March 17, 2014

BORN-DIGITAL U.S. FEDERAL GOVERNMENT INFORMATION: PRESERVATION AND ACCESS

PREPARED FOR THE CENTER FOR RESEARCH LIBRARIES
GLOBAL RESOURCES COLLECTIONS FORUM

BY JAMES A. JACOBS

CENTER FOR RESEARCH LIBRARIES
GLOBAL RESOURCES COLLECTIONS FORUM
April 24-25, 2014

LEVIATHAN

Libraries and Government Information
in the Age of Big Data

CONTENTS

BORN-DIGITAL U.S. FEDERAL GOVERNMENT INFORMATION

PRESERVATION AND ACCESS

I. INTRODUCTION

Libraries, and more specifically depository libraries, and, most importantly, Federal Depository Library Program (FDLP) libraries, have successfully preserved an important part of the public record of our democracy for 200 years (McGarr). Although some librarians have questioned whether or not preservation was either an intentional goal of the FDLP or an objective of the participating libraries (Shuler 2004), it is undeniable that the Program has successfully preserved millions of volumes, even if that was a byproduct of other intentions.

But the migration of government information from print to digital has introduced new problems into the challenge of preserving government information. Very little government information is being deposited in FDLP libraries. In 2013 the Government Printing Office (GPO) estimated that 97% of federal government information was born-digital and current GPO policy limits FDLP deposit of digital information to so-called "tangible" objects such as CD-ROMs and DVDs (GPO 2006), which create their own preservation problems (Gano). While libraries played an essential role in preservation of government information in the print era, most born-digital government information is not held, managed, organized, served, or preserved by libraries.

GPO's own role in preservation has changed over time. In the print era, GPO was able to rely on FDLP libraries for preservation. It even relinquished its role of preserving print entirely at one point, turning its print collection over to the National Archives (Russell). During the early years after the passage of the *Government Printing Office Electronic Information Access Enhancement Act of 1993,* GPO attempted to assume sole responsibility of preserving born-digital information in its purview (GPO 2007, 2009). In the last few years, GPO has actively embraced preservation partnerships inside and outside the government (USDocs private LOCKSS network, GPO 2014).

GPO's preservation activities are, today, overwhelmingly focused on Congress. Although GPO provides no statistics on the quantity of its holdings, more than half of the "FDsys Collections" are explicitly Congressional (see Appendix A).

GPO began including federal court opinions in FDsys in a pilot project in 2011 and has expanded that project to include more than 600,000 opinions of some, but not all, federal appellate, district, and bankruptcy courts, dating back to 2004 (GPO 2011, Administrative Office of the U.S. Courts, GPO United States Courts Opinions).

Although there are no precise statistics available, it is clear that an increasing amount of government information from executive agencies that would have once been routinely routed through GPO to FDLP libraries is now not even gathered by GPO (Koontz). GPO's official purview is limited by U.S. Code Title 44, by the Paperwork Reduction Act, and by Office of Management and Budget directives that allow executive agencies to avoid complying with even the limited scope of Title 44 (GAO).

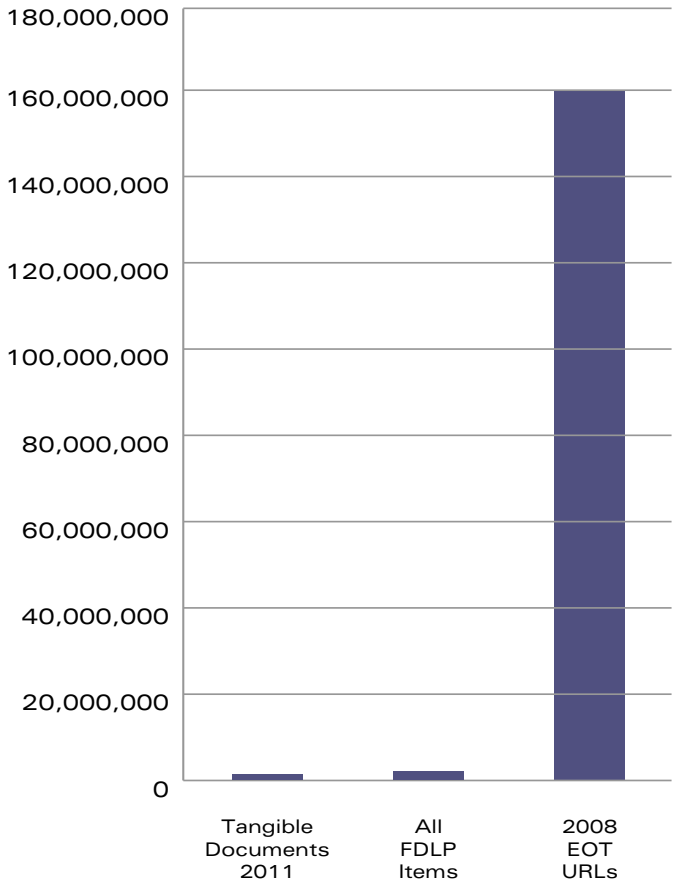
While a few information-producing executive agencies such as the Energy Information Agency (EIA) see information preservation as part of their mission (Johnson), there is rarely if ever, an

explicit legislative requirement for agencies to preserve that information. Legislative preservation requirements for agencies are largely limited to "Records" (not publications or websites). Definitions of Records is subject to interpretation by both NARA and the agency and charging fees for access to records is explicitly permitted by legislation. (See, for example, the actual legislative requirements for the EIA in the Department of Energy Organization Act and 5 U.S. Code § 552).

Even GPO's legislative mandate is limited to providing "an electronic storage facility" and "a system of online access" without any explicit mention of historical documents or long-term preservation (44 USC 41).

II. SCOPE OF THE PRESERVATION CHALLENGE

The size of the digital-preservation challenge is difficult to measure for a number of reasons, but the scale can be illustrated by comparing the 10,200 items distributed by GPO to FDLP libraries in one year (GPO 2012), to all 2.3-3 million items estimated to be held in the Federal Depository Library Program (Burger, Peterson), to the 160 million URLs harvested in the 2008 End of Term Crawl (Hartman).



Although the above figures are raw and subject to interpretation, we can certainly conclude that the *production* of born-digital government information is very, very much greater than the earlier production of printed government information. One might reasonably estimate that there are more born-digital government information items produced *in a single year* than all the two or three million non-digital government information items accumulated in the FDLP *over 200 years*.

This leads to many questions such as: What portion of born-digital government information is being preserved? How much government information evades web-harvesting because it is buried in the deep web? Do large-scale web harvesting projects such as the End of Term Crawls and the work done by the Internet Archive adequately capture the information output of the federal government?

In the following sections we will examine the scope of the challenge of preserving born-digital government information, look at some well-known preservation projects, and draw some conclusions.

DEFINING GOVERNMENT INFORMATION

The first step in identifying the scope of the challenge of preserving born-digital government information is to define where it is, who is producing it, and how much of it there is. Unfortunately, there is no comprehensive directory or index or catalog of born-digital government information. The simple fact is that no one knows how much born-digital U.S. Federal government information has been created or where it all is. The very nature of born-digital information also raises other questions. In the print era, it was relatively easy to distinguish between a "document" intended for the public (because it was published in bulk, often by or through GPO) and "records" of government, which were usually not published or made publicly available. Today, that distinction is blurred since agencies can easily "publish" information to the web without the expense involved in printing and binding and distributing paper. At the same time many public records of government are stored in databases that may not be directly or completely available to the public.

FINDING GOVERNMENT INFORMATION ON THE WEB

One might define government information as that information that the government posts on its public websites. But there is not even a list of all government websites, and certainly no list of all the information posted on those sites.

There are counts of internet .gov and .mil domains (like state.gov), but a domain is an internet address and that address may or may not host a website. After 9/11, the General Services Administration claimed that releasing a list of government domains presented a security risk (Claburn). One frequently cited estimate says that there are approximately 2,000 top-level federal .gov domains and an estimated 24,000 websites (Phillips). This figure is, however, based on an incomplete and narrow measurement of the government web. A more comprehensive estimate is closer to 16,015 government and military domains and 135,215 websites with government information.

The most official count we have, from the U.S. General Services Administration, is the new periodic listing of "Federal Executive Agency Internet Domains." The February 2014 edition lists 1,228 domains, all of which are two-level addresses (e.g., americorps.gov). This count is

a bit misleading, however, because it includes more than 300 addresses that only redirect to actual sites in the list (e.g., from americore.gov to americorps.gov). It is also incomplete because there are obvious omissions. For example, although the list includes 21 State Department two-level domains (e.g., state.gov, usconsulate.gov), it does not include three-level domains such as keystonepipeline-xl.state.gov. It is also limited to .gov domains, excluding .mil and other top-level domains (e.g., .org, .edu) where some government agencies (goarmy.com) and quasi-government agencies (si.edu) reside.

The 2013 *United States Government Manual* provides another incomplete official list. It lists 246 agencies and 527 unique web addresses.

The January 2014 Internet Systems Consortium (ISC) domain survey (which surveys domain addresses -- essentially any machine on the Internet, not just publicly accessible websites), found 2,050 two-level .gov domains (twice the official count above) and 678,254 three-level .gov domains. The survey also found 191 two-level and 104,295 three-level .mil domains.

A baseline inventory of Federal Executive Branch websites in 2011 found that nearly a fifth of federal .gov domains had gone inactive and that agencies reported plans to eliminate most of the non-functioning domains. Several agencies reported that they did not know the answers to basic questions about their inventory of websites. Nearly all agencies reported that decision-making with regard to specific domains/websites happens within operating units and not at an agency level. (.gov Reform Task Force).

The most accurate count we currently have is probably from the 2008 "end of term crawl." It attempted to capture a snapshot "archive" of "the U.S. federal government Web presence" (Hartman) and, in doing so, revealed the broader scope of the location of government information on the web. It found 14,338 .gov websites and 1,677 .mil websites. These numbers are certainly a more comprehensive count than the official GSA list and more accurate as a count of websites than the ISC count of domains. The crawl also included government information on sites that are not .gov or .mil. It found 29,798 .org, 13,856 .edu, and 57,873 .com websites that it classified as part of the federal web presence. Using these crawl figures, the federal government published information on 135,215 websites in 2008.

DEFINING BORN-DIGITAL GOVERNMENT INFORMATION

The difficulties in accurately and comprehensively identifying even the sources of government information is a well known problem. It was well described more than ten years ago in the California Digital Library report, *Web-Based Government Information: Evaluating Solutions for Capture, Curation, and Preservation* (Cruse). That report said that "the domain of web-based government information is hard to define, constantly expanding, and highly volatile" and that "a high percentage of the content is hidden within the deep web."

A combination of insecure funding, changing political priorities, and information embedded deeply in websites and sometimes hidden from search engines and harvesting-robots in the "deep web" of databases behind websites can cause information to be lost. One example of this was the closing of the National Biological Information Infrastructure (NBII) in 2012 (U.S. Geological Survey). Its website and its associated node sites were all shut down and CyberCemetery was not able to capture all of its data. More recently, the USGS announced that its National Atlas of the United States will be removed from service in September and some of its services will no longer be available. GPO has captured "the content" but says that some of the functionality may be lost (Diaz).

In addition to web-based born-digital information, the government has also released DVDs, CD-ROMs and even floppy disks. Cruse reported in 2003 that between 1995 and 2002 GPO distributed 4,890 titles on such disks. The usability of these disks today is unknown.

Tables 1 - 4 present other counts of government websites. Some surveys count "domains," others count "websites." These different ways of counting can obscure both quantity and complexities of information production since a single domain, like state.gov, could actually have many different websites (like keystonepipeline-xl.state.gov).

The ambiguity in these various counts demonstrates more than the absence of a definitive catalog of born-digital government information and its creators. It also shows the difficulty of defining what to preserve. Cruse noted that "Dot-gov web page boundaries are often ambiguous, with links directly to external content and quasi-governmental sites" and that "Many government sites do not use the dot-gov domain, or have such an ambiguous status that it isn't clear if they are a government entity." Additionally, defining what is in scope for preservation is more subjective than objective:

> *[I]t is unlikely that any one definition of the government domain is ever likely to be agreed upon by those who set out to archive it.... [O]rganizations that preserve web-based government information do so under a variety of very different circumstances. They serve different audiences and exist in very different political, financial, and technical regimes. Together, these influences shape very different selection policies about what to collect (Cruse).*

Such variety in scope is a strength of libraries, not a weakness, but it complicates our ability to determine a single scope of born-digital government information.

GOVERNMENT INFORMATION VS. GOVERNMENT SERVICES

The preservation challenge is complicated further as the federal government moves toward an e-government model of information-as-service. E-government is a *service* (Shuler 2010). As a service, it is like the gate at a national park that protects the resource and provides access to it — but is not the resource itself. Government information is a *resource* like a national park. Information services can make information more easily accessible by individuals (Marks), but limits access to the whims of government (Shuler 2014). When the underlying information resource is not available for preservation outside the government, preservation of the resource is also left to the whims of government and to the subjective determination by government alone of what is worth preserving.

Although preservation activities have always focused on the resource rather than the service itself, the proliferation of e-government creates two important preservation issues. First, it makes it more difficult to determine if the information resource is being preserved. It may even make it difficult for those outside the government to evaluate the resource to determine its value and the need to preserve it. It also may make the raw information unavailable for preservation outside the government agency. These are increasingly important questions to answer because the service is visible to the public, but the information resource is not. Second, some librarians believe that the existence of government information services makes it unnecessary for libraries to have collections of the government information resources (Shuler 2010, Rossmann).

TABLES

Tables 1 through 4 list different kinds of counts of the government on the web dating from 1997 to 2014. These show different ways to estimate the scope of the preservation challenge. Sources for the tables are listed on p. 17.

Table 1. Counts of Top-level Government Domains, 1997-2014

NUMBER	KIND	YEAR	SOURCE
4244	Federal web sites	1997	GAO
1,585	.gov domains	2001	registrar.nic.gov
1,952	.gov domains	2002	registrar.nic.gov
2,410	government websites	2009	End of Term 2008-09
4,541	seeds crawled	2013	End of Term 2012
5,957	.gov websites	2013	Stanford Univ.
1,228	Executive Agency Domains	2014	data.gov

Table 2. Distribution of Government information by Top-Level Domain, 2009, 2013

TLD	2009	2013
.gov	14,338	7,019
.mil	1,677	945
.org	29,798	12,798
.edu	13,856	5,599
.com	57,873	36,309
.net		2,601
.us		1,413
other		3,528
total	135,215	70,212

Table 3. Two-level and three-level government domains, 2012

35,424	2-level domains
58,912	3-level domains

Table 4. Government Information: URL counts 2009, 2013

160,211,356	2009
32,837,215	2013

III. PRESERVATION ACTIVITIES

There is no central registry or directory or catalog of preserved information, or preservation activities or projects, or websites that preserve born-digital U.S. Federal government information. There is no standard for reporting the scope and coverage or contents of projects. The data that we do have varies from project to project in its detail and metrics of reporting.

This section provides descriptions of each of several well-known digital preservation projects. The individual projects are grouped into somewhat arbitrary categories by size and method.

One problem in analyzing the information we have about preservation activities is the lack of a clear and consistent unit of measurement of preservation. Ideally, producers would instantiate information in preservable packages of some kind, either an end-user package, such as a PDF file, or a package intended to be preserved by an archive, such as a Submission Information Package (Consultative Committee for Space Data Systems). When they don't, we are left with inconsistent units of measure (PDFs, URLs, HTML pages, databases, files, etc.). A "website" can have its own domain (Deserttortoise.gov), or it can be subsumed under a two-level domain name (flu.gov/pandemic), or it can have a three-level domain name: (canada.usembassy.gov). It may contain a few pages or thousands of documents. What we might have once called a "title" or a "book" might now exist on the web as a single PDF file, or as several PDF files, or as an HTML page with multiple URLs of images, or as dynamically-created responses to user-queries, or as deep-web entries in a database, or any combination of these. A single HTML "page" may include many URLs (e.g., for images to display on that page); it may have nothing more than links to the actual information (PDFs, other pages) or be a PDF itself. Some archives store (and count) every individual URL separately; others bundle multiple URLs that comprise a single "page" into archival packages such as WARCs. Providing useful metrics in this environment is difficult at best and almost always inconsistent.

Scope Note: This section focuses on projects that preserve born-digital U.S. Federal Government information. There are other projects that focus on preserving digitized paper documents (Digitization Projects Registry); state documents (Inventory of Projects Preserving State Government Information); government information of other countries (International Internet Preservation Consortium), and related activities. Some of the projects listed below have multiple missions, but the annotations describe only the preservation of born-digital U.S. Federal Government information.

A. GOVERNMENT REPOSITORIES

FDsys

http://www.fdsys.gov
Institutions: Government Printing Office
Scope: Congress (bills, hearings, laws, etc.). Presidential documents, Federal Register, GAO reports, Supreme Court decisions. 151 "government authors."
Dates: 1994- . Some older documents
Size: unknown
Functionality / Access: Search and browse and download of individual documents and (for a few collections) bulk download in XML format.
Integrity: Designed to be OAIS compliant.
Partnerships: LOCKSS-USDOCS. NARA.

NARA Electronic Records Archive

http://www.archives.gov/electronic_records_archives/index.html
Institutions: NARA, San Diego Supercomputer Center
Scope: Electronic records. May include web sites, but also includes institutional "records" in addition to "documents." Includes records from the George W. Bush White House, many Federal agencies, and Congress.
Dates: Initiated 2005
Size: As of January 2012: over 131 TB
Functionality / Access: Some, but not all searchable with the Online Public Access (OPA) system.
Integrity: unknown
Partnerships: NARA has established partnerships with other organizations such as the and the Government Printing Office and the University of North Texas' CyberCemetery. These partnerships allow the partners to preserve and provide access to government information while NARA retains the responsibility for legally accessioning the records as part of the Archives.
Notes: Uses separate systems to preserve different types of records and the processes and documentation required for each type.

B. REPLICATIONS OF GOVERNMENT REPOSITORIES

Bulk.Resource.org

https://bulk.resource.org/gpo.gov/
Institutions: Public.Resource.Org
Scope: Replicates in bulk selected troves of government information. Includes a replica of GPO Access (predecessor to FDsys). Also: archived or in progress: house.gov, law.gov, uscourts.gov, uspto.gov, change.gov, copyright.gov, gao.gov, gpo.gov, justice.gov, ntis.gov, sec.gov, si.edu
Dates: varies
Size: example: 5,177,003 PDFs from GPO Access
Functionality / Access: Mostly bulk access only. No search
Integrity: unknown

USDocs private LOCKSS network

http://lockss-usdocs.stanford.edu
Institutions: 36 institutions provide a replication of GPO's Federal Digital System (FDsys).
Scope: 44 collections in FDsys
Dates: 1994- . Some older documents
Size: 1.3TB
Functionality / Access: dark archive, there are plans to make the archive more publicly available.
Integrity: Designed to be OAIS compliant.
Partnerships: GPO, Library of Congress, various FDLP libraries.

C. LARGE-SCALE WEB-HARVESTS

Internet Archive

https://archive.org/details/USGovernmentDocuments
Institutions: The Internet Archive (IA)
Scope: Unknown. Broad mandate to crawl the web.
Dates: 1996-
Size: [totals unknown. Checking 924 "Agency Internet Domains as of 02122014" using the IA API and found 12.5% of those URLs *not* in IA.]
Functionality / Access: URL-access. Some search.
Integrity: unknown

NARA End of Term Crawls

http://www.webharvest.gov/
Institutions: National Archives and Records Administration (NARA)
Scope: NARA has initiated and been a partner in creating one-time snapshots of agency public websites at end of Congressional terms. Two- and three-level federal .gov and .mil.
Dates: Agency crawls: 2001 and 2004. Congressional web sites in 2006, 2008, 2010 and 2012.
Size: example: 2004: about 75 million web pages.
Functionality / Access: URL and URL search. Apparently not indexed by Google.
Integrity: unknown.
Notes: In January 2005, NARA issued "Guidance on Managing Web Records," which addresses agencies' responsibilities for identifying, managing and scheduling web materials they identify as Federal records. Accordingly, each agency is now responsible, in coordination with NARA, for determining how to manage its web records, including whether to preserve a periodic snapshot of its entire web page.

2008 End of Term Crawl Project

http://eotarchive.cdlib.org/index.html
Institutions: Library of Congress, Internet Archive, California Digital Library, University of North Texas, U.S. Government Printing Office.
Scope: Public U.S. Government Web sites at the end of the presidential administration. Also: intended to document federal agencies' presence on the Web during the transition of Presidential administrations.
Dates: 2008-2009
Size: 160 million URIs from 3,300 websites. 16TB
Functionality / Access: Browse and Search
Integrity: Unknown.

2012 End of Term Crawl

http://crawls.archive.org/collections/eot2012/
Institutions: California Digital Library, Internet Archive, Library of Congress, University of North Texas, GPO.
Scope: any U.S. Federal Government domains
Dates: 2012-2013
Size: 32 million web pages. 12 TB
Functionality / Access: URL browsing
Integrity: Unknown.

D. FOCUSED COLLECTIONS OUTSIDE THE GOVERNMENT

Archive-It

https://archive-it.org/explore?show=Collections&fc=meta_Subject%3AGovernment-usfederal

Institutions: Internet Archive provides web harvesting and collection-building services for a fee.
Scope: 112 collections focus on U.S. Federal Government
Dates: mostly 1991-
Size: About 4500 files. Size of collection varies from 1 to more than 500 files.
Functionality / Access: browse and search
Integrity: unknown
Partnerships: 19 organizations use this service including GPO, U.S. Department of Health and Human Services, San Francisco Public Library, Stanford University Libraries, and the Wisconsin Historical Society

California Digital Library Web Archiving Services

http://webarchives.cdlib.org/archives

Institutions: University of California
Scope: Three collections focus on federal government information: Federal Regional Agencies in California Web Archive, USDA Economic Research Service, and USDA ERS Publications.
Dates: 2011- 2014
Size: 15 websites
Functionality / Access: browse, search, filter by filetype
Integrity: unknown
Partnerships: subscribing institutions
Notes: Web Archiving Service (WAS) subscribing institutions can build collections using the CDL WAS tools.

CyberCemetery

http://digital.library.unt.edu/explore/collections/GDCC/

Institutions: University of North Texas, GPO
Scope: Government agencies that have ceased operation (usually websites of defunct government agencies and commissions that have issued a final report).
Dates: 1990-
Size: 96 collections
Functionality / Access: Full text search. Browse
Integrity: unknown.

E. OTHER SPECIALIZED COLLECTIONS

Census 2000

http://library.case.edu/ksl/census/
Institutions: University Library of Case Western Reserve University, Census Bureau
Scope: Census 2000 data issued by the Census Bureau in comma-delimited ASCII format.
Dates: 2000 Census
Size: Census "Summary Files" 1-4 and Redistricting Data. multiple files for each state.
Functionality / Access: download compressed files
Integrity: unknown

CIC Floppy Disk Project

http://www.indiana.edu/~libgpd/mforms/floppy/floppy.html
Institutions: Indiana University-Bloomington Libraries and GPO on behalf of the Committee on Institutional Cooperation (CIC).
Scope: publications that were distributed to federal depository libraries on floppy disk.
Dates: roughly 1980-1996
Size: 117 titles, 335 "zip" archives, nearly 8000 files.
Functionality / Access: browse by title, download files.
Integrity: unknown

Cornell Legal Information Institute

http://www.law.cornell.edu/lii/get_the_law/our_legal_collections
Institutions: Cornell University with publishers, legal scholars, computer scientists, government agencies, and other collaborators.
Scope: Extensive collections of legal information, including Federal law, Constitution, U.S. Code, Code of Federal Regulations, Supreme Court, Federal Rules. LII's mission is "to ensure that the law remains free and open to everyone."
Dates: current, with some collections 1990-
Size: unknown
Functionality / Access: browse and search
Integrity: unknown

Department of State Foreign Affairs Network (DOSFAN)

http://dosfan.lib.uic.edu/ERC/
Institutions: Richard J. Daley Library, University of Illinois at Chicago and the Department of State.
Scope: U.S. Department of State, U.S. Arms Control and Disarmament Agency, and the U.S. Information Agency.
Dates: 1990 through 1997
Size: unknown
Functionality / Access: browse
Integrity: unknown
Partnerships: GPO

Library of Congress Minerva, September 11, 2001, Web Archive

http://lcweb2.loc.gov/diglib/lcwa/html/sept11/sept11-overview.html
Institutions: Library of Congress
Scope: Includes U.S. and non-U.S. government sites; press, corporate/business, portal, charity/civic, advocacy/interest, religious, school/educational, individual/volunteer, professional organizations sites; and other sites.
Dates: September 11, 2001- December 1, 2001
Size: 2,313 websites
Functionality / Access: search and browse
Integrity: unknown
Notes: This is an example of a collection that includes, but is not limited to U.S. Government websites.

OpenCRS

https://opencrs.com/
Institutions: Center for Democracy and Technology
Scope: Congressional Research Service Reports
Dates:
Size: estimate: more than 19,000 reports
Functionality / Access: full text search.
Integrity: unknown

USDA Economics, Statistics and Market Information System (ESMIS)

http://usda.mannlib.cornell.edu/MannUsda/
Institutions: Albert R. Mann Library at Cornell University and several agencies of the U.S. Department of Agriculture.
Scope: U.S. and international agriculture and related topics. Most reports are text files that contain time-sensitive information. Most data sets are in spreadsheet format and include time-series data that are updated yearly.
Dates: “current and historical data”
Size: 2500 reports and datasets
Functionality / Access: search and browse.
Integrity: unknown
Partnerships: agencies: http://usda.mannlib.cornell.edu/MannUsda/aboutAgency.do

IV. CONCLUSIONS

SCOPE OF CHALLENGE

Just as ten years ago when the California Digital Library report was written (Cruse), the extent of born-digital government information is still hard to define, constantly expanding, and highly volatile. Congressional information is relatively well and redundantly preserved (FDsys, LOCKSS-USDOCS) but preservation of executive agency information varies widely. Most government born-digital information is in dire straits of being lost and some is being preserved in a relatively stable and consistent, if imperfect, way. More than ever before, most born-digital information fits into that broad category of what was once called "fugitive documents" --documents that do not go through GPO--and so are not in FDLP libraries, or FDsys, or LOCKSS-USDOCS and are therefore at greatest risk of being lost.

Although there is much that we do not know, we can draw three general conclusions:

- **We lack adequate means to identify and measure what is being produced and what is being preserved.**
- **Using the measurements we do have, the scope of born-digital government information being produced far outpaces what is being preserved.**
- **We do not have a unified approach to identifying and preserving born digital government information.**

SCOPE OF ISSUES

The community's experience in digital preservation also allows us to characterize some of the key preservation issues.

- **Versioning**
  Although some government information is intended to be static, the nature of digital information makes it easy to change. Such changes may be intentional or unintentional, substantive or not. They may motivated by politics or policy or economics. Being able to identify and preserve different versions of "documents" over time is important -- both in order to preserve unique content, and in order to minimize preserving the same content many times unnecessarily.
- **The need for persistent URLs**
  The extent of "link-rot" has been well documented and increases over time. Link-rot does necessarily mean that the information is not preserved (it may be that a document has simply been moved to a new URL). But some link-rot is attributable to information that is no longer available. Link-rot is always an indication that the information is, at best, harder to find and identify (Chesapeake Digital Preservation Group, Zittrain). Adequate preservation could help solve the problem of missing information and appropriate techniques could ensure that broken URLs redirect to preserved copies.

- **The need for temporal context**
  This relates both to versioning and link-rot. Providing temporal context means preserving the context of a document at the time it was created. Discreet digital "documents" (and most web-based information) often rely on, or refer to, other digital information. Digital information also changes over time (through link-rot, and what Herbert Van de Sompel calls "content decay"), which means that the document that you refer to today may not be the document at that URL tomorrow. Users of preserved information need a way of referring to, locating, and using information in its original context (Ainsworth).

- **E-government issues.**
  The move to e-government delivery of information provides many potential barriers to preservation. The information resource target-of-preservation may not be available for preservation actions to anyone outside the agency. The information may be also stored in databases, which have their own preservation challenges. Information stored in databases may be updated without preserving previous versions. If libraries do not preserve information but choose instead to point to e-government services and rely on the government to deliver information "just in time," there will be no way of guaranteeing against information loss.

- **Fragility of relying on government for preservation and free access**
  As we saw most notably in the recent government shutdown, when government is the only source of information, government decisions affect the availability of information (Shuler 2014). Even GPO, which has a primary mission of providing access to government information, cannot guarantee long-term preservation. Although the current GPO administration operates with the intention of preserving and making information freely available (Vance-Cooks), this has not always been the case (James). Title 44 section 4102 of the U.S. Code specifically authorizes GPO to charge fees for access. As recently as 2013, the National Academy Of Public Administration recommended instituting fees for access in order to fund digital preservation. In 2012, the Congressional Research Service noted that Title 44 is "silent on GPO's retention and preservation responsibilities for digital information" (Petersen).

- **Selection.**
  As noted above, part of the challenge of preserving born-digital government information is created by the fact that different people will (legitimately) define the scope of what needs to be preserved differently. This can be seen as an opportunity. It is a strength of libraries to be able select information for their own designated user communities and build collections that fit the needs of those communities. If libraries rely only on issuing agencies to preserve their own information, they will be relinquishing to those agencies the decision as to what is worth preserving.

- **Collections need Services**
  As Paul Conway pointed out, "In the digital world, the concept of access is transformed from a convenient byproduct of the preservation process to its central motif." Selecting and preserving bits is only the first of many steps. Organizing and describing those bits and making them discoverable and usable is an essential component of preservation. In the twenty-first century, just dumping a web crawl into WARC files will increasingly be seen as a very primitive service. While providing advanced services for preserved content adds expense to projects, it also adds value to the information preserved and, in turn, to the library providing the services. This can be seen as an opportunity for libraries to provide services that the issuing agency either cannot (because of statutory limitations) or does not provide.

SCOPE OF SOLUTIONS

Because there is still no catalog of what born-digital government information is being preserved, the extent to which it is being preserved is largely unknown. Some of the preservation projects themselves are part of the unindexed deep-web.

MODELS OF PRESERVATION

The existing preservation projects do provide at least three models for preserving born-digital government information.

1. **Government assumes sole responsibility for preservation.**

   Examples: NARA, some preservation-focused agencies.

   Advantages: Closest to the information. Costs are born by the government, not libraries.

   Disadvantages: Scope of preservation is defined by government information creators, not user-communities. Responsibility for preservation, sustainability, and succession planning (Center for Research Libraries, 2007) resides in a single institution, putting information at risk of technological, economic, or political loss or alteration (intentionally or unintentionally).

2. **Government partnership with non-government institutions.**

   Examples: GPO / LOCKSS-USDOCS, GPO / CyberCemetary

   Advantages: Varies with partnership. Can, potentially, provide increased security, redundancy, increased avenues of access, more reliable sustainability and succession planning.

   Disadvantages. Potential advantages are not guaranteed. If only a single-institution holds the information, this model has the same disadvantages as government assuming sole responsibility.

3. **Non-government projects.**

   Examples: Internet Archive.

   Advantages: Can be focused or broad; several institutions can work cooperatively. Does not require government agency approval or participation.

   Disadvantages. Difficult to be accurate, complete. Cannot easily keep up with rapidly changing web content.

METHODS OF SELECTION

The projects listed in section III above demonstrate three different methods of selecting government information for preservation.

1. **Broad web harvesting.**
   Examples: Internet Archive and End of Term Crawls.

2. **Focused selection**
   These can be very specific one-document-at-a-time selections, such as the Chesapeake project and the Stanford Everyday Electronic Materials project (Kott), or broader projects that select entire agencies, like the CyberCemetery, or focused web harvests like the crawls defined in Archive-It and WAS.

3. **Digital Deposit.**
   These are characterized by partnerships between government agencies and non-government memory organizations and involve the agency actively transferring content to the organization for the explicit purpose of preservation. The most notable example is the USDOCS-Lockss project.

SOURCES FOR TABLES

Table 1

- GAO: http://www.gao.gov/archive/1997/gg97086s.pdf
- registrar.nic.gov 2001 https://web.archive.org/web/20100513035720/http://www.thememoryhole.org/govt/dotgov-domains2.htm
- registrar.nic.gov 2002 https://web.archive.org/web/20100513033403/http://www.thememoryhole.org/govt/dotgov-domains.htm
- End of Term 2008-09 http://eotarchive.cdlib.org/search?browse-all=yes
- End of Term 2012 http://wbgrp-svc040.us.archive.org/collections/eot2012/stats/
- Stanford Univ. http://diglib.stanford.edu:8091/~testbed/doc2/WebBase/site_lists/gov-03-2013.tx
- data.gov https://explore.data.gov/d/ku4m-7ynp

Table 2

- End of Term 2008-09 http://research.library.unt.edu/eotcd/w/images/8/8d/LG-06-09-0174-09_UNT_Feb2013_FINAL.pdf
- End of Term 2012 http://wbgrp-svc040.us.archive.org/collections/eot2012/stats/Fall_2012/FINAL/hosts-report.txt

Table 3

- End of term 2012 http://wbgrp-svc040.us.archive.org/collections/eot2012/stats/

Table 4

- End of Term 2008-09 http://research.library.unt.edu/eotcd/w/images/8/8d/LG-06-09-0174-09_UNT_Feb2013_FINAL.pdf
- End of Term 2012 http://wbgrp-svc040.us.archive.org/collections/eot2012/stats/

APPENDIX A

FDSYS COLLECTIONS
HTTP://WWW.GPO.GOV/FDSYS/BROWSE/COLLECTIONTAB.ACTION

CONGRESSIONAL COLLECTIONS

Congressional Bills | XML Bulk Data (House)
Congressional Calendars
Congressional Committee Prints including Ways and Means Committee Prints
Congressional Directory
Congressional Documents
Congressional Hearings including House and Senate Appropriations Hearings
Congressional Pictorial Directory including New Member Pictorial Directory
Congressional Record (Bound)
Congressional Record (Daily)
Congressional Record Index (Daily)
Congressional Reports including Conference Reports
History of Bills
House Practice
House Rules and Manual
Independent Counsel Investigations
Journal of the House of Representatives
Precedents of the U.S. House of Representatives
Public and Private Laws
Riddick's Senate Procedure
Senate Manual
United States Code
United States Statutes at Large

OTHER COLLECTIONS

Additional Government Publications
Budget of the United States Government
Bulk Data
Coastal Zone Information Center
Code of Federal Regulations | XML Bulk Data
Commerce Business Daily Bulk Data
Compilation of Presidential Documents
Constitution of the United States of America: Analysis and Interpretation
Economic Indicators
Economic Report of the President
Education Reports from ERIC
Federal Register | XML Bulk Data | FR 2.0
GAO Reports and Comptroller General Decisions
Internal Revenue Cumulative Bulletin to the Treasury Department
List of CFR Sections Affected
Privacy Act Issuances
Public Papers of the Presidents of the United States | XML Bulk Data
Supreme Court Decisions (FLITE) Bulk Data
Treasury Department
United States Courts Opinions
United States Government Manual | XML Bulk Data
United States Government Policy and Supporting Positions (Plum Book)

END NOTES

.gov Reform Task Force. 2011. State of the Federal Web Report. http://www.usa.gov/webreform/state-of-the-web.pdf.

Administrative Office of the U.S. Courts. "About CM/ECF" http://www.uscourts.gov/FederalCourts/CMECF/AboutCMECF.aspx

Ainsworth, Scott G. 2013. "Browsing and Recomposition Policies to Minimize Temporal Error When Utilizing Web Archives." Bulletin of IEEE Technical Committee on Digital Libraries 9 (2). http://www.ieee-tcdl.org/Bulletin/v9n2/papers/ainsworth.pdf.

Brown, Geoffrey. 2007?. "Virtualizing the CIC Floppy Disk Project: an Experiment in Digital Preservation Using Emulation." http://www.cs.indiana.edu/~geobrown/jcdl.pdf

Burger, John. 2013. Federal Documents as Content for DPLA? Association of Southeastern Research Libraries (ASERL), DPLA Content and Scope Workstream

(February 28, 2013) Washington, DC https://docs.google.com/presentation/d/1bK4Vb9fFQxInIsmwKvQplnoP_fHMxg-8mSryaBYXuMs/edit?pli=1#slide=id.p17

Center for Democracy and Technology. 2007. "The Importance of OpenCRS." https://www.cdt.org/blogs/ross-schulman/importance-opencrs

Center for Research Libraries. 2013. Wayback Machine [review]. *EDesiderata*. http://edesiderata.crl.edu/resources/wayback-machine

Chesapeake Digital Preservation Group. http://cdm266901.cdmhost.com/

Chesapeake Digital Preservation Group. 2013. *Link Rot and Legal Resources on the Web: A 2013 Analysis by the Chesapeake Digital Preservation Group*. Chesapeake Digital Preservation Group. http://cdm16064.contentdm.oclc.org/ui/custom/default/collection/default/resources/custompages/reportsandpublications/2013LinkRotReport.pdf.

Claburn, Thomas. 2009. "Government Keeping Its .Gov Domain Names Secret." *Information Week*. (3/2/2009) http://www.informationweek.com/regulations/government-keeping-its-gov-domain-names-secret-/d/d-id/1077160?

Consultative Committee for Space Data Systems. 2002. *Reference Model for an Open Archival Information System* (OAIS). Blue Book, Issue 1. CCSDS Publications 650.0-B-1. Washington D.C.: CCSDS Secretariat, National Aeronautics and Space Administration. http://public.ccsds.org/publications/archive/650x0b1.pdf.

Conway, Paul. 1996. "Preservation in the Digital World" Council on Library and Information Resources, Pub62 (March 1996). http://www.clir.org/pubs/reports/conway2/

Cruse, Patricia, Charles Eckman, John Kunze, and Heather Christenson. 2003. *Web-Based Government Information: Evaluating Solutions for Capture, Curation, and Preservation*. An Andrew W. Mellon Funded Initiative. Berkeley, CA :: California Digital Library. http://www.cdlib.org/services/uc3/docs/Web-based_archiving_mellon_Final.pdf.

Department of Energy Organization Act, Public Law 95-91. http://www.gpo.gov/fdsys/pkg/STATUTE-91/pdf/STATUTE-91-Pg565.pdf

Diaz, Carlos. 2014. "Re: The National Atlas." Govdoc-l. (Mar 4, 2014). http://lists.psu.edu/cgi-bin/wa?A2=ind1403A&L=GOVDOC-L&F=&S=&P=9215

Gano, Gretchen, and Julie Linden. 2007. "Government Information in Legacy Formats." *D-Lib Magazine* 13 (7/8). doi:10.1045/july2007-linden. http://www.dlib.org/dlib/july07/linden/07linden.html.

Hartman, Cathy Nelson, Kathleen Murray, and Mark Phillips. 2013. Classification of the End-of-Term Archive: Extending Collection Development Practices to Web Archives. Final Report. Denton, TX: University of North Texas, UNT Libraries. IMLS Award Final Report. http://research.library.unt.edu/eotcd/w/images/8/8d/LG-06-09-0174-09_UNT_Feb2013_FINAL.pdf.

International Internet Preservation Consortium. Member Archives. http://netpreserve.org/resources/member-archives

Internet Archive. 2013. hosts-report.txt. http://wbgrp-svc040.us.archive.org/collections/eot2012/stats/Fall_2012/FINAL/hosts-report.txt

Internet Systems Consortium, Inc. 2014. "Internet Domain Survey, January, 2014: Distributions by Top-Level Domain Name (by name)." http://ftp.isc.org/www/survey/reports/2014/01/byname.txt

James, Bruce. 2004. Exchange between Greta Marlatt and Bruce James, "Summary, 2003 Fall Meeting Depository Library Council October 19-22, 2004 [sic] Washington, DC Fall Council/Depository Library Conference Meeting October 19 - 22, 2003," Arlington, VA. *Administrative Notes: Newsletter of the Federal Depository Library Program* 25(3) February 25, 2004.

Johnson, Nicholas. 2008. Comment. http://freegovinfo.info/node/1799

Koontz, Linda D. 2004. "Government Printing Office: Technological Changes Create Transformation Opportunities". GAO-04-729T. Washington, D.C.: U.S. General Accounting Office. http://www.gao.gov/products/GAO-04-729T.

Kott, Katherine B. 2010. "Everyday Electronic Materials in Policy and Practice." Project Briefings Presented at the CNI Fall 2010 Membership Meeting. http://www.cni.org/topics/digital-curation/everyday-electronic-materials-in-policy-and-practice/

Marks, Joseph. 2011. "New Adobe platform would personalize an agency website for individual users," *NextGov* (06/20/2011). http://www.nextgov.com/nextgov/ng_20110620_4684.php

McGarr, Sheila M. 1994. "Snapshots of the Federal Depository Library Program." *Administrative Notes*, August 15. http://www.access.gpo.gov/su_docs/fdlp/history/snapshot.html.

Memorandum Of Understanding (MOU) Between The Government Printing Office And The National Archives And Records Administration. 2003. http://www.gpo.gov/help/naramemofinal.pdf

Miller, Jason. 2005. "NARA Web site harvest yields 75 million pages." *GCN*. http://gcn.com/Articles/2005/01/21/NARA-Web-site-harvest-yields-75-million-pages.aspx

Murray, K. & Hartman, C. 2012. "Classifying the end-of-term archive." In *Archiving 2012 Final Program and Proceedings* (pp. 84-87). Springfield, VA: Society for Imaging Science and Technology.

National Academy Of Public Administration. 2013. *Rebooting The Government Printing Office: Keeping America Informed in the Digital Age*, A Report by a Panel of the National Academy Of Public Administration for the U.S. Congress, Congressional Research Service, and the Government Printing Office. National Academy Of Public Administration, Washington, DC (January 2013). http://www.napawash.org/wp-content/uploads/2013/02/GPO-Final.pdf

NDSA Content Working Group. 2012. *National Digital Stewardship Alliance Web Archiving Survey Report* (June 19, 2012) http://www.digitalpreservation.gov/ndsa/working_groups/documents/ndsa_web_archiving_survey_report_2012.pdf

Petersen, R. Eric, Jennifer E. Manning, and Christina M. Bailey. 2012. "Federal Depository Library Program: Issues for Congress". R42457. CRS Report for Congress. Washington, D.C.: Congressional Research Service. http://www.fas.org/sgp/crs/misc/R42457.pdf.

Phillips, Macon. 2011. "TooManyWebsites.gov." *White House Blog*. (June 13, 2011) http://www.whitehouse.gov/blog/2011/06/13/toomanywebsitesgov

Rossmann, B. (2005). On the Range: A Response to "The Once and Future Federal Depository Library Program." *DttP: A Quarterly Journal of Government Information Practice & Perspective*, 33(2), 8–9.

Russell, Judy. 2003. "Future Directions of the Depository Library Program." Remarks by the US Superintendent of Documents 142nd ARL Membership Meeting, Federal Relations Luncheon, May 15, 2003 Association of Research Libraries http://www.arl.org/resources/pubs/mmproceedings/142mmrussell.shtml

Shuler, John A. 2004. "New Economic Models for the Federal Depository System—Why Is It So Hard to Get the Question Answered?" Journal of Academic Librarianship, 30 (3): 246.

Shuler, John A., Paul T. Jaeger, and John Carlo Bertot. 2010. "Implications of Harmonizing the Future of the Federal Depository Library Program within E-Government Principles and Policies." *Government Information Quarterly* 27 (1): 9–16. doi:10.1016/j.giq.2009.09.001.

Shuler, John A., Paul T. Jaeger, and John Carlo Bertot. 2014. "Editorial: E-government without government." Government Information Quarterly, v31,n1. p1-3. http://dx.doi.org/10.1016/j.giq.2013.11.004

Theimer, Kate. 2008. "NARA and the web harvest: a discussion of the issues." *ArchivesNext*. http://www.archivesnext.com/?p=137

U.S. Code. Title 5, § 552, Public information; agency rules, opinions, orders, records, and proceedings. http://www.law.cornell.edu/uscode/text/5/552

U.S Code. Title 44 Chapter 41, "Access To Federal Electronic Information."

U.S. Department of Defense. Department of Defense Websites. http://www.defense.gov/RegisteredSites/RegisteredSites.aspx

U.S. Department of Defense. Social Media Sites. http://www.defense.gov/RegisteredSites/SocialMediaSites.aspx

U.S. General Accounting Office. 2001. "Electronic Dissemination of Government Publications". GAO-01-428. GAO Report. Washington, D.C.: U.S. General Accounting Office. http://www.gao.gov/assets/240/231303.pdf.

U.S. General Services Administration. Office of Governmentwide Policy. 2014. "Federal Executive Agency Internet Domains as of 02122014." https://explore.data.gov/Federal-Government-Finances-and-Employment/Federal-Executive-Agency-Internet-Domains-as-of-02/ku4m-7ynp https://explore.data.gov/d/ku4m-7ynp

U.S. Geological Survey. 2014. "What's going to happen to the products and services of the National Atlas?" http://nationalatlas.gov/status.html

U.S. Government Printing Office. FDsys Collections. http://www.gpo.gov/fdsys/browse/collectiontab.action

U.S. Government Printing Office. United States Courts Opinions. http://www.gpo.gov/fdsys/browse/collection.action?collectionCode=USCOURTS

U.S. Government Printing Office. 2006. Dissemination/Distribution Policy for the Federal Depository Library Program (SOD 301) http://www.fdlp.gov/file-repository/about-the-fdlp/policies/567-disseminationdistribution-policy-for-the-federal-depository-library-program-sod-301/file

U.S. Government Printing Office. 2007. "GPO LOCKSS Pilot: Final Analysis, Executive Summary" April 12, 2007 http://www.fdlp.gov/file-repository/about-the-fdlp/gpo-projects/lockss/611-executive-summary-1/file

U.S. Government Printing Office. 2009. *Federal Depository Library Program Strategic Plan, 2009 - 2014.* 01/15/2009 http://www.fdlp.gov/file-repository/about-the-fdlp/strategic-plan-for-the-fdlp/661-fdlp-strategic-plan-2009-2014-draft-from-januanry-2009/file

U.S. Government Printing Office. 2011. "GPO & Federal Judiciary Enhance Public Access To Federal Court Opinions" [press release] No. 11-23 (May 4, 2011). http://gpo.gov/pdfs/news-media/press/11news23.pdf

U.S. Government Printing Office. 2012. *Budget Justification, Fiscal Year 2013.* January 25, 2012. http://gpo.gov/pdfs/congressional/Budget_Justification_FY2013.pdf.

U.S. Government Printing Office. 2013. "GPO's Federal Digital System Reaches 500 Million Retrievals." Press Release May 1, 2013 No. 13-19. http://gpo.gov/pdfs/news-media/press/13news19.pdf

U.S. Government Printing Office. 2014. Partnerships. http://www.fdlp.gov/about-the-fdlp/partnerships

U.S. National Archives and Records Administration. "ERA Status and Accomplishments." http://www.archives.gov/era/about/status-accomplishments.html

U.S. National Archives and Records Administration. 2006. "University of North Texas Libraries Joins NARA/GPO Partnership." http://www.archives.gov/press/press-releases/2006/nr06-99.html

U.S. National Archives and Records Administration. Office of the Federal Register. 2013. *United States Government Manual.* http://www.gpo.gov/fdsys/pkg/GOVMAN-2013-11-06/pdf/GOVMAN-2013-11-06.pdf http://www.gpo.gov/fdsys/pkg/GOVMAN-2013-11-06/xml/GOVMAN-2013-11-06.xml

USDocs private LOCKSS network. http://www.lockss.org/community/networks/digital-federal-depository-library-program/

Van de Sompel, Herbert, Robert Sanderson, and Michael Nelson. "Thoughts on Referencing, Linking, Reference Rot". Memento. http://mementoweb.org/missing-link/.

Vance-Cooks, Davita. 2013. Letter, March 20, 2013, from Acting Public Printer, to Bernadine Abbott Hoduski. https://archive.org/details/LetterToMsBernadineHoduskiOcr

WARC File Format (ISO 28500) - Information, Maintenance, Drafts. 2009. http://bibnum.bnf.fr/WARC/.

Wikipedia. List of Internet top-level domains. http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains

Zittrain, Jonathan, and Kendra Albert. 2013. "Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations". SSRN Scholarly Paper ID 2329688. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2329688.

FURTHER READING

Ainsworth, Scott G., Ahmed Alsum, Hany SalahEldeen, Michele C. Weigle, and Michael L. Nelson. 2011. "How Much of the Web Is Archived?" In Proceeding of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries, 133–36. JCDL '11. New York, NY, USA: ACM. doi:10.1145/1998076.1998100. http://doi.acm.org/10.1145/1998076.1998100.

"Cathy Hartman and CyberCemetery." 2014. Digital Preservation Pioneers. Accessed January 17. http://www.digitalpreservation.gov/series/pioneers/hartman.html.

CBSNews AP. 2009. "Cyber Cemetery Keeps Gov't Web Sites Alive." September. http://www.cbsnews.com/news/cyber-cemetery-keeps-govt-web-sites-alive/.

Feldman, Emily. 2008. "Partners Join Together to Preserve Government Web Sites." AALL's Washington Blawg. August 18. http://aallwash.wordpress.com/2008/08/19/partners-join-together-to-preserve-government-web-sites/.

Glenn, Valerie D. 2007. "Preserving Government and Political Information: The Web-at-Risk Project." First Monday 12 (7). http://www.firstmonday.org/ojs/index.php/fm/article/view/1917.

Goethals, Andrea, and Wendy Marcus Gogel. 2014. "Election 2012 Web Archive." Free Government Information (FGI). Accessed January 17. http://freegovinfo.info/node/author/eotarchive.

Gomes, Daniel, Sérgio Freitas, and Mário J. Silva. 2006. "Design and Selection Criteria for a National Web Archive." In Research and Advanced Technology for Digital Libraries, edited by Julio Gonzalo, Costantino Thanos, M. Felisa Verdejo, and Rafael C. Carrasco, 196–207. Lecture Notes in Computer Science 4172. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/11863878_17.

Grotke, Abbie. 2012. "An Abundant Crop: The End of Term Harvest | The Signal: Digital Preservation." The Signal. http://blogs.loc.gov/digitalpreservation/2012/11/an-abundant-crop-the-end-of-term-harvest/.

Masanes, Julien. 2005. "Web Archiving Methods and Approaches: A Comparative Study." Library Trends 54 (1): 72–90. doi:10.1353/lib.2006.0005. http://muse.jhu.edu/journals/library_trends/v054/54.1masanas.html.

McKinney, Richard J. 2011. "An Overview of the Congressional Record and Its Predecesessor Publications: A Research Guide." Law Librarians' Society of Washington, D.C. October. http://www.llsdc.org/congressional-record-overview#20.

Murray, Kathleen. 2010. "Metrics For Web Archives". Focus Group Report IMLS Award Number LG-06-09-0174-09. Classification of the End-of-Term Archive: Extending Collection Development Practices to Web Archives. Denton, TX: University of North Texas, UNT Libraries. http://research.library.unt.edu/eotcd/w/images/0/0f/eotcd_metrics_fg_final_rpt_krm_12augl2010.pdf.

Murray, Kathleen, and Cathy Hartman. 2012. "Classifying the End-of-Term Archive." In Archiving 2012 Final Program and Proceedings, 84–87. Springfield, VA: Society for Imaging Science and Technology. http://research.library.unt.edu/eotcd/w/images/0/0e/murray_classifying_the_endofterm_archive_ist_2012.pdf.

Phillips, Margaret E. 2005. "What Should We Preserve? The Question for Heritage Libraries in a Digital World." Library Trends 54 (1): 57–71. doi:10.1353/lib.2006.0007. http://muse.jhu.edu/journals/library_trends/v054/54.1phillips.html.

Seneca, Tracy. 2009. “The Web-at-Risk at Three: Overview of an NDIIPP Web Archiving Initiative.” Library Trends 57 (3): 427–41. doi:10.1353/lib.0.0045. http://muse.jhu.edu/journals/library_trends/v057/57.3.seneca.html.

Seneca, Tracy, Abigail Grotke, Cathy Nelson Hartman, and Kris Carpenter. 2012. “It Takes A Village To Save The Web: The End Of Term Web Archive.” Documents to the People (DttP) 40 (1). http://digital.library.unt.edu/ark:/67531/metadc84373/m1/1/.

Stevenson, John A. 2004. “Letter to Judith C. Russell, Superintendent of Documents U.S. Government Printing, from GODORT, Comment on the Discussion Draft of ‘Collection of Last Resort (CLR)’”, July 29. http://wikis.ala.org/godort/images/3/38/CLRfinal.pdf.

Szydlowski, Nick. 2010. “Archiving the Web: It’s Going to Have to Be a Group Effort.” The Serials Librarian 59 (1): 35–39. doi:10.1080/03615260903534908.

U.S. Government Printing Office. 2013. “Web Harvesting Pilot Project.” March 1. http://beta.fdlp.gov/all-newsletters/featured-articles/1493-web-harvesting-pilot-project.

U.S. National Archives and Records Administration. 2001. “Memorandum to Chief Information Officers: Snapshot of Agency Public Web Sites.” http://www.archives.gov/records-mgmt/basics/snapshot-public-web-sites.html.

———. 2005. “NARA Guidance on Managing Web Records.” http://www.archives.gov/records-mgmt/pdf/managing-web-records-index.pdf.

———. 2008. “MEMORANDUM TO FEDERAL AGENCY CONTACTS: End-of-Administration Web Snapshot.” http://www.archives.gov/records-mgmt/memos/nwm13-2008.html.

———. 2014. “Attachment to Memorandum to CIOs, January 12, 2001 GUIDELINES TO AGENCIES ON PRESERVING A SNAPSHOT OF THEIR WEB SITES AT THE END OF THE CLINTON ADMINISTRATION.” http://www.archives.gov/records-mgmt/basics/snapshot-public-web-sites-attachment.html.