Government information: everywhere and nowhere

···

James R. Jacobs and James A. Jacobs • GPLNE, 10.24.2017

FGI

freegovinfo.info

Agenda

Introduction: disappearing govt info has *always* been a problem.

The rise of the internet as a publishing platform.

Current coping mechanisms for the preservation of born-digital government information.

Comprehensive strategy for preservation and access.

Thanks Andrew for inviting Jim and I to speak to GPLNE and for GPO for virtually hosting our talk. Today we'll talk about 4 things:

1. The building of the national collection and Fugitive documents
2. The rise of the internet as a publishing platform and its impact on govt information
3. A couple of current projects focused on the preservation of born-digital government information.
and
4. our thoughts on building a comprehensive strategy for preservation and access.

Venn diagram of government information
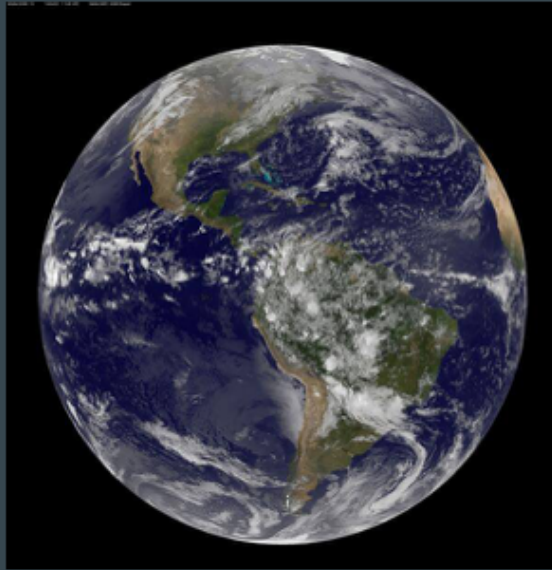
The world of govt information is global and all encompassing, and has *inherent* value as a record of democracy and record of data collected and maintained by the govt about the country. The goal of the Federal Depository Library Program (FDLP) has since 1813 been to build and maintain the "national collection." The FDLP's scope consists of a large swath of published materials from all three branches of government, from the 440 or so executive agencies and commissions, Congressional bills, committee hearings, committee prints, House and Senate documents and reports, and publications, and reports and opinions of the federal courts. GPO states that the scope of the FDLP includes, "publications having public interest or educational value."

The FDLP has generally worked well as both an access and long-term preservation mechanism for 200+ years because it consists of a relatively simple collaborative process: 1) GPO receives documents from government entities; 2) catalogs and indexes them – from 1895 to 2004 in the Monthly Catalog and currently online in the Catalog of Government Publications (CGP) and govinfo.gov – and then prints and distributes or "deposits" them in libraries; and 3) 1100+ libraries receive documents and provide access and public services for them to the public. The act of collecting, describing, giving access to and preserving documents remains critical to building the national collection.

Where Documents Are Obtainable

When the office of the Superintendent of Documents has been fully organized and all the provisions of law relating to it have become operative, it is expected that it will be able to supply all ordinary demands for the purchase of the current documents and most demands for copies of the older documents. It will, however, be some time before this expectation can be entirely realized.

At present the State Department sells the Statute Laws, and the Bureau of American Republics, attached to the State Department, sells its own publications.

The Treasury Department sells the Catalogue of Copyrights, the Monthly Statements of Finance, Commerce, and Immigration by the Bureau of Statistics, and the publications of the Coast and Geodetic Survey.

The War Department sells the Rebellion Record.
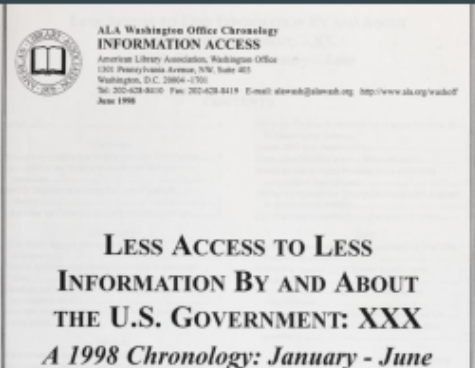
The Post-Office Department sells the post-route maps.

The Navy Department sells the Naval War Record and the publications of the Nautical Almanac Office and the Hydrographic Office.

The Interior Department sells the publications of the Census Office; the Decisions and Maps of the General Land Office; the reports, maps, monographs, and bulletins of the Geological Survey; the Decisions of the Pension Office, and the numerous issues of the Patent Office. The average number of newly issued patents for which specifications are printed each week is 450. Copies of these specifications are sold at 10c each. A list of them appears weekly in the Patent Office Gazette, for which the subscription price is $5.00 a year.

The national collection has always run into pitfalls and so-called "fugitive" documents have always been a fact of life. "Fugitives" are those documents that fall within scope of the FDLP, but for various reasons were not distributed to libraries through the program. Fugitives are therefore less accessible (or inaccessible) to the General Public AND at much higher risk of permanent loss.

While government information librarians tend to think of fugitives as a few random documents that have mistakenly fallen through GPO's cataloging and indexing and/or distribution nets, the reality is not as clear-cut as that and is much more prevalent. It may come as a surprise to some, but GPO has never had a monopoly on government printing, despite the agency's authority under Title 44. There are whole classes of fugitives, all of which were not distributed to FDLP libraries by GPO. (quick shameless plug: I just wrote an article about fugitives which will come out in the december issue of Against the Grain)

The Monthly Catalog (MoCat) is littered with examples of agency publications not included in the FDLP though clearly in scope. Some were not printed due to bureaucratic issues. Many agencies through the 20th century ran their own printing presses to save $$ or get around irksome Congressional "censorship." Besides the million or more fugitive agency documents, there are also ones like CRS reports, which to this day are considered by LC as "privileged communication" and commoditized documents like NTIS technical reports.

ALA Washington Office Chronology
INFORMATION ACCESS
American Library Association, Washington Office
1301 Pennsylvania Avenue, NW, Suite 403
Washington, D.C. 20004-1701
Tel: 202-628-8410 Fax: 202-628-8419 E-mail: alawash@alawash.org http://www.ala.org/washoff
June 1998

LESS ACCESS TO LESS INFORMATION BY AND ABOUT THE U.S. GOVERNMENT: XXX

A 1998 Chronology: January - June

Less Access to Less information by and about the US Government
https://freegovinfo.info/less_access

Controversial tobacco document pulled from congressional Web site

After mounting thousands of secret tobacco industry papers on their public Web site, the House Commerce Committee removed a controversial 104-page memo after R.J. Reynolds Tobacco Company officials objected it being made public.

The memo described how RJ R's lawyers suppressed research on the health hazards of smoking over a 30-year period. The Committee held back 400 documents from the public in response to tobacco industry pleas that the papers contained trade secrets or were entitled to confidentiality because they had been prepared by outside lawyers now working for

While others were not distributed (or worse were pulled out of libraries or off the Web after the fact) because of politics. If you've not had a chance to read ALA Washington Office's important series "Less Access to Less Information by and about the US Govt" which ran from 1981 - 1998, please check out the series on FGI. "Less Access" tracked and made the community aware of the many times the govt limited access to information.

Fugitives scope

"The number of fugitive print documents has been estimated as about 50% of the universe of Federal printing, but this estimate may be conservative."

Gil Baldwin, "Fugitive Documents – On the Loose or On the Run," Administrative Notes: Newsletter of the Federal Depository Library Program 24, no. 10 (August 15, 2003): 4–8, http://web.archive.org/web/20160321083457/ http://www-personal.umich.edu/~graceyor/govdocs/adnotes/2003/241003/an2410d.htm

"The Superintendent of Documents recently stated that 85% of these non-GPO publications fail to appear in the Monthly Catalog due to the fact that the issuing agencies do not provide copies of them to GPO for cataloging."

GODORT Federal Documents Task Force, "Suggestions to GPO. A Letter to the Superintendent of Documents. February 5, 1973," DttP 1(3) (May 1973): 21–28. https://searchworks.stanford.edu/view/wf053wt9145

Various people over the last 50 years have tried to get their heads around the fugitive issue. and have estimated that the number of fugitives range anywhere from 50% to 85% of govt publications.

Fugitives scope (cont'd)

Cynthia Bower study:

"No one knows" … "varies by agency and type"

fugitive documents outnumber depository documents by an average ratio of eight to one
43% of docs in American Statistics Index

Bower, Cynthia. "Federal Fugitives, DND, and other Aberrants: a Cosmology."
DttP v17 n3 (September, 1989).

Figure 3. Comparison of EPA Reports Cited in *EPA Publications Bibliography* With Those Cited in the *Monthly Catalog*, 1980-1988
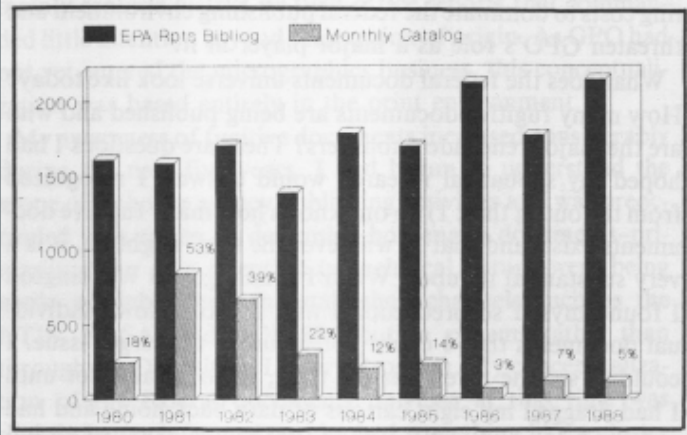
I'd especially like to note Cynthia Bower's research findings published in Documents to the People (DttP) in 1989. She has some fascinating data that seems all too familiar even though it is more than 25 years old. She began by saying that no one knows the scope of the problem.

She found that the fugitive problem varied by agency and type, so she dug into some of the specifics, finding for example, that 43% of documents indexed in the American Statistics Index were fugitives.

And (as you see on the chart at the bottom of this slide) she found that EPA publications (the black bars) became less and less likely to be listed in the Monthly Catalog (grey bars) over the decade she studied.

Examples

- Air quality benefits of alternative fuels.

- Greenhouse effect, sea level rise and coastal wetlands.

- Reagan administration regulatory achievements

- A Report to the Secretary on homeless & emergency shelters

- Report, Task Force on Women in the Military.

And here are some example fugitives noted by Bower's study.

Ms. Bower's paper is a really good one and I recommend it to you for its scope and scholarship, but also for the clear and insightful way she defines the problem.
The subjects are still relevant and current today. her study highlights the fact that when we talk about fugitives, we're not talking about ephemeral publications but important documents that need to be captured and preserved because they provide the historical record of government actions and understanding over time.

Past strategies

Institutional. DocEx service (LoC) 1946-2004. Libraries subscribed & received fugitives tracked down in D.C. (See Shaw, 1966 http://bit.ly/shaw-library-trends-1966)

Technical. GPO microfiche copies of documents printed elsewhere. Also https://www.everycrsreport.com/)

Individual. Librarians captured fugitives by scouring agency newsletters and press releases and local newspapers and by cultivating agency contacts. (see http://lostdocs.freegovinfo.info/)

Legal. Attempts to revise Title 44

There's a range of strategies that have been used or attempted over the years. As you can see, they include institutional and individual strategies; technical and legal strategies.

I want to highlight the DocEX project for you too and an excellent article on that project by Thomas Shaw in Library Trends.

The project was librarian-driven. It was created and sponsored by ALA, the Association of Research Libraries, the Special Libraries Association, and the American Association of Law Libraries. DocEx not only provided copies of fugitives to subscribing libraries, it also provided a copy to GPO for listing in the *Monthly Catalog*. Some of the important fugitive series that DocEx saved included Congressional "committee prints" and the *Daily Report* of the Foreign Broadcast Information Service.

The key takeaways here are that The FDLP has long guaranteed both "short-term accessibility" AND "preservation and long-term accessibility," there have historically been gaps and problems due to the lack of strong overarching public policy directives and political machinations, but librarians have worked long and hard to iteratively fill those gaps. now I'll hand it over to Jim to talk about the internet as a publishing platform.

Rise of the Internet as publishing platform

The Good

- Easy access to much more information than ever before.

Today, as you know, almost all government information is "born-digital" and made available on the web.
This information is much more readily and easily accessible than printed books were in the past.
This is a Good Thing!

The Bad

- Most born-digital govt. information is *not* being systematically preserved.

- This information can also be altered, moved, or deleted without notification, indication, or any record of change.

Unfortunately, this short-term "accessibility" MASKS the problems of long-term preservation, access, and usability of this information.
The result is that, in a very real sense, almost all born-digital government information is "fugitive"!
When we think about "fugitives" in the digital age, we have to ask if all this information that is accessible today will be accessible tomorrow.
Unfortunately, the answer is usually "No."
There are two reasons for this:
First, Most born-digital govt. information is not being systematically preserved.
Second, the government can alter, move, or delete any of this information without notification or indication of changes, and without any record of those changes.
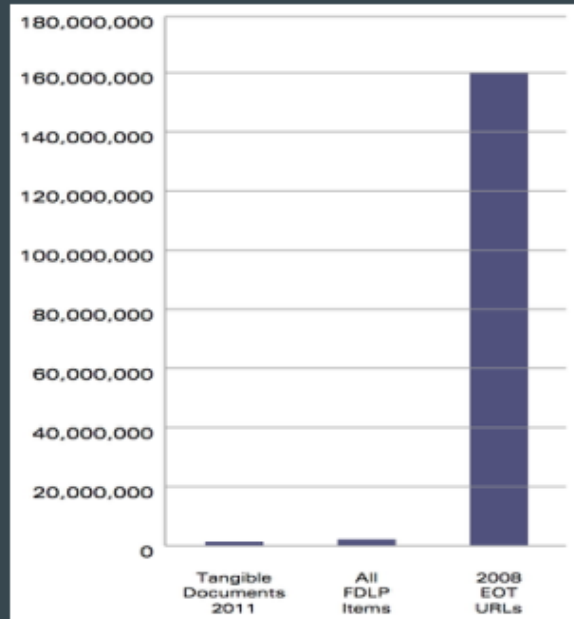
James A. Jacobs, *Born-Digital U.S. Federal Government Information: Preservation and Access*, March 2014. Prepared for Leviathan, the Center for Research Libraries Global Resources Collections Forum. http://www.crl.edu/leviathan

●To understand the scope of this problem, consider this chart.
●The first column is the number of items distributed by GPO to FDLP libraries in 2011. (about 10,000 items.)
●The second column is the total number of items distributed by GPO to FDLP over the last 200 years. (about 3 million items.)
●The third column is the number of URLs harvested by the 2008 End of Term crawl. (about 160 million items!)
●This demonstrates 3 things:
●• First, Most of those 160 million URLs can be considered "fugitives." (Although captured by the End of Term Crawl, they have not been selected and organized and described into usable collections and are not really available for creation of services or for use or re-use.)
●• Second, Even **this** probably **under-**estimates the scope of the born-digital fugitive problem. (The End of Term crawl was only a snapshot and it undoubtedly missed things that were already gone at the time of the snapshot. Also, it only captured things that had URLs: no databases, etc. And automated crawls like this can miss a lot.)
●• Third, and Perhaps most importantly, Look again at that second column (which represents the entire FDLP Historic paper collection) and that third column (which represents at most 4 years of born-digital information). This tells us that the digital-fugitive problem is much larger **in a single year** than the paper-fugitive problem was over the entire **200 year** history of the FDLP.

Access is not preservation.

Short-term access ≠ Long-term access

"Digital preservation is access …
in the future."

So, when we try to understand the problem of "disappearing government information" let's remember that having "easy access" today tells us nothing about access tomorrow.

The point is that "Digital preservation" is not something esoteric. It will not happen magically. And it is intimately and inherently linked to long-term access.

"Access in the future": David Brunton, quoted by Ed Summers in "The Web as a Preservation Medium" inkdroid.org Nov 27, 2013

http://inkdroid.org/journal/2013/11/26/the-web-as-a-preservation-medium/

OAIS: the international standard for preservation

Information must be

- not just preserved, but *discoverable* [2.2.2]

- not just discoverable, but *deliverable* [2.3.3]

- not just deliverable as bits, but *readable* [2.2.1]

- not just readable, but *understandable* [2.2.1]

- not just understandable, but *usable* [4.1.1.5]

So, if we want to know if government information is "disappearing," we have to ask if it is being preserved.
That means we have to ask what "digital preservation" means.
Luckily, there is an international standard for preservation, "The Reference Model For An Open Archival Information System" or OAIS, which defines what "digital preservation" is.
By the way, OAIS is the basis for certifying Trusted Digital Repositories. (GPO intends to get TDR certification for FDsys/govinfo.gov.)
OAIS tells us that just storing files or making them "accessible" is not the same thing as preserving them.
To preserve information you have to ensure that the information is all these things...

...

It is worth mentioning here, that GPO's govinfo.gov serves an important, indeed essential, role in preserving and providing systematic, consistent access to a large subset of government information.
But it is also worth noting that what GPO does is not enough.

First, almost everything it is doing is voluntary and beyond its legal mandate in Title 44 of the U.S. Code. It is great that it is doing this, but its actions are very fragile. they are subject to the budgetary whims of Congress and the policy whims of the next administration of GPO.

Second, GPO provides generic "access," which is good, but insufficient in the digital age. As this this list implies, users need, not just access to single files, but better discovery tools -- not just to files, but to parts of files, and data and facts embedded in databases. And users need better mechanisms for acquiring files groups of files and parts of files and data. And they need better mechanism for using and reusing and analyzing digital information....

This means that FDLP libraries need to do things that GPO is not or cannot do. We could wait, and hope the private sector will do this for us, but then we would have to pay them over and over again. or we can do these things ourselves.

Those numbers in brackets refer to sections of the OAIS standard.
https://public.ccsds.org/Pubs/650x0m2.pdf

Ensuring long-term access requires:

- Intentional, ongoing, active attention.

- Commitment.

- Resources.

If we want to ensure that digital govinfo does not disappear and if we want to guarantee long-term, free public access to it, we have to have organizations that are committed to this and that have sufficient resources to engage in intentional, ongoing preservation activities.

Without this, we will lose information, or have it commodified by the private sector, or both, even if it is temporarily freely "accessible" today.

The Ugly

Different laws for paper docs and digital docs:

- **Paper**: *44 USC Chapter 19* mandates distribution, preservation, and free public access. It covers all "government publications."

- **Digital**: *44 USC Chapter 41* has no mandate for distribution to libraries and no mandate for preservation. It allows fees and has a narrow scope (2 titles)

That brings us (after the good and the bad) to The Ugly.

Who will preserve government information? To answer that question we have to ask what the law requires.

Today there are two different laws that cover long-term access to government information. Both are in Title 44 of the U.S. Code:

Chapter 19 covers paper documents and mandates that they be distributed to FDLP libraries. It gives to those libraries the control to preserve those documents and mandates that they make those documents freely available to the General Public. The scope of this chapter is very broad and covers all "government publications."

Chapter 41 covers digital government information. (This is the law that enables govinfo.gov.) It does not require distribution of this information to FDLP libraries. It does not require GPO to preservation digital information. It allows GPO to charge fees for access to digital government information. And, it has a narrow scope: it only mentions 2 titles (The Congressional Record and the Federal Register)!

As I mentioned, the current GPO administration goes beyond the mandate of Chapter 41 to encompass its values of preservation and free access and more titles. But past GPO administrations have not shared these values, and these values could easily

change overnight with a new GPO administration. And Congress could easily reduce funding to limit GPO's ability to do more than the law mandates.

Coping mechanisms for preserving born-
digital government information

Now let's talk about digital preservation coping mechanisms. I use the term "coping mechanism" purposefully because, to date, there are only 2 digital preservation efforts specifically targeting born-digital govt information on a large scale. So let me talk for a few minutes about LOCKSS-USDOCS and End of Term web archiving project.

LOCKSS-USDOCS

36 libraries in a collaborative preservation network

Collecting and preserving copies of all collections on https://Govinfo.gov

Added assurance to information published by GPO

Primarily Congressional information

Preserved, but *not* discoverable

http://lockss-usdocs.stanford.edu

LOCKSS-USDOCS, which started in 2008, is a collaborative project among 36 libraries and the GPO. The project has created a distributed preservation network of federal government information based on FDLP principles of distribution, deposit and preservation, using the award-winning, open source LOCKSS preservation software. USDOCS collects and preserves copies of all documents published by GPO on its FDsys/Govinfo content management system. This includes bills and statutes, Congressional committee materials like hearings, documents and reports, the Congressional Record, Federal Register and Code of Federal Regulations, as well as a smattering of executive branch and judicial branch documents and reports. It is overwhelmingly Congressional in focus, and most content is from the mid-1990s to present. GPO has put up a LOCKSS permission statement and configured its site maps so that the LOCKSS software can collect all documents and metadata on the site.

USDOCS is critical to the FDLP community's ongoing preservation efforts because it assures that many digital copies of GPO published information are held outside the government domain, giving local control of content to libraries, and that those copies are in digital preservation environments maintained by the LOCKSS software for the long-term.

However, there are 2 drawbacks to the project that cause me to label it a coping mechanism. 1) As I said earlier, the project is overwhelmingly Congressional, largely leaving out executive branch agencies and the federal judiciary from any community-

driven born-digital preservation efforts. and 2) LOCKSS is primarily a preservation system. LOCKSS was built so that access would only happen after a "trigger event" like an organization going out of business. The LOCKSS team is working on the access piece, allowing LOCKSS content to be available via a links resolver service like SFX, and we're talking about transfering a copy of the USDOCS cache into the Internet Archive, but those pieces are not yet in place. So, in OAIS parlance, USDOCS content is preserved but not discoverable.

.Gov Web archiving efforts

web archiving of the .gov domain at various levels and scopes.

LOC: targeted .gov and congressional, election, other
GPO: agency sites, often ephemeral documents
NARA: congressional web harvest every 2 years
IA: global, national domain & curated crawls of all sorts
Agency-level: NIH/NLM, DOE, DOL, HHS, CMS, others, using Archive-It or other tools to preserve their own domains
UNT, Stanford & Others: Topical and targeted .gov collecting

now let's shift to Web archiving, the library world's go-to digital collecting industrial sized bulldozer. Over the last several years, web archiving of the .gov domain has begun to happen at various levels and scopes, including from govt agencies like LC, GPO, NARA and several executive agencies as well as from non-gov libraries like the Internet Archive, UNT, Stanford and others.

EOT 2016 partners!

So, as you can see, there was already a lot of activity going on in this space. Several institutions decided in 2008 to come together every 4 years (at the end of each presidential term) to bring together a community of experts and interests around the effort of crawling the .gov/.mil domain, in order to expand all of our collections and leverage economies of scale.

So, in the fall of 2016 a group of intrepid librarians, archivists and technologists from the Internet Archive, Library of Congress, GPO, CA Digital Library, and libraries from the University of North Texas, Stanford University, and George Washington University came together for the 3rd End of Term crawl to preserve a snapshot of the federal government web. The group once again came together with the goals of identifying, harvesting, preserving and providing access to a snapshot of the federal government web presence both as a way of documenting the changes caused by the transition of elected officials in the executive branch of the government and to provide a broad snapshot of the federal domain that is replicated among a number of organizations for long-term preservation.

Volunteer contributions

End of Term Presidential Harvest 2016

Project Home   About This Project   Project Reports   Feeds ▾   Add A URL

Capture the Following Site

Website URL: *
The URL you wish to nominate for capture

http://

2008: 457 from 26 nominators
2012: 1476 from 31 nominators
2016: 15,000+ from 400+ nominators (via UNT form)
Plus!: Over 100,000 from DataRescue/EDGI events/tools

The way end of term works is that we've collected a list of top-level .gov/.mil domains (EPA.gov, State.gov etc) from official sources like USA.gov and the GSA. Then we also gather "seed" or url nominations from the public as a way to hone the crawler so that it's net goes the depth and breadth of the .gov Web space.

This time around, the number of seed nominations exploded due to the public's fears and concerns over the loss of information and data brought on by the actions of the new Trump administration (look for silver linings where you can find them :-)). So besides the many more individual seed nominations received this time around, we also received over 100,000 nominations from DataRescue events, the Environmental Data and Governance Initiative (EDGI), ClimateMirror and other grass-roots actions and groups.

EOT 2016 results

- ~300 TB data total
  - ~110 TB web crawls + ~130 TB of gov ftp site archiving + social media
- 310,000,000 web URLs + 12,000,000 ftp files
- UNT PDF metadata project http://bit.ly/eot-metadata-guide

https://web-beta.archive.org/details/collection-eot2016-waybacksummary
http://eotarchive.cdlib.org
http://bit.ly/eot-200tb

Because of these efforts, we were able to expand our collection targets to include data sets on FTP servers and social media accounts from twitter, facebook, youtube, flickr etc.

We collected almost 300TB of data over the 6 months of crawling, including 130TB of extremely important FTP datasets which had never been collected before because the Internet Archive's crawlers generally target information via HTTP/HTTPS protocol. We collected over 310 million web urls and 12 million FTP files. We collected every type of file imaginable from PDF, text, video, audio etc including even a handful of semantic Web Turtle files (Terse RDF Triple Language)!

We're still processing metadata and doing full-text indexing so the 2016 archive will be available in the next few weeks from the CDL's site. Additionally, UNT has pulled out a subset of PDFs (assuming that these are most likely to be "publications") and geared up a volunteer metadata project to enhance the metadata and access to crawled materials. If you're interested, please see the EOT metadata guide linked from this slide.

UNT PDF metadata project http://bit.ly/eot-metadata-guide

Issues and challenges with web archiving

Indescriminate + Unorganized = snapshot in time

Databases + dynamic content + robots.txt = oh my

Access and usability issues

Loss of provenance

National collection vs National haystack

Volunteer-run, no long-term institutional, funded support

Web archiving is of course not a perfect science. There are issues and challenges with Web archiving which make it also a "coping mechanism." I could spend an entire webinar talking about the well-known and well-documented issues with Web archiving. But suffice it to say that EOT project was a snapshot in time, done once every 4 years, so a LOT can and does change and disappear over that amount of time. The Web is also becoming much more dynamic, with databases, javascript and streaming content making it more difficult to crawl and collect. Besides the obvious technical difficulties inherent in Web crawling, perhaps the biggest issue is that this massive effort is done completely by volunteers and infrastructural donations, meaning there is no long-term, funded, institutional support for this critical effort.

https://freegovinfo.info/node/9559

https://www.cendi.gov/projects/Public_Access_Plans_US_Fed_Agencies.html

https://science.gov

The grassroots have clearly rallied around efforts at born-digital preservation. And we're beginning to see some movement within the government in that direction as well.

In the last few years, there are efforts within federal agencies which have started to take shape. For example, GPO is actively collaborating with library efforts and working toward a trusted digital repository audit for FDsys/govinfo.

GPO, NARA and LC and a couple of others have come together to form the Federal Web archiving working group to share information and best practices.

In 2013, the Office of Science Technology Policy (OSTP) published a memo entitled "Expanding Public Access to the Results of Federally Funded Research." This memo has resulted in the inter-agency CENDI federal STI managers group pushing for and publishing public access plans for 12 scientific agencies including the Depts of Defense, Agriculture, Education, Energy, Homeland Security, NSF, and NASA. Some but not all of these public access plans mention preservation as part of their efforts toward information access.

And there are a number of Federal agency portals coming online like science.gov.

But until there is a comprehensive strategy or strategies among and between the government and libraries to actively collect and preserve born-digital govt information,

there will continue to be loss and erosion of the national collection.

That brings us to our last segment, and I'll turn it over to Jim now.

Comprehensive strategy

Points to remember

- Short term access to digital masks long term problem.

- "Access" to static documents is no longer enough.

- Discovery, acquisition, and *functionality* all need to be tailored to communities of users.

- "Digital preservation is access ... in the future."

I hope by now that you agree with us that we need a comprehensive plan to preserve digital govinfo and to provide long-term, free public access to it.
Let's summarize what we know that can inform such a Comprehensive strategy.

Short-term and Long-term Strategies

- Librarians can use existing tools to preserve government information today.

- But we must also *lead* a movement for a long-term, comprehensive plan for the *life-cycle* of government information.

I know that "digital preservation" may sound big and scary to many of you. I know that your day-to-day responsibilities probably do not allow you much time to think of big issues and long-term strategies.

But that doesn't mean you are powerless.

So what can we do individually, every day?

First, we can use existing preservation tools. There are many that are easy to use and free or inexpensive.

This is necessary in order to complement and extend what GPO can do on its own with its limited mandate and inadequate funding.

Second, we can take the lead in building a movement for a long-term, comprehensive plan for the life-cycle of government information. There are lots of small steps we can take every day to do that.

We cannot assume that government agencies will do this on their own. They do not have the legal mandate or the resources to do so. And they lack the knowledge that we as librarians have of the needs of our different communities of users. We must drive this change!

Advocate for Information and Communities

- Librarians must be advocates of the information itself because of its inherent, long-term value -- regardless of the amount of use it gets.

- And we must be advocates for the communities that need this information.

- We can do both by building digital collections and digital services that meet the needs of our communities.

Indeed, librarians have a unique position that is different from that of the government agencies that produce information and from the communities that use that information.

We must be advocates of the information itself because of its inherent, long-term value -- regardless of the amount of use it gets.

And we must be advocates for the communities that need this information. Different communities use digital information differently. Each community looks for, and gets, and uses, and analyzes, and re-uses digital information it its own way. As librarians, we must do more than point at files that exist on government servers.

We can do both these things by building digital collections and building digital services for those collections: Services that meet the needs of the communities we support.

When we do this, we increase the value of the library to our users.

Short-term Strategies

- **Keep track** of your favorite agency's publications/data. Make sure those urls are in the Internet Archive's WayBack Machine.
- **Share** the fugitives you find with GPO and *lostdocs.freegovinfo.info*.
- **Save** documents to your library's web servers; upload them to the Internet Archive.
- **Build**: Digital collections that support the needs of communities you support.
- **Create DOIs**. Create and use Digital Object Identifiers for every Digital Object you control.
- **Create and re-use Metadata** in your DOIs, your library catalog, your lib-guides, in The Open Library, OCLC, the IA, Wikipedia.
- **Demonstrate Value**: Track and report what you learn.(e.g. Chesapeake Group http://cdm16064.contentdm.oclc.org/cdm/linkrot2015)
- **Join and participate**: PEGI, DLF, EDGI, Data Rescue.

Let's review some of the short term strategies that we can employ today.

We don't all have to use every one of these strategies.

But if every FDLP library employed one of these strategies every day, imagine how much further along we would be at the end of a single year! We could literally preserve hundreds of thousands of documents that might otherwise disappear.

Here are some:

• Keep track of your favorite agency's publications/data. Make sure those urls are in the Internet Archive's WayBack Machine.

• Share the fugitives you find with GPO and lostdocs.freegovinfo.info.

• You can Save digital documents.

If your library has an institutional repository or a digital assets management system or a digital preservation facility, you should investigate how you can use it to store and describe and provide access to government information that you can acquire and add to your local system.

• You can build Digital collections that support the needs of communities you support.

This means identifying communities of users that your library serves and identifying their needs and making sure that you are building digital services based on digital collections that you control.

• Create and use Digital Object Identifiers for every Digital Object you control.

DOIs have the wonderful capacity to point to more than one copy of the same digital object. So: you can point to the original copy, the IA copy, and a copy you acquire and

control.
•	You can Create and re-use Metadata: Put it in your DOIs, your library catalog, your lib-guides, in The Open Library, OCLC, the Internet Archive, and Wikipedia. We call this "seeding the cloud"!
Rich metadata is the key to providing digital services and functionality. every bit of metadata you create today is an investment in your future services to your communities.

•	You can Demonstrate the Value of what you are doing by Tracking and reporting what you learn.

The Chesapeake group has been preserving digital documents for its communities for years now and, in addition to doing so, it has reported regularly on how many of those documents are no longer available from the producer: either the original has changed, or been moved, or has been permanently deleted. Tracking this and reporting it helps library administrators and library users understand the difference between short-term access and long-term preservation and access; the difference between "pointing" and "collecting."
•	You can join and participate in any of the many groups that working hard to understand the need for preservation of government information and that are developing strategies for preservation.

Long-term Strategies

- **Support** GPO

- **Participate**: Build digital FDLP depository collections with *digital deposit*!

- **Join** LOCKSS-USDOCS. http://lockss-usdocs.stanford.edu

- **Act**: Build *collections* and *services* for *your communities*.

- **Help reform** Title 44. http://bit.ly/title44-petition

- **Advocate Policy reform**: OMB Information Management Plans (IMPs) http://freegovinfo.info/node/11741

- **Demand** that govt agencies produce Preservable Digital Objects (PDOs)

●Those short-term strategies will help a lot, but we also need long-term strategies.
●We need to think and plan and act in way that will change the whole life-cycle of government information from creation to use and re-use.
●• Support GPO
●

●Everything FDLP does will build on and complement what GPO does. If congress cuts gpo funding, or further commercializes its printing, or cuts its buget even more than it has so far, GPO will not be able to support FDLP.
●• If GPO will agree, and we can get Title 44 changed, we can start building digital FDLP depository collections with digital deposit!
●• supporting lockss is important because that support enables the maintenance of what it has already done and the expansion and further development of the LOCKSS software.
●• Even without changes to Title 44, you can Build your own digital collections and services for your communities.
●• You can Help reform Title 44. There is a link here to a petition that advocates many of the things we have described here today. the petition goes to the committee revising Title 44 and to GPO. If you haven't already, i urge you to consider signing it today.
●• We are trying to persuade the Office of Management and Budget to require Information Management Plans of all executive agencies. This could have as much impact on government-produced information as the policies mandating Data

Management Plans have had on government funded data collection.
●• agencies are, rightly, focused on services, but they are much less focused on preserving the content behind those services. we need to educate them on the long-term value of their content and the need to build their services on preservable digital objects which can be easily preserved by gpo or nara or libraries -- even after the services change or are discontinued.

In summary...

- Participate
- Learn
- Educate
- Advocate
- Lobby
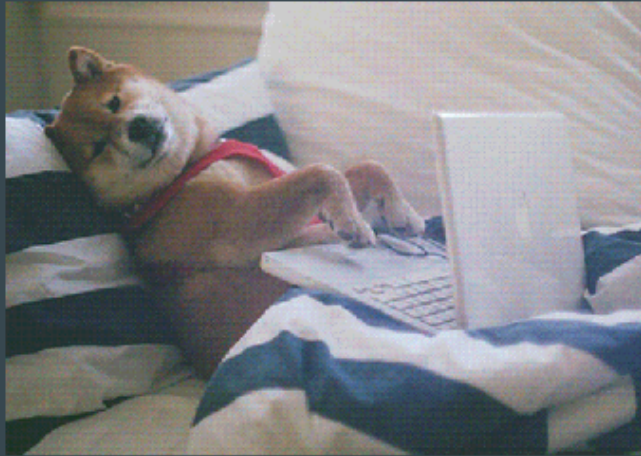
Thanks Jefferson Bailey for the cute dog GIF!

These lists of strategies are just examples of what we can do every day. There are others and new opportunities will arise.
To summarize, we recommend that you have a personal strategy that includes regular small activities that, together, will add up to big change.
You can:
•   Participate in long-term preservation and access.
•   Learn about the digital preservation and digital services...
•   ...so that you can educate producers and consumers of information and educate your own library managers and administrators.
•   Advocate for appropriate policies within your own library and within GPO and FDLP.
•   and Lobby for laws and regulations that will facilitate long-term preservation and access and not impede it.