

The LOCKSS Program
Digital Federal Depository Library Program
Frequently Asked Questions

Q: Why does the LOCKSS-USDOCS program call itself the “digital FDLP?”

A: Our use of this term is purely descriptive and not meant to imply that LOCKSS-USDOCS is part of the Federal Depository Library Program, see our disclaimer, below. There are three characteristics of the program that taken together, add up to a truly “digital FDLP”:

- LOCKSS-USDOCS is the *only* library program that is systematically dealing with the preservation of born-digital government documents and doing so in a collaborative, distributed way similar to the way the FDLP is currently organized. For more on the FDLP, see <http://fdlp.gov>.
- LOCKSS-USDOCS is a collaborative network that is "tamper evident" -- a critical piece of the paper FDLP transferred to the digital realm.
- LOCKSS-USDOCS is a program that empowers libraries to work collaboratively on behalf of their constituent citizens.
- Disclaimer: The term “digital FDLP” is purely for descriptive purposes to contextualize LOCKSS-USDOCS within the historic work of the Federal Depository Library Program (FDLP). LOCKSS-USDOCS is a grassroots effort in which the Government Printing Office (GPO) is fully cooperating. LOCKSS-USDOCS is not an official program of the GPO, as no legal paperwork or memorandum of understanding (MOU) has been signed.

Q: How big is LOCKSS-USDOCS?

A: LOCKSS-USDOCS has about three dozen partners in the US and Canada and the program is growing! The list of partners is at <http://snipurl.com/lockss-usdocs-partners>. Please contact James Jacobs (jrjacobs AT stanford DOT edu) if your institution is interested in participating.

Q: Is LOCKSS USDOCS sanctioned by the Government Printing Office (GPO)?

A: LOCKSS-USDOCS, with the express written consent of and in full partnership with GPO, is currently harvesting and preserving ALL collections hosted on FDsys (<http://www.gpo.gov/fdsys/browse/collectiontab.action>) as well as their bulk data repository. LOCKSS-USDOCS is much more than a simple mirror site for FDsys -- something that the depository community has been asking GPO to provide for well over 5 years! Participating libraries have independent custody of and control over the content, thus assuring tamper evidence of depository materials in the same manner that paper materials are protected by the 1250 FDLP libraries. Future collections outside of FDsys (but still within scope of the FDLP) will be added based on collective decisions. As noted above, even though GPO is an active participant in the project, we have not as yet signed a memorandum of understanding (MOU) making LOCKSS-USDOCS an official GPO partner (for more on GPO partnerships, see <http://www.fdlp.gov/outreach/partnerships>).

Q: What sort of files does LOCKSS-USDOCS preserve and how much disk space does it take?

A: LOCKSS-USDOCS is preserving all available file types, including metadata from FDsys and XML from the bulk data repository. The current collection -- covering content from 1991 - present -- is approximately 1TB and growing. Because born digital files are smaller than digital files of scanned paper (typically .tiff files), the space requirements for future years may not be as

For more information - LOCKSS-USDOCS web site (lockss-usdocs.stanford.edu) and contact James Jacobs (jrjacobs AT stanford DOT edu).

The LOCKSS Program
Digital Federal Depository Library Program
Frequently Asked Questions

high as the requirement for past years of digitized files. This and the trend of ever decreasing hardware and storage costs make the storage requirements for a serious collection affordable.

Q: Just 1 TB? I thought GPO had way more stuff than that! Are you sure?

A: Technically, there is currently about 18TB of content in FDSys; but this includes, "a lot of duplicates (a copy of content in the AIP, in the ACP, and in the ACP cache) as well as backups". Much of this content is not publicly available and therefore is not needed by the Digital FDLP, any more than the paper FDLP needs to be concerned about the size of GPO's warehouses and printing plants.

The content that is on the "public side" is that which this project is concerned. FDSys' public content currently is between 1.1 to 1.4TB. FDSys contains four versions of each document, plus metadata. Participating libraries have collectively decided that it is sufficient to collect two of these four versions plus metadata to assure flexibility and redundancy.

Q: How much will the FDSys collections grow in the future?

A: GPO staff state, "growth estimates for current FDSys content collections are about 380 GB per year, but this doesn't include growth from the addition of new collections." LOCKSS-USDOCS members will work on a consensus model for the addition of new collections. If you'd like to suggest other collections within scope of the FDLP but outside FDSys, please contact James Jacobs (jrjacobs AT stanford DOT edu).

Q: Do members get to select what content they want to preserve?

A: The FDLP works via a process whereby FDLP libraries may select items of importance to their local communities (for more see <http://www.fdlp.gov/home/about/462-selection-mechanisms>). This idea has been discussed for LOCKSS-USDOCS. However, at this time, participants must select all content in order to assure redundancy. As the number of participants grows, the idea of "item selection" may be re-opened for discussion.

Q: What sort of hardware and storage will I need and how much should my institution expect to pay?

A: We recommend that participants have the equivalent of a system with 4 2TB disks for 8TB. Future hardware configurations will allow us to expand storage to meet the needs of the program. Some Universities, (e.g., University of Michigan) are running their system in a virtual server environment. The cost of the hardware varies tremendously with approach and configuration, but on average partners are re-using and upgrading older machines. New machines built to LOCKSS Program specifications are approximately \$1,500.

Q: Are there other costs that my institution should know about?

A: LOCKSS alliance libraries participate in the program at no additional cost. Several participants are non-LOCKSS alliance libraries, and they are asked to pay a token support fee of \$1250/yr or \$750/yr if they bring an additional library into the program. Hardware/software upkeep is very minimal (less than 1hr/month for 1 IT staff). Staff time for documents librarians is less than 15hrs/year, which includes periodic conference calls and selection of new archival

For more information - LOCKSS-USDOCS web site (lockss-usdocs.stanford.edu) and contact James Jacobs (jrjacobs AT stanford DOT edu).

The LOCKSS Program
Digital Federal Depository Library Program
Frequently Asked Questions

units as they are harvested. In short, participation in terms of hardware, software and staffing is extremely cost-effective, flexible, efficient, and highly scalable.s