



# Bridging the Divide

## Preserving Government Information in All Formats

- David Walls, Preservation Librarian, LSCM, GPO
- Thomas Mills, Director of Notre Dame, Kresge Law Library
- Jana Ronan, Interim Regional Government Documents Librarian, Chair, Humanities & Social Sciences Library, University of Florida, Library West
- Gwen Sinclair, Head of Government Documents & Maps, University of Hawai'i at Mānoa Library
- Chelsea Dinsmore, Director of Digital Production Services, University of Florida, George A. Smathers Libraries
- Dory Bower, Senior Archival Specialist, LSCM, GPO
- Abbie Grotke, Lead Information Technology Specialist, Web Archiving Team, Library of Congress



## Preservation Stewards

Suzanne Ebanues, Senior Planning & Development Specialist, Partnership Program Coordinator, GPO

- Thomas Mills, Director of Notre Dame, Kresge Law Library
- Jana Ronan, Interim Regional Government Documents Librarian, University of Florida Library
- Gwen Sinclair, Head of Government Documents and Maps, University of Hawaii



## Digital Content Contributors

David Walls, Preservation Librarian, GPO

- Guidance Documentation for Digital Content Contributor Partners
- Chelsea Dinsmore, Director of Digital Production Services, University of Florida Library



## Archiving Federal Web Content

Dory Bower, Senior Archival Specialist, GPO

Abbie, Grotke, Lead Information Technology Specialist, Web Archiving Team, Library of Congress



## Notre Dame Law School, Kresge Law Library



**Thomas Mills**  
Director, Kresge Law Library  
[tmills@nd.edu](mailto:tmills@nd.edu)



## University of Florida, Library West



**Jana Ronan**  
**Interim Regional Government Documents Coordinator**  
**Chair, Humanities & Social Sciences Library**  
**[jronan@ufl.edu](mailto:jronan@ufl.edu)**



## The University of Hawai'i at Mānoa experience



**Gwen Sinclair**  
**Head of Government Documents & Maps**  
**University of Hawai'i at Mānoa Library**  
**[gsinclair@Hawaii.edu](mailto:gsinclair@Hawaii.edu)**



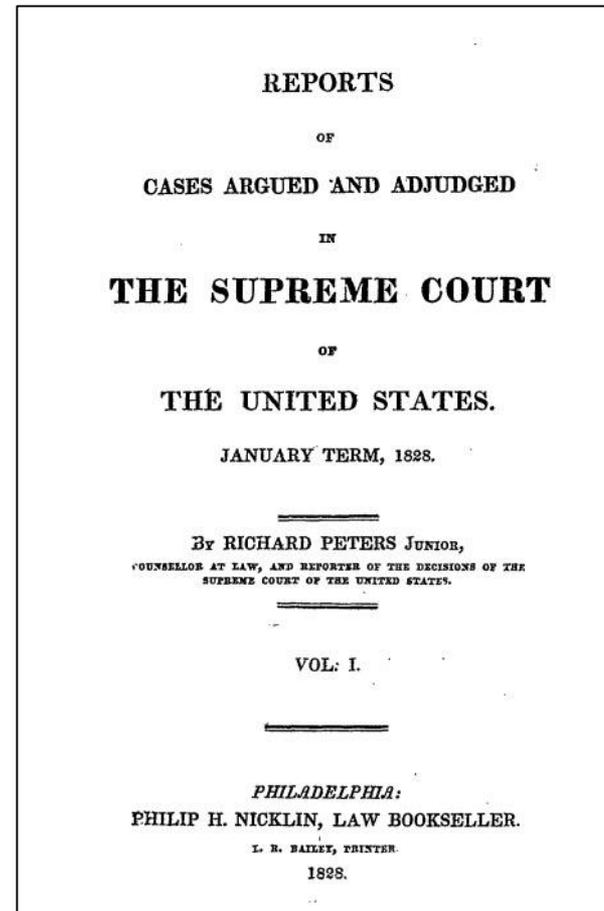
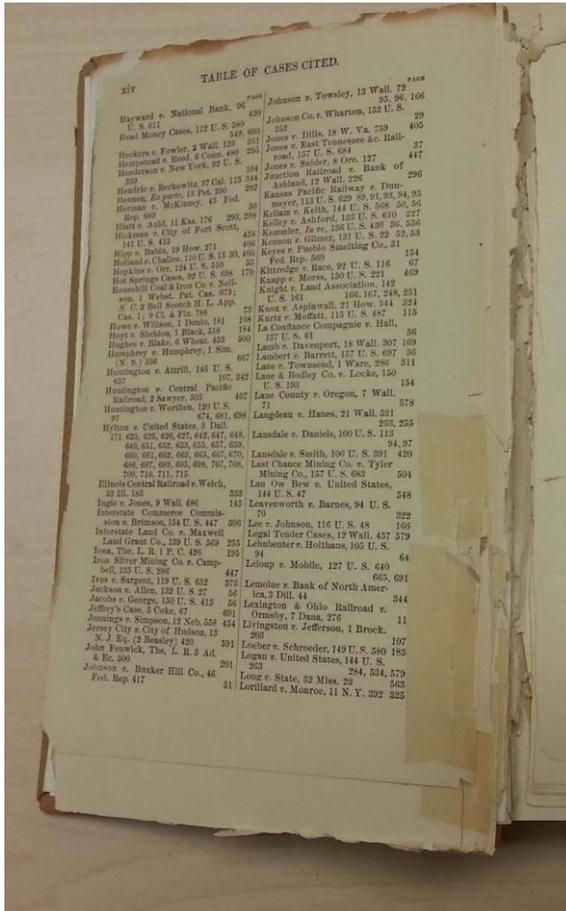
## The decision

- Support regional discard policy by agreeing to preserve selected titles
- *United States Reports* was one of the few titles for which we have a complete run



## The process

- Inventory of holdings
- Assessment of condition
- Cataloging, linking, barcoding, property stamping
- Determination of which volumes were issued by GPO or were considered officially government issued



Damaged 1894 volume

1828 privately published volume

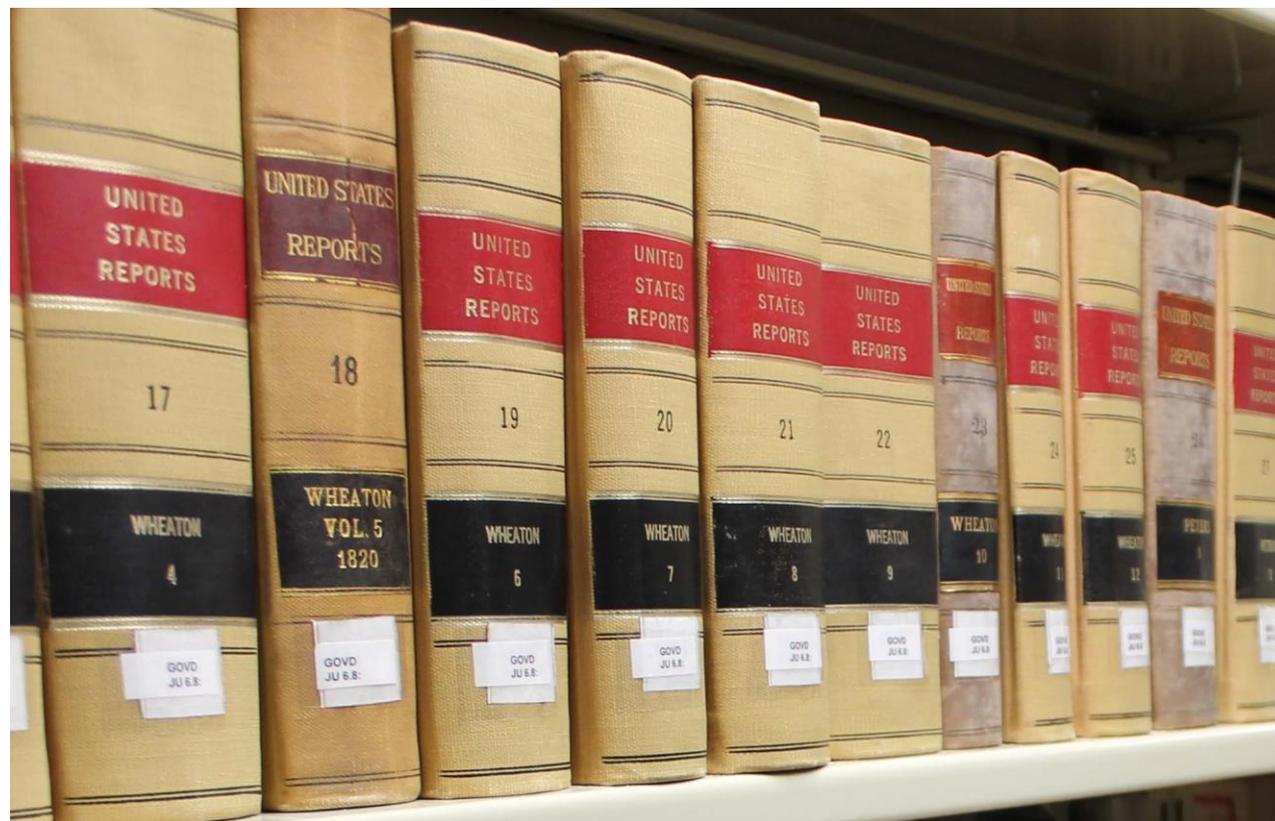


# The result

- 566 volumes total (1751-present)
- 480 volumes (1876-present) in scope (funded by the Government) and covered by MOA
- 81 volumes were reprints or were not in Good or Fine condition and we are seeking replacements
- Kept in open stacks but non-circulating
- Most volumes had never been depository stamped or property stamped by UHM



# Thank you!





# Scanning the Canal

## Becoming a Preservation Steward for Document Collections



**Chelsea S. Dinsmore**  
**University of Florida**  
**George A. Smathers Libraries**



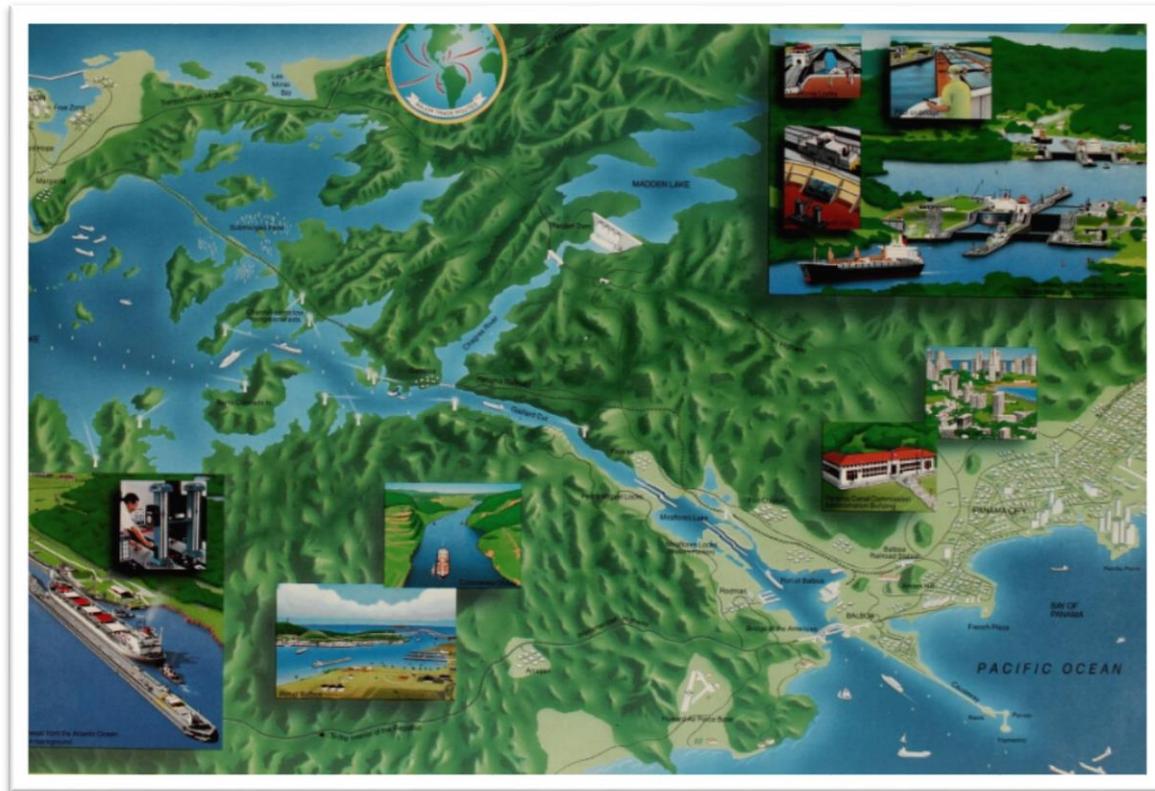
## ASERL's Collaborative Federal Depository Program (CFDP)

### Centers of Excellence:

- actively replace damaged or lost pieces,
- seek to fill holes in their collections, as necessary, via Needs and Offers lists,
- provide active preservation for the collection,
- catalog the pieces that they own on OCLC and identify them as the archival copy,
- participate in GPO's program to set holdings on OCLC for Regionals free of charge, if eligible,
- retain content permanently.



## Panama Canal Documents



<http://ufdc.ufl.edu/ufpancan>



## Contributing to FDsys

- Process was reasonable
  - Create folders
  - Insert JP2s, thumbnails, PDFs, and XML
  - Used FTP to deliver files
- Return was excellent
  - Files loaded into national collection
  - Authenticated PDFs provided back to UF collections



## Contributing to FDsys

- Increased accessibility for FDLP librarians and researchers
- GPO reviewed catalog records
- Allowed digital files to be authenticated





Thank you!

Questions?

Chelsea Dinsmore  
chelseaz@ufl.edu



# FDLP Web Archive

**Dory Bower**  
**Archives Specialist**  
**U.S. Government Publishing Office**



# FDLP Web Archive: What and Why?

- Since the 1990s much has changed with dissemination in electronic format.
- 1998: begin manually harvesting electronic Government publications and use of PURLs for permanent access
- Agencies disseminating information on own websites, not through FDLP or informing the SuDoc, as stated in Title 44, U.S.C.
- How do we continue mission of *Keeping America Informed*, especially considering huge increase in online content and ephemeral nature?
- 2011: website level archiving, became Archive-It partner



## July 2012

You are viewing an archived web page, collected at the request of [Federal Depository Library Program Web Archive](#) using [Archive-It](#). This page was captured on 12:56:33 Jul 18, 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](#) [Enable QA](#)

### UNITED STATES HOLOCAUST MEMORIAL MUSEUM

Entire Site Search SEARCH

MUSEUM EDUCATION RESEARCH HISTORY REMEMBRANCE GENOCIDE SUPPORT CONNECT

DONATE



## RESPONDING TO GENOCIDE TODAY

How Will You Meet the Challenge?

Take action

### HOLOCAUST HISTORY

Introduction to the Holocaust  
Holocaust Encyclopedia

Français  
Español  
Italiano  
Русский  
Türkçe  
Português (BR)

عربي  
فارسي  
زرتي

Ελληνικά  
Bahasa Indonesia  
简体中文  
한국어

Mapping Initiatives  
Online Exhibitions

### THE MUSEUM

Plan a Visit  
Museum Shop  
Exhibition Information

### FEATURED



The Holocaust in Hungary  
Eminent historian Randolph L. Braham addressed the destruction of Hungarian Jewry during the Holocaust in the 2012 Weinmann Lecture.

[MORE »](#)



Join the 20th Anniversary Legacy Challenge  
In honor of the Museum's 20th anniversary in 2013, a generous couple will match up to 10 percent of each new planned gift to the endowment.

[MORE »](#)



Museum Hosts DC Premiere of *La Rafle*  
Join director Rose Bosch and actress Mélanie Laurent for the July 18 screening of this French drama about the 1942 Vél d'Hiv Roundup.

[MORE »](#)



Risk and Resistance in the Netherlands  
Hear the extraordinary story behind the forging materials a Dutch couple used to save countless Jews during the Holocaust.

[MORE »](#)



Ratko Mladic Trial in The Hague  
Museum Goldfarb Fellow and journalist Michael

### PREVENTING GENOCIDE

Take action to confront genocide and related crimes against humanity today.

[MORE »](#)

### CONFRONTING ANTISEMITISM

Learn about the history and the continuing problem of antisemitism.

[MORE »](#)

### RESCUING THE EVIDENCE

View highlights from the Museum's collections and the stories they bring to life.

[MORE »](#)



## May 2014

You are viewing an archived web page, collected at the request of [Federal Depository Library Program Web Archive](#) using [Archive-It](#). This page was captured on 14:08:05 May 19, 2014, 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 Enable QA](#)

[SITE](#) [EVENTS](#) [PLAN YOUR VISIT](#) [SUPPORT THE MUSEUM](#) [CONNECT](#) [DONATE](#)



UNITED STATES  
HOLOCAUST  
MEMORIAL  
MUSEUM

[Museum Information](#)

[Resources for Academics and Research](#)

[Resources for Educators](#)

[Resources for Professionals and Student Leaders](#)

### Never Again. What You Do Matters.

Through our comprehensive campaign, "Never Again. What You Do Matters," led by honorary chair Elie Wiesel, the Museum seeks to secure the resources to keep Holocaust memory alive as a force for change in today's world.

LEARN ABOUT MUSEUM PROGRAMS & EVENTS NEAR YOU

E-mail Address

Zip Code

SIGN UP





## March 2017

You are viewing an archived web page, collected at the request of [Federal Depository Library Program Web Archive](#) using [Archive-It](#). This page was captured on 16:06:57 Mar 10, 2017, 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 Enable QA](#)

LANGUAGE PLAN YOUR VISIT UPCOMING EVENTS SUPPORT THE MUSEUM DONATE



MENU

UNITED STATES  
**HOLOCAUST  
MEMORIAL  
MUSEUM**

**Learn About**  
THE HOLOCAUST

**Remember**  
SURVIVORS AND VICTIMS

**Confront**  
GENOCIDE | ANTISEMITISM



SEARCH

## EXPLOITING DIVISIONS

Nazi propaganda sought to tap US racial tensions to win the war





## Collection Development

Priorities for website level collection:

- Must be within scope of FDLP
- Not distributing through print
- Government information disseminated through web and not cataloged
- Avoid duplication of effort with other institutions
- Worked with the Acquisitions staff to help determine needs



## Collection Development: to Date

### Collection to date

- SuDoc Y3 sites: commissions, committees, independent agencies
- Non-standard Government websites: *flu.gov*, *choosemyplate.gov*
- Special Collections
  - Native American Resources
- Notification from community
- Maintaining Regular frequency of crawls (resource intensive)

All while monitoring size of account



## Collection Development

### Notification from community

- Document Discovery (Govt Agencies)  
(<http://usgpo.wufoo.com/forms/document-discovery/>)
- askGPO  
<http://www.gpo.gov/askgpo/>
- Lost Docs Reporting Form  
<http://www.fdlp.gov/collection-tools/lostdocs>
- Monitor listservs



## Collection Development

### Moving forward

- Working with Collection Development Librarian
- Move from smaller to larger sites
- Special Collections: Themes
- Google Analytics, extensive monitoring
  - more info (24 minutes in)

<https://archive.org/details/AccessToArchiveItCollections01Video>



## FDLP Web Archive

Collection size:

- 12.2 TB, 99 million URLs crawled
- 141 collections in AIT
- 181 records on CGP (analytical cataloging)
- FDLP Project page

<http://www.fdplp.gov/377-projects-active/2020-web-archiving>

Using web archives for reference:

- FDLP Academy webinar: [goo.gl/vGx17E](http://goo.gl/vGx17E)



## Access

### Two locations for Access

- Archive-It
  - Search for “GPO” or “FDLP”

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

- Catalog of U.S. Government Publications (CGP)
  - Search FDLP Web Archive Catalog

<https://catalog.gpo.gov/>



## Collaboration

### Federal Web Archiving Working Group

- With nominations came questions: What are others doing?
- Priorities are to maximize use of account: stick to other priorities if being done by others
- Began talks with Library of Congress who organized inaugural meeting with Library of Congress, GPO, and NARA, and attained management approval



## Collaboration

- What now?
- Detailed presentations on each agency's web archiving program, what and how
- Outreach: blog  
<https://blogs.loc.gov/thesignal/2015/02/introducing-the-federal-web-archiving-working-group/>
- Increase participation: Smithsonian, NLM, HHS, Education



## Collaboration

What we have done or discussed?

- Bi-monthly updates from each agency keeping the group well informed of new activities
- Developed an internal wiki to post information on agency programs, outreach, etc.
- Developed contracts and RFI's
- Processes for ingest and transfer of WARC files
- Policy topics, sharing policies
- Technical topics, such as how social media is being archived, which is constantly changing



## Collaboration

Where we want to go

- More outreach: Increase participation and knowledge of group
- More relationship with content creators
- Guidance documentation
- Web presence

http://

BROWSE HISTORY

# Web Archiving at the Library of Congress

Abbie Grotke

Web Archiving Team Lead

Digital Collections Management and Services Division

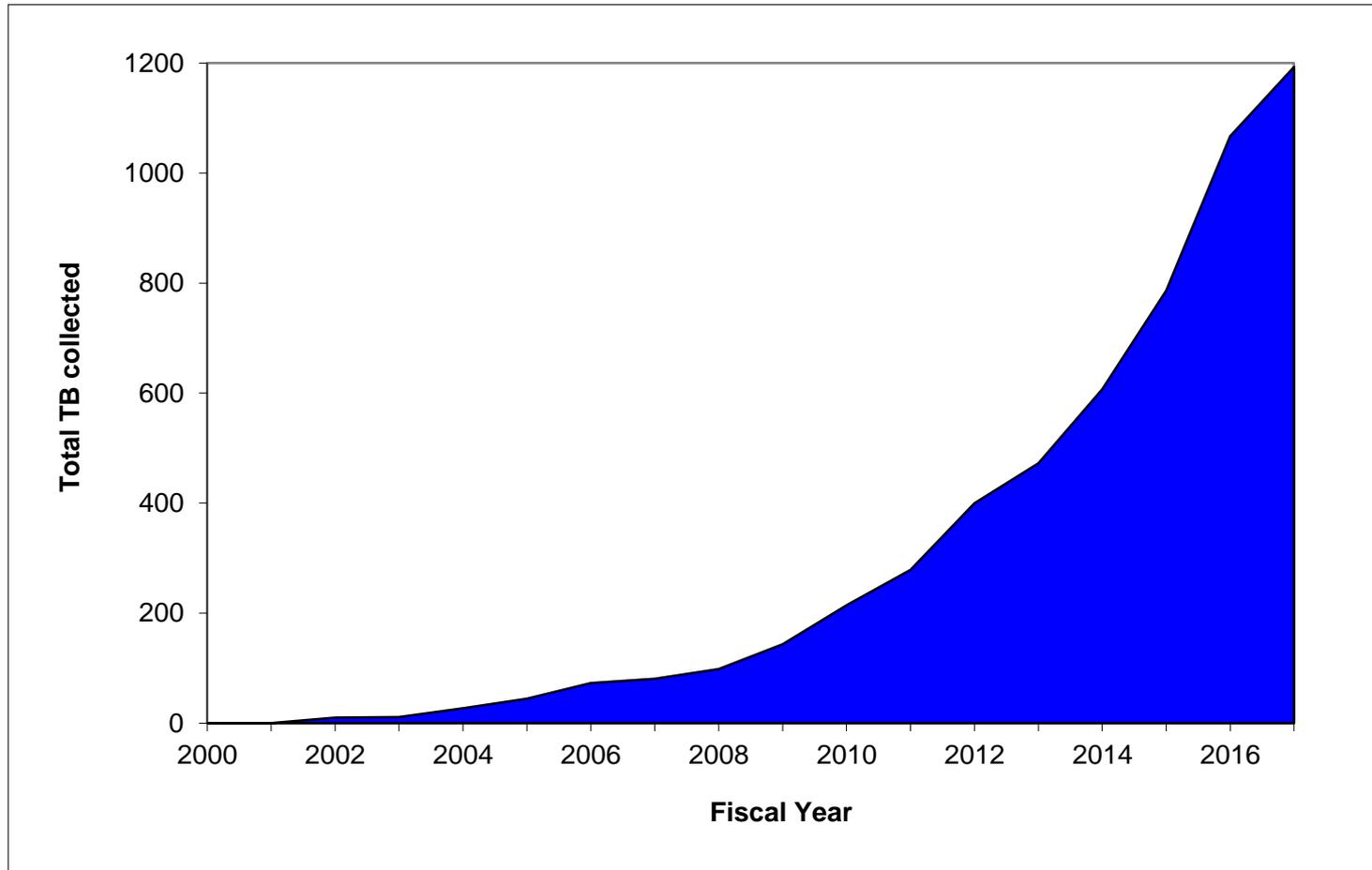
Library of Congress

@agrotke [abgr@loc.gov](mailto:abgr@loc.gov)

DLC ~ October 17, 2017

# Library of Congress Web Archives: A Snapshot

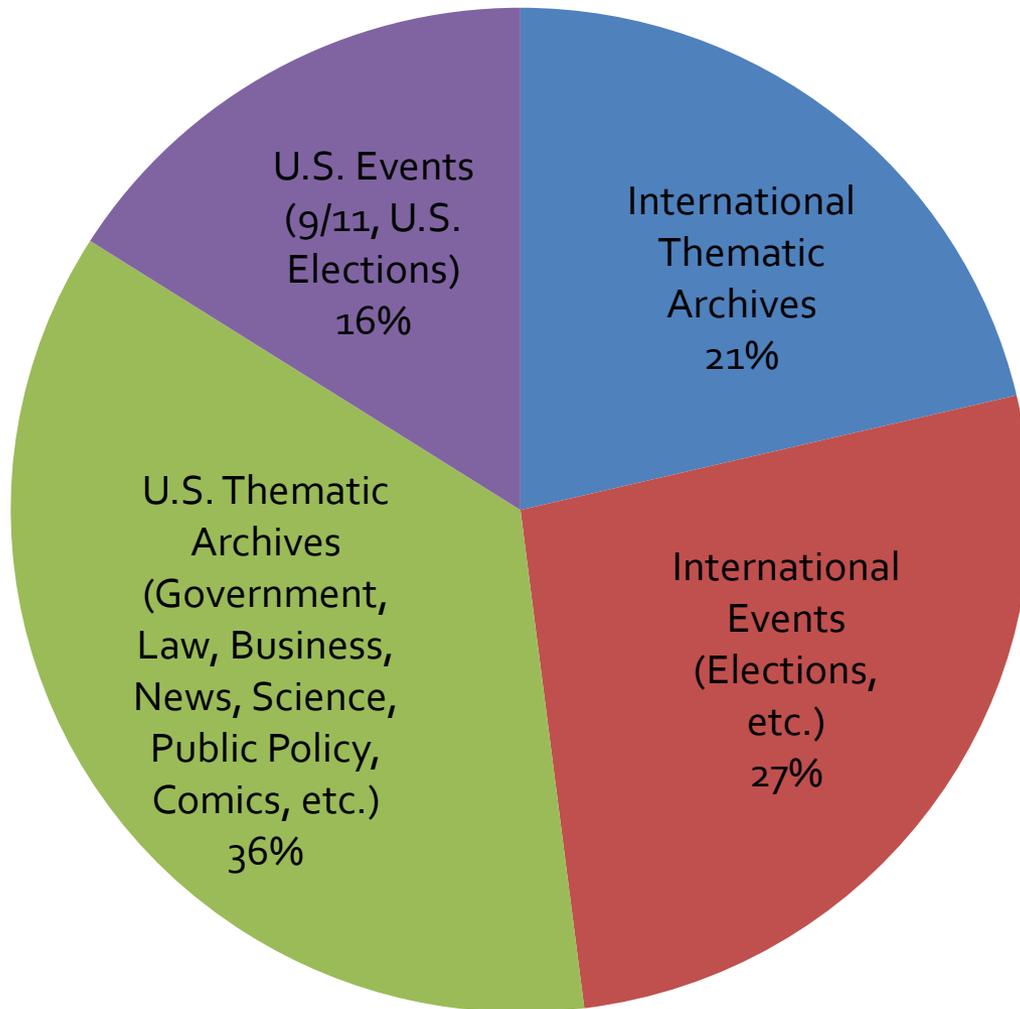
- **1.193 petabytes collected;** and over **14.9 billion documents** as of June 2017



# Tools and Processes at LC

- Staffing: 5 FTs on Web Archiving Team, plus many more around LC
- Workflow
  - Inhouse-developed curatorial tool , Digiboard, to manage nomination permissions, quality review, etc.
- Crawling
  - Heritrix crawler <https://webarchive.jira.com/wiki/display/Heritrix/Heritrix>
    - Umbra, a browser automated tool, is also used to improve results <https://github.com/internetarchive/umbra>
    - Experimenting with Webrecorder <https://webrecorder.io/>
  - Contract out large-scale effort (~20-25 TB/mo) to Internet Archive
  - Inhouse crawling: ~10 TB/year using Heritrix
- Access
  - Open Wayback Machine <https://github.com/iipc/openwayback/wiki>
  - Metadata Object Description Schema (MODS) <https://www.loc.gov/standards/mods/>
- Preservation
  - WARC (Web ARChive file format) <http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml>
  - Use Bag-it to transfer data over Internet2 from contractor to LC
  - Two local copies, one in long-term storage, another access copy

# LC Web Archive Collections



- Permissions-based, selective harvesting of web content published U.S. and 127 other countries
- Over 33,000 Seed URL nominations; over 31,000 have been archived in ~90 collections. Full list: <http://www.loc.gov/webarchiving/collections.html>
- 32 **active** collections, with over 7,100 seed URLs currently in crawl status, crawled at various frequencies
- Collection Development Policies: <http://www.loc.gov/acq/devpol/webarchive.pdf>

FORMAT

## Archived Web Site

[Search Archived Web Sites](#) [Collections with Archived Web Sites](#)

### Featured Content



Official Website of General (Rtd.) Sarath Fonseka



Dinosaur Comics!



Belief.net Pope John Paul II



The Library of Congress



LRI-Lembaga Riset Informasi

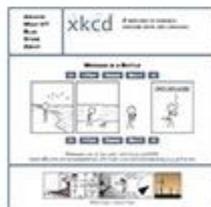
## Collections with Archived Web Sites

View   Sort By



COLLECTION  
**Winter Olympic Games 2002 Web Archive**

Collection Period: February 8, 2002 to February 27, 2002. The 2002 Winter Olympics were held in Salt Lake City, Utah from February 8 through February



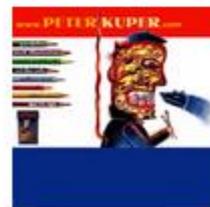
COLLECTION  
**Webcomics Web Archive**

Collection Period: June 4th, 2014 to ongoing. This collection focuses on comics created specifically for the web and supplements the Library of Congress' extensive holdings in both



COLLECTION  
**Web Cultures Web Archive**

Collection Period: June 4th, 2014 to ongoing. The Web Cultures Web Archive includes sites documenting the creation and sharing of emergent cultural traditions on the web. The mission of



COLLECTION  
**Visual Image Web Archive**

Collection Period: April 5th, 2008 to November 30th, 2008. Like books and magazines in the Library's general collections, this set of websites complements the original and historical visual



COLLECTION  
**Vietnamese General Election 2011 Web Archive**

Collection Period: July 2011 to December 2011. Elections in this single-party country provide opportunity to understand upcoming leaders and changes in overall

# Archived Web Site

Search Archived Web Sites Collections with Archived Web Sites

Results: 1-25 of 11,734 | Refined by: Original Format: Archived Web Site

## Refine your results

Available Online 11,734

All Items 11,734

## Original Format

Archived Web Site

## Online Format

Web Page 11,668

PDF 29

Online Text 1

Audio 1

## Date

2000 to 2099 11,723

1900 to 1999 2

## Location

United States 8,188

Afghanistan 1,052

California 550

Texas 394

Florida 313

New York 307

Michigan 251

Illinois 220

Sudan 219

Pennsylvania 212

[More Locations](#)

## Search Archived Web Sites

View List  Sort By Select



ARCHIVED WEB SITE  
**Official Campaign Web Site - Arthur William Myatt**  
 Web Site. electronic | Electronic (Form).  
 Contributor: Myatt, Arthur William  
 Date: 2006-11-09



ARCHIVED WEB SITE  
**Official Campaign Web Site - Richard Frankhuizen**  
 Web Site. electronic | Electronic (Form).  
 Contributor: Frankhuizen, Richard  
 Date: 2002-07-01



ARCHIVED WEB SITE  
**Official Campaign Web Site - Jeffery Robert Barnett**  
 Web Site. electronic | Electronic (Form).  
 Contributor: Barnett, Jeffery Robert  
 Date: 2010-10-08

# Dinosaur Comics!



[View Captures](#)  
 Captured from August 20, 2010 to May 4, 2016

## About this Item

**Title**  
Dinosaur Comics!

**Summary**  
 "Dinosaur Comics is a comic about talking dinosaurs! It appears online and also in print in some newspaper magazines!" This widely read webcomic by Ryan North has been publishing since 2003.

**Contributor Names**  
North, Ryan, 1980-

**Created / Published**  
United States.

**Subject Headings**  
 - Webcomics  
 - Art and Art History

**Genre**  
web site



LIBRARY OF CONGRESS | WEB ARCHIVES

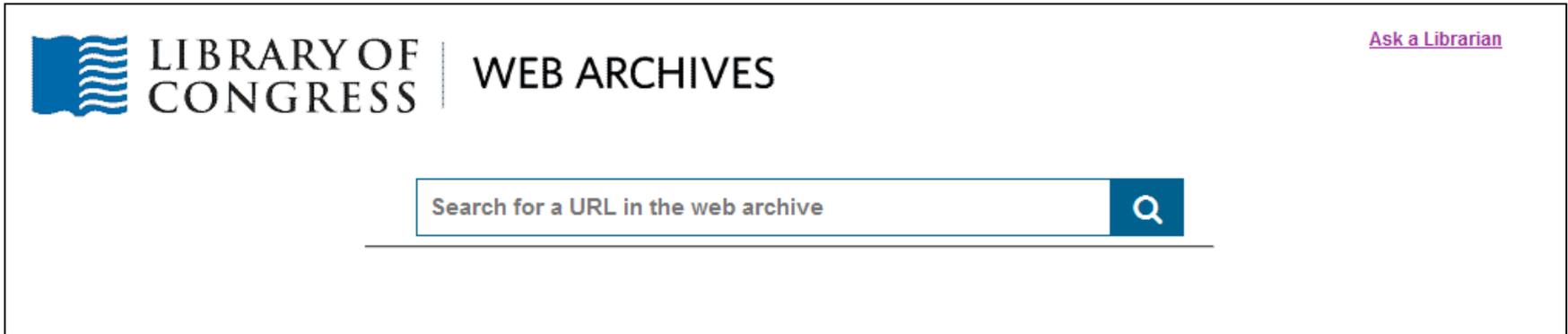
[Ask a Librarian](#)

<http://www.qwantz.com/index.php>  
 Saved 91 times between August 20, 2010 and August 3, 2016.

JAN							FEB							MAR							APR								
				1	2					1	2	3	4	5	6			1	2	3	4	5					1	2	
3	4	5	6	7	8	9	7	8	9	10	11	12	13	6	7	8	9	10	11	12	3	4	5	6	7	8	9		
10	11	12	13	14	15	16	14	15	16	17	18	19	20	13	14	15	16	17	18	19	10	11	12	13	14	15	16		
17	18	19	20	21	22	23	21	22	23	24	25	26	27	20	21	22	23	24	25	26	17	18	19	20	21	22	23		
24	25	26	27	28	29	30	28	29	27	28	29	30	31	24	25	26	27	28	29	30									
31																													
MAY							JUN							JUL							AUG								
			4	5	6	7				1	2	3	4					1	2					1	2	3	4	5	6
8	9	10	11	12	13	14	5	6	7	8	9	10	11	3	4	5	6	7	8	9	7	8	9	10	11	12	13		
15	16	17	18	19	20	21	12	13	14	15	16	17	18	10	11	12	13	14	15	16	14	15	16	17	18	19	20		
22	23	24	25	26	27	28	19	20	21	22	23	24	25	17	18	19	20	21	22	23	21	22	23	24	25	26	27		
29	30	31	26	27	28	29	30	24	25	26	27	28	29	30	28	29	30	31											
31																													
SEP							OCT							NOV							DEC								
1	2	3	1	2	3	4	5	1	2	3	4	5	1	2	3														



# URL search at [webarchive.loc.gov](http://webarchive.loc.gov)\*



*\*everything publicly and offsite accessible up through September 1, 2016  
(we are under a one-year embargo)*



# Collaborations



## Federal Government Web Archiving Working Group

A screenshot of the End of Term Web Archive website. The browser address bar shows 'eotarchive.cdlib.org/index.html'. The page title is 'End of Term Web Archive' with the subtitle 'US Federal Web Domain at Presidential Transitions'. There are navigation tabs for 'Home', 'Search Full Text', and 'Browse Web Archive'. The 'Search Full Text' tab is active, showing search results for 'Ukraine Conflict'. The results list includes a search bar, a 'Search Page Text' button, and a list of results. The first result is titled 'Appeal to international leaders and the world community | Іншіативна група «Першого грудня»' with a URL starting with 'http://1-12.org.ua/2014/09/10/3224'. The description mentions it was captured between Sep 15, 2014 and Oct 20, 2014. The contributor is listed as 'Library of Congress'.

Project Background

The End of Term Web Archive captures and saves U.S. Government websites at the end of presidential administrations. Beginning in 2008, the EOT has thus far preserved websites from administration changes in 2008 and 2012, and is currently preparing for the 2016 electoral season.

Project Partners

Browse by Timeframe:

- 2008-2009
- 2012-2013
- Browse All
- End of Term 2016

EOT at a Glance

Crawl dates:  
September 2008 to May 2009  
September 2012 to March 2013

Number of websites captured:  
2008: 3,305 websites  
2012: 3,247 websites

Terabytes of data captured:  
2008: 16 TB  
2012: 21 TB

Archive Scope

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

A screenshot of the Internet Archive website. The page title is 'Ukraine Conflict' and it is collected by 'Internet Archive Global Events'. The description states that the collection seeks to document conflict in Ukraine and includes news outlets, social media, blogs, and government websites. The subject is 'Government, Spontaneous Events'. Below the description, there is a 'Narrow Your Results' section with a search bar and a list of results. The first result is titled 'Appeal to international leaders and the world community | Іншіативна група «Першого грудня»' with a URL starting with 'http://1-12.org.ua/2014/09/10/3224'. The description mentions it was captured between Sep 15, 2014 and Oct 20, 2014. The contributor is listed as 'Library of Congress'.

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 >> Internet Archive Global Events >> Ukraine Conflict

INTERNET ARCHIVE

**Ukraine Conflict**  
Collected by: Internet Archive Global Events

Archived since: Feb, 2014

Description: This collection seeks to document conflict in Ukraine. Contributions to this collection were made by the Archive-IT team and subject matter experts from Stanford University, Library of Congress, Global Investigative News Network, and Harvard Ukrainian Research Institute in the fields of Investigative Journalism, Russian, and Eurasian studies. Content includes news outlets, social media, blogs, and government websites. Sites are in English, Ukrainian, Russian, and other languages.

Subject: Government, Spontaneous Events

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.

Enter search terms here [Search] [Clear]

Group [Sort By: Count] [A-Z]

- 1 (100)
- 2 (245)
- 3 (407)
- 4 (331)
- 5 (650)

More

Contributor [Sort By: Count] [A-Z]

- Library of Congress (697)
- Harvard Ukrainian Research Institute (458)
- Hoover Institution, Stanford University (41)

Type [Sort By: Count] [A-Z]

- Article (402)
- Website (97)

Page 1 of 67 (6,616 Total Results) [Next Page]

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

Title: Appeal to international leaders and the world community | Іншіативна група «Першого грудня»

URL: http://1-12.org.ua/2014/09/10/3224

Description: Appeal to international leaders and the world community | Іншіативна група «Першого грудня»

Captured 1 times between Sep 15, 2014 and Oct 20, 2014

Videos: 1 Videos Captured

Language: English

Contributor: Library of Congress



Results of a Survey of Organizations Preserving Web Content  
February 2017

## More information

- Library of Congress Web Archives: [www.loc.gov/websites](http://www.loc.gov/websites) and [loc.gov/webarchiving](http://loc.gov/webarchiving)
- IIPC : <http://www.netpreserve.org>
- End of Term Archive: <http://eotarchive.cdlib.org/> and <https://archive.org/details/EndOfTerm2016WebCrawls>

# Questions?

Abbie Grotke

[abgr@loc.gov](mailto:abgr@loc.gov)

202-707-2833

@agrotke



## Contact Us

David Walls, [dwalls@gpo.gov](mailto:dwalls@gpo.gov),  
[preservefedinfo@gpo.gov](mailto:preservefedinfo@gpo.gov)

Suzanne Ebanues, [sebanues@gpo.gov](mailto:sebanues@gpo.gov),  
<https://www.fdlp.gov/partnership-inquiry-form>

Dory Bower, [dbower@gpo.gov](mailto:dbower@gpo.gov),  
[fdlpwebarchive@gpo.gov](mailto:fdlpwebarchive@gpo.gov)