



A Time Machine for Federal Information

Using Web Archive content in government
information reference work

June 13, 2017

Ashley Dahlen

Senior Outreach Librarian

U.S. Government Publishing Office



Agenda

- Overview of FDLP Web Archive
- When to use web archive content in reference work
- Examples and demo
- Drawbacks
- How libraries can take advantage of the web content being saved



Overview of the FDLP Web Archive



- Captures website content in native interface
- Websites regularly crawled, indexed and searchable on Archive-It, cataloged in CGP
- Point-In-Time captures or 'snap shots'
 - Work around link rot
 - Over time, snapshots capture changes made to websites

Types of file formats and sites captured



Capital Punishment in the United States, 2013 - Statistical Tables December 31, 2013, and persons executed in 2014.

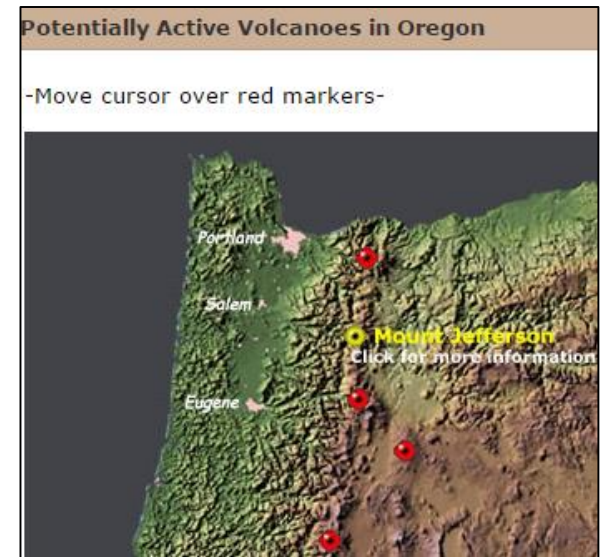
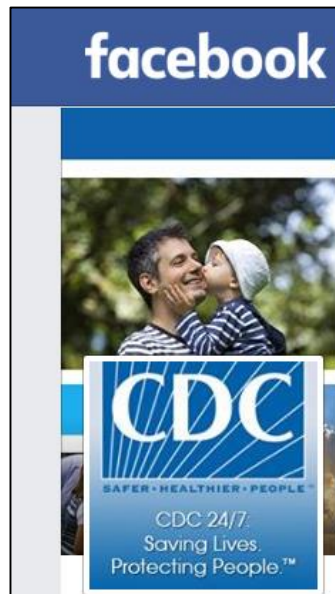
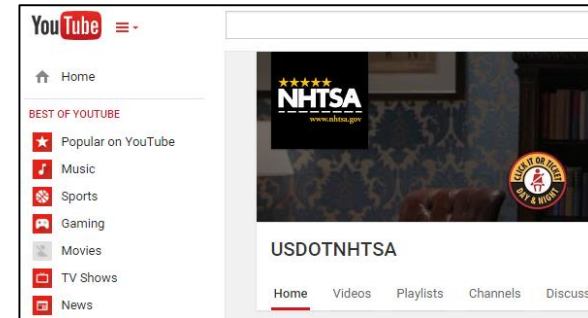
PDF (1M) | ASCII file (34K) | Comma-delimited format (CSV) (Zip)
Part of the *Capital Punishment Series*

Capital Punishment, 2012 - Statistical Tables (Revised) Presenting data for December 31, 2012, and persons executed in 2012.

PDF (765KB) | ASCII file (14KB) | Zip format (28KB)
Part of the *Capital Punishment Series*

Capital Punishment, 2011 - Statistical Tables Presents characteristics of persons executed in 2011.

PDF (1.1M) | ASCII file (14K) | Comma-delimited format (CSV) (Zip)
Part of the *Capital Punishment Series*





How GPO captures Federal websites



Take a look at the LSCM Workflow in a 2014 webinar:

[Web Archiving for the FDLP](#)

<http://login.icohere.com/public/topics.cfm?cseq=1172>



What to capture and what to skip:

- Must be within scope of FDL P
- Not distributing through print
- Content less likely to be cataloged
- Avoid duplication of effort with other institutions
- What is important to FDL P community

Examples:

- SuDoc 'Y3 sites': commissions, committees, independent agencies
- Non standard government sites or jointly managed sites like www.benefits.gov

Nominate websites for the FDLP Web Archive

Three ways to nomination a site:

- [Document Discovery](#)
- [AskGPO](#)
- fdlpwebarchiving@gpo.gov

Please do not nominate content already in FDsys or already in Archive-It!

Avoiding Duplication - Collaboration



GPO does not capture Congressional Committee websites because that is captured in:

- **Congressional Web Harvest**
 - Done once by NARA in 2004, since done collaboratively by EOT group every two years (end of every congressional term)
 - www.webharvest.gov/
- **End of Term Web Archive**
 - Jointly done every 4 years (end of every presidential term)
 - <http://eofarchive.cdlib.org/>

FDLP Web Archive Statistics

- 132 collections on Archive-It
- 9.2 terabytes (TB) of data (total)
- Over 70,320,000 URLs
- 152 Records in CGP



The Library of Congress has over **290 TB** of data. For **FY 2016**.



Access and Discoverability

1) **Catalog of U.S. Government Publications**
(CGP) <https://catalog.gpo.gov/F>

2) **Archive It** interface <https://archive-it.org/>



Caution – boring looking slide!!!

1) CGP basic search



CATALOG OF U.S. GOVERNMENT PUBLICATIONS (CGP)

Catalogs to Search:
[Congressional Serial Set](#) [Historic Shelflist](#) [FDLP Web Archive](#) [Serials](#)
[Congressional Publications](#) [Electronic Titles](#) [Gov eBooks](#)

My Options:
[Bookshelf](#) [Results List](#)
[Preferences](#) [Previous Searches](#)

[Catalog of U.S. Government Publications Home Page](#) > [National Bibliography of U.S. Government Publications](#)

Welcome to the Catalog of U.S. Government Publications (CGP)!

The CGP is the finding tool for federal publications that includes descriptive information for historical and current publications as well as direct links to the full document, when available. Users can search by authoring agency, title, subject, and general keywords, or click on "Advanced Search" for more options.

The catalog offers you the option to find a nearby [Federal Depository Library](#) that has a particular publication or that can provide expert assistance in finding and using related U.S. government information. Click on the title of interest from your search results list. Then click on the Locate in a Library link within the displayed record.

Also try our federated search [across multiple U.S. Federal Government databases](#)

type [webarch](#) + keyword in search box

[Bookshelf](#) - [Preferences](#) - [Results List](#) - [Previous Searches](#) - [Advanced Search](#) - [FDLP Sign In](#)

Example: wildlife "environmental impact statement" NOT draft

Keyword(s): ▼

Catalogs: ▼



Basic

Advanced

Expert

Browse

New Titles

Help

About

FDLP Login

MetaLib

CATALOG OF U.S. GOVERNMENT PUBLICATIONS (CGP)

Catalogs to Search:

- [Congressional Serial Set](#)
- [Historic Shelflist](#)
- [FDLP Web Archive](#)
- [Serials](#)
- [Congressional Publications](#)
- [Electronic Titles](#)
- [Gov eBooks](#)

My Options:

- [Bookshelf](#)
- [Results List](#)
- [Preferences](#)
- [Previous Searches](#)

[Catalog of U.S. Government Publications Home Page](#) > National Bibliography of U.S. Government Publications

Advanced Search: Full Catalog

Search for word/"phrase" in...

Enter word/"phrase"

Keyword(s)

Keyword(s)

Keyword(s)

holocaust AND

AND

or limit search by:

[Search Hints](#)

Limit search by:

Year/s: to: yyyy (Use ? for truncation when not using from/to)

Format:

Language:

Catalogs:

- National Bibliography of U.S. Government Publications
- Congressional Serial Set Catalog
- Congressional Publications

- Search hints:
- Fill in as little or
 - Use quotation
 - For combination searches using a phrase and a word, put your phrase in quotation marks then use a Boolean Operator (AND, OR or NOT) w

FDLP Web Archive

er your search.

connect your term on each line to further narrow a search. Example: line 1 type "public law" NOT resolution. Select operator drop down NOT

CGP Advanced Search

Quick link on the CGP page home page



The screenshot shows a web browser window with the address bar displaying <https://catalog.gpo.gov/F>. The page title is "Entire CGP Catalog - Basic". The navigation menu includes "Basic", "Advanced", "Expert", "Browse", "New Titles", "Help", "About", "FDLP Login", and "MetaLib". The main heading is "CATALOG OF U.S. GOVERNMENT PUBLICATIONS (CGP)". Below the heading, there are several circular icons representing different search and navigation functions. The "Catalogs to Search:" section lists "Congressional Serial Set", "Historic Shelflist", "FDLP Web Archive" (highlighted with a red box), "Serials", "Congressional Publications", "Electronic Titles", and "Gov eBooks". The "My Options:" section includes "Bookshelf", "Results List", "Preferences", and "Previous Searches". The breadcrumb trail at the bottom reads "Catalog of U.S. Government Publications Home Page > National Bibliography of U.S. Government Publications".

FDLP Web Archive

Welcom

The Federal Depository Library Program (FDLP) Web Archive is comprised of selected U.S. Government Web sites, harvested and archived in order to create working "snapshots" of the Web sites at various points in time. The aim is to provide permanent public access to Federal Agency Web content.



Type word or phrase

Field to search

New FDLP Web Archive Records:

[Last 7 days](#)

[Last Two Weeks](#)

[This Month](#)

[Last Month](#)

[Last Three Months](#)

[Please click here](#) to view all current FDLP Web Archive titles in the CGP. For Help information, please see the [CGP Help page](#).

Example of Search Result

Records 1 - 1 of 1 (maximum display and sort is 2000 records)

Jump to #

INTERNET in SuDoc number means it is a part of the FDLP Web Archive

Previous Page Next Page

| # | Title | Year | Author | SuDoc Number | Internet Access |
|----------------------------|--|------|--|---------------------|--|
| 1 <input type="checkbox"/> | United States Holocaust Memorial Museum. | 20uu | United States Holocaust Memorial Museum, | Y 3.H 74:2 INTERNET | http://purl.fdlp.gov/GPO/gpo27784 http://purl.fdlp.gov/GPO/gpo73232 |

Previous Page Next Page

Full View of Record

Choose format: [Standard](#) | [Short](#) | [MARC](#)

Record 1 out of 1

| | |
|------------------------|---|
| Title | United States Holocaust Memorial Museum. |
| Format | online resource |
| Internet Access | <p>→ Archived site http://purl.fdlp.gov/GPO/gpo27784</p> <p>→ Collection page seed list http://purl.fdlp.gov/GPO/gpo73232</p> |
| Author | United States Holocaust Memorial Museum, author. |
| Published | [Washington, DC] : [United States Holocaust Memorial Museum] |
| Distributed | [Washington, D.C.] : Government Printing Office, -December 16, 2014. [Washington, D.C.] : Government Publishing Office, December 17, 2014- |
| SuDoc Number | Y 3.H 74:2 INTERNET |

Search Result

| | |
|------------------------|---|
| Subject | United States Holocaust Memorial Museum. |
| Subject - LC | Holocaust, Jewish (1939-1945) -- Museums -- Washington (D.C.) |
| | Holocaust, Jewish (1939-1945) |
| | Holocaust survivors -- Services for -- United States. |
| | Holocaust survivors' families -- Services for -- United States. |
| | Genocide -- Prevention. |
| Added Entry | Federal Depository Library Program, collector. |
| | United States. Government Printing Office, distributor. |
| Linking Field | Current version: United States Holocaust Memorial Museum (OCoLC)44312956 |
| URL | URL at time of capturehttp://www.ushmm.org/ |
| CGP Record Link | https://catalog.gpo.gov/F/?func=direct&doc_number=000919212&local_base=WEBARCH |
| Holdings | All items |
| holdings (2) | Item (no year) |
| OCLC Number | (OCoLC)869749005 |
| System Number | 000919212 |



CGP records may also include

PURLs to other agency web archives

Example: National Institute of Nursing Research



PURL: <http://purl.fdlp.gov/GPO/gpo37375>
CGP system number: 901666

2) Archive-It Interface

<https://archive-it.org/home/FDLPwebarchive>

- Search for “**GPO**” or “**FDLP**” from main search box to find FDLP Web Archive content
- A FDLP Web Archive ‘collection’ is an agency or specific website resource, along with all of associated sub-domains (Twitter, Facebook, etc.)
- Search through FDLP Web archive or within a specific collection
 - You are searching all the website text that has been indexed and can search URLs.



How can you use web archive content?

- When you cannot find what you know used to be on an agency website
- When the Government shuts down or when a website goes down
- When an agency or commission ceases to exist
- When a researcher is studying a point-in-time – See what has been captured
- When a researcher finds a reference to something likely to be on a website, but only for a short period of time (e.g. something time specific like the Olympics)

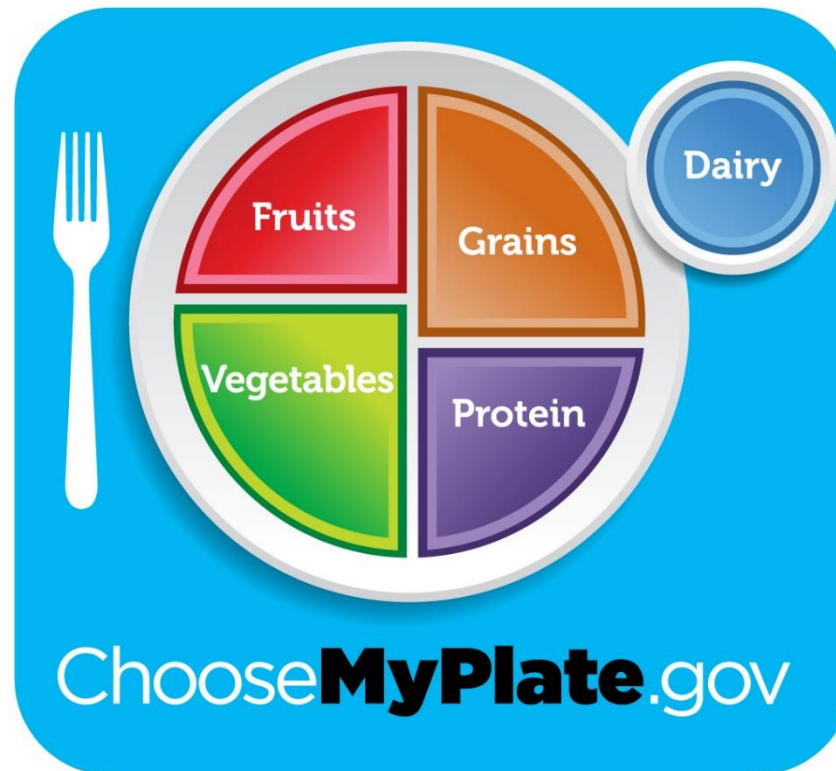


How can you use web archive content?

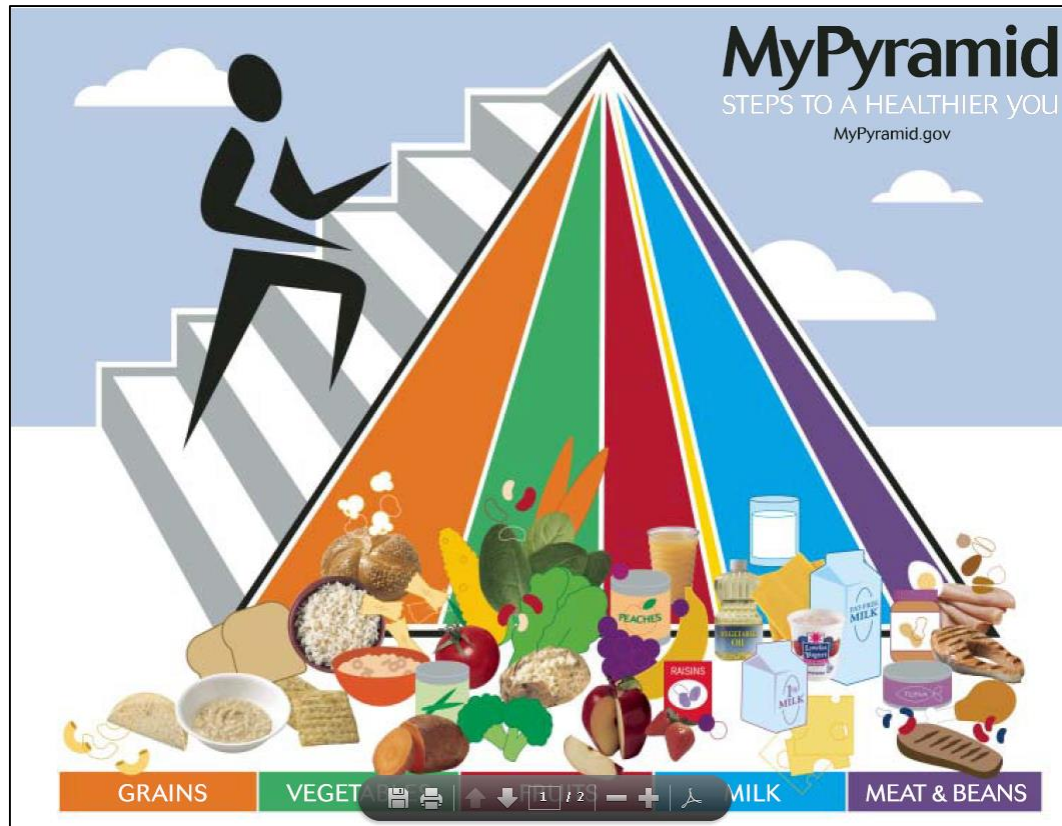
- When a researcher wants to find all iterations of something or track the evolution of something over time
- When a researcher is trying to gauge how an agency collaborated with the public or engaged in eServices
- Researching a recurring Federal information topic
- When searching for older titles, tutorials, & guides
- When the content was not picked up in the FDLP (i.e. it is 'fugitive')
- When the content is likely to be overwhelmed by non-official or secondary resources in a normal Internet search



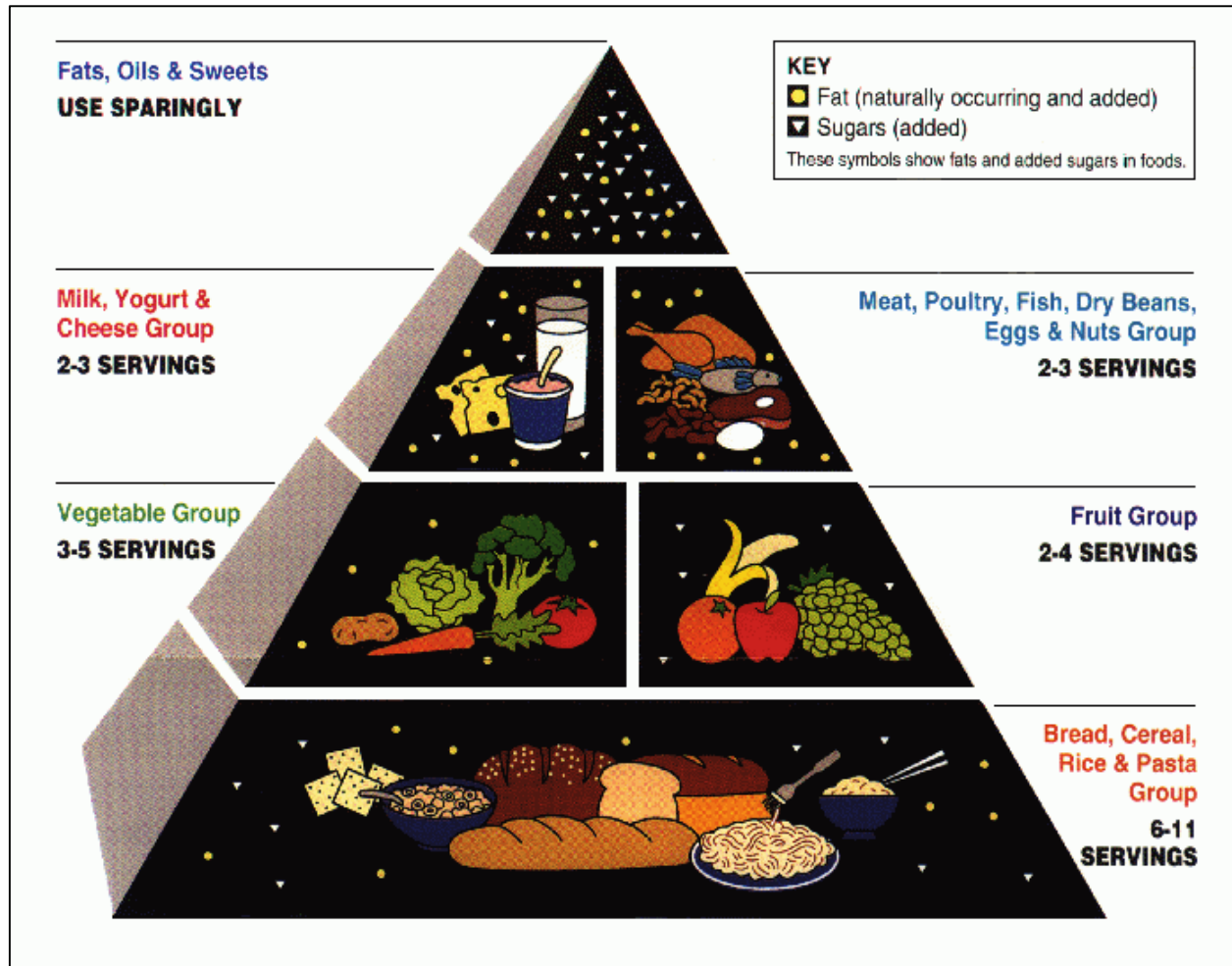
Example one – seen this?



Remember this?

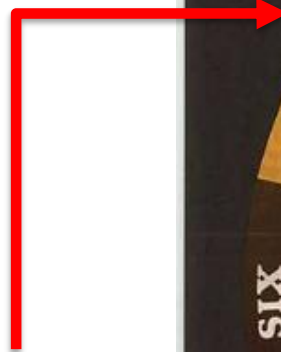


How about this?

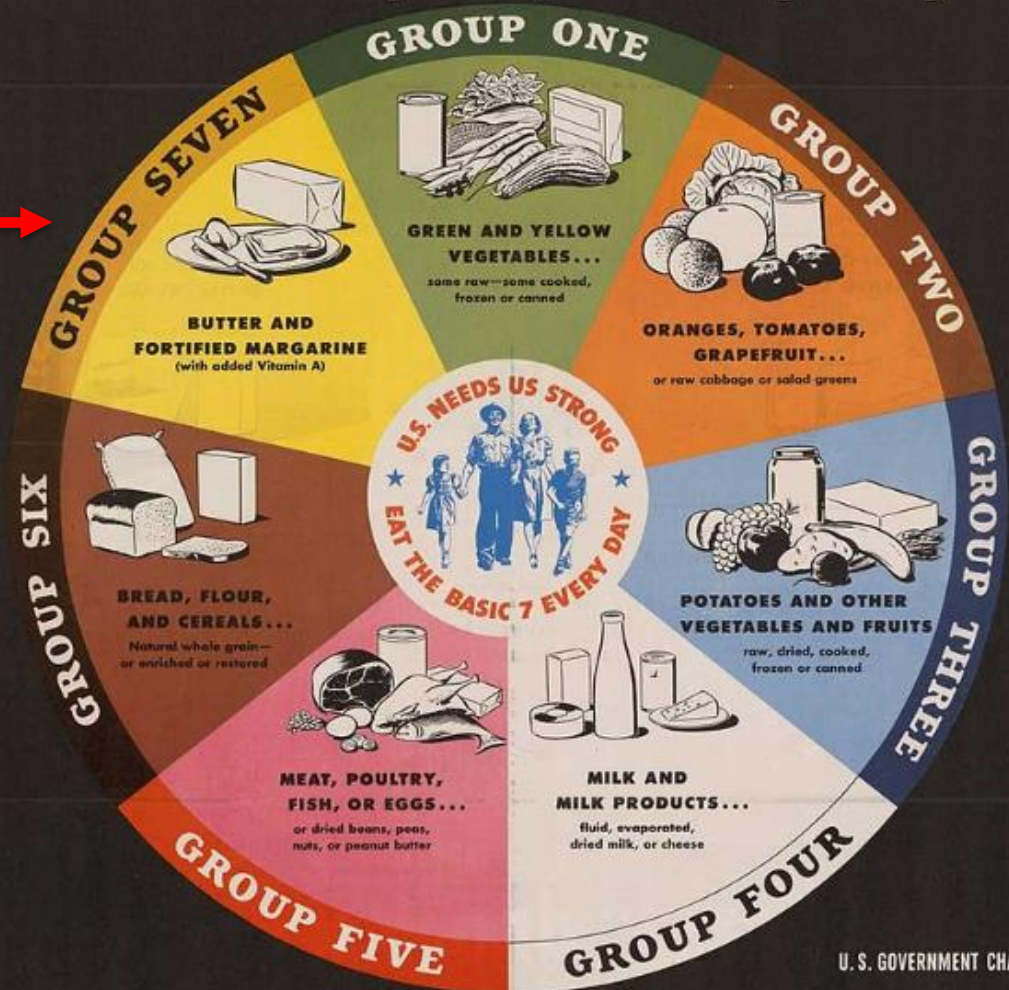


And from
WWII...

*For Health...eat some food
from each group...every day!*



*Have you consumed your
recommended daily amount
of butter and fortified
margarine?*



**IN ADDITION TO THE BASIC 7...
EAT ANY OTHER FOODS YOU WANT**

Example two – reference found to traveling exhibit

NAZI OLYMPICS BERLIN 1936

| | | | | | | |
|-----------------------------|-------------------------|------------------------|-------------------------|---------------------------|------------------------------|-------------------------------|
| AUGUST 1936 | GERMANY | SPORTS | BOYCOTT | TO BERLIN | THE OLYMPICS | THE AFTERMATH |
|-----------------------------|-------------------------|------------------------|-------------------------|---------------------------|------------------------------|-------------------------------|

Languages: [English](#) | [简体中文](#) | [Español](#) | [عربي](#) [Home](#) | [Site Menu](#)

August 1936 [Previous](#) | [Next](#)



▲ The last of 3,000 runners who carried the Olympic torch from Olympia, Greece, arrives in the Lustgarten in Berlin to light the Olympic Flame and start the 11th Summer Olympic Games.
—USHMM #21674/Bettmann/CORBIS

For two weeks in August 1936, Adolf Hitler's Nazi dictatorship camouflaged its racist, militaristic character while hosting the Summer Olympics. Minimizing its antisemitic agenda and plans for territorial expansion, the regime exploited the Games to impress many foreign spectators and journalists with an image of a peaceful, tolerant Germany. Having rejected a proposed boycott of the 1936 Olympics, the United States and other western democracies missed the opportunity to take a stand that contemporary observers claimed might have restrained Hitler and bolstered international resistance to Nazi tyranny. After the Olympics, Germany's expansionism and the persecution of Jews and other "enemies of the state" accelerated, culminating in World War II and the Holocaust.



Tracking it down

- Step 1 – Use any URL found in the news article in a regular search engine (e.g. google)
- Step 2 – Use any URL found in the news article in the Internet Archive’s [Wayback Machine](https://archive.org/web/) (<https://archive.org/web/>)
- Step 3 – Identify if the agency website has been captured in a web archive



An honest assessment about access...

1) Catalog record:

- a) Browse archived collections (if in CGP)
- b) Author is typically the website agency
- c) Subject headings, e.g. **United States Holocaust Memorial Museum:**
 - Holocaust, Jewish (1939-1945) -- Museums -- Washington (D.C.)
 - Holocaust, Jewish (1939-1945)
 - Holocaust survivors -- Services for -- United States.
 - Holocaust survivors' families -- Services for -- United States.
 - Genocide -- Prevention.
- c) Keyword search in notes field

An honest assessment about access...

2) The Archive-It User eXperience (UX)

- Archive-It is very responsive to feedback about the tool.
- They are focused on the content management aspect.
- User interfaces can be developed by the institution.

Example of secondary interface portal/site

Columbia University - Human Rights Web Archive:

- Locally-hosted: <https://hrwa.cul.columbia.edu/>
- Archive-It hosted: <https://archive-it.org/collections/1068>



LEARN MORE | NOMINATE A SITE | CONTACT

The Human Rights Web Archive @ Columbia University is a searchable collection of archived copies of human rights websites created by non-governmental organizations, national human rights institutions, tribunals and individuals. Collecting began in 2008 and has been ongoing for active websites. New websites are added to the collection regularly.

The HRWA is an initiative of the **Center for Human Rights Documentation & Research** and is a key focus of the Columbia University Libraries' **Web Resources Collection Program**. The HRWA was made possible by generous support from the Andrew W. Mellon Foundation. [Learn More](#)

FEATURED WEBSITES | SUBJECTS | PLACES | LANGUAGES

A selection of sites relating to the subject **Ombudspersons**.

Defensoría del Pueblo de la República...

[Site Details](#)

Ombudsman Republike Srpske--zaštitt...

[Site Details](#)

Volksanwaltschaft

[Site Details](#)

ARCHIVE-IT | HOME | EXPLORE | LEARN MORE | CONTACT US

The leading web archiving service for collecting and accessing cultural heritage on the web. *Built at the Internet Archive*

Explore >> Columbia University Libraries >> Human Rights

Center for Human Rights Documentation & Research

Human Rights
Collected by: Columbia University Libraries
Archived since: May, 2008
Description: An initiative of CUL's Center for Human Rights Documentation and Research, the Human Rights Web Archive is an effort to preserve and ensure access to freely available human rights resources created mainly by non-governmental organizations, national human rights institutions, and individuals.
Subject: Society & Culture, Human rights, Non-governmental organizations, Human rights workers, National human rights institutions, Web archives
Collector: Columbia University Libraries, Center for Human Rights Documentation and Research

Narrow Your Results

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

Group **Count** | (A-Z)

- Amnesty International sections (54)
- Blogs by individuals (12)
- National human rights institutions (97)
- Non-governmental organizations (524)
- Truth commissions, tribunals, and courts (22)

Subject **Count** | (A-Z)

- Human rights (564)
- Human rights advocacy (197)
- National human rights institutions (95)
- Civil rights (57)
- Democracy (52)

Enter search terms here

Sites

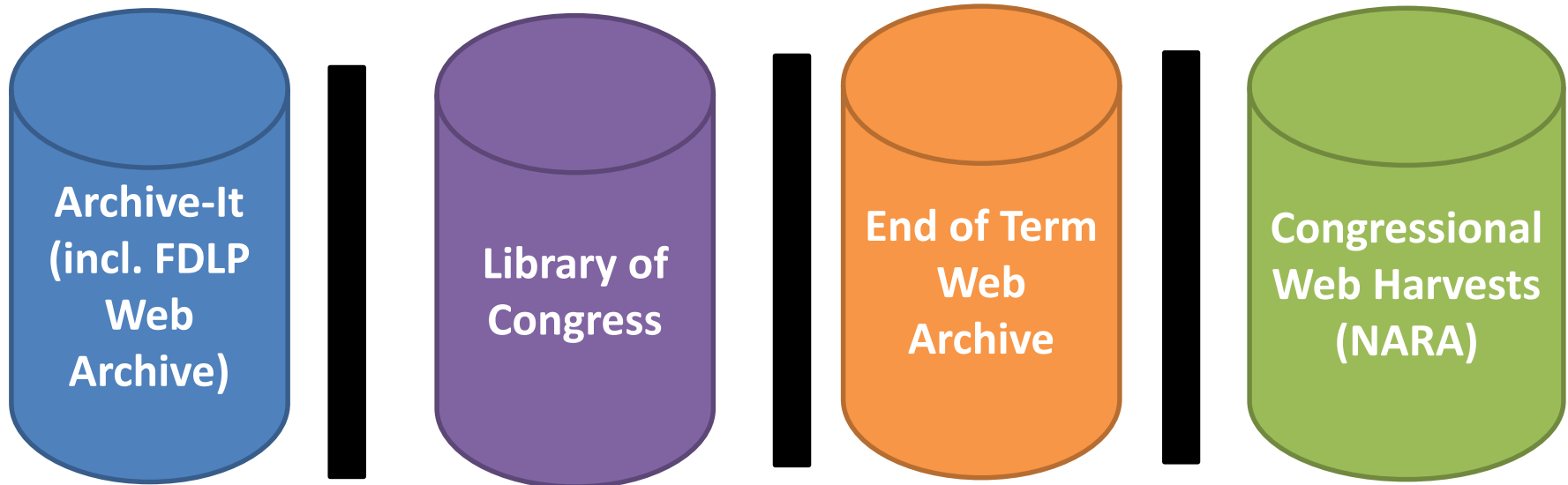
Page 1 of 8 (729 Total Results)

Sort By: **Title (A-Z)** | Title (Z-A) | URL (A-Z) | URL (Z-A)

Title: Women's Initiatives for Gender Justice
URL: <http://4genderjustice.org/>
Description: "The Women's Initiatives for Gender Justice (Women's Initiatives) is an international women's human rights organisation that advocates for gender justice through the International Criminal Court (ICC) and through domestic mechanisms, including peace negotiations and justice processes. We work with women most affected by the conflict situations under investigation by the ICC, and collaborate with more than 6,000 grassroots partners, associates and members across multiple areas of focus."

Free Reference Tip - Explore Other Web Archive Projects

Web archive sites are silos. There is NO meta search.



Silos make me





Silo: End of Term Web Archive

Goal is to capture:

- Agency websites (legislative, executive, judicial)
- Congressional member websites (including social media)

...before the new administration comes in

- Partners have varied over the years
- 2008, 2012, & 2016 captures are valuable information about the Federal Government at those time periods. (2004 captured in [NARA's Presidential Term Web Harvest](#))

<http://eotarchive.cdlib.org/>

Silo: United States Congressional WebArchive (Library of Congress)

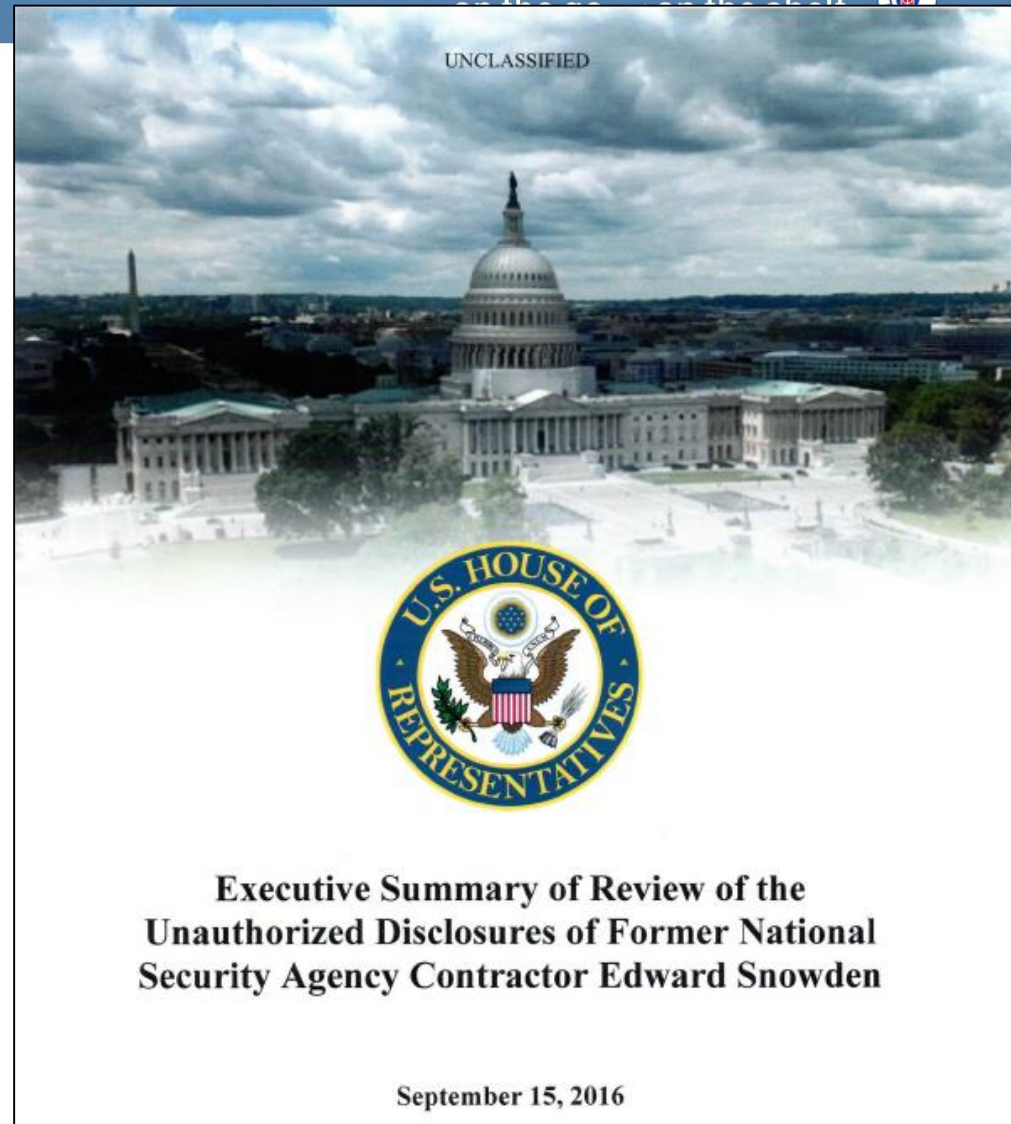
Goal is to capture:

- Member websites of from the House of Representatives and Senate, beginning with the end of the 107th Congress
 - Committee websites on a monthly basis
-
- Uses same crawling technology (Heritrix) used by GPO, but stores it in BagIt Library and replays the content via a local installation of the Wayback Machine.

<https://www.loc.gov/collections/united-states-congressional-web-archive/about-this-collection/>

Discovering fugitive content in other web archives

- ‘Fugitive’ from the FDLP
 - Never ordered to be printed or put on FDsys
- Will hopefully be captured in EOT Web Archive and NARA’s Congressional Web Archive



House Permanent Select Committee on Intelligence:
http://intelligence.house.gov/uploadedfiles/hpsci_snowden_review_-_unclass_summary_-_final.pdf



Example of what can be found in Google versus web archives

On September 11, 2012,
four Americans, including two diplomats,
were killed in Benghazi, Libya...

Question:

“What was the State Department’s official response?”

Google:

“Department of State” response diplomat death Benghazi site:.gov

Results:

- Archived press releases
- Numerous reports

Note numerous PDFs

About 3,710 results (1.07 seconds)



[PDF] Benghazi: Where is the State Department Accountability?
<https://foreignaffairs.house.gov/.../HFAC%20Majority%20Staff%20Report%20on%20...> killing U.S. Ambassador Christopher Stevens, and U.S. officials Sean Smith, Tyrone Chairman Royce introduced H.R. 2848, the Department of State Operations and Embassy The ARB that convened in response to the 1998 attacks on.

Denied: US Diplomats in Libya Requested More Security - United ...
<https://oversight.house.gov/denied-us-diplomats-in-libya-requested-more-security/> Oct 2, 2012 - Denied: US Diplomats in Libya Requested More Security ... 2012 terrorist attack that killed U.S. Ambassador J. Christopher Stevens. ... U.S. Department of State ... April 6, 2012, BENGHAZI – Two Libyans who had been fired from a ... Ambassador Stevens' murder please prepare a written response to the ...



[PDF] Accountability Review Board (ARB) Report - US Department of State
<www.state.gov/documents/organization/202446.pdf> The Benghazi attacks represented the first murder of a U.S. ambassador The Board found the Libyan government's response to be profoundly lacking on the ...

Remarks on the Deaths of American Personnel in Benghazi, Libya
<www.state.gov> > ... > Remarks by Secretary Clinton: September 2012 Yesterday, our U.S. diplomatic post in Benghazi, Libya was attacked. ... that took place at our Embassy in Cairo yesterday, as a response to inflammatory material posted on the internet.



[PDF] Securing U.S. Diplomatic Facilities and Personnel Abroad ...
<fpc.state.gov/documents/organization/230753.pdf> by A Tiersky - 2014 - Cited by 7 - Related articles
Jul 30, 2014 - The deaths of Ambassador Christopher. Stevens and three other U.S. Department of State Actions in Response to the Benghazi Attack .



[PDF] Bureau of Diplomatic Security - US Department of State
<www.state.gov/documents/organization/225841.pdf> the U.S. compound in Benghazi, ... a terrorist murder squad appeared suddenly and attempted to blast and shoot its way ... back up the initial DS response.

What google will not return:

State Department Home Page (captured September 15, 2012)



Web archive content is not typically indexed by Internet browsers – it would confuse users

Missing from Google:

- Blog posts on the State Dept. blog DipNote

(No I am not making that title up)

DIPNOTE

U.S. DEPARTMENT of STATE OFFICIAL BLOG

HOME | ABOUT DIPNOTE | VIDEOS | PHOTOS | ARCHIVE


Honoring Those Lost in Attacks in Libya

POSTED BY DIPNOTE BLOGGERS / SEPTEMBER 14, 2012



On September 14, 2012, President Barack Obama and Secretary of State Hillary Rodham Clinton delivered remarks at the Transfer of Remains Ceremony to honor those lost in attacks in Benghazi, Libya. Secretary Clinton began her [remarks](#) by saying, "...Today we bring home four Americans who gave their lives for our country and our values. To the families of our fallen colleagues, I offer our most heartfelt condolences and deepest gratitude."

Missing from google: State Dept. Tumblr account in EOT

The video was not
successfully captured 

But now you know
a video exists...

= your “Gov Doc”
sleuthing skills are
needed

Benghazi, [Libya](#), [Maghreb](#).' and 'Source:youtu.be'." data-bbox="434 206 980 847"/>

U.S. Secretary of State Hillary Rodham Clinton comments on Benghazi during her remarks on Democratic Transitions in the Maghreb at the Center for Strategic and International Studies (CSIS) in Washington, D.C., October 12, 2012. [Go to <http://video.state.gov> for more video and text transcript.]

Posted 5 days ago 5 notes · View comments

Tagged: [Benghazi](#), [Libya](#), [Maghreb](#). Source:youtu.be

Tip – in YouTube search **Benghazi StateVideo**



Gov Docs Job Security

Experts are still needed to apply facets of government information to track it down:

- What agency produced the information?
- When was it produced?
- What medium/channel/resource name would the information be disseminated through?
- Stop pounding your head on your keyboard – think of it as a challenge!
- Where can that information resource be found?

Questions?

FDLPOutreach@gpo.gov

202-512-1119

