

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

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 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.

Established procedures exist today for selective federal depository libraries to withdraw unneeded items from their collections. However, the UC Libraries are in a unique position to offer ourselves as a model to GPO of 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 would seek to reach an agreement with the U.S. Government Printing Office (GPO) for blanket approval for the sheet-fed scanning of federal documents and for the withdrawal of unneeded duplicates of titles included in the archive. We would also plan to coordinate our activities with the California State Library, our regional depository.

This project will also allow 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.

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 SuDocs 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.