

Preservation for all: the future of government documents and the “digital FDLP” puzzle

James R. Jacobs
jrjacobs@stanford.edu
lockss-usdocs.stanford.edu
Ohio GODORT Spring Meeting
Friday June 3, 2011 10 - 11:30

agenda

- library principles and best practices
- case studies:
 - Everyday Electronic Materials (EEMs) “*Water droplets*”
 - Archive-it “*Oceans*”
 - lockss-usdocs “*Waterfalls*”
 - Collaboration: delicious, state agency databases “*Reservoirs*”
- reflection of projects based on principles
- [[slides available at slideshare.net/freegovinfo]]

I'd like to thank Ohio GODORT President Andrea Peakovic, Tom Adamich, Audrey Hall, Sarajeon Petite and the other organizers for inviting me to talk with you today and facilitating a smooth and comfortable trip. I hope it'll be worth your while :-)

We're at the very beginning of the digital era where tools, policies, best practices, etc are all in flux. In many ways, we're at the age of new metaphors needed to describe what it is that we as librarians do on a daily basis.

Librarians ...

... Explore

... Collect

... Describe

... Share

... Preserve

mood survey

- how many of you have:
- run across a site or a PDF or a database that you wanted to add to your collection?
- sent in a fugitive document to the GPO?
- have created subject guides/pathfinders to govt information sites?
- have (or want to have) a digital repository?

how many of you have:
run across a site or a PDF that you wanted to add to your collection?
ever sent in a fugitive document to the GPO?
have created subject guides/pathfinders to govt information sites?
have (or want to have!) a digital repository?

I've answered each of these in the positive, and that's what has led me to work on and advocate for digital documents collections and the digital FDLP.

Stanford University Library has been a federal depository library since 1895. Given that collections are still a vital aspect of libraries and govt document collections are becoming hybrid paper/digital, the question becomes: how do government documents librarians deal with the shift to digital formats and continue to build robust collections that serve our local user communities? Printing out or downloading digital documents to our desktops doesn't even begin to answer that question. When everything's ephemeral, how do we collect, organize, give access to, and preserve government documents?

Today my talk will be centered around what I'm calling the "digital FDLP." I use this term not as an accepted project name but simply as a metaphor to contextualize the work that I'm doing within the historic work of the federal depository library program (FDLP). I must add a disclaimer at the outset that, although GPO is participating in the locks-usdocs program, the work that I'm doing is largely outside of GPO/FDLP structure and no paperwork or MOU has been signed and therefore has no official status.

principles

- Forward democratic ideals
- Serve public interest / public access / public control / public preservation
- Serve the information needs of the community
- Forward the long-term institutional viability of libraries
- Promote and leverage collective action

So with that disclaimer out of the way, I'd like to talk about a few case studies in digital government information. But I'd like to talk about more than simply the technical aspects of these projects. Because of the nature and history of libraries as memory organizations, we also must deal with the social aspects and impacts on our practices. The social aspects of libraries are our fundamental *raison d'être*.

So in thinking about the projects of which I'm about to show, I kept coming back to the fundamental principles of libraries. Because those principles are, at the end of the day, the criteria for judging whether or not a project, a workflow, institutional energy should be considered to be a success and whether your practices can be seen as "best."

So I'd like to first talk about library ideals. For me, running through this checklist helps me evaluate my work or a specific project. For instance, if I'm evaluating a project that seems to be valuable, but uses proprietary software or control of the content is not in the hands of the library (a trusted non-commercial entity!), or the goal of the project is profit over public interest, then this leads me to have questions about that project.

So as a reminder, I'd first like to enumerate some of the library principles or values that I use as a checklist as I go about my work. If you have others, please let me know:

Are you:

- forwarding democratic ideals?
- serving public interest / public access / public control / public preservation?
- serving the information needs of the community?
- forwarding the long-term institutional life of libraries?
- promoting and leveraging collective action?

These are the principles that we as documents librarians (and librarians in general!) hold dear. best practices = practices in which these principles are embedded.

Actions in support of values:

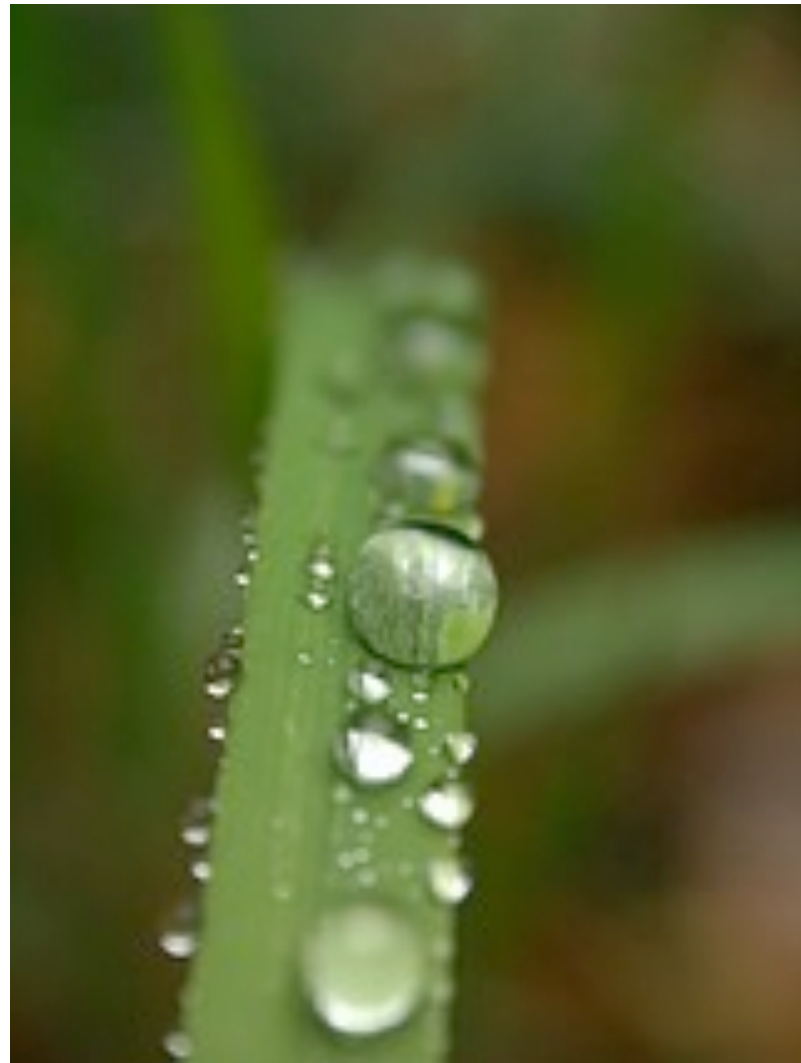
- libraries as memory organizations
- local control of collections (print libraries resist attack and are self-healing e.g. just yesterday I got an email from the American Samoa bar assn. Samoan national library had lost much of collections in the 2009 tsunami including the entire run of the Samoan Pacific Law Journal and found that Stanford had holdings of the journal. So we're trying to figure a way to help them rebuild their collections. U of HI Manoa is another case where collections were lost due to natural disaster and the FDLP community helped to rebuild their collections.)
- distributed system to meet local needs (spread responsibility for content among various locations and administrations)
- public interest (affirms FDLP libraries' role in ensuring permanent public access!)
- value of library community
- forward democratic ideals
- A community, sharing preservation responsibilities

While I talk about the following projects, please keep these principles and ideals in mind.

EEMs

- Everyday Electronic Materials
- serendipitous collection
- Collecting the Web a drop at a time

- Flickr photo by Elle Is Oneirataxic. Attribution-NonCommercial-ShareAlike 2.0 Generic Creative Commons license



EEMs – or Everyday Electronic Materials – is a Mellon funded project to build infrastructure and a workflow to support the collection, description, preservation and public access of digital objects by bibliographers and subject specialists.

EEMs are those digital materials that are distributed by posting on Web sites, or through email notification to scholars and bibliographers; those items that selectors come across in the course of doing their everyday work. The project has been a successful collaboration between Public Services, Technical Services, and Digital Library Systems and Services and has produced results that may be adopted and adapted for use by other libraries including:

My colleague Katherine Kott wrote a report on the project for the fall 2010 CNI meeting. The policies and procedures for collecting and processing EEMs that Katherine laid out are:

- A clear framework for managing copyright issues associated with digital material distributed via the Web, and for applying access policies that are consistent with redistribution rights
- Training events and material for selectors and technical services staff
- A Web-based tool to support selector and staff processing of EEMs via a lightweight workflow
- Integration with the current integrated library system (ILS) and traditional ILS-based processes
- Integration with other components of Stanford's digital library infrastructure, including its preservation repository, discovery systems and "digital stacks" delivery environment

**From Katherine Kott's CNI report on the project. See end slide for citation.

Subject specialist workflow:

1. identify the document (only pdf at this time)
2. drag url of doc to the EEMs browser widget
3. determine copyright status. Request permission to harvest/preserve if need be
4. describe the document (title, author, rights status, comments)
5. submit to acq and cataloging workflow.

Agencies tracked for EEMs

- Bureau of Land Management CA field office: <http://www.blm.gov/ca/st/en/info/publications.html>
- Department of Justice: http://www.justice.gov/05publications/05_3_a.html
- Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) (including Minerals Management Service): <http://www.boemre.gov/>
- NOAA: <http://www.noaa.gov/>
- National Cancer Institute: <http://www.cancer.gov/>
- National Institutes of Health: <http://www.nih.gov/>
- USDA: <http://www.usda.gov/wps/portal/usda/usdahome>
- OMB: <http://www.whitehouse.gov/omb/>
- **Harvesting with archive-it:
 - EPA: <http://www.epa.gov/>
 - GAO: <http://gao.gov/>
 - Census current industrial reports: <http://www.census.gov/manufacturing/cir/index.html>



Friday, June 3, 2011

My use of the EEMs workflow and tool grew out of 2 other projects that I'd like to briefly mention: delicious IAdeposit and lostdocs.freegovinfo.info.

The delicious project was an attempt to build a collaborative effort toward collecting digital fugitive documents (those documents that should be distributed via the FDLP but are not). The original intent of that project was that anyone could tag documents of interest with the delicious tag IAdeposit and a group of us would periodically upload those tagged documents to the Internet Archive's US documents collection (<http://www.archive.org/details/USGovernmentDocuments>). I continue to upload documents when I come across them, but there hasn't been enough structure to the project for many people to take it up and run with it unfortunately.

A second, more successful project is Lostdocs.freegovinfo.info. Lostdocs is a blog and community effort that tracks fugitive document submissions to the GPO in order to provide a public listing of fugitive documents.

Through the work of the lostdocs blog, I've been able to target 8 agencies that generally are the worst offenders in terms of fugitive documents:

- Bureau of Land Management CA field office (can also check OR and WA field offices) <http://www.blm.gov/ca/st/en/info/publications.html>
- Department of Justice http://www.justice.gov/05publications/05_3_a.html
- Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) (including Minerals Management Service) <http://www.boemre.gov/>
- National Oceanic and Atmospheric Administration (NOAA) <http://www.noaa.gov/>
- National Cancer Institute <http://www.cancer.gov/>
- National Institutes of Health <http://www.nih.gov/>
- USDA <http://www.usda.gov/wps/portal/usda/usdahome>
- OMB <http://www.whitehouse.gov/omb/>

We also found that 3 other agencies that were top fugitive offenders published too many documents to make the EEMs workflow feasible. So I'm harvesting the following 3 agencies with Archive-it:

- EPA <http://www.epa.gov/>
- GAO
- Census current industrial reports <http://www.census.gov/manufacturing/cir/index.html>

I have a staff person working 1-2 hrs per week on this project. She checks the agency publications page for new publications; Checks the CGP (<http://catalog.gpo.gov>) to see if the document has made it into the catalog of govt

Search Title

SearchWorks

Space + Policy + of + the + United + States + of + America
cted)

Selected items (none)

< Previous | 2 of 76 | Next >

Select thisCite ThisSend to

Online

- purl.stanford.edu

Other libraries

Find it at other libraries via WorldCat

Language: English

Imprint: [Washington, D.C.] : Executive Office of the President, [2010]

Format: Book

1 online resource (14 p.)

Note: Title from PDF title page (Whitehouse.gov, viewed June 29, 2010).
Attributed to: Barack Obama. "June 28, 2010." Downloaded from http://www.whitehouse.gov/sites/default/files/national_space_policy_6-28-10.pdf on Aug 31, 2010.

Contents:

- Introduction
- Principles
- Goals
- Intersector guidelines
- Sector guidelines.

Summary: The National Space Policy expresses the President's direction for the nation's space activities. The policy articulates the President's commitment to reinvigorating U.S. leadership in space for the purposes of maintaining space as a stable and productive environment for the peaceful use of all nations.

EEM: <http://searchworks.stanford.edu/view/8707790>

Through the EEMs workflow, we’ve been able to collect over 300 documents like this one (notice the Stanford PURL), preserve them locally in the Stanford digital repository (SDR) and give access to them through our catalog, searchworks. Think what we could do if 100 libraries instituted this workflow? Collectively, we could cover all federal agencies to assure that no document within scope of the FDLP falls through the cracks and becomes fugitive.

Archive-it

- collecting the Web in bulk

- Archive-it.org

- Fotopedia image by Marcus Revertegat. Creative Commons Attribution 3.0 Unported license.



Archive-it is a subscription service from the Internet Archive. It's an easy collection building tool whereby you give the software a list of urls (called "seeds"), schedule the crawler to harvest the seeds, and then give public access to the content collected. It's a good way to contextualize or make sense of the ocean of content on the open Web.

Since 2007:

Documents Crawled: 41,176,614

Data Archived: 3,683.4 GB

Web archiving @ Stanford University Library

Search...

Congressional Research Service (CRS) Reports

Go



Welcome to the Web archiving project of Stanford University Libraries Social Sciences Resource Center (SSRG). On the right, you will see the growing list of collections that we're currently harvesting. You can search within collections by using the search box above or by clicking on the title of a collection on the right for both a search box and a list of Websites that are currently being harvested within each collection. Other groups that are working with Archive-it include the [Islamic and Middle Eastern Collection](#) (currently harvesting Iranian blogs) and the [Stanford Humanities Lab](#) (focusing on digital games, virtual worlds and Massive Multiplayer Online Games (MMOs)).

The SSRG supports research, instruction and collection development within the social sciences, including subject areas like political science, economics, sociology, anthropology, psychology, communication, environmental and public policy. The SSRG also includes CA state, local, US Federal and International government information, data and statistics. SSRG's Archive-it collections will focus on San Francisco bay area governments, IGO/NGO working papers, and other subject areas of interest to Stanford social scientists.

- [Stanford Social Sciences Test Collection](#)
- [Bay Area Governments](#)
- [Freedom of Information \(FOIA\)](#)
- [California Department of Education Curriculum and Instruction](#)
- [Climate change and environmental policy](#)
- [Congressional Research Service \(CRS\) Reports](#)
- [Stanford Knowledge](#)

SULAIR archive-it home:
<http://www.archive-it.org/home/SSRG>



Friday, June 3, 2011

What I'm collecting with Archive-It:

- CRS Reports
- FOIA
- Fugitive US documents (shout-out to lostdocs.freegovinfo.info)
- Bay Area governments
- Climate change and environmental policy
- G-20
- CA Dept of education curriculum and instruction
- US budget
- FRUS

- <http://countingcalifornia.cdlib.org/crs/> - **Congressional Research Service reports (1994-1999)**
Collection of Congressional Research Service reports on the California Digital Library's Counting California Web pages.
- <http://digital.library.unt.edu/govdocs/crs/> - **Congressional Research Service reports, hosted by UNT (University of North Texas) Libraries**
Collection of Congressional Research Service reports on the Web site of the University of North Texas Libraries. A variety of topics are covered.
- <http://fpc.state.gov/fpc/c18185.htm> - **Congressional Research Service reports (CRS) and issue briefs**
Collection of Congressional Research Service reports maintained on the Web site of the U.S. State Department. Topics include U.S. foreign polic
- <http://fpc.state.gov/fpc/c18193.htm> - **Congressional Research Service reports (CRS) and issue briefs by topic**
Collection of Congressional Research Service reports maintained on the Web site of the U.S. State Department. On this page, the reports are org
- <http://italy.usembassy.gov/policy/crs/> - **U.S. policy issues, Congressional Research Service reports**
Collection of Congressional Research Service reports on the Web site of the U.S. Diplomatic Mission to Italy. Reports cover U.S. foreign relations
- <http://italy.usembassy.gov/policy/crs/default.asp>
- <http://ncseonline.org/NLE/CRs/> - **Congressional Research Service reports at the National Library for the Environment**
Collection of Congressional Research Service reports on environmental issues and policy compiled by the National Library for the Environment.
- <http://www.fas.org/sqp/crs/index.html> - **FAS Congressional Research Service reports**
Collection of Congressional Research Service reports on national security, foreign policy, and related topics on the Web site of the Federation of
- http://www.firstamendmentcenter.org/about.aspx?item=crs_reports - **CRS reports for Congress on First Amendment topics**
Collection of Congressional Research Service reports about First Amendment issues compiled by the First Amendment Center.
- <http://www.globalsecurity.org/military/library/report/crs/> - **Congressional Research Service (CRS) reports**
Collection of Congressional Research Service reports on the GlobalSecurity.org Web site. Topics include U.S. national security and foreign polic
- <http://www.iwar.org.uk/news-archive/crs/> - **Congressional Research Service (CRS) reports and issue briefs**
Collection of reports by the Congressional Research Service on the Information warfare site. Reports cover national security, U.S. foreign relation
- <http://www.law.umaryland.edu/marshall/crsreports/> - **CRS reports in homeland security, terrorism, and health law & policy**
Collection of Congressional Research Service reports, created by the Thurgood Marshall Law Library. The reports cover the topics of national se
- <http://www.nationalaglawcenter.org/crs/> - **Congressional Research Service reports**
Collection of Congressional Research Service reports about agriculture and food issues on the Web site of the National Agricultural Law Center.
- <http://www.opencrs.com/> - **Open CRS Network—CRS Reports for the People**
Web site created by the Center for Democracy and Technology that brings together collections of reports produced for the United States Congres
- http://www.rules.house.gov/archives/crs_reports.htm - **Congressional Research Service reports—legislative process in the House**

Collection seeds

<https://archive-it.org/public/collection.html?id=1078>

building collections of urls or seeds
metadata creation: Metadata: assistance from a cataloger to do Dublin core metadata at the collection and seed level. Archive-it allows for metadata at the document level, but we have not done that.
crawl reports

Basic Search | [Advanced Search](#)

"energy efficiency"

Search

[Help with Search](#)

☒ just this collection ☐ all collections

Text Results

1 - 20 of about 3,000 for "energy efficiency" found in 1.8 seconds

1 2 3 4 5 6 7 8 9 10 next»

WIKILEAKS - Congressional Research Service - Energy Efficiency Policy: Budget, Electricity Conservat

Research Service Report RL33599 Energy Efficiency Policy: Budget, Electricity Conservation, and Fuel... Efficiency Program, which is conducted by the Office of Energy Efficiency and Renewable Energy (EERE...), and Industry programs. The energy efficiency provisions in the Energy Policy Act of 2005 (P.L. 109-58, H.R. 6) and other bills of the 109th Congress are discussed in the "Energy Efficiency in the 109th... Order Code RL33599 Energy Efficiency Policy: Budget, Electricity Conservation, and Fuel Conservation.../w... 3-RL33599 Energy Efficiency Policy: Budget, Electricity Conservation, and Fuel Conservation Issues Summary Energy efficiency issues include research and development priorities, funding..., regulation of vehicle fuel efficiency, and electricity industry ratemaking for energy efficiency...

Document type: PDF - 389.2 KB - crawled once Feb 10, 2009

wikileaks.org/leak/crs/RL33599.pdf - [more results from wikileaks.org](#)

search and discover
<http://snipurl.com/crs-energyefficiency>



Friday, June 3, 2011

access (full text search IA, archive-it site, databases page, embeddable search form, Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) + plans to index collection/seed metadata in SUL catalog)

Paste this into your HTML:

```
<form action="http://www.archive-it.org/public/search">  
  <input type="hidden" name="collection"  
    value="***COLLECTIONID***" />  
  <input type="text" name="query" />  
  <input type="submit" name="go" value="Go" />  
</form>
```

COLLECTIONID = 1078 (CRS reports collection)

add search to other pages
</gratuitous_code>

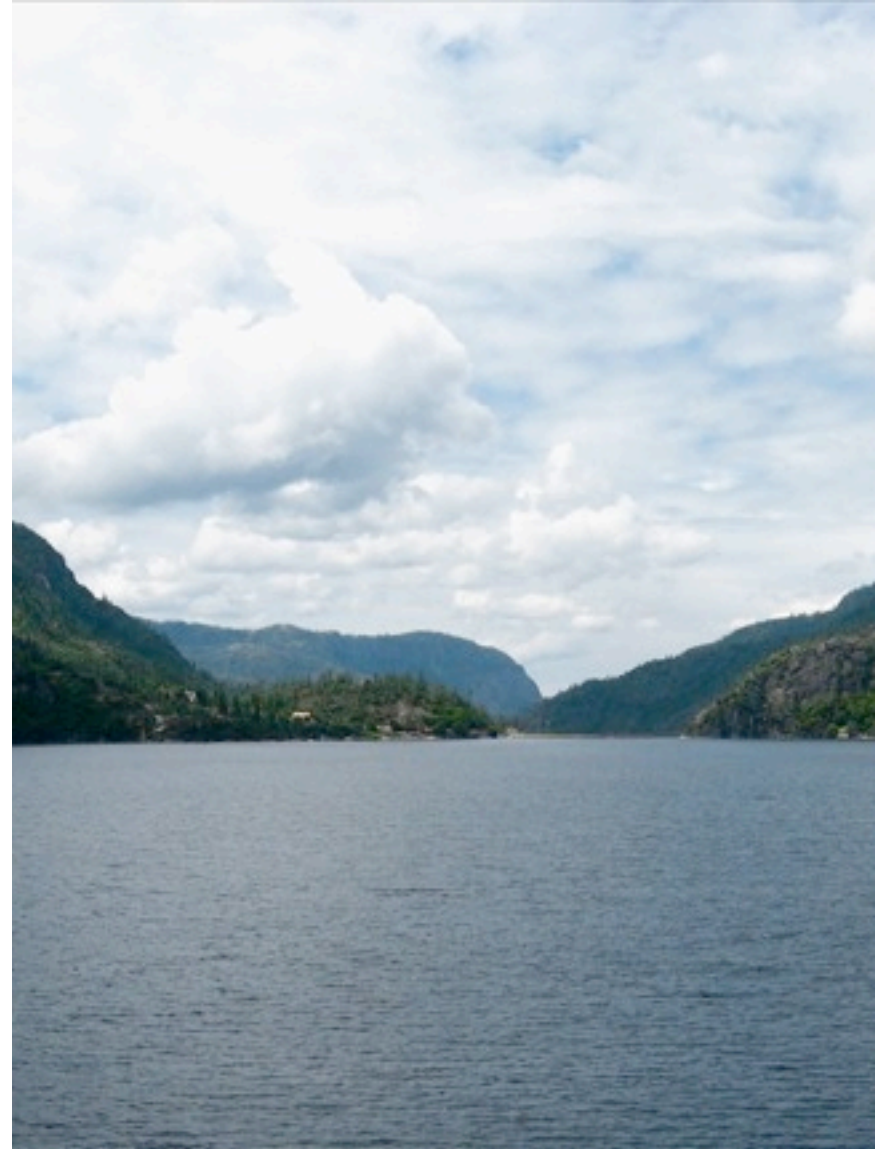
You can also, with a small snippet of javascript code, embed the search box into other html pages.

Archive-it utility:

- easy to use (but takes some time and effort to tune and need better analysis tools)
- Harvests can be fed into LOCKSS caches
- Archive staff knowledgeable and interested in making the tool better
- growing network of different types of libraries with very different needs.

Collaboration

- Delicious IAdeposit
- State agency databases:
[wikis.ala.org/godort/index.php/
State_Agency_Databases](http://wikis.ala.org/godort/index.php/State_Agency_Databases)
- Seeding wikipedia with digital documents.
See “Using Wikipedia to Extend Digital
Collections” D-Lib.
www.dlib.org/dlib/may07/lally/05lally.html
- Yosemite's Hetch Hetchy reservoir flickr photo by Random Curiosity. Attribution-NonCommercial-ShareAlike 3.0 creative commons license.



I've already mentioned the delicious IAdeposit project, but I'd like to also briefly mention the far more successful collaborative GODORT project on - what else!? - state agency databases organized by my freegovinfo colleague Daniel Cornwall and kept up to date by 50+ volunteers. This project has been a huge success in collecting various public databases at the state level. I have to give a shoutout here to Audrey Hall, reference librarian in Government Information Services at the State Library of Ohio. Audrey's built a robust listing for OH state agencies (<http://wikis.ala.org/godort/index.php/Ohio>) Thanks Audrey!!

State Agency Databases
http://wikis.ala.org/godort/index.php/State_Agency_Databases

LOCKSS-USDOCS

- Targeted Web collection and distributed preservation
- **L**ots **o**f **C**opies **K**ee**p** **S**tuff **S**afe
- lockss-usdocs.stanford.edu
- Flickr waterfall picture by discordia1967. That's actually me at Hanakapi`ai falls in Kauai :-)



LOCKSS is ...

- Distributed Digital Preservation System
- Open source peer to peer (P2P) software
- Standards based: OAIS, OpenURL, HTTP, WARC
- content migrator to new formats as required “on the fly” at point of access
- bits and bytes are continually audited and repaired

LOCKSS – Lots of Copies Keep Stuff Safe – began at Stanford in 1999. The LOCKSS software was built to solve the problem of long-term preservation of digital content. It is an open-source distributed digital preservation system based on open standards (OAIS, OpenURL, HTTP, WARC (web archive file format). Originally LOCKSS was focused on journal literature -- and today CLOCKSS is going strong with 81 libraries and 30 journal publishers participating! – but over the last 10 years has been used by other projects focusing on government information, theses and dissertations, numeric data etc.

The goals of LOCKSS is to spread out the economic cost of digital preservation and use off the shelf hardware, so that libraries and content publishers can easily and affordably create, preserve, and archive local electronic collections and readers can access archived and newly published content transparently at its original URLs.

Think of LOCKSS boxes as digitally distributed bookshelves!

- Distributed Digital Preservation System
- Open source peer to peer software on Linux OS.
- Standards: OAIS, OpenURL, HTTP, WARC
- Migrates content to new formats as required “on the fly” at point of access

Private LOCKSS Networks

US Government Documents



Besides the LOCKSS global network and CLOCKSS, the software is used by other projects focusing on government information, theses and dissertations, numeric data etc. showing LOCKSS is flexible, reliable, efficient and highly scalable.

- The Alabama Digital Preservation Network (ADPN)
- Arizona State Library, Archives and Public Records' Persistent Digital Archives and Library System (PeDALS)
- Council of Prairie and Pacific University Libraries (COPPUL) Consortium
- Data Preservation Alliance for the Social Sciences (Data-PASS)
- Digital Commons - Berkeley Electronic Press
- MetaArchive Cooperative Project

LOCKSS is funded by...

- LOCKSS alliance library members
- LOCKSS has received funding and in-kind support from:
 - Andrew W. Mellon Foundation
 - National Science Foundation
 - Library of Congress
 - The UK's Joint Information Systems Committee
 - Sun Microsystems
 - HP Labs
 - Intel Research Berkeley
 - Stanford University Libraries and Academic Information Resources
 - Stanford Computer Science Dept.
 - Harvard Computer Science Dept.

sustainable funding is always an issue. LOCKSS is primarily funded by libraries participating in the LOCKSS alliance and has also received major funding and in-kind support from several other organizations. Non-LOCKSS alliance members may participate in LOCKSS-USDOCS for a token support fee of \$1250/yr or \$750/yr if they bring an additional library into the program.

Web stronger and more viable than a silo

LOCKSS Permission Statement

Digital Dissemination of Access Content Packages

Interest has been expressed by customers and partners in creating their own digital collections of content. This includes building digital collections by accepting digital files and metadata disseminated by GPO's Federal Digital System (FDsys).

GPO will make available Access Content Packages (ACP's) to download and store on local systems. For instance, for Federal depository libraries (FDL's), this may assist in efforts to build digital collections at their libraries.

GPO will continue to maintain responsibility for managing Federal content within scope of the Federal Depository Library Program (FDLP), and providing free and permanent access to this information.

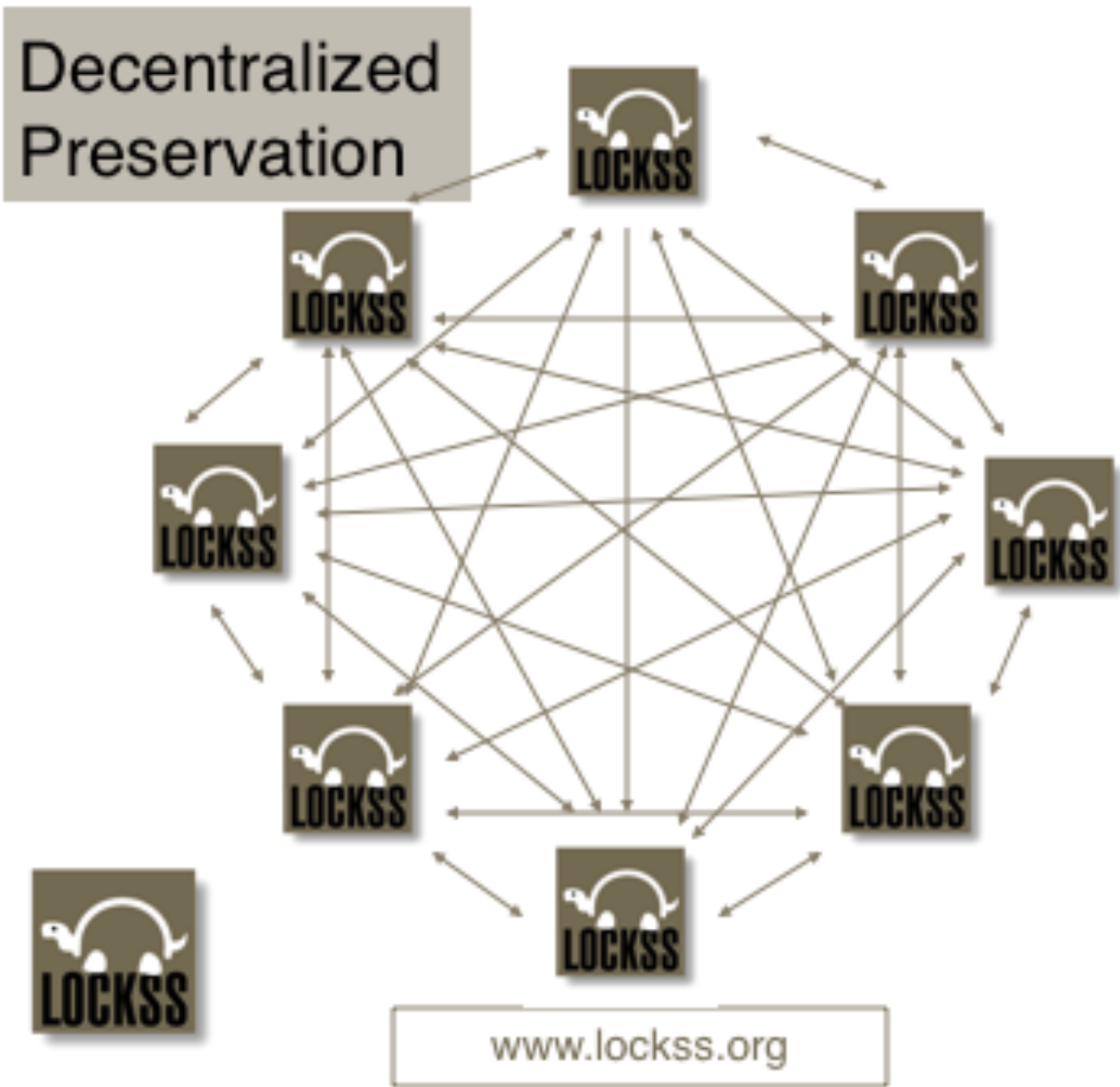
LOCKSS system has permission to collect, preserve, and serve this Archival Unit.

<http://www.gpo.gov/fdsys/bulkdata/FR/resources/lockss.html>

There are 2 parts to the LOCKSS software: harvest and content collection; and content checking and replication.

1) any site that gives lockss permission to harvest can be collected by the LOCKSS harvester the state of the art in Web harvesting!

next slide for #2 ...



2) and this is the cool part: locks goes through a process of checking and polling all digital content in all of the locks boxes on a network. If 1 box has content that is different from all of the other boxes, the software will fix the content, assuring that all content in the whole network is exactly the same. It is for all intents and purposes injecting stem cells into the network to replicate and fix content that's become corrupted over time.

That's it. LOCKSS is elegant in its simplicity and proven effective in keeping digital content safely preserved over time. This is as close to the unix maxim of "doing one thing, doing it well." In the digital world, this is as close to perfect as one could get.

LOCKSS-USDOCS

- LOCKSS for US Documents
- Replicates FDLP in the digital environment
- “digital deposit” (for more on “digital deposit,” see <http://freegovinfo.info/taxonomy/term/3>)
- Tamper evident

So now you can see why some of us in the documents world are so excited about LOCKSS and why we decided to implement LOCKSS-USDOCS.

Lockss-usdocs replicates key aspects of the FDLP in the digital environment – a network of 1250 libraries supporting access to and long-term preservation of govt documents – and furthers the concept of "digital deposit," an essential component of the digital FDLP. for more on “digital deposit” see <http://freegovinfo.info/taxonomy/term/3>

In the paper environment, the decentralized FDLP is a tamper evident system. When someone tried to alter or withdraw a paper document from the system, the librarians were alerted. They had a chance to react, and frequently persuaded the government to take different actions. Using the LOCKSS software we are re-implementing a tamper evident preservation system for digital documents. Rather than a central silo on a .gov server, digital govt documents reside on 36 servers at 36 different libraries (and counting!).

Preserving

- GPOaccess content (1991 - 2007) harvested from <http://bulk.resource.org/gpo.gov/>
- All current and future FDsys collections <http://www.gpo.gov/fdsys/browse/collectiontab.action>

LOCKSS-USDOCS is ...

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!!



Friday, June 3, 2011

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, etc. 40 collections plus bulk data repositories for the federal register and the code of federal regulations.

Process is simple:

Join the project and the discussion list

Set up a lockss box with at least 3TB hard drive (some libraries recycle older hardware, some run in a virtual server environment, some purchase new boxes from a vendor who has worked closely with the lockss staff to build to lockss specifications)

Sit back and watch it fill up



LOCKSS-USDOCS participants
<http://snipurl.com/lockss-usdocs-partners>

What's next for LOCKSS-USDOCS?

- More participants
- Expand collections
- Make project participant-driven

In the 2008 Blue Ribbon Task Force on Sustainable Digital Preservation and Access, Abby Smith Rumsey wrote, “Access to valuable digital materials tomorrow depends upon preservation actions taken today; and, over time, access depends on ongoing and efficient allocation of resources to preservation.”

Do these projects ...

- ... forward democratic ideals?
- ... serve public interest / public access / public control / public preservation
- ... serve the information needs of the community?
- ... forward the long-term institutional viability of libraries?
- ... promote and leverage collective action?

Librarians ...

... Explore

... Collect

... Describe

... Share

... Preserve

If you're like me, you have a passion for documents. You want to explore, collect, describe, share, and preserve government information. Formats have unique properties and unique issues with doing these things. But format is beside the point. We must continue to explore, collect, describe, share and preserve government documents.

“...let us save what remains: not by vaults and locks
which fence them from the public eye and use in
consigning them to the waste of time, but by such
a multiplication of copies, as shall place them
beyond the reach of accident.”
— Thomas Jefferson, February 18, 1791

Digital Strategy:

These projects are part an overall digital strategy. It runs the gamut from saving individual documents, to small collaborative projects (IADeposit), to large-scale harvesting and preservation efforts. I want to stress that this strategy DOES NOT preclude paper documents. In many respects our historic collections are what drive our raison d'etre.

Farmington Plan Redux (collaborative collections)

Technological tools are there. But there's a real, critical need for a "Farmington Plan redux":
The Farmington Plan, which lasted from 1948 – 1972, was an innovative ARL program of collaborative collection development whereby subscribing libraries would have responsibility for collecting and cataloging research materials in certain subject and/or linguistic areas and would then distribute records (in the form of cards) to the National Union Catalog. We need the same kind of plan in the FDLP. ASERL's proposed "collections of excellence" starts to do that, but only focuses on historic collections.



©2004 P. S. Mueller From Your Belief System Is Shot-2004 Jones Books

Thanks!

Further reading

- 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).
<http://freegovinfo.info/system/files/lockssusdocs-dtt38%283%29.pdf>
- Everyday Electronic Materials in Policy and Practice. Coalition for Networked Information (CNI) project briefing. Fall 2010. Katherine Kott.
<http://www.cni.org/tfms/2010b.fall/Abstracts/PB-everyday-kott.html>
- A Guide to Distributed Digital Preservation. K. Skinner and M. Schultz, Eds. (Atlanta, GA: Educopia Institute, 2010). <http://www.metaarchive.org/GDDP>
- Several technical articles on LOCKSS at D-Lib Magazine www.dlib.org
- “Digital Deposit” <http://freegovinfo.info/taxonomy/term/3>
- <http://lockss-usdocs.stanford.edu>