outlined in section 1 above when those cost are incurred as part of participation in the Fed Doc Arc project.
- Fed Doc Arc costs need to be differentiated from regular operations costs. Separate Statements of Work and Costs Spreadsheets need to be submitted to CDL.



**STATEMENT OF WORK**

PROJECT TITLE: FedDocArc Mass Digitization

CAMPUS: UC Riverside

DATE: 02/24/2016

PERIOD OF PERFORMANCE: 01/04/2016 – 06/30/2017 (18 months)

TOTAL COSTS PROJECTED FOR PERIOD OF PERFORMANCE: \$62,102.45

**SCOPE OF WORK:**

The UCR Library is commencing participation in the Mass Digitization Project with Google. UCR will be making available approximately 2,300 federal documents every four weeks to Google for sheet-fed (destructive) scanning. The project is scheduled to continue at least through December 2017. CDL will fund direct operational staff and activities according to existing guidelines.

This work encompasses the following:

1. Pull from shelves, prepare, and ship documents to Google on agreed-upon schedule.
2. Maintain statistics for the shipments
3. Participate in regular operations review meetings
4. Work with Google and CDL staff to resolve problems

COSTS: See attached UCR Campus Costs Template for weekly breakdown.

Projections for cost for GBS for UCR Library				
	Hours/% of Time	Salary	Benefits	Total
Student Staff	3,330 hrs.	\$ 34,695.00	0%	\$34,695.00
Project Manager	74 hrs./2.5%	\$ 2,750.58	19%	\$3,273.19
Supervisor	148 hrs./5%	\$ 3,475.04	19%	\$4,135.30
Library Assistant	888 hrs./30%	\$ 16,578.96	19%	\$19,728.96
		\$	%	\$
<b>TOTAL - Monthly</b>				<b>\$ 3,450.13</b>

**PROJECT CONTACTS:**

1. Paul Fogel, Acting Manager, Mass Digitization, CDL
2. Vincent Novoa, (UC Riverside) FedDocArc Project Manager

## FedDocArc Proposed Campus Schedule

### OPTION A

2017

UCR monos (pilot)

December 2017 - May 2018

UCR serials

UCSD monos

June 2018 - December 2018

UCSD serials

UCSB monos

January 2019 - May 2019

UCSB serials

UCLA monos

June 2019 - December 2019

UCLA serials

UCB monos

January 2020 - May 2020

UCB serials

UCD monos

June 2020 - December 2020

UCD serials

UCI monos

January 2021 - May 2021

UCI serials

UCSC monos

June 2021 - September 2021

UCSC serials

UCM monos

October 2021 - January 2022

UCM serials

UCSF monos

February 2022 - July 2022

UCSF serials

### OPTION B

#### Monographs

To November 2017: UCR (pilot)

December 2017 - March 2018: UCSD

April 2018 - July 2018: UCSB

August 2018 - November 2018: UCB

December 2018 - March 2019: UCLA

April 2019 - July 2019: UCD

August 2019 - November 2019: UCI

December 2019 - February 2020: UCSC

March 2020 - May 2020: UCM

June 2020 - August 2020: UCSF

#### AND

#### Serials

December 2017 - May 2018: UCR

June 2018 - December 2018: UCSD

January 2019 - May 2019: UCLA

June 2019 - December 2019: UCB

January 2020 - May 2020: UCD

June 2020 - December 2020: UCI

January 2021 - March 2021: UCSC

April 2021 - June 2021: UCM

July 2021 - September 2021: UCSF