

UC Federal Documents Archive Project Update

December 11, 2014

Elizabeth Dupuis, Project Lead

This update summarizes progress on the UC FedDocArc project since October 2014, and includes two requests for action.

Implementation Oversight Team Charged and Appointed

An implementation oversight team has been formed to officially begin work on Phase One and Phase Two of the creation of the archive. The charge is attached.

Selective Housing Agreement (SHA) – CoUL action requested

CoUL asked the original project team to choose whether to use a SHA or MOU to document the persistence agreement. The project team opted for a SHA and has created that document following the FDLP template. The document was reviewed by members of the original project team and SAG3 and is ready for each UL to sign. Once all UL signatures are gathered the agreement should be returned to Elizabeth Dupuis to formally file with the California State Library and GPO. Confirmation of this document is the first step toward confirming the California State Library's support of our disposition plans.

Advocacy for Full Text Access in HathiTrust – CoUL action requested

The disposition lightning team noted: “establishing a UC print archive with digital access expands the availability of federal documents and serves a valuable public service. To emphasize this strength of the digital archive and to successfully position FedDocArc as a publicly-accessible digital archive of our federal documents, the Team highly recommends that CoUL and UC Libraries advocate strongly for free and open viewing and full PDF download of all federal documents in HathiTrust.” This issue has been raised by several member institutions and, while it may take time to fully realize, it is important for the representatives for the UC Libraries to continue to champion this issue.

Proposed Disposition Process

A lightning team of GILS members was tasked to explore and document options for a streamlined disposition process able to handle the large-scale collection review we envision. As part of the FDLP agreement, we are expected to offer unneeded copies of federal documents to our regional depository and other state depositories; it is likely that many will be desirable since the UC collections are both broad and deep. Their report outlined three options. The approach preferred by the majority of members of the lightning team, original project team, and SAG3 captures four essential principles:

- 1) supports the needs/offer model at the heart of the depository program,
- 2) focuses emphasis on meeting the needs of the depositories in our state,
- 3) offers an efficient web-based approach to managing requests/confirmations, and
- 4) includes pre-determined periods of time for review for disposition decisions.

The disposition lightning team identified the open source software developed by ASERL for the disposition of federal documents (<http://apps.uflib.ufl.edu/GovDocs/>) as a model for further exploration. Creating a UC Libraries' instance is possible though further investigation of this tool is needed before a recommendation is possible. The FedDocArc Implementation Oversight Team will need to balance the desire to get a functional system in place immediately to proceed with Phase One, with the interest in exploring how such a

system might support similar needs/offers purposes for the UC Libraries. Unless discouraged by CoUL to pursue this direction, the new team will investigate the options and document the rationale, impact, and resources needed for the recommended approach for approval by SAG3 and CoUL.

Google Sheet-fed Digitization Pilot

The UC Libraries – especially CDL, NRLF and SRLF – have extensive experience with Google digitization. The FedDocArc project still has much analysis to do before we are ready to pull items from the RLFs to send to Google for sheet-fed digitization in spring 2015. To better understand the workflow and project the quantity possible to send, we are piloting the approach by sending items from the Berkeley campus starting in January 2015. Google provided a candidate list of Berkeley items not yet scanned; we filtered out those with a copy already stored at an RLF and are developing criteria for any other considerations to withhold the copy. Berkeley is prepared to send 500 volumes/month for the next year. Before the end of 2015, the FedDocArc project should be prepared to send items from the RLFs for digitization. During this pilot process we are documenting time spent on various steps to inform our estimation of costs for scaling up and sustaining this work.

Proposed Disposition of Items Sent to Google

After one copy of a title is adopted into the print archive, we will offer another copy to Google for sheet-fed scanning to fill gaps in the digital holdings available. These items will be disbound before scanning, hence impractical for use or re-shelving. By not requiring these to be returned from the Google scanning facility we will save needless shipping costs and save staff time in managing these items. With the commitment to identify the print copy into the archive first, the California State Library has indicated tentative support for this approach. With the signed SHA in hand, we will confirm a decision with the California State Library.

Metadata Comparison of RLF Records

With shared commitment of library staff at UCB/NRLF and UCLA/SRLF we have begun gathering the records at the RLFs. From this set we will eventually develop lists to determine duplication and adoption into the shared print archive. As an initial step, the FedDocArc project is undertaking a sampling the relevant government documents collections in the RLFs for a confirmation of the availability and condition of the items. Based on the number of federal government documents at NRLF and SRLF, we estimate the sample size of ~600 items. As pilot sampling project is planned at NRLF so we can gauge the resources needed to implement such a project. The goal is to see how accurately the RLF record statuses reflect the actual status of the items. We hope to have the first results to report in January 2015 and, assuming the staff time required is manageable, would finish the sampling at NRLF and SRLF early in 2015.