

Preservation for all: LOCKSS-USDOCS and our digital future

James Jacobs and Victoria Reich, Stanford University Libraries
Documents to the People (DttP) Volume 38:3 (Fall 2010)

Leading libraries, in partnership with the United States Government Printing Office (GPO) are transforming the almost 200-year-old geographically distributed Federal Depository Library Program (FDLP) from the print to web environment. Working with the Stanford University Libraries *Lots of Copies Keep Stuff Safe* (LOCKSS) team (www.lockss.org), libraries will protect government documents published via GPO's Federal Digital System (FDsys, www.fdsys.gov) in a distributed digital preservation system called **LOCKSS-USDOCS**.¹ These actions will ensure that current and future citizens will have access to authentic and authoritative documents including the *US Budget*, *Federal Register*, *Code of Federal Regulations* (CFR), *US Code*, *Congressional Record*, Congressional reports, hearings and documents, Congressional Bills, *Public Laws*, *Papers of U.S. Presidents*, *GAO Reports* etc.

This work builds upon the community of libraries who are already using the LOCKSS system to preserve GPO documents harvested from GPOaccess (1991 – 2007).² **LOCKSS-USDOCS** maintains libraries' vital role as players in the digital information infrastructure. The preservation of federal documents is too important to be left to the federal government alone. **Please join us!**

At a time when some depository libraries are considering dropping out of the FDLP, **LOCKSS-USDOCS** provides an alternative. While many libraries do not have sufficient space to house their print collections and are moving to free up space by moving their print collections off-site or deaccessioning them altogether, **LOCKSS-USDOCS** offers them a secure way to free up space while ensuring access to essential government information resources. Rather than abandoning their vital societal role of preserving government information for future generations, libraries will be able to be full participants in the digital government information future.

Libraries have been extremely effective in providing access and long-term preservation to print materials in spite of the fact that the management and preservation of print collections are difficult and time-consuming tasks. Distributed print collections protected government information from inadvertent loss and from attempts to change or censor the historical record. In the print world, libraries have many copies of most things. Lots of physical copies are a hedge against physical calamities and attacks. With copies scattered around the world, held under different legal, administrative, and financial regimes, it was and still is, practically speaking, impossible to destroy or alter all copies and it is very hard to destroy or alter all copies. It is not enough to have copies in multiple locations; it is also necessary for those different copies to be independently managed. Paper collections, in this sense, are what we call TAMPER EVIDENT.

Authenticity, a critical feature to have in any trusted government information infrastructure, is enhanced and strengthened with a distributed collection. In the current, non-distributed collection environment, digital government information has been altered without notice.³ While there are no documented instances of this happening to GPO content, the potential is there as long as GPO's servers continue to be the exclusive source for government information. Multiple copies on geographically disparate servers allow possible alterations to be detected and corrected, thus protecting against deliberate tampering. **LOCKSS-USDOCS** explicitly does this. Research suggests

¹ For a full list of LOCKSS-USDOCS libraries, see http://lockss.org/lockss/Government_Documents_PLN.

² GPOaccess documents were harvested by Carl Malamud and hosted publicly at <http://bulk.resource.org/gpo.gov/>.

³ Althaus, S. & Letaru, K. (2008, November). *Airbrushing History, American Style*. http://www.clinecenter.uiuc.edu/airbrushing_history/. Accessed June 8, 2010.

that only a large-scale network attack lasting months could successfully change content stored in a LOCKSS network.⁴ A Web is much stronger than a silo.

Fortunately, digital collections don't take up much space and the cost of digital storage continues to decline. Foresighted librarians are continuing to build and preserve collections by collecting and preserving digital content. They are ensuring their libraries role in society and are building a tamper evident library system just as the FDLP has provided for almost 200 years.

The FDLP world of distributed physical collections is tamper evident. In order to withdraw a publication from depository collections, GPO must notify the holding libraries of the item to be withdrawn and order them to either return the publication to GPO or destroy it. Sometimes withdrawal is appropriate and libraries comply.⁵ But in some instances, publications are withdrawn needlessly or explicitly to protect the government's reputation. In these instances, depository librarians have been known to create a loud hue and cry that usually results in the withdrawal order being cancelled.⁶

In the paper and ink world, libraries have played a key role in democracy, particularly with regard to government information, by making government publications available and tamper resistant for the long-term. In the digital age, libraries will be able to play the vital role of protecting the digital public record by building a tamper evident preservation network using the LOCKSS system. NO centralized preservation model – even if that model has physically distributed mirror or backup copies – provides tamper evident protection. In a centralized archive, no public process need be followed to alter or remove information. A simple delete command is all it takes.

In addition to tamper evidence, there are myriad reasons why a distributed digital preservation system for government information is necessary. Among them are: protection from natural disasters, server outages, etc.; assurance of authenticity; prevention of surreptitious withdrawal or tampering of information; and building local services for local collections.

While the government itself plays a vital role of creating government information it cannot guarantee the preservation of federal documents by itself. This job is also too important to outsource to private services that rely on profit to select what to preserve. There are NO private services that provide distributed digital preservation and none that are tamper evident.

Here's an opportunity to work with your library colleagues, with the GPO and other government officials, and activists to preserve government documents in a distributed digital preservation network for current and future citizens.

Join Us! Contact James Jacobs (jrjacobs@stanford.edu) for information; there are no additional costs for LOCKSS Alliance members. Special pricing is available for libraries wishing to participate in LOCKSS-USDOCS project and host a US government LOCKSS box.

⁴ Giuli, P., et al. (2005) Attrition Defenses for a Peer-to-peer Digital Preservation System. Proceedings of the 2005 USENIX Annual Technical Conference, April, 2005.

<http://www.eecs.harvard.edu/~mema/publications/hotel.pdf>. Accessed June 8, 2010.

⁵ For a list of government documents withdrawn from the FDLP 1981 – 2009, see <http://snipurl.com/fdlp-recalls-1981-2009>. Accessed June 8, 2010.

⁶ Department of Justice Rescinds Order to Have Libraries Withdraw/Destroy Publications (2004, August) Posting on Resource Shelf. <http://www.resourceshelf.com/2004/08/02/dept-of-justice-rescinds-order-to-have-libraries-withdrawdestroy-publications/>. Accessed June 8, 2010.