

The University of California's Federal Documents Archive: 5 Years and 200,000+ Documents Later



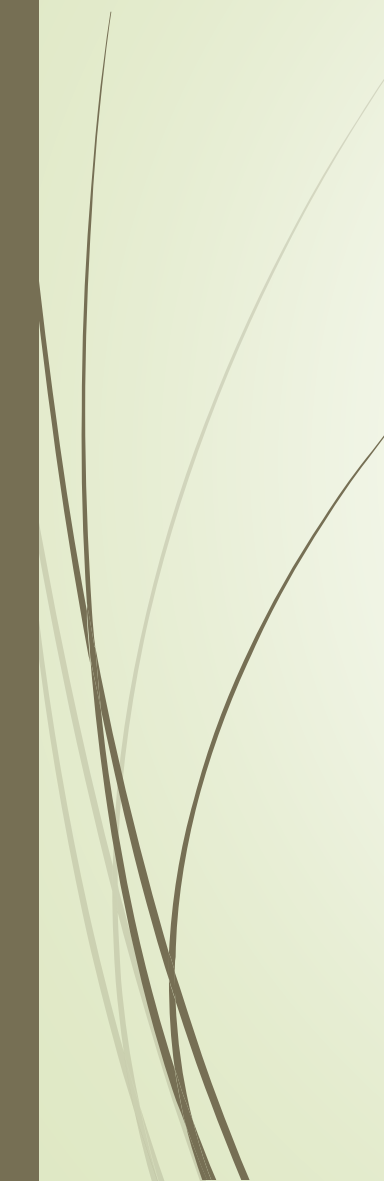
Jesse Silva, UC Berkeley

Heather Christenson, HathiTrust

(with help from Vince Novoa (UCR) and Kelly Smith (UCSD))



Agenda

- What is UC's Federal Documents Archive?
 - UC's Process
 - Issues Encountered
 - Campus Experiences
 - Accomplishments
 - Plans for the Future
 - HathiTrust's Role
- 

What is the UC Federal Documents Archive?



- Federal Documents Archive (FedDocArc)
- UC is carefully consolidating our historical documents collection and will share 1 print copy of every document across the 10 campuses
- Digital access and full-text searching is accessible via HathiTrust
- Each campus decides what it may retain and what it may weed. *No campus is being forced to weed!*
- FedDocArc is one of several shared collection initiatives UC is working on: JACS, WEST, JSTOR

UC Libraries Shared Infrastructure



- Shared catalog – Melvyl
- Shared off-campus storage facilities (NRLF and SRLF)
- Shared efficient ILL program
- CDL coordinates several internal programs, projects, and shared purchases
- Important Shared Policies
 - RLF Persistence Policy
 - Shared Print in Place Policy
- UC has been digitizing our collection through Google for inclusion in Hathi.
- Strong relationship with our regional



Brief History of FedDocArc

- Preliminary discussions began in 2011
- Planning team met throughout 2013
- Implementation team began in 2015
 - Representatives from UCB, CDL, UCLA (SRLF), UCR, and UCSD
 - Implementation Team membership changes
- Worked closely with our regional (a big thank you to the California State Library!) and GPO

Additional Infrastructure to Support FedDocArc

The Federal Depository Library Program



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.

- Selective housing agreement signed by all 10 campuses
- Flexibility from the California State Library
- Commitment from the University Librarians across the UC system

FedDocArc Process pt. 1

Duplication Between RLF's

- Started at the RLF's (2015-2018)
- Comparing holdings by OCLC Number (even comparing multiple OCLC numbers)
- Decisions made on duplicates -- MANY hours with spreadsheets!
- Serials decisions on duplicates required enlisting documents librarians from around UC to help out


| Decisions | Notes | OCLC Number | RECORD .(BIBLIO) | 245 | 086 | 110 | 710 | 260 c | Primary_ Barcode | Secondary_ Barcode | RECORD .(ITEM) | LOCATION | VOLUME | NRLF | SRLF | Other UC | Hathi Trust |
|-----------|-------|-------------|------------------|----------------------------|-----|-----|------------------------------|------------|------------------|--------------------|----------------|----------|-------------|------|------|-------------|-------------|
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | 5C 227 356 | | | i1239340 | ncscu | v. 3 1967 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | 5C 227 355 | | | i1239340 | ncscu | v. 2 1966 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | 5C 227 355 | TE0018148 | | i1239339 | ncscu | v. 1 (1965) | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | C 2 790 7 | C0069423 | | i1027696 | nbmau | 1971 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | C 2 790 7 | C0069423 | | i1027696 | nbmau | 1969 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | C 2 790 7 | C0069422 | | i1027696 | nbmau | 1967 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | C 2 790 7 | C0069422 | | i1027696 | nbmau | 1966 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | C 2 790 7 | C0069422 | | i1027696 | nbmau | 1965 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | C 2 790 7 | C0069423 | | i1027696 | nbmau | 1970 | nrlf | srif | CUY,CUI,CRU | |
| | | 1566050 | b119975282 | Defense industry bulletin. | | | United States. Defense Suppl | C 2 790 7 | C0069423 | | i1027696 | nbmau | 1968 | nrlf | srif | CUY,CUI,CRU | |



FedDocArc Process pt. 2

Campus Participation

- ▶ Each participating campus has a local team of people involved in FedDocArc—minimum of 2 years of work
- ▶ Three lists generated over approx. 12-18 months
 - ▶ In FedDocArc but not in HT
 - ▶ Not in FedDocArc but in HT
 - ▶ Neither FedDocArc nor HT
- ▶ Monographs are first, then serials and serial like items (multi-volume monographs, serials, etc.)
- ▶ Same decision tree for duplicate and non-duplicate titles.
- ▶ UCR was the first campus and has piloted a number of decisions. UCSD is currently looking at monographs. UCLA is next.
- ▶ Updated lists of monographs committed to FedDocArc and in HathiTrust shared with UC librarians to make duplicate decisions. First list was over 190K titles.



FedDocArc Process pt. 3

Decision Tree on Duplicates

- ▶ Titles not in the archive are deemed part of the FedDocArc print archive if the owning library chooses to do so.
 - ▶ Print titles in the archive can be held in an RLF or on-campus
 - ▶ Titles could be retained by the campus and not be included in FedDocArc.
- ▶ Duplicate titles checked for inclusion in Hathi.
 - ▶ If not in Hathi, sent for destructive scanning or non-destructive scanning if the campus wants to retain.
 - ▶ If in Hathi, title could be offered or retained by the campus
- ▶ Titles with no duplicates will be scanned non-destructively for inclusion in Hathi

Issues Encountered: Cataloging and Enumeration

| 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 |



Issues Encountered: Serials and Serial Like Items

- ▶ Hathi has a large number of serial and serial-like titles scanned
- ▶ Few are complete runs.
- ▶ Labor intensive-checking individual titles for completeness and contributing missing issues, volumes, parts, etc to Google for digitization and inclusion in Hathi.



Issues Encountered: Staffing

- ▶ Much of the work involved cannot be done programmatically—need actual humans to do the work
- ▶ This is not a specially-funded project. FedDocArc is in addition to the other work everyone involved is doing. It's a slow project!
- ▶ One temporary half-time programmer was hired by UCB to develop a database to track of all FedDocArc decisions.



Issues Encountered: Outside UC Response

- UC is moving to a model that not every depository, nor every library, needs to retain everything. We want to share collections.
- This has been difficult for some outside UC.
- Emails received:
 - “UC is going to destroy the FDLP”
 - “Berkeley’s great docs collection is being dismantled”
- Change can be difficult. Fear should be identified and discussed.

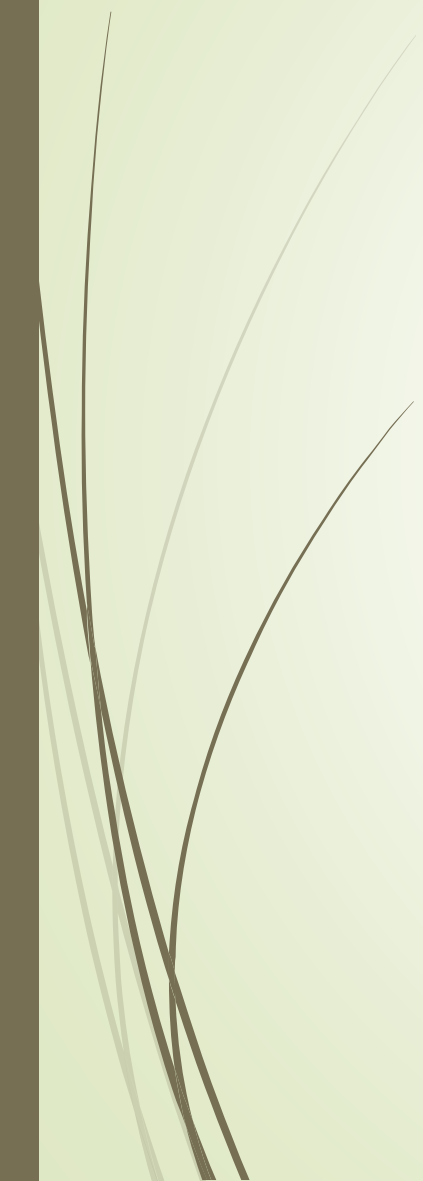


UCR's Experience

- FDLP library since 1963
- Wanted a smaller footprint for their docs collection
- Have become a large contributor to Hathi
- Documents with no print circulation now have usage via Hathi
- UCR is retaining print docs that are important for teaching and research and/or the local community.
- Cataloging for items didn't have any cataloging before.



UCSD's Story

- Statement from Kelly Smith, UCSD Federal Documents Librarian.
- 



Project Highlights

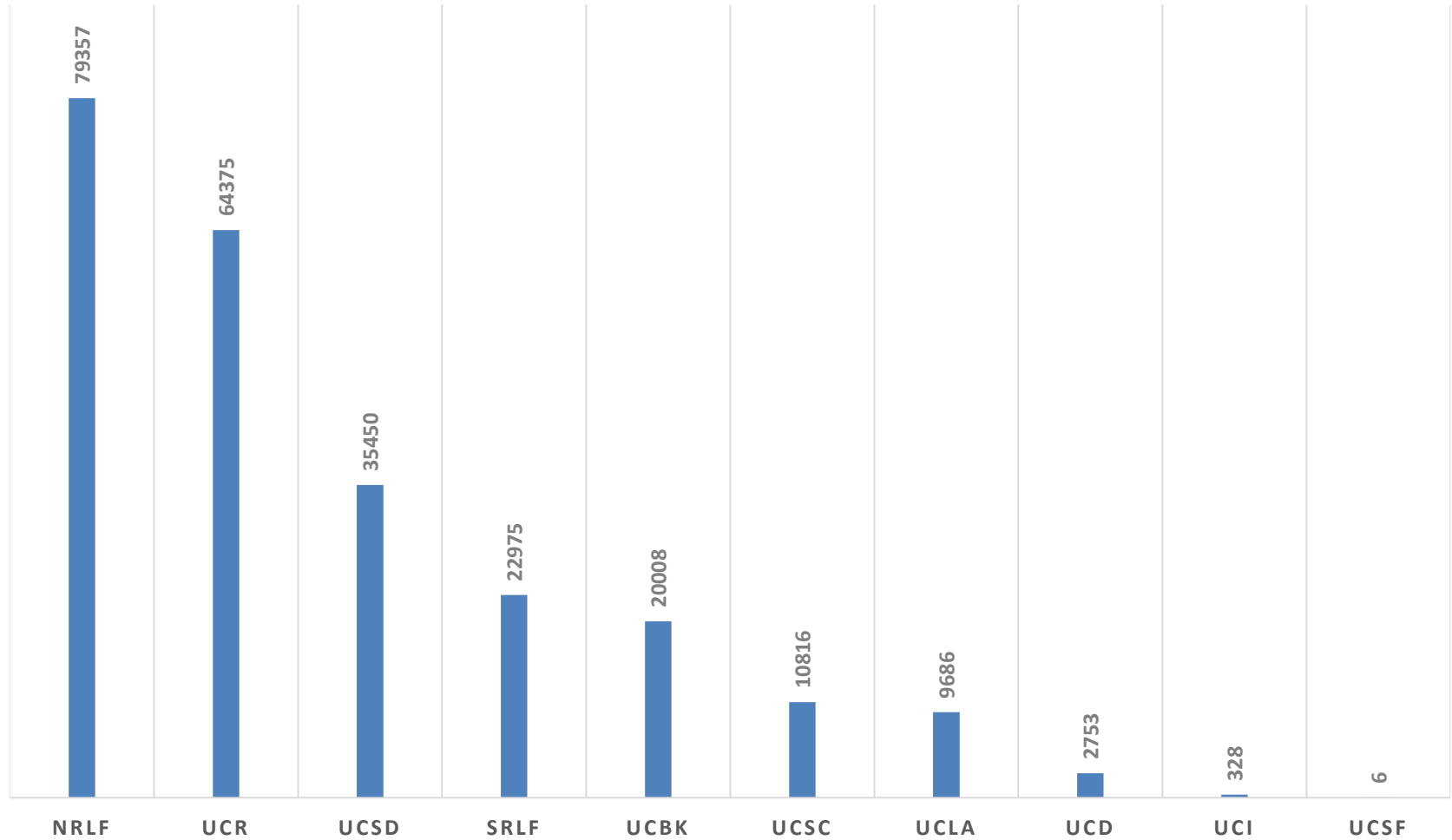
- Increased public access to FDLP documents
- Access like never before via HathiTrust -- Titles with no print circulation are being used in Hathi
- Increased and better cataloging of FDLP documents
- Over 200K titles have been committed to FedDocArc, 190K+ available in both print and HathiTrust
- Rescanning foldouts (Soil Surveys, USGS Professional Papers, 1970 Census Tract and Block Group Maps, etc)
- Created a standard for shared monographs (no standard existed before)
- UCB's ERIC microfiche is committed to FedDocArc.



Future Plans & UC Questions

- Continue doing holdings comparisons.
- Continue sharing lists of what has been committed
- Continue advocating for HT to allow full pdf download of federal documents by the public
- How can we be more efficient?
- What to do about e-only publications?
- What to do about the other legacy formats (microfiche, CDs, etc)?
- How to divide up collecting U.S. docs among 10 campuses, is this something we even want to do?

FEDERAL DOCUMENTS IN HATHITRUST BY UC CAMPUS/RLF



HATHITRUST

- 130 member institutions, providing access to 185 campuses
- Large scale digital collection with services based on shared infrastructure
- Supported entirely by members. Members contribute to governance and programs
- Preservation and access
- U.S. Federal Documents Program

Not possible without FDLF & Depository Libraries!



HATHITRUST COLLECTION

By the numbers

16,751,398 total volumes

8,129,135 book titles

449,206 serial titles

6,288,161 vols (37%) open for reading (public domain & CC-licensed)

5.86 billion pages, **~2.5** trillion words indexed / tokens computable

1,286,418 U.S. federal documents (volumes)

489,995 monographic titles

31,436 serial titles



HATHITRUST U.S. FEDERAL DOCUMENTS PROGRAM

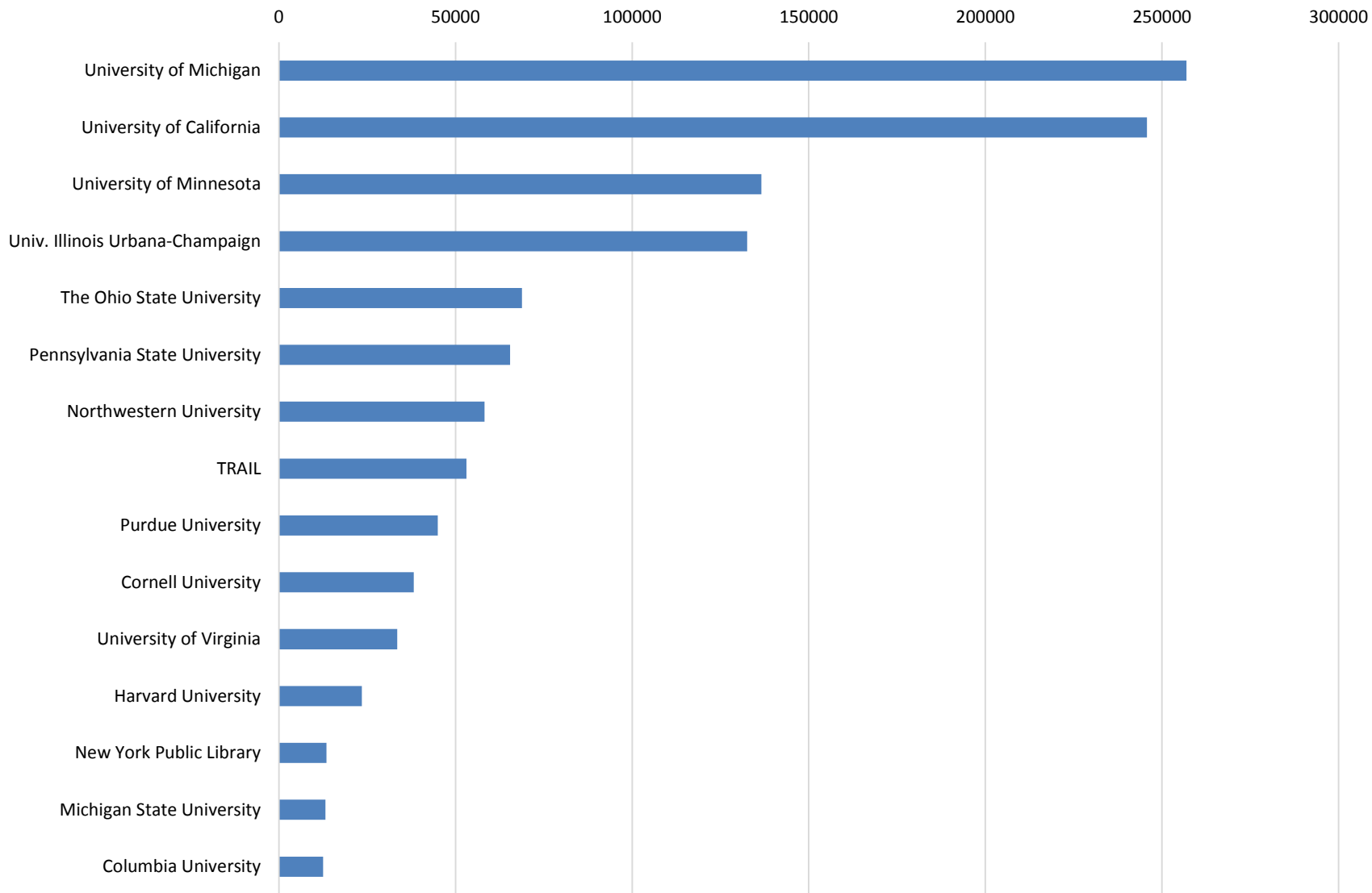
Program Activities

- Building a comprehensive collection
 - [U.S. Federal Documents Registry](#)
 - Identification of our [federal documents collection](#)
 - Intentional collection development, including “gap filling”
- Improving discovery of and access to federal documents
 - Building sub-collections
 - Internal work within metadata management
 - Investigating user needs
 - Community Engagement

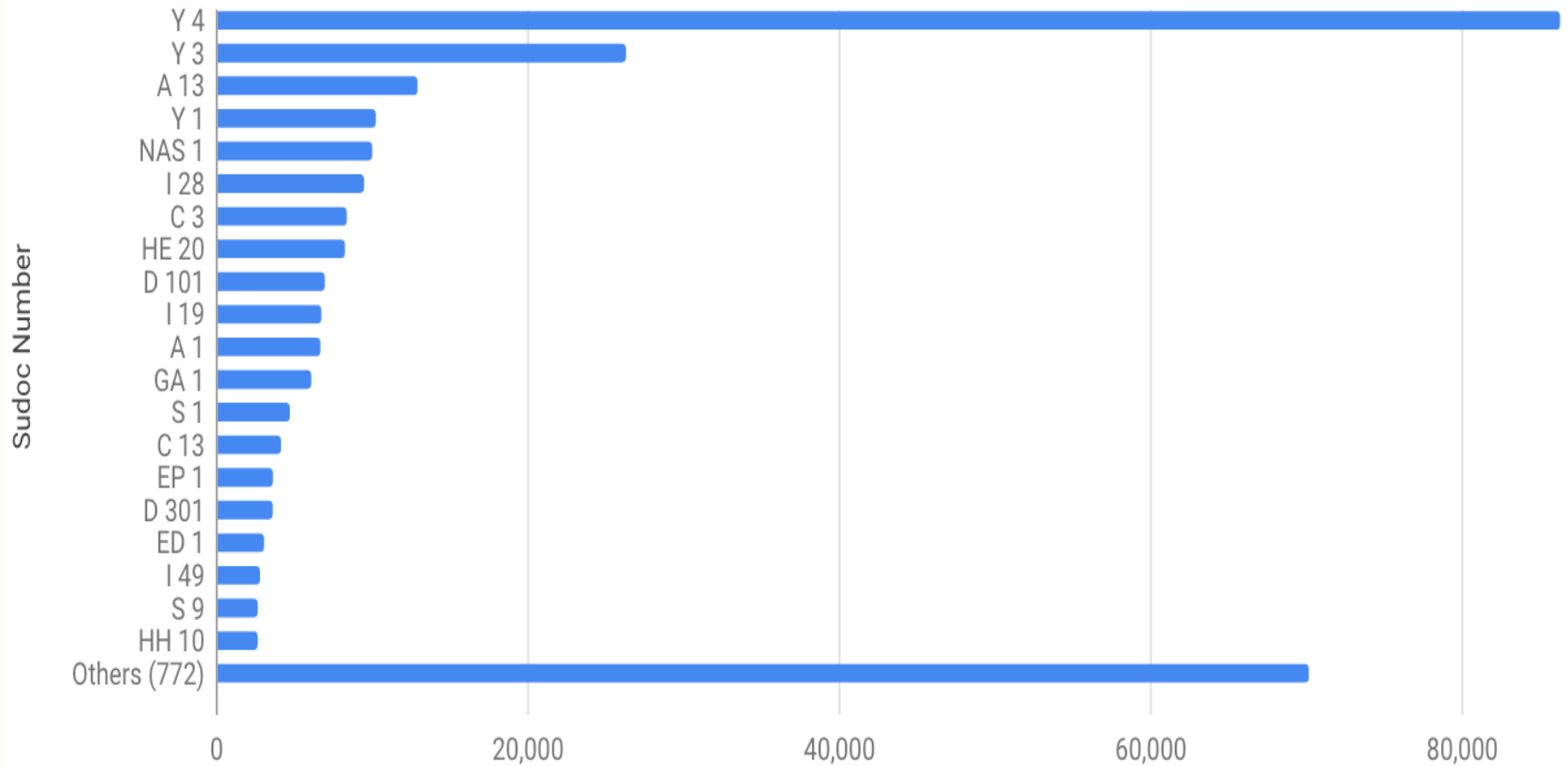
<https://www.hathitrust.org/usgovdocs>



TOP 15 CONTRIBUTORS OF FEDERAL DOCUMENTS IN HATHITRUST



SUDOC NUMBERS IN HATHITRUST



HathiTrust Collection Profile

https://www.hathitrust.org/usdocs_registry/collection_profile



U.S. FEDERAL DOCUMENTS COLLECTIONS FRAMEWORK

Acquire comprehensive runs of essential titles:

- Congressional Serial Set
- Congressional Record [bound]
- Statistical Abstract of the United States
- Economic Report of the President
- Foreign Relations of the United States
- United States Reports
- Large statistical yearbooks, such as Agricultural Statistics, etc.
- Agency annual reports
- All hearings

Acquire key agencies:

- EPA
- NOAA
- NASA
- Central Intelligence Agency
- Bureau of Indian Affairs

Acquire agencies within:

- Department of Transportation
- Department of the Interior
- Department of Agriculture

<https://www.hathitrust.org/hathitrust-federal-documents-collection-framework>



HATHITRUST U.S. FEDERAL DOCUMENTS COLLECTIONS

- [U.S. Federal Documents](#)
- [U.S. Congressional Serial Set](#)
- [Bureau of Indian Affairs publications](#)
- [U.S. Environmental Protection Agency publications](#)
- [Foreign Relations of the United States](#)
- [Statistical Abstract of the United States](#)

<https://www.hathitrust.org/u-s-federal-documents-collections>



HATHITRUST RESEARCH CENTER: U.S. FED DOCS WORKSETS

- [EPA Publications](#)
- [Bureau of Indian Affairs publications](#)
- [Foreign Relations of the United States](#)
- [US Congressional Serial Set](#)

More information

- 2018 FDL Conference presentation: *Digging Deeper into Government Information: An Introduction to Text Mining*, given by Eleanor Dickson Koehl of the HathiTrust Research Center, and Valerie Glenn, previously of HathiTrust
- <https://analytics.hathitrust.org/>



HATHITRUST'S ROLE

Library Stewardship

- Digital copies are preserved in shared infrastructure supported by all the HathiTrust members
- Shared Print Program

Federal Documents Are Discovered and Used!

- Bibliographic and full text search
- Viewing access
- Permanent URLs for volumes and pages enable embedding in discovery systems, LibGuides, syllabi, and social media
- Computational use is available via the HathiTrust Research Center
- User Support is provided by HathiTrust staff and team of 30+ from across HathiTrust member libraries



FUTURE & SHARED CHALLENGES

Improved HathiTrust infrastructure and processes to support collaborative work

- Coordinated digitization & “gap-filling”
- Serials: the final frontier
- Metadata improvement & enrichment
- Improve user-facing collection-building tools

Community roles

- Born digital?
- Coordination with GPO & other organizations



THANK YOU!

Heather Christenson

Program Officer for Federal
Documents & Collections

HathiTrust

christeh@hathitrust.org

More information:

<https://www.hathitrust.org/usgovdocs>

Jesse Silva

Librarian for Federal & State
Government Information,
Political Science, Public Policy
and Legal Studies

UC Berkeley

jsilva@library.berkeley.edu

