

## **UC Federal Documents Archive Project Update**

*November 2016*

*Elizabeth Dupuis, Project Lead*

This update summarizes progress on the UC FedDocArc project since May 2015.

### **Confirmed Disposition Process**

The ASERL Documents Disposition database, previously identified as a possible model for us, is built with separate databases called upon to provide the search and claim capabilities. In 2015 after a sustained investigation and drawing upon consultations with ASERL staff, we concluded that the ASERL Document Disposition Database could not be modified to support the needs of our project. Because we required a solution that 1) could be developed within a couple weeks, 2) would require little funding for development and maintenance, and 3) would allow for transparent and easy communication with UC librarians, the California State Library (CSL), and other state selective depositories, 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.

### **Finalized Offers Process**

In 2015 we identified monograph titles at SRLF determined to be duplicates of items also shelved at NRLF and already available digitally in HathiTrust. We piloted the steps for pulling the designated items from the shelves, tracking the freed shelf space and barcodes for reuse, publishing the list of items through the offer database to other state depositories, fulfilling claims from other state depositories, withdrawing the records from the RLF and owning campus catalogs, and properly disposing of the remaining items.

To date, we have offered:

- 22,360 volumes from the RLFs (monographs and serials)
- 12,474 from UCR (monographs)

Of all the offers made, 102 items were transferred to other federal depository libraries in California.

The California State Library (CSL) has recently agreed to let UC Libraries check our future offers against offers lists previously made to them. This will streamline the offers process with the State Library (CSL will not have to check titles multiple times) and will help us meet our timelines. We have also agreed to share our growing list of offers with CSL to help them with checking titles from other libraries (CSL's collection is not cataloged).

### **NRLF Trial Pulling for Google Sheet-fed Scanning**

In summer 2015 staff from CDL, UC Berkeley, and NRLF adapted the steps developed from the sheet-fed scanning pilot at Berkeley to fit special circumstances of the RLFs. The project identified a set of titles for which we had copies at SRLF and NRLF; the SRLF copy will be designated as the shared print archive copy. These items were not yet available in HathiTrust, so the NRLF copy was designated for scanning. For this first set of 6,111 items, a small group of UCB Library employees

spent two days at NRLF pulling the designated items from the shelves, checking out items to a special Google account, and packing the items for shipment. In the pilot NRLF and Library Systems staff also developed a process for tracking the freed shelf space and barcodes for reuse, creating the shipment manifests and batch of MARC records, shipping the items to Google, and withdrawing the records from the RLF and owning campus catalogs. Per the agreement with the California State Library and Google, once disbound and scanned, the items will be disposed at the Google scanning facility. From our experience, within a few months of scanning, the items are viewable in HathiTrust.

### **RLF Serials Review**

Serials require human review for an 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. From our initial work reviewing serials, we developed guidelines for reviewing records and marking action decisions. Members of the UC Government Information Librarians 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 serving 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 proved inadequate to make an accurate assessment about duplication.

### **Print Archive Disclosure**

The FedDocArc has begun officially identifying items as part of the shared print archive in item records for monographs and holdings records for serials shelved at the RLFs. Records of all items in 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. 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. Currently we are discussing the parallel implementation for SRLF. Once this process is complete, we will be able to announce the number of items officially disclosed as part of the UC Federal Documents Archive and a subset of UC Government Information Librarians is ready to begin training UC Library staff on identifying these records.

### **Pilot of Campus Federal Documents Holdings Analysis**

In response to interest from individual campuses to review the 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. The project extended an invitation to UC Riverside as the first campus to work out the approach we can use with the 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 the archive so far for the batches eligible to will go to digitization and offers to other depositories, identifying all other federal documents through a range of techniques for the batches eligible to become part of the print and/or digital archive. We hope to collect data about the time and staff required to complete each of these steps.

### **Pilot of In-House Scanning at Berkeley**

Our preference is to send federal documents through the Google scanning partnership, but some items do not meet the Google criteria. This 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 which have too many pullouts for Google to accept. These will be scanned with an anticipated completion date in spring 2017.

### **Project Oversight**

Documentation about the UC FedDocArc project is available on UC Libraries wiki:

<https://wiki.library.ucsf.edu/display/UCLFDA/FedDocArc+Project>

As of April 2016, the FedDocArc Implementation Oversight Team is:

- Elizabeth Dupuis (Project Lead, DOC, UCB)
- Jesse Silva (Collection Lead, GILS, UCB)
- Lynne Grigsby (Systems Lead, UCB)
- Paul Fogel (Google Books liaison, CDL)
- Erik Mitchell (NRLF)
- Tin Tran (SRLF)
- Vincent Novoa (UCR)
- Emily Stambaugh (Shared Print, CDL)

As of August 2016, the HathiTrust Federal Documents Advisory Committee announced refreshed membership including Elizabeth Dupuis (UCB) and Ivy Anderson (CDL).