

Saving democracy: The LOCKSS USDocs Network and the digital FDLP



James R. Jacobs
US Government
Information Librarian
jrjacobs@stanford.edu



Snowden Becker
LOCKSS Community
Manager
snowdenb@stanford.edu



Mary-Ellen Petrich
Digital Preservation
Analyst
mpetrich@stanford.edu

The LOCKSS Program



Stanford | LIBRARIES

Committed to initiating and developing open-source technologies and services such as **Samvera**, the **International Image Operability Framework (IIIF)**, **Spotlight**, and **Mirador**



LOCKSS Program

A key initiative of Stanford Libraries' Digital Library Systems and Services, providing collaborative open-source technologies and services for **high-confidence, resilient, secure digital preservation**

What is LOCKSS?



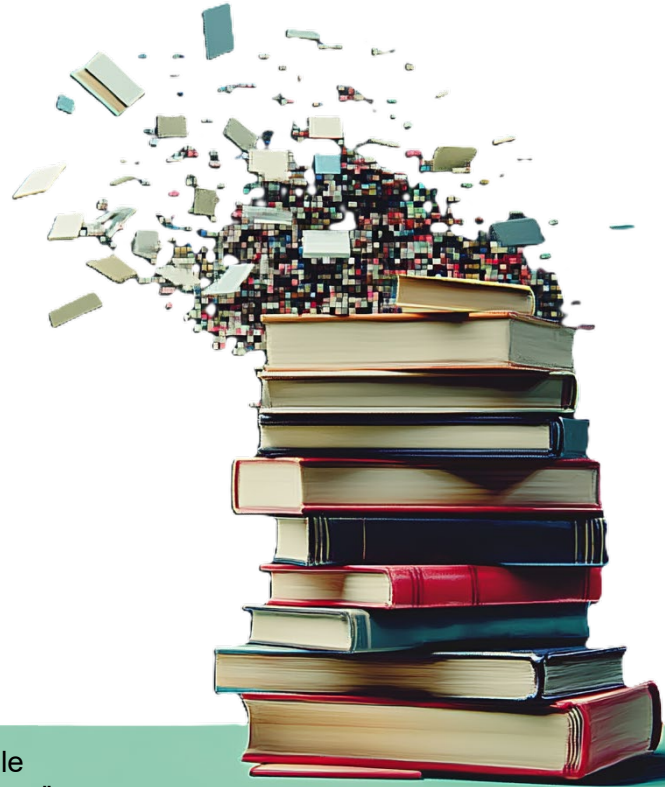
LOCKSS began as...

“Libraries are in an odd place now. They have never before owned so little of the content they manage.”

Paper publishing → e-journals and books

Local ownership → remote access

Authoritative copies → ephemeral formats



Schneider, Karen G. “Lots of Librarians Can Keep Stuff Safe: Libraries Are Able to Safeguard Content with LOCKSS, Open Source Digital Preservation Software.” *Library Journal* 132, no. 13 (2007): 30–31.

LOCKSS is now...



CLOCKSS



MDPN
MICHIGAN DIGITAL
PRESERVATION
NETWORK



THE ALABAMA DIGITAL PRESERVATION NETWORK
PRESERVING ALABAMA'S DIGITAL RESOURCES

SAFE **PLN**
SAFE Archive FEderation



What's with the turtle?

The LOCKSS Program logo reflects our core principles of community, decentralization, and durability. You may see it as:

A tortoise — In folklore tortoises represent deliberateness, determination, and longevity

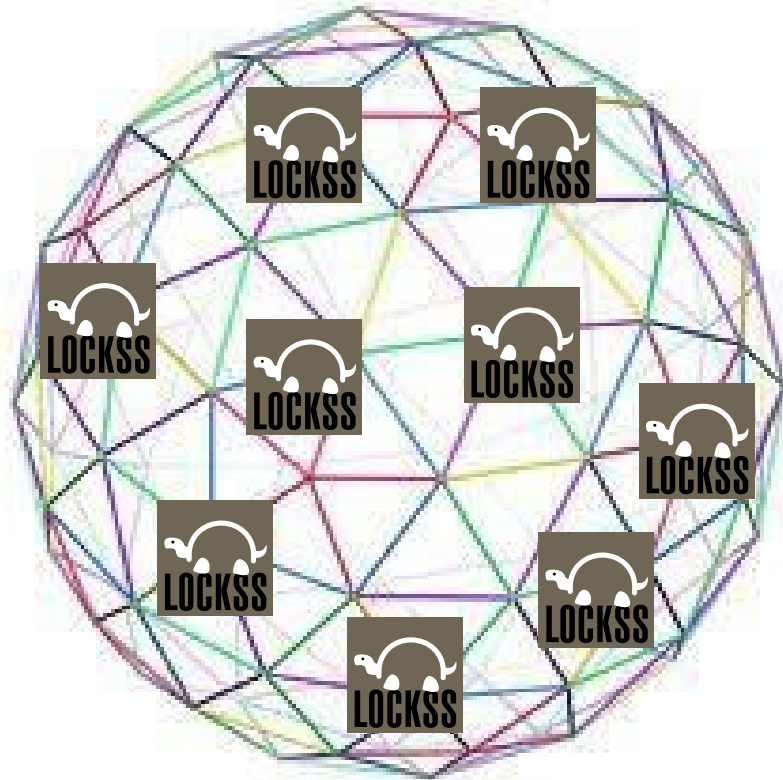
Peers seated around a table — Digital preservation is most sustainably situated in a community with clear lines of communication and a shared commitment to saving content

Nodes in a network — LOCKSS is digital preservation in which multiple systems are all working together



LOCKSS Principles

- Libraries have local control of assets
- Perpetual access is guaranteed
- Preserve original authoritative version
- Decentralized, distributed preservation
- Affordable and sustainable



Principles of the FDLP

- Distribution / "deposit" of government information
- Geographic participation
- No-fee permanent public access
- Tamper-evident, redundant system
- Collaborative preservation
- Forward democratic ideals
- Serve public interest / public access / public control / public preservation



The USDocs Private LOCKSS Network (PLN)

USDocs (lockss-usdocs.stanford.edu) is a vital initiative within the LOCKSS program, focused on the long-term preservation and accessibility of U.S. government publications in digital format from **govinfo.gov**.

- Requires **community cooperation** among libraries and GPO to create a decentralized preservation network
- Aims to make preserved documents readily accessible to researchers, academics, and the public, supporting **transparency and informed citizenship**
- Employs peer-to-peer networks to store multiple copies of publications, enhancing longevity and resilience against **data manipulation and loss**
- Replicates FDLP in the digital environment
- “digital deposit” (for more on “digital deposit,” see <https://freegovinfo.info/node/13361/>)

History of the USDocs PLN

- **Established Network (16 Years and Counting)**
 - Initiated to create a robust preservation solution independent of federal control, ensuring the integrity of digital government publications
- **Widespread Participation**
 - At various times, 36 libraries have actively participated in the network, contributing to its collective strength and diversity
- **Tamper-Evident Preservation**
 - Designed as a decentralized network, LOCKSS USDOCS provides a tamper-evident solution that safeguards the authenticity and reliability of preserved documents
- **Commitment to Community Cooperation**
 - Continuously fosters collaboration among libraries and government agencies, reinforcing the need for a shared commitment to preserve crucial government information

LOCKSS-USDOCS includes ...

The logo for GovInfo, featuring the text "GovInfo" in a white, sans-serif font on a dark gray rectangular background.

- Federal Register
- Code of Federal Regulations
- Congressional Record
- Congressional Bills
- Congressional Reports
- US Code
- Public & Private Laws
- Public Papers of the President
- Historic Supreme Court Decisions
- US Statutes at Large
- GAO Reports
- US Budget
- ... and more!!

<https://www.govinfo.gov/browse-a-z>

A day in the life of a LOCKSS web crawler

- Discover
 - Find content through site maps on GovInfo.Gov
- Document
 - Create simplified catalog records
- Test
 - Test crawls of content to ensure software compatibility
- Evaluate
 - Configure tools for best preservation
- Release to the network
 - Most recently 210 items 9/25



CC0 1.0 Universal by Bernard Spragg
<<https://www.flickr.com/photos/volvob12b/9606387860>>

A day in the life of a LOCKSS web crawler

- Crawl
 - Documents are collected from GovInfo.Gov
- Poll
 - A quorum of machines compares successful crawls
- Hash
 - Some sections are ignored for comparison purposes
- Repair
 - Disagreements are resolved with a repair from the original source if available

A day in the life of a LOCKSS web crawler

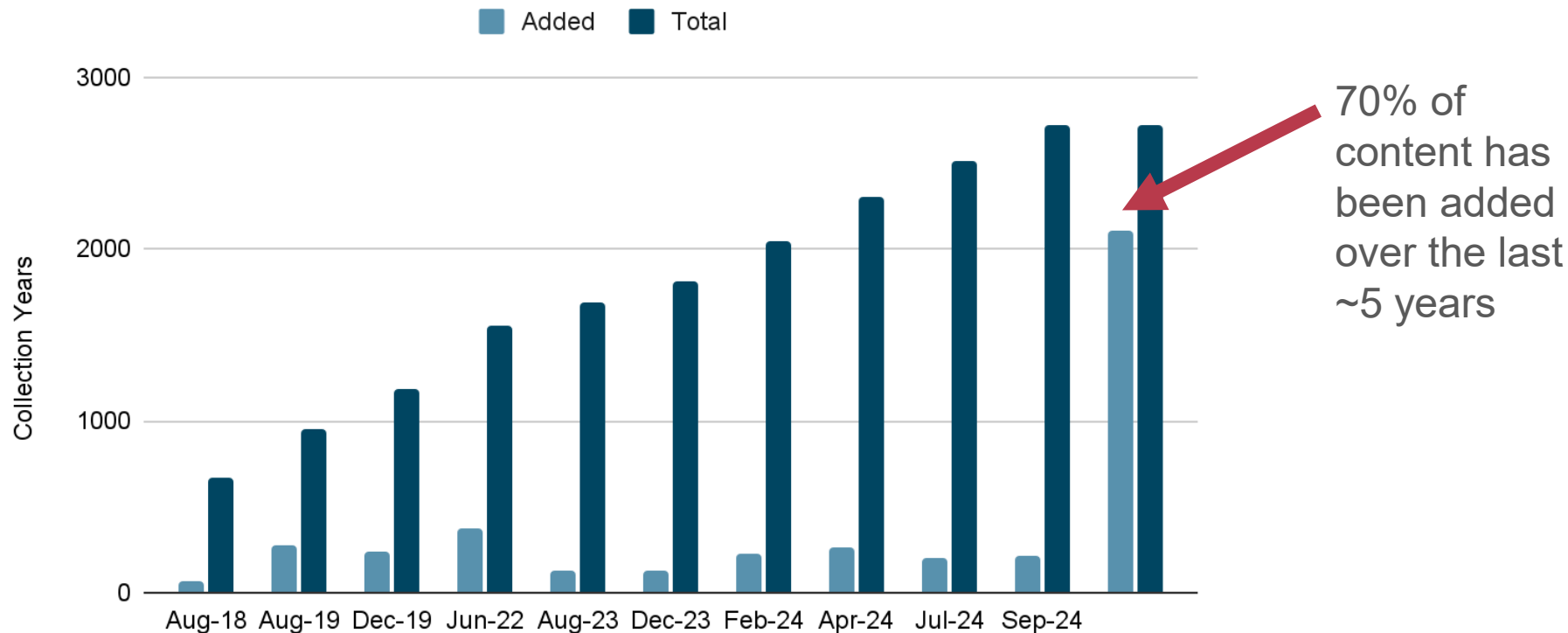
- Collections
 - From BUDGET to USCOURTS
 - Including Serial Set, Congressionally Mandated Reports, and 2024 Volumes
 - Content is divided into Collection-Years (CY)
- Open: 2046 CY
 - ~50% are USCOURTS
- Closed: 776 CY
 - Not currently collecting
 - Pre 2019
- In Progress: 1031 CY
 - 90% USCOURTS



CC BY-NC-SA 2.0 by Rami™ <<https://www.flickr.com/photos/rhk313/3800471907>>

Hugely increased harvesting from govinfo.gov

Growth of USDocs content



A day in the life of a LOCKSS web crawler

- GPO Support and Commitment
 - Quarterly meetings, Rapid response to issues, information sharing
- Updates: last 12 months
 - Increased crawl speed by 8x
 - Increased releases by 4x
 - Increased size by 2x (now 13.6 TB)
- Updates: next 12 months
 - Continue new content releases
 - Re-release closed volumes
 - Improved Reporting
 - Improved Access

Past praise for USDocs:

“Although this quiet network is likely not on public services librarians’ radar, LOCKSS-USDOCS is proving to be one of the most common sense ways to ensure the integrity of digital federal government documents, with its simple, open-source platform, low maintenance for each institution, and use of existing content and systems.”

Flynn, Kian A., and Cassandra J. Hartnett. “Cutting through the Fog: Government Information, Librarians, and the Forty-Fifth Presidency.” *Reference & User Services Quarterly* 57, no. 3 (March 16, 2018): 208–16.
<https://doi.org/10.5860/rusq.57.3.6608>.

How you can support USDocs

- Already a LOCKSS user? USDocs is **free** for you to join!
- Not using LOCKSS yet? USDocs is affordably priced!
 - \$1250/year standalone
 - “Bring a friend” discount
 - GLN pricing based on annual subscription spending
- Hardware: BYOS (Bring Your Own Storage)
 - 12TB recommended capacity
 - <https://docs.lockss.org/projects/manual/en/latest/introduction/>
- Staff time: minimal following setup
- Already in the USDocs network? Help us spread the word and meet our goals!
 - One(+) node in every state (19/50)
 - Better coverage among the Regional depositories (only 10/46)
 - Panels and info sessions for government librarians - CoSA, NAGARA, etc.
 - Share your success stories, lessons learned, suggestions, wish lists?

Questions?



James R. Jacobs

US Government
Information Librarian

jrjacobs@stanford.edu



Snowden Becker

LOCKSS Community
Manager

snowdenb@stanford.edu



Mary-Ellen Petrich

Digital Preservation
Analyst

mpetrich@stanford.edu

Thank you!

Please stay in touch and let us know if you'd like further
information, news, and updates on USDocs.

lockss-usdocs.stanford.edu