
 **any specialized services**
contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES 


Web Archiving for the FDLP

GPO's Web Harvesting and Archiving of Federal Web Content

Depository Library Conference
October 18, 2012

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdlp.gov> | <http://fdensguide.gpo.gov> | <http://bookstore.gpo.gov>


 **any specialized services**
contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES 

Where it all began


1895: Federal Depository Library Program created by an act of Congress; made part of GPO Library program now at the source of all authentic government information

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdlp.gov> | <http://fdensguide.gpo.gov> | <http://bookstore.gpo.gov>


 **any specialized services**
contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES 

Information model 1895-



U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://lensguide.gpo.gov> | <http://bookstore.gpo.gov>

 **any specialized services**
contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES 

Information model 1993-

“Government Printing Office Electronic Access Enhancement Act” 1993


“This important step forward in the electronic dissemination of Federal information will provide valuable insights into the most effective means of disseminating all public Government information.”
President Bill Clinton

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://lensguide.gpo.gov> | <http://bookstore.gpo.gov>

For any specialized services, contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Accessing government information 2012 -



U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://hensguide.gpo.gov | http://bookstore.gpo.gov


For any specialized services, contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES


Government "publications" 2012 -



U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://hensguide.gpo.gov | http://bookstore.gpo.gov


 ny specialized services
 contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES




Web archiving for the FDLP

1. Manually harvest individual publications when that is the only content within scope on a site.
2. Harvest full sites using automation (Heritrix) at determined depth levels when that content is within scope
3. Create facsimile copies of web sites and preserve as much original functionality as is technically possible
4. Classify and catalog all harvested content; both publications and sites and create records and links in the CGP

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdlp.gov> | <http://fdlsguide.gpo.gov> | <http://bookstore.gpo.gov>


 ny specialized services
 contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES



Web archiving for the FDLP

Acquire web content through:

- Contacts with agency web masters and librarians
- Outreach to identify offline and fugitive agency publications
- Develop formal partnerships for formats not easily harvested

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdlp.gov> | <http://fdlsguide.gpo.gov> | <http://bookstore.gpo.gov>


 For any specialized services, contact GPO today.




Web archiving for the FDLP



IIPC
netpreserve.org

**INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM**



**LIBRARY OF
CONGRESS**



**NATIONAL
ARCHIVES**






**THE
NATIONAL
ARCHIVES**
ARCHIVES.GOV



UNT Libraries
THE POWER OF IDEAS STARTS HERE

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov | http://hensguide.gpo.gov | http://bookstore.gpo.gov


 For any specialized services, contact GPO today.

Continued harvesting

ARCHITECTURAL AND TRANSPORTATION BARRIERS COMPLIANCE BOARD (ACCESS BOARD)

Y 3 B 27.1 Annual Reports (Annual) (EL) 1063-D-01

Y 3 B 27.2 General Publications 1062-D-03

Y 3 B 27.3-2 Bulletin (P/EL) 1062-D-02

Y 3 B 27.5 Laws (P) 1062-D-04

Y 3 B 27.8 Handbooks, Manuals, Guides 1062-D-05

Y 3 B 27.16 Access Currents (Bi-monthly) (EL) 1061-D-03

UNITED STATES CHEMICAL SAFETY AND HAZARD INVESTIGATION BOARD

Y 3 C 42.2.2 General Publications 1093-A

Y 3 C 42.2.3 Safety Bulletin (Completely Irregular) (P) 1093-A-03

Y 3 C 42.2.15 Electronic Products (misc.) (E) 1093-A-01

Y 3 C 42.2.16 Investigations and News (EL) 1093-A-02

Y 3 C 42.2.17 Case Study (Items/Issues) (Completely Irregular) (EL) 1093-A-04


COMMODITY FUTURES TRADING COMMISSION

Y 3 C 73.5.1 Annual Report (Annual) (EL) 0039-B

Y 3 C 73.5.1.9 FY ... President's Budget & Performance Plan (Annual) (EL) 0039-B-03

- Y 3. of SuDoc Classification
- Collection of small, independent federal agencies, commissions, review boards and committees.
- Contains publications from agencies/commissions that predominately publish in electronic formats

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov | http://hensguide.gpo.gov | http://bookstore.gpo.gov



 ny specialized services
 contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES

Workflow of automated archiving

- Scope determination
 - Create Seed list to tell crawler what is in scope
 - a. Add the main website url
 - example: <http://www.ushmm.org/>
 - All subsequent directories should be captured
 - example: <http://www.ushmm.org/education/>
 - b. Search site for any sub-domains
 - example: <http://resources.ushmm.org/>
 - Sub-domains are divisions of a larger site named to the left of the host name and will not be captured automatically.

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://lensguide.gpo.gov | http://bookstore.gpo.gov



 ny specialized services
 contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES


Workflow of automated archiving

- Run test crawl
 - Allows us to see data without archiving it
- Contact agencies
 - 48 hour notice to Webmaster, CIO
- Quality Review
 - Review all reports available on Archive-It, check urls, file size, anything blocked

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://lensguide.gpo.gov | http://bookstore.gpo.gov


 ny specialized services
 contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES




Workflow of automated archiving

- Make changes
 - Change scoping, go behind a robot for a site
- Run crawl
 - Meanwhile, we begin scoping other sites
- Quality review on “Wayback”
 - Can view how actual site will look
- Challenges
 - Java, database functionality

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov | http://lensguide.gpo.gov | http://bookstore.gpo.gov


 ny specialized services
 contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES




8 Sample sites complete


- Accessible in two locations
 - Catalog of Government Publications (CGP)
 - Archive-It (Internet Archive)
<http://www.archive-it.org>
- Feedback welcome
 fdlpwebarchiving@gpo.gov

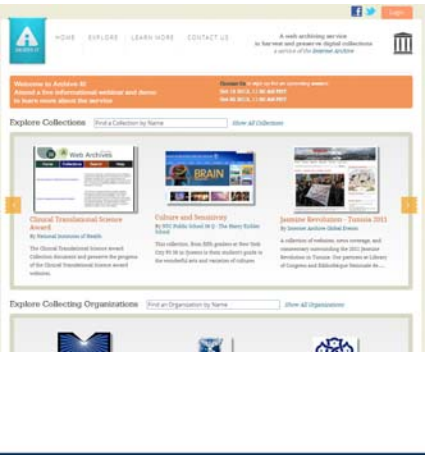
U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov | http://lensguide.gpo.gov | http://bookstore.gpo.gov


 any specialized services, contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES



Access locations





U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>


 any specialized services, contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES


Searching options

- Catalog of Government Publications (CGP)
 - keyword title searching
 - Expert search mode WCAT=web archiving
- Archive-It
 - “Find Collection by name” type “GPO”

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>

any specialized services, contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES

Title Keyword= Holocaust Museum Look for "Internet" in SuDoc Number

Results for W-Titles= holocaust museum
 Sort options: | [View/Title](#) | [View/Title](#) | [Title/Keyword](#) | [SuDoc](#) | [SuDoc](#) | [SuDoc](#)
 Display options: [Expanded](#) | [Brief](#)
 Records 11 - 19 of 19 (maximum display and sort is 2000 records)
[Jump to #](#)

[Previous Page](#)
[Next Page](#)

#	Title	Year	Author	SuDoc Number	Internet Access
11	The United States Holocaust Memorial Museum site	1990		Y 3 H 74 I 2 H 74	
12	Holocaust Memorial Book Museum	1999		Y 3 H 74 I 2 H 744	
13	You can help shape the United States Holocaust Memorial Museum's call for artifacts	1999		Y 3 H 74 I 2 H 744	
14	The National Learning Center presents "Remember the children", an exhibit for children about the Holocaust - Center Children's Museum, 200 Third Street N.E., Washington, D.C.	1999		Y 3 H 74 I 2 C 43	
15	A National commitment to remembrance - official proceedings, United States Holocaust Memorial Museum - Days of Remembrance 1998, national civil commemoration	1998		Y 3 H 74 I 2 C 730	
16	Days of remembrance 1993 - national Holocaust commemoration ceremony - presentation of the key to the United States Holocaust Memorial Museum and Education Center	1993		Y 3 H 74 I 2 C 3990	
17	U.S. Holocaust Memorial Museum newsletter	1994		Y 3 H 74 I 10	
18	U.S. Holocaust Museum - 50th anniversary of the Wannsee Conference, persisat.org	1994		Y 4 N 914 I 10-24	
19	United States Holocaust Memorial Museum	uuuu	United States Holocaust Memorial Museum	Y 3 H 74 I 2 INTERNET	http://www.gpo.gov/olapo27734

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>

any specialized services, contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES


Expert Search WCAT=Web Archiving



Results for W-956 &= web archiving
 Sort options: | [View/Title](#) | [View/Title](#) | [Title/Keyword](#) | [SuDoc](#) | [SuDoc](#) | [SuDoc](#)
 Display options: [Expanded](#) | [Brief](#)
 Records 1 - 8 of 8 (maximum display and sort is 2000 records)
[Jump to #](#)

[Previous Page](#)
[Next Page](#)


#	Title	Year	Author	SuDoc Number	Internet Access
1	911.gov - The National 911 Office	uuuu	National 911 Program,	TD 8 2-INTERNET1	http://www.fdsys.gov/olapo27739
2	Administrative Conference of the United States	uuuu	Administrative Conference of the United States,	Y 3 AD 6-2INTERNET	http://www.fdsys.gov/olapo27738
3	Consumer Financial Protection Bureau	uuuu	CFPB, Consumer Financial Protection Bureau,	FR 2-2-INTERNET	http://www.fdsys.gov/olapo27737
4	Distraction.gov - official US Government web site for distracted driving	uuuu	National Highway Traffic Safety Administration,	TD 8-2-INTERNET2	http://www.fdsys.gov/olapo27736
5	National Highway Traffic Safety Administration (NHTSA)	uuuu	NHTSA, National Highway Traffic Safety Administration,	TD 8-2-INTERNET	http://www.fdsys.gov/olapo27735
6	NHTSA EMS	uuuu	United States National Highway Traffic Safety Administration	TD 8-2-INTERNET3	http://www.fdsys.gov/olapo27734
7	United States Access Board	uuuu	Architectural and Transportation Barriers Compliance Board	Y 3 B 27-2 INTERNET	http://www.fdsys.gov/olapo27731
8	United States Holocaust Memorial Museum	uuuu	United States Holocaust Memorial Museum	Y 3 H 74 I 2 INTERNET	http://www.fdsys.gov/olapo27734

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>


 any specialized services, contact GPO today.





Archive-It Search Collection="GPO"




HOME EXPLORE LEARN MORE CONTACT US

A web archiving service to harvest and preserve digital collections a service of the Internet Archive



Expires: Government Printing Office



Government Printing Office
Archive 8 Partner Since: Nov. 2011
 Organization Type: National Government
 Organization URL: <http://www.gpo.gov>

Narrow Your Results

Subject

- Government - US Federal (3)
- Administrative procedure - United States (2)
- Barrier-free design - Standards - United States (2)
- Consumer credit - United States (1)
- Consumer protection - United States (1)

Creator

- CFPB, Consumer Financial Protection Bureau (1)
- NHTSA, National Highway Traffic Safety Administration (1)
- U.S. National Archives and Records Administration (1)
- United States Access Board (1)
- United States Holocaust Memorial Museum (1)

Sites and collections from this organization are listed below. Narrow your results at left, or enter a search query below to find a collection, site, specific URL, or to search the text of archived webpages.

Enter search terms here

Collections Sites Search Page Text

Page 1 of 1 (8 Total Results)


Sort By: [Collection Name \(A-Z\)](#) | [Collection Name \(Z-A\)](#)



Administrative Conference of the United States (ACUS)
Archived since: Jul, 2012

Description: Describes the mission and activities of this independent federal agency that comprises a coalition of federal and public sector officials and academics who evaluate agency policymaking processes and recommend improved procedures. Features case studies of model agency initiatives, a best practices forum, and the Conference's formal recommendations.


Subject: Administrative procedure - United States, Delegated legislation - United States - Evaluation

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>



 any specialized services, contact GPO today.

Search results



United States Holocaust Memorial Museum Web Archive (Government Printing Office)



Enter Web Address:

Searched for: <http://www.ushmm.org> 3 Results

[Link on USG](#) in general Internet Archive web collection [Printable View](#)

* denotes when page was updated

Found 3 Captures between Jul 18, 2012 - Aug 15, 2012	
Year	Pages
2012	3 pages
Jul 18, 2012 *	
Aug 15, 2012 *	
Aug 15, 2012 *	

[Home](#) | [Internet Archive](#)

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>

U.S. GOVERNMENT PRINTING OFFICE: 2011
 GPO: any specialized services, contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES

Archived site

You are viewing an archived web page, collected at the request of Government Printing Office using Archive-It. This page was captured on 19 11 01 Aug 15, 2012, and is part of the United States Holocaust Memorial Museum collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#)

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://bensguide.gpo.gov | http://bookstore.gpo.gov


U.S. GOVERNMENT PRINTING OFFICE: 2011
 GPO: any specialized services, contact GPO today.



GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES

Metadata creation, curation, and management

- The Dublin Core Metadata Set, version 1.1 used in the Archive-It application
- Metadata created for different granule levels (collection, seed, and document)
- RDA-Compatible

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://bensguide.gpo.gov | http://bookstore.gpo.gov


 any specialized services, contact GPO today.

DC record from the Archive-It Site (1)

Dublin Core

Title
 Add
 United States Holocaust Memorial Museum **Remove**

Creator
 Add
 United States Holocaust Memorial Museum **Remove**

Subject
 Add
 Holocaust survivors--Services for--United States **Remove**
 Holocaust survivors' families--Services for--United States **Remove**
 Holocaust, Jewish (1939-1945) **Remove**
 Genocide--Prevention **Remove**

Description *
 Add
 A complete orientation to the services, collections, symposia, exhibitions, and publications of the U.S. Holocaust Museum. Includes coverage of services of the Holocaust **Remove**

Publisher
 Add
 United States Holocaust Memorial Museum **Remove**


Contributor
 Add



Date
 Add

Type
 Add
 online resource **Remove**
 text **Remove**
 still image **Remove**
 two-dimensional moving image **Remove**
 sounds **Remove**

Format
 Add

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://bensguide.gpo.gov | http://bookstore.gpo.gov


 any specialized services, contact GPO today.

DC record from the Archive-It Site (2)

Identifier *
 Add
 http://url.fdp.gov/GPO/g27784 **Remove**

Source
 Add
 http://wayback.archive-it.org/2011/* **Remove**
 http://www.saltion.org/ **Remove**

Relation
 Add

Coverage
 Add
 2012-present **Remove**

Rights
 Add
 The content is provided by the U.S. Government Printing Office in accordance with Title 44 of the U.S. Code **Remove**

Collector
 Add
 U.S. Government Printing Office **Remove**

Language
 Add

- English **Remove**
- French **Remove**
- Spanish **Remove**
- Italian **Remove**
- Russian **Remove**
- Turkish **Remove**
- Portuguese **Remove**
- Arabic **Remove**
- Farsi **Remove**
- Urdu **Remove**
- Greek **Remove**

Custom Fields

Name


Value

Item Number:
 1070-K-01 **Remove**

Supt. of Documents Number:
 Y3.H74:2/INTERNET **Remove**

Subject/Corporate Body
 United States Holocaust Memorial Museum **Remove**

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://bensguide.gpo.gov | http://bookstore.gpo.gov



[GPO.gov](#) | [Specialized services](#) | [Contact GPO today.](#)

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES

Archived content: Catalog of U.S. Government Publications (CGP)

- CGP is the national bibliography of the publications of the United States Government
- Cross-walk and convert metadata from DC to MARC
- Metadata Mapping Profile

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>



[GPO.gov](#) | [Specialized services](#) | [Contact GPO today.](#)

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES


MARC record in the CGP

<pre> #MARC# LDR# 00000w k0 001 00078949 005 20121005142122.0 006 121004 500099890 eng 035 1 (DGPO) 040 1a GPO : rda : j : GPO 042 1a dc 074 1a 1070-K-01 (online) 0860 1a Y 3 H 742 : INTERNET 1182 1a United States Holocaust Memorial Museum : s : creator 24510 1a United States Holocaust Memorial Museum. 260 1a [Washington, DC] : s : United States Holocaust Memorial Museum. 336 1a text : 2 : rdacontent 337 1a 2D image : 2 : rdacontent 338 1a two-dimensional moving image : 2 : rdacontent 339 1a sounds : 2 : rdacontent 357 1a computer file : 2 : rdacontent 333 1a online resource : 2 : rdacontent 500 1a United States Holocaust Memorial Museum. 520 1a A complete orientation to the services, collections, symposia, exhibitions, and publications of the U.S. Holocaust Museum, Inc. 520 1a complete of services of the Holocaust Survivors and Victims Resource Center and The Center for Advanced Holocaust Studies 500 1a The content is provided by the U.S. Government Printing Office in accordance with Title 44 of the U.S. Code. </pre>	<pre> 546 1a English 546 1a French 546 1a Spanish 546 1a Italian 546 1a Russian 546 1a Turkish 546 1a Portuguese 546 1a Arabic 546 1a Farsi 546 1a Urdu 546 1a Oroski 500 1a Coverage: 2012-present. 610 1a United States Holocaust Memorial Museum 65000 1a Holocaust survivors--Services for--United States. 65000 1a Holocaust survivors' families--Services for--United States. 65000 1a Holocaust, Jewish (1939-1945) 65000 1a Genocide--Prevention. 7102 1a United States Holocaust Memorial Museum : s : publisher 7101 1a United States. : s : Government Printing Office : s : collector 85640 1a http://fulf.fdsys.gov/GPO/gpo27784 : s : Archived site 8564 1a http://wayback.archive-it.org/3111/http://www.ushmm.org/ 856 13 (online) : s : http://text.catalog.gpo.gov/fdp/fdocate.jsp?itemNumber=1070-K-01&SYS=100078949 955 1a Web Archiving 20121004 </pre>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://bensguide.gpo.gov> | <http://bookstore.gpo.gov>


 ny specialized services
 contact GPO today.


GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES




Implications

- Workflow adjustment, collaboration across different units, new skill sets, training needs
- LC's Bibliographic Framework Transition Initiative, linked data on Semantic Web
- Reuse and repurpose metadata to reduce duplicate work and avoid redundancy
- RDA: New cataloging standard designed for digital world

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://lensguide.gpo.gov | http://bookstore.gpo.gov


 ny specialized services
 contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
 FEDERAL DEPOSITORY LIBRARIES



Next steps

- Announcement through FDLP website
- Feedback on SuDoc numbers, nominating sites, etc.
- Expand staff and integrate the process into the regular workflow
- Fine tuning the cross-walking process from DC to MARC via XML feeds
- Batch loading records from CGP to OCLC for resources sharing with the wider community

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | http://www.fdp.gov | http://lensguide.gpo.gov | http://bookstore.gpo.gov

gpo ny specialized services
contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES



Contact Us:

Dory Bower, Fang Gao, David Walls

fdlpwebarchiving@gpo.gov

U.S. GOVERNMENT PRINTING OFFICE | www.gpo.gov | www.fdsys.gov | <http://www.fdp.gov> | <http://hensguide.gpo.gov> | <http://bookstore.gpo.gov>



Web Archiving at the Library of Congress and around the world

Abbie Grotke
Web Archiving Team Lead

abgr@loc.gov

<http://www.loc.gov/webarchiving/>

Overview

- What is web archiving?
- How does it work? What are some of the challenges?
- Archiving Around the Globe
- Archiving at the Library of Congress
- Collaborative Archives
- Questions



LIBRARY OF CONGRESS

What is Web Archiving?



LIBRARY OF CONGRESS

web harvesting
web crawling
web capture
web preservation
web archiving

Web Archiving...

- The process of creating an archival or preservation copy of a website
- A snapshot of how the original site looked at a particular point in time and how users interacted with it (aka “look and feel”)
- Often captures multiple copies of a site over a set period of time
- Captures and preserves as much as possible from the original site, including text, images, audio, videos, and PDFs

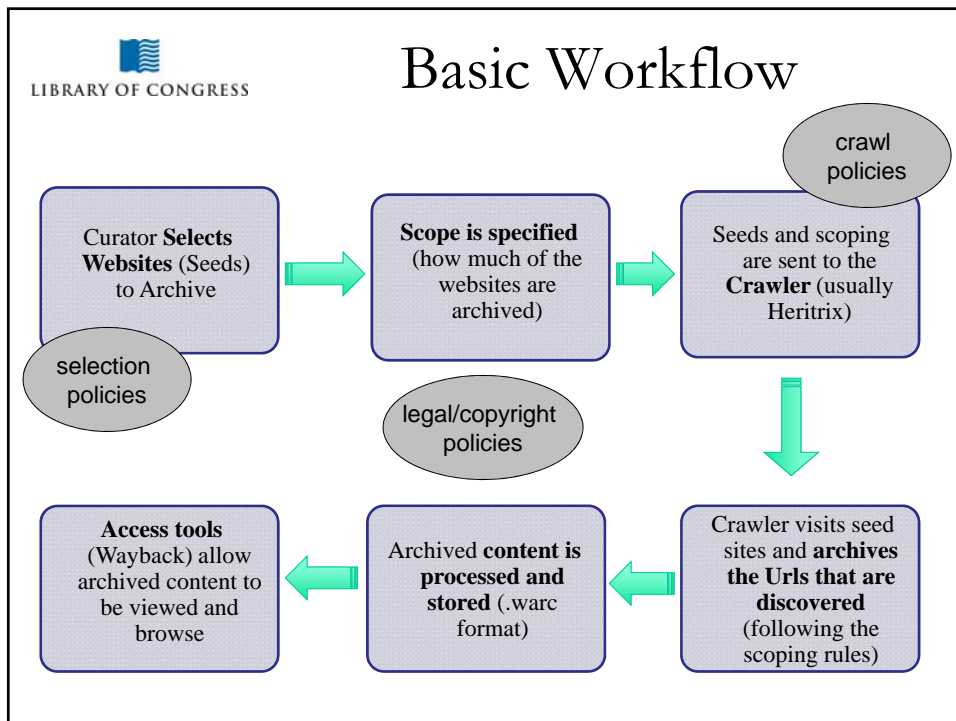


Why Archive the Web?

- Document events unfolding on the web
- Document changes on the web
- Document the changes and record of governments
- Capture the creative output of citizens
- Capture entire domains of some countries to meet legal deposit mandates



How does it work?
What are some of the challenges?



Web Archive Curator Tools

The collage displays several web archiving interfaces. On the left, the Library of Congress 'Web Archiving Digiboard: Nominations' page shows a table with columns for 'url', 'collection', and 'Actions'. In the center, a 'WEB CURATOR TOOL' window is overlaid on a desktop environment, showing a sidebar with 'In Tray' (13 tasks, 201 Applications), 'Harvest Authorizations' (20 Harvest Authorizations), 'Targets' (4 Targets), 'Target Instances' (20 Scheduled instances, 27 ready for Quality reviews), and 'Groups' (1 Target Groups). On the right, the 'web archiving service' page features a search bar, a 'Search Archives' section, and a list of collections including 'Mexican Cooking Blogs Web Collection', 'Rio Grande Watershed', and 'Wikileaks Documents Release Collection'.

Technical Challenges

This collage highlights technical issues. At the top, a 'DAVE KAISER 2012 REGISTER' page is shown with a prominent 'REGISTER' button. Below it, a 'Videos from NOLA.co' player displays an error: '// Javascript wayback retrieval error: // // Title: Resource Not In Archive // Message: The Resource you requested is not in this archive.' To the left, a social media sidebar lists 'Connect' options: Facebook, Twitter, Tumblr, RSS, YouTube, Flickr, and SMS. At the bottom, a Flickr photostream for 'naralprochoiceamerica' is visible, showing a grid of photo thumbnails.

Legal Challenges

- Legal deposit legislation allows some national libraries to collect and preserve web resources -- but not all...
- Section 108 of the U.S. Copyright Act provides library exceptions but doesn't yet address digital preservation and web archiving
- Varying Approaches taken:
 - Notify of crawling and/or access
 - Ask Permission to Crawl
 - Ask Permission to Provide Office Access
 - “Respect” robots.txt exclusions (or decide not to)
 - Fair use arguments

Archiving Around the Globe

49 member organizations:

national libraries & archives,
universities, service providers

harvesting, preservation, &
access working groups



**INTERNATIONAL
INTERNET
PRESERVATION
CONSORTIUM**

training & educational programs

collaborative collections & access

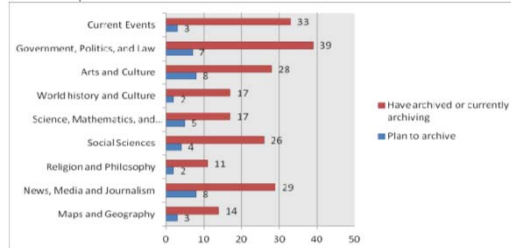
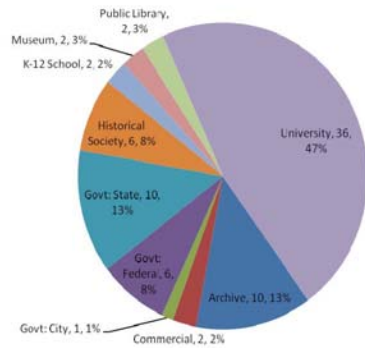
member archives: <http://netpreserve.org/about/archiveList.php>

The image is a screenshot of the 'The Web Archives Timeline' website. The top left corner features the 'webarchivists' logo and the title 'The Web Archives Timeline' with the subtitle 'how people around the world are archiving the Internet since 1991'. Below this is the URL 'timeline.webarchivists.org'. The main content area shows a timeline entry for 1996 titled 'Pandora, Australia's Web Archive'. The entry includes a colorful, abstract graphic of a globe and text describing Pandora as a growing collection of Australian online publications established in 1996. A search bar at the bottom of the page shows a search for 'Pandora, Australia's Web Archive'. In the bottom left corner, there is a world map with several countries highlighted in green, indicating the global reach of web archiving initiatives. At the bottom of the screenshot, the URL 'http://en.wikipedia.org/wiki/List_of_Web_Archiving_Initiatives' is provided.



National Digital Stewardship Alliance 2011 U.S. Web Archiving Survey

Survey Participant Institution Type



More:

<http://blogs.loc.gov/digitalpreservation/2012/07/the-ndsa-web-archiving-survey/>



LIBRARY OF CONGRESS

Web Archiving at the Library of Congress

LIBRARY OF CONGRESS

ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS


Search Search Loc.gov GO

The Library of Congress > Web Archiving

Print Subscribe Share/Save

WEB ARCHIVING

Web Archiving



The Library of Congress preserves and provides enduring access to the nation's cultural artifacts. This includes digital materials such as websites.

History of Web Archiving at the Library of Congress

In 2000, the Library of Congress established a pilot project to collect and preserve websites. A multidisciplinary team of Library staff studied methods to evaluate, select, collect, catalog, provide access to, and preserve these materials for future generations of researchers.

The Library has developed thematic web archives on such topics as the United States National Elections, the Iraq War, and the events of September 11. More about these collections plus many other available collections can be found at the [Library of Congress Public Web Archives](#) website.

In July 2003, the Library and the national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden, the British Library (UK), and the Internet Archive (USA) acknowledged the importance of international collaboration for preserving Internet content for future generations and formed the [International Internet Preservation Consortium](#). The goals of the Consortium include collecting a rich body of Internet content from around the world and fostering the development and use of common tools, techniques and standards that enable the creation of international archives.

In 2004, the Library's Office of Strategic Initiatives created a Web Archiving Team to support the goal of managing and sustaining at-risk digital content. The team is charged with building a Library-wide understanding and technical infrastructure for capturing web content. The team, in collaboration with a variety of Library staff, and national and international partners, is identifying policy issues, establishing best practices and building tools to collect and preserve web content.

Resources

- > [Digital Preservation](#)
- > [Library of Congress Web Archives](#)
- > [LC Digital Collections & Programs](#)

Web Archiving Home

- Technical Information
- Collections
- FAQs
- Contact

LIBRARY OF CONGRESS

Approaches to Archiving

- Large-Scale Collections
 - Documenting a theme or event, on a particular topic collected during a defined time frame or ongoing basis
 - Captured repeatedly over time: weekly, monthly, semi-yearly or yearly
 - Crawling done by contract
- Collaborative Archives
- In-House Archiving
 - Limited capacity
 - Testing and special projects
 - loc.gov and loc.gov/staff archiving
 - Emergency archiving



**LC Web Archive stats
as of October 1 2012**

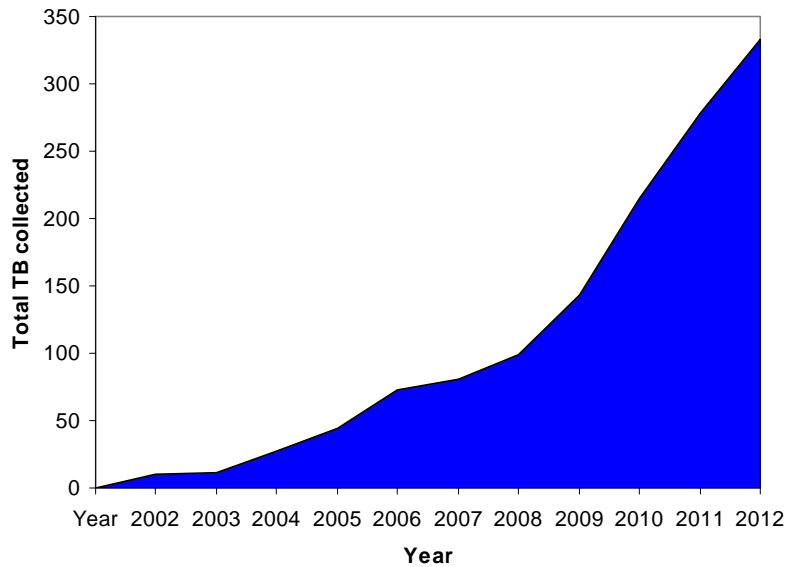
55 collections (13 active)

23,000+ nominated websites

6.985 billion documents

374 terabytes

Growth of Web Archive Data



Selection

- LC's collections are highly selected
- Collection topics developed and websites selected by Recommending Officers in Library Services and the Law Library
- Collection Policy Guidelines: <http://www.loc.gov/acq/dev/pol/webarchive.pdf>
- Supplemental guidelines made available

Web Archiving Selection
The purpose of these wiki pages is provide guidance to Recommending Officers on one hand more context and information about the topic generally while on the other
<p>Contents (page)</p> <ul style="list-style-type: none"> 1 Collection-based selection statements 2 Guidance for Web Archiving Selection 3 Other Sources of Information 4 Brief Background Statement 5 Why a Wiki for this information
<p>Collection-based selection statements</p> <ul style="list-style-type: none"> • Selection Statement for Public Policy Topics
<p>Guidance for Web Archiving Selection</p> <ul style="list-style-type: none"> • Types of Web Archiving • How Selection of Web Sites for Archiving is Like Selecting Books • Examples of Good Sites to Archive • The Portal Problem • Information (Content) vs Information About the Site Owner/Sponsor • Compilation Sites • Digital Libraries • Duplicating Other Organizations' Archiving • Interactive Database • News Sites • Sites in Foreign Languages • A Site that is a Single Document • Frequency of Capture/Harvest • Sites Generally Excluded • Relationship of Permissions to Selection of Sites • Other Issues for Consideration • Justification in the DigBoard Record

u.s. national elections

public policy

congress, government, & law

elections & events in overseas offices

disasters & emergencies (911, katrina)

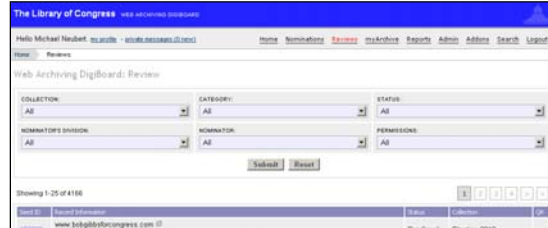
special collections: visual image, manuscript division

civil war and war of 1812 anniversaries

Full list here: <http://www.loc.gov/webarchiving/collections.html>

The DigiBoard Tool

- Current Functions:
 - URL nominations
 - Permissions management
 - Scoping and seed list creation
 - Quality Review
 - Administrative functions & reports



Researcher access: loc.gov/lcwa

The Library of Congress >> More Online Collections

Library of Congress Web Archives *Minerva*

LC Web Archives

Web Archives Available:

- [Crisis in Darfur, Sudan Web Archive, 2006](#)
- [Indian General Elections 2009 Web Archive](#)
- [Indonesian General Elections 2009 Web Archive](#)
- [Iran War 2003 Web Archive](#)
- [Law Library Legal Blawgs Web Archive](#)
- [Library of Congress Manuscript Division Archive of Organizational Web Sites](#)
- [Papal Transition 2005 Web Archive](#)
- [September 11, 2001 Web Archive](#)
- [Single Sites Web Archive](#)
- [United States 107th Congress Web Archive](#)
- [United States 108th Congress Web Archive](#)
- [United States Election 2000 Web Archive](#)
- [United States Election 2002 Web Archive](#)
- [United States Election 2004 Web Archive](#)
- [United States Election 2006 Web Archive](#)
- [United States Election 2008 Web Archive](#)
- [Visual Image Web Sites Archive](#)



The Library of Congress Web Archives (LCWA) is composed of collections of archived web sites selected by subject specialists to represent web-based information on a designated topic. It is part of a continuing effort by the Library to evaluate, select, collect, catalog, provide access to, and preserve digital materials for future generations of researchers. The early development project for Web archives was called MINERVA.

LC Web Archives

The Library of Congress >> More Online Collections
August 5, 2011

The Library of Congress >> More Online Collections

Library of Congress Web Archives *Minerva*

LC Web Archives >> Election 2008 browse >> Browse names

Browse by » [Subjects](#) (1806) [Names](#) (1298) [Titles](#) (2189)

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

<ul style="list-style-type: none"> Abbond, Robert Abercrombie, Neil Abrams, Thomas D. Acayado-Vila, Anibal Ackerman, Gary L. Ackerson, Nels Adams Jr., Eddie Adkins, Will Adler, John H. Agrilla, Christopher M. Akin, William Todd Aldrich, Charles Alesch, Steven J. Alexander, Lamar 	<ul style="list-style-type: none"> Alexander, Robert Darwood Alexander, Rodney Alexander, Stewart A. Allen, Donald Kenneth Allen, Janice Lee Allen, Jonathan Edward Allen, Joseph William Allen, Susan Allen, Thomas H. Altmire, Jason Amata, Aumua Amundson, Gene Andal, Dean E. Anderson, Bob
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

LC Web Archives >> Search all >> Search results

You searched: [-obama]

< Previous | 1 to 12 | Next >

- 2008 Election ProCon.org -- Which candidate would make the best U.S. President?**
2008 Presidential Election ? candidate positions and quotes in a pro and con format on many issues and topics. From: [Election 2008 Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]
- Albert J Franzen for Illinois Senate**
From: [Election 2004 Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]
- BarackObama.com**
Official Website of Barack Obama 2008 Presidential Campaign From: [Election 2008 Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]
- China Hearsay: China law, business, and economics commentary**
From: [Legal Blaws Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]
- Christopher Balding for Congress**
Christopher Balding for Congress in the 34th District of California From: [Election 2008 Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]
- Crack the Bell**
A progressive blog by Tim Kelly, covering state and local politics from our South Carolina base. From: [Election 2008 Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]
- Orlehaus for Congress**
Democratic candidate for United States Congress in Ohio's 1st District. From: [Election 2008 Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]
- Election Projection: 2008 Elections - Polls, Projections, Results**
Your best source for election tracking. State-by-state presidential races, Senate, House and gubernatorial contests - get the lowdown on all of them. Do navigate and fun to watch. From: [Election 2008 Web Archive](#).
[[bibliographic information](#)] [[archived web site](#)]

The Library of Congress >> More Online Collections
August 6, 2011

Title: BarackObama.com

Abstract: Official Website of Barack Obama 2008 Presidential Campaign

Date Captured: March 2, 2007 - November 26, 2008
[Archived Site](#)

Subject(s): United States--Democratic Party
Presidents--United States--Election
Democratic Party--United States
Obama, Barack
Political candidates--United States
Elections--United States
United States--Politics and government--2001-

Language(s): English

Genre: web site

Access Condition: Access restricted to on-site users

URL at time of capture: www.barackobama.com/

Citation ID: http://hdl.loc.gov/loc.natlib/mrva0008.0885

Record ID: mrva0008.0885

Collection Title: [Election 2008 Web Archive](#)

The Library of Congress >> More Online Collections

Library of Congress Web Archives *Minerva*

BROWSE | SEARCH | TECHNICAL INFORMATION

home >> previous page >> Election 2008 Web Archive Collection Resource Page



Election 2008 Web Archive

<http://www.barackobama.com/>

2008 2007

January 02*, 09*, 16*, 23*, 30*	February 06*, 13*, 20*, 27*	March 05*, 12*, 19*, 26*	April 02*, 09, 10, 16, 17, 23*, 30*	May 07*, 14*, 21, 22, 28*	June 05*, 11, 12, 10, 19, 25, 26
July 02, 03, 09, 10, 16, 17, 23, 24, 30, 31	August 06, 07, 13, 14, 20, 21, 27, 28	September 03, 04, 11*, 17, 20, 24*	October 01, 02, 08, 09, 15, 16, 22, 23, 29, 30	November 06*, 12*, 19*, 26*	December

96 archived captures for 2008

* denotes more than one capture for that date

[< Back to collection page](#)

Wayback 1

home >> previous page >> Election 2008 Web Archive Collection Resource Page

The Library of Congress >> More Online Collections

Contact Us

LIBRARY OF CONGRESS

Note: External links, forms and search boxes may not function within this collection

Election 2008 Web Archive Collection

This is an archived Web site from the Library of Congress
<http://www.barackobama.com/index.php>
 Archived: 02/07/2008 at 00:48:54

EN ESPAÑOL

OBAMA [Back to previous page](#) [First \(07/01/2007\)](#) [Previous \(#28 of 70\)](#) [Next](#) [Last \(11/27/2008\)](#)

OBAMA '08

GET INVOLVED
SIGN UP FOR EMAIL UPDATES

Email Address:

Zip Code: SIGN UP

LEARN ISSUES MEDIA ACTION PEOPLE STATES BLOG STORE

[DONATE NOW](#)

MAKE A DIFFERENCE

- [Find Events](#)
- [Make Calls](#)
- [Register to vote](#)
- [Volunteer](#)

OBAMA BLOG

FEBRUARY 6, 2008

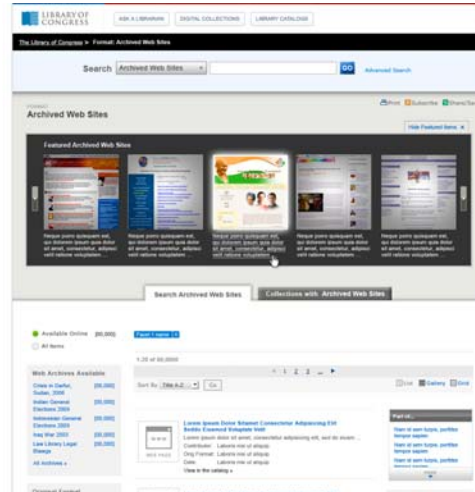
Obama Wins Super Tuesday: Wins Most States, Wins Most

OTHER STORIES

Feb 09 Where Obama Stands: On Service

Future: Access Improvements

- More seamless integration with other types of Library content
- Better presentation of sites archived over time (when URL changes, or if they appear in multiple collections)
- Better search



Collaborative Archives



LIBRARY OF CONGRESS

Benefits of Collaborative Archiving

- React more quickly to rapidly unfolding events
- Archive more content
- Variety of partners bring their own expertise to table: subject, technical, languages
- Learn from others, share with others



LIBRARY OF CONGRESS

Examples

- September 11 Web Archive
- Hurricanes Katrina and Rita
- End of Term Government Archive
- Earthquake in Haiti
- Jasmine Revolution
- North Africa and Middle East (Arab Spring)
- Japanese Earthquake
- Olympics 2012

End of Term Archive

The screenshot shows the 'End of Term Web Archive' website. At the top left is the Library of Congress logo. The main title is 'End of Term Web Archive' with the subtitle 'US Federal Government Websites 2008-2009'. Below the title are navigation tabs: Home, Search Full Text, Site List, and Explore Data. On the left side, there are links for Overview, Project Background, Project Partners, and End of Term 2012. The central content area features a screenshot of a government website titled 'Committee on Natural Resources, Republican Site' with a date of 'Sep. 15, 2008'. To the right of this screenshot, there is text explaining the archive's purpose and a list of tweets related to the project.

EOT Project Goals

- Work collaboratively to preserve public U.S. government websites at the end of the Bush presidential administration ending January 19, 2009.
- Document federal agencies' presence on the web during the transition of Presidential administrations.
- To enhance the existing research collections of the five partner institutions.

EOT Partners

- Library of Congress
- Internet Archive
- California Digital Library
- University of North Texas
- US Government Printing Office

EOT 2012-2013

- Actively soliciting nominations of U.S. federal government domains from anyone willing to help (contact eotproject@loc.gov)
- Crawl began this month
- Related Election 2012 project, with Harvard
- Read more:
<http://blogs.loc.gov/digitalpreservation/2012/08/collaborating-to-identify-government-or-election-related-websites-to-preserve/>

Questions?

The screenshot shows the Library of Congress website's 'Web Archiving' page. At the top, there is a navigation bar with 'LIBRARY OF CONGRESS', 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY CATALOGS'. A search bar is on the right. Below the navigation, the page title is 'The Library of Congress > Web Archiving'. The main content area is titled 'WEB ARCHIVING' and includes a sub-header 'Web Archiving'. There are three columns of content: a left sidebar with links like 'Web Archiving Home', 'Technical Information', 'Collections', 'FAQs', and 'Contact'; a middle column with a collage of web pages and a 'History of Web Archiving at the Library of Congress' section; and a right column with a 'Resources' section. The 'History' section contains three paragraphs detailing the program's start in 2000, international collaboration in 2003, and the creation of a Web Archiving Team in 2004.

LIBRARY OF CONGRESS ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS Search

The Library of Congress > Web Archiving

WEB ARCHIVING Print Subscribe Share/Save

Web Archiving

Resources

- Web Archiving Home
- Technical Information
- Collections
- FAQs
- Contact

History of Web Archiving at the Library of Congress

In 2000, the Library of Congress established a pilot project to collect and preserve websites. A multidisciplinary team of Library staff studied methods to evaluate, select, collect, catalog, provide access to, and preserve these materials for future generations of researchers.

The Library has developed thematic web archives on such topics as the United States National Elections, the Iraq War, and the events of September 11. More about these collections plus many other available collections can be found at the [Library of Congress Public Web Archiving](#) website.

In July 2003, the Library and the national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden, the British Library (UK), and the Internet Archive (USA) acknowledged the importance of international collaboration for preserving Internet content for future generations and formed the [International Internet Preservation Consortium](#). The goals of the Consortium include collecting a rich body of Internet content from around the world and fostering the development and use of common tools, techniques and standards that enable the creation of international archives.

In 2004, the Library's Office of Strategic Initiatives created a Web Archiving Team to support the goal of managing and sustaining at-risk digital content. The team is charged with building a Library-wide understanding and technical infrastructure for capturing web content. The team, in collaboration with a variety of Library staff, and national and international partners, is identifying policy issues, establishing best practices and building tools to collect and preserve web content.

<http://www.loc.gov/webarchiving>
abgr@loc.gov