The State of US Government Information: Toward a Sustainable Ecosystem

James R. Jacobs
jrjacobs@stanford.edu
freegovinfo.info
Govinfo Day 2018
Friday April 20, 2018
Agenda

1. scope/introduction
2. access breakdowns
3. projects of note
4. conclusion: building the govt information ecosystem
ACCESS BREAKDOWN POINTS

1. Libraries
2. Internet infrastructure
3. Political/economic issues
Scope of the problem

I. Libraries

1. Ignoring born-digital collections
2. Dismantling historic collections

Fugitives: When we depend on pointing instead of collecting
http://freegovinfo.info/node/3900

"Issued for Gratuitous Distribution' The History of Fugitive Documents and the FDLP"
https://freegovinfo.info/node/12735
"I spent a long time thinking about how to design a system for long-term organization and storage of subject-specific informational resources without needing ongoing work from the experts who created them, only to realized I'd just reinvented libraries." Digital Resource Lifespan (10/30/2017) https://xkcd.com/1909/ This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License.
II. INTERNET INFRASTRUCTURE
1. Link Rot

2. Content Drift

Dodging the memory hole. http://freegovinfo.info/node/10087
3. Lack of preservation infrastructure

GPO’s Govinfo public database https://govinfo.gov
III. POLITICAL/ECONOMIC ISSUES

What was first seen as an emerging trend in April 1981 when the American Library Association Washington Office first started this chronology of items which came to our attention, had by December 1987 become a continuing pattern of federal government to restrict government publications and information dissemination activities. A policy has emerged which is less than sympathetic to the principles of freedom of access to information as librarians advocate them. A combination of specific policy decisions, the Reagan Administration's interpretations and implementation of the 1980 Paperwork Reduction Act (PL 96-511, as amended by PL 99-500), implementation of the Grace Commission recommendations, and agency budget cuts have significantly limited access to public documents and statistics.

–Less Access to Less information by and about the US government [http://freegovinfo.info/less_access](http://freegovinfo.info/less_access)
1. The EPA Under Siege
2. Changing the Digital Climate
3. Removal of Climate Action Plans and Links from the National Park Service’s Climate Friendly Parks Program Website
4. Removal of Substantial Portion of Content and Reduction in Access to Resources on the USDA’s Climate Hubs Website

And more!

https://envirodatagov.org
Access requires ...

- Thoughtful collection and curation
- Systematic preservation
- Human expertise
- Understanding political economic context
PROJECTS OF NOTE
Raising awareness for the preservation of government information

https://www.pegiproject.org
https://digital.library.unt.edu/ark:/67531/metadc1115862/
United States Federal Government Documents Registry

The HathiTrust US Federal Government Documents Registry is a database of metadata representing the comprehensive corpus of U.S. federal documents produced from 1789 to the present. It is not yet complete.

For more detailed information on the Registry, see Creating a Registry of US Federal Government Documents.

https://www.hathitrust.org/usdocs_registry

Archive Scope

The End of Term Web Archive contains federal government websites (.gov, .mil, etc) in the Legislative, Executive, or Judicial branches of the government. Websites that were at risk of changing (i.e., whitehouse.gov) or disappearing altogether during government transitions were captured. Local government websites, or any other site not part of the federal government domain were out of scope.

http://eotarchive.cdlib.org
Volunteer contributions

End of Term Presidential Harvest 2016

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
EOT 2016 results

• ~300 TB data total!

  ■ ~110 TB web crawls + ~130 TB of gov ftp site archiving
  ■ 310,000,000 web URLs + 12,000,000 ftp files
  ■ 9,000+ social media accounts (scrape of gov SM registry API) 44% FB, 37% TW, 10% YT
Digital Federal Depository Library Program

The Digital Federal Depository Library Program, or the "USDocs" private LOCKSS network replicates key aspects of the United States Federal Depository System. The content is held in geographically distributed sites and replicated many times. Citizens have oversight and responsibility for the long-term care and maintenance of the content. All these characteristics mean the content will be preserved so that any alteration of the content (either deliberate or accidental) will be detected and repaired. For all documents, preservation in a tamper-evident environment is important, but for government documents, this is essential.

Our current focus is harvesting and preserving collections hosted on the The U.S. Government Publishing Office's (GPO) Federal Digital System (FDsys). See also, Digital Federal Depository FAQ

https://lockss-usdocs.stanford.edu
https://www.govinfo.gov
Public Policy: the source code of government

1. H.R. 4631: Access to Congressionally Mandated Reports Act
   https://www.govtrack.us/congress/bills/115/hr4631
2. Congressional Research Service (CRS) Reports
   https://www.everycrsreport.com
3. H.R. 5305 the FDLP Modernization Act
   https://www.govtrack.us/congress/bills/115/hr5305
1. Federal Web Archiving Group https://freegovinfo.info/node/9559
GOVT INFORMATION ECOSYSTEM
DILBERT

THE ELECTRONIC SUGGESTION BOX PROJECT IS HALF-WAY DONE.

THE ORIGINAL DESIGN CALLED FOR A BOX THAT SCANS AND DIGITIZES SUGGESTIONS WRITTEN ON PAPER AND EMAILS THEM TO THE APPROPRIATE MANAGER.

THEN THE DEVICE SHREDS THE ORIGINAL PAPER SUGGESTION TO MAKE ROOM FOR MORE.

I ALREADY BUILT THE BOX AND THE SHREDDER.

I’LL NEED ADDITIONAL FUNDING TO FINISH THE SCANNING PART.

WE DON’T HAVE ANY FLEXIBILITY IN OUR BUDGET. LET’S JUST DEPLOY WHAT YOU HAVE.

ALL I HAVE IS A BOX THAT SHREDS SUGGESTIONS BEFORE ANYONE READS THEM.

DON’T LET PERFECT BE THE ENEMY OF GOOD.
Guiding principles of government information

Information must be:

Not just *preserved*, but *discoverable*.
Not just *discoverable*, but *deliverable*.
Not just *deliverable* as bits, but *readable*.
Not just *readable*, but *understandable*.
Not just *understandable*, but *usable*.*

*Open Archival Information System (OAIS) principles

HT to Ranganathan's 5 laws of library science!
DCC Curation Lifecycle Model

 Already happening:
LOCKSS (C, P)
Hathitrust (A, M)
GOVINFO (O, A)
EOT (C, P)
Memento (A)
Cobweb (C)
Perma.cc (A)
DPLA (A, M)
Science.gov (O, A)

Needs:
Public Policy Funding!
DOI/Purl Versioning
(Git/wiki)
Open standards
Open APIs
Interoperability
Provenance
Redundancy
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.

“The only kinds of fights worth fighting are those you are going to lose because somebody has to fight them and lose and lose and lose until someday, somebody who believes as you do wins. In order for somebody to win an important, major fight 100 years hence, a lot of other people have got to be willing -- for the sheer fun and joy of it -- to go right ahead and fight, knowing you're going to lose. You mustn't feel like a martyr. You've got to enjoy it.”

--I.F. Stone
THANKS!