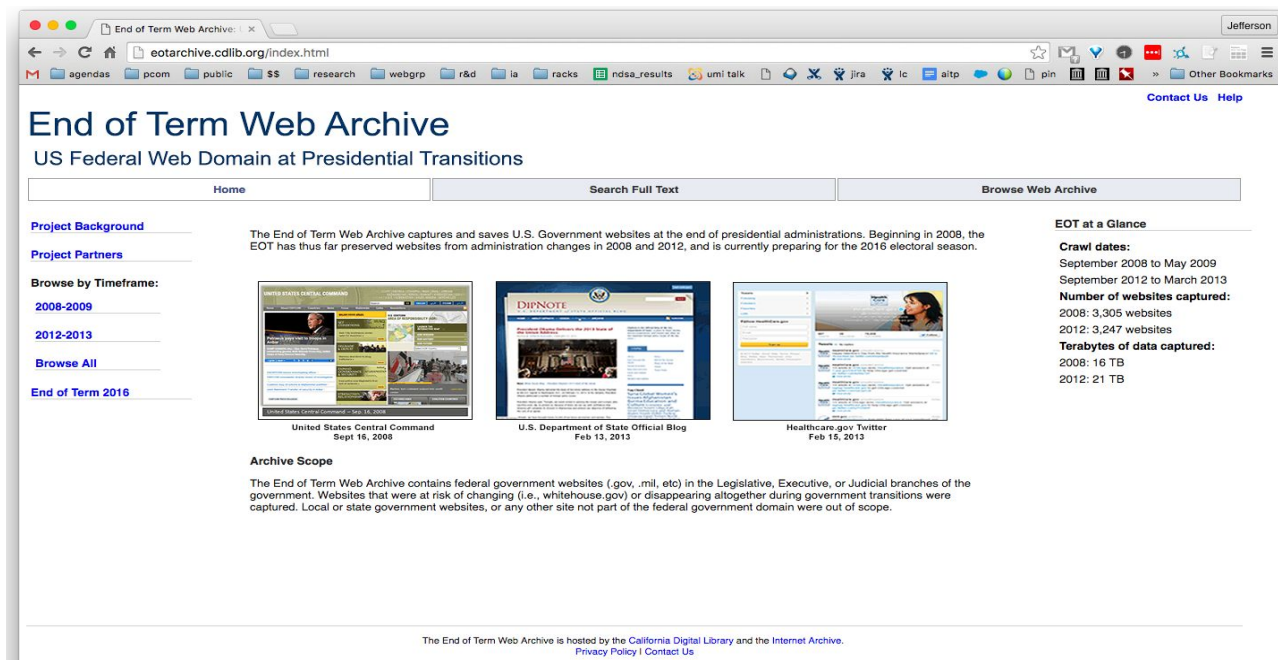


The End of Term Archive: Archiving the U.S. Government Web



Abigail Grotke, Web Archiving Team Lead, Library of Congress
@agrotke | abgr@loc.gov

MLTW | Dec. 5, 2017



It all began a long, long, time ago, in a far away...



<https://flic.kr/p/4N2jHU>



<https://flic.kr/p/4JNkLE>

National Library of Australia

nla.int-nl39859-al1-v

Goals of the End of Term Project



United States Central Command
Sept 16, 2008



U.S. Department of State Official Blog
Feb 13, 2013



Healthcare.gov Twitter
Feb 15, 2013

- work collaboratively to preserve public U.S. Government websites
- document federal agencies' presence on the web during the end of Presidential terms
- enhance the existing research collections of the partner institutions
- raise awareness about the need for preservation
- engage with researchers and subject experts

Extant .gov web archiving efforts

Capture, Preservation, & Access

- LOC: legislative branch, some executive
- GPO: agency sites, often ephemeral
- NARA: congressional web harvest every 2 years
- IA: global & curated crawls
- Agency-level: NIH/NLM, DOE, DOL, HHS, CMS, others, using Archive-It or other tools
- UNT & Others: Topical .gov collecting
- Internal agency guidelines

Community Efforts

- Federal Web Archiving Group
- Research Initiatives
 - academic
 - NGO or watchdog
- Citizen Driven
 - grassroots efforts, e.g. Data refuge/rescue
- End of Term
 - focused but large-scale multi-institutional project

Original End of Term Web Archive Partners

INTERNET ARCHIVE



LIBRARY OF
CONGRESS



University of California

CDL

California Digital Library



for 2008/2012 - all IIPC & NDIIPP/NDSA partners

EOT 2016: more partners!



STANFORD UNIVERSITY LIBRARIES



Federal Government Web Archiving
Working Group

EOT Collaborative Roles

- Project management and coordination (All, rotates)
- Nomination tool development (UNT)
- URL nomination (All + community/public)
- Crawling (IA, UNT, LC, GW)
- Preservation of full copy (IA, LC, UNT)
- Access: portal, full-text search, metadata, research & support (IA, CDL)
- Outreach, press, twitter account (primarily IA, LC, UNT, Stanford)

The 2016 project brought more partners, capacity for better, distributed crawling, and more community and researcher engagement

The screenshot displays the 'Add a Site/URL' form on the EOT project website. The form is divided into two main sections: 'Add a Site/URL' and 'Metadata'. The 'Add a Site/URL' section has a header 'Capture the Following Site (Required)' and a row of 15 stars for rating. It contains a 'Website URL:' label, a text input field with 'http://' and a 'view' button, and a small note 'The URL you wish to nominate for capture'. The 'Metadata' section includes a 'Branch:' label with the instruction 'Which branch of government', a 'Select All' button, and three radio button options: 'Executive', 'Judicial', and 'Legislative'. At the bottom, there is a 'Comment:' label and a large text area for additional input.

Project Home About This Project Project Reports Add A URL

Add a Site/URL

Capture the Following Site (Required)

Website URL:

http:// view

The URL you wish to nominate for capture

Metadata

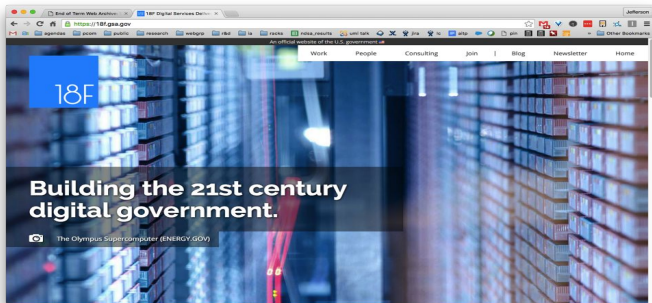
Branch:
Which branch of government

Select All

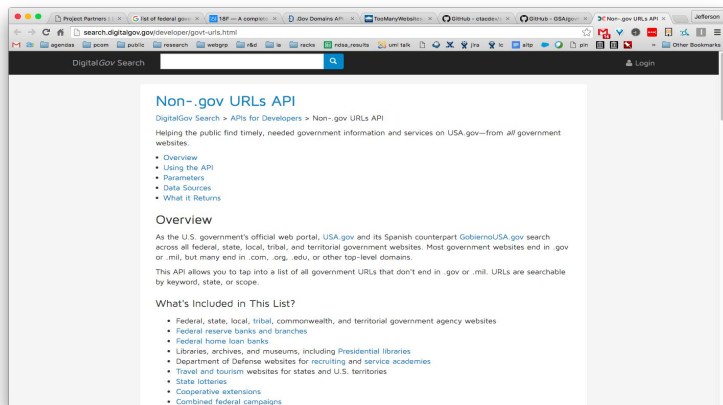
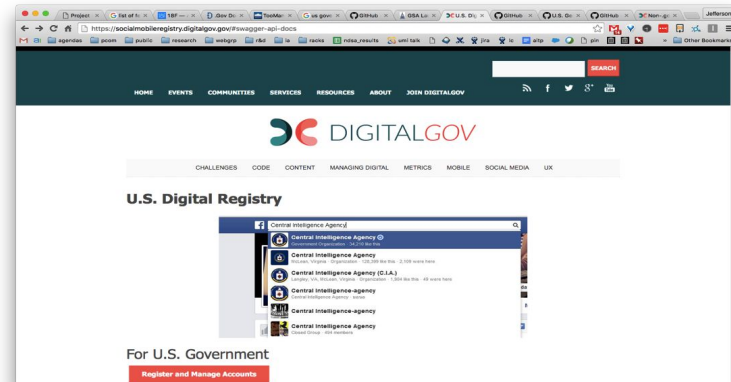
☐ Executive
☐ Judicial
☐ Legislative

Comment:

Defining the “government web presence”



Stanford WebBase Project



```
4719 gov, noaa, nodc)
4720 gov, noaa, nohrsc)
4721 gov, noaa, nos, ccoma)
4722 gov, noaa, nos, co-ops)
4723 gov, noaa, nos, co-ops, gobi)
4724 gov, noaa, nssl)
4725 gov, noaa, nssl, blog)
4726 gov, noaa, nwrsc)
4727 gov, noaa, nwrfc)
4728 gov, noaa, nws)
4729 gov, noaa, nws, iwin)
4730 gov, noaa, nws, md1)
4731 gov, noaa, oceanexplorer)
4732 gov, noaa, oceanservice)
```



```
4751 gov, noaa, spc)
4752 gov, noaa, srh)
4753 gov, noaa, srh, gis)
4754 gov, noaa, srh, innovation)
4755 gov, noaa, ssd)
4756 gov, noaa, swpc)
4757 gov, noaa, swpc, helios)
4758 gov, noaa, swpc, helios-www)
4759 gov, noaa, swpc, services)
4760 gov, noaa, sidesandcurrents)
4761 gov, noaa, volunteer)
4762 gov, noaa, vws)
4763 gov, noaa, wdtb)
4764 gov, noaa, weather)
4765 gov, noaa, wrf, ferret)
4766 gov, noaa, wrh)
4767 gov, noaa, wrh, inws)
4768 gov, noaa, wrh, raws)
4769
```

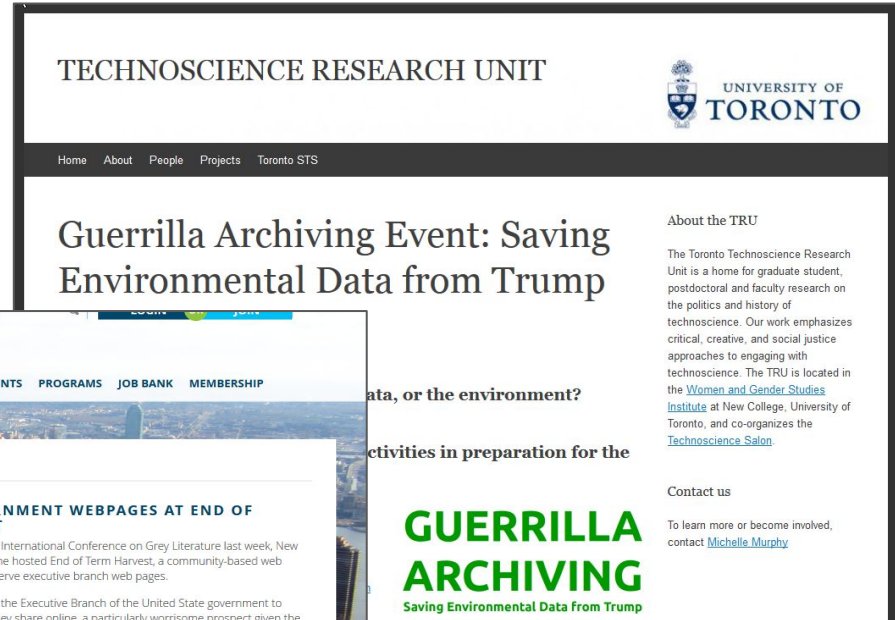
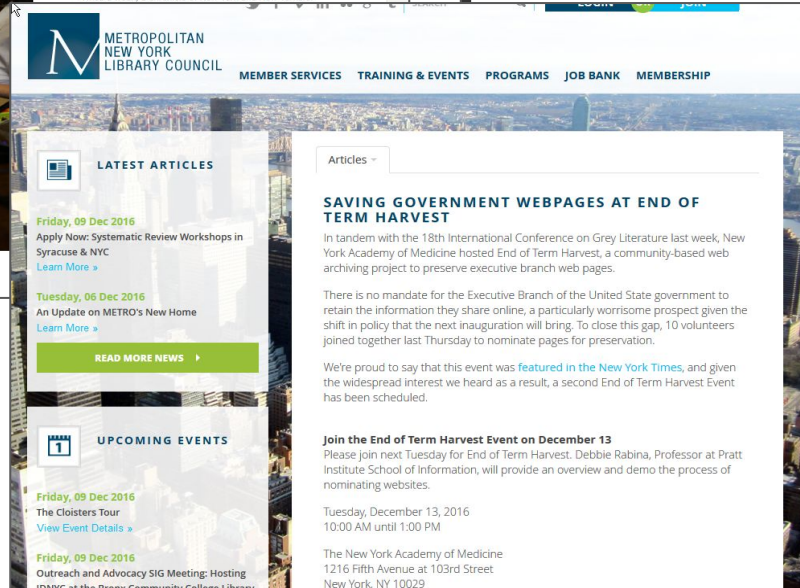


Community Engagement



Events: NYAM, U-Toronto, U-Penn, UC-Riverside, more

Research: CMU, Georgetown, U-Washington



Volunteer contributions

	Nominations	Volunteers
2008	457	26
2012	1476	31
2016	15,000+	400+

End of Term Presidential Harvest 2016

[Project Home](#)

[About This Project](#)

[Project Reports](#)

[Feeds](#) ▾

Add A URL

Capture the Following Site

Website URL: *

The URL you wish to nominate for capture

http://



Plus over 150,000 from DataRescue/EDGI events/tools!

Size of Archive

	URLs	Size	Comments
2008	~160 million	17.95 TB	Multiple crawls, deduplicated
2012	~120 million	18.60 TB	More focused crawls, deduplicated. Notable for media richness, uniqueness, density.
2016	over 429 million	293.91 TB	Includes ~150 TB of FTP crawls

EOT 2016 Content

- **Content includes:**
 - 9,000+ social media accounts (scrape of gov SM registry API) 44% FB, 37% TW, 10% YT
 - ~190K total domain, subdomains, gov and non-gov sites
 - more crowdsourced, curatorial nominations

gov,dontserveteens)
gov,dot)
gov,dot,adfs)
gov,dot,fastlane)
gov,dot,fhwa)
gov,dot,fhwa.borderplanning)
gov,dot,fhwa.collaboration)
gov,dot,fhwa.cfl)
gov,dot,fhwa.environment)
gov,dot,fhwa.fhwapap04)
gov,dot,fhwa.flh)
gov,dot,fhwa.international)
gov,dot,fhwa.mutcd)
gov,dot,fhwa.nhi)
gov,dot,fhwa.ops)
gov,dot,fhwa.safety)
gov,dot,fhwa.wfl)
gov,dot,fhwa.wwwcf)
gov,dot,fmcsa)
gov,dot,fmcsa.ai)
gov,dot,fmcsa.cms)
gov,dot,fmcsa.csa)
gov,dot,fmcsa.csa2010)
gov,dot,fmcsa.li-public)
gov,dot,fmcsa.mrb)
gov,dot,fmcsa.nrcme)
gov,dot,fmcsa.safer)
gov,dot,fra)
gov,dot,fra.safetydata)
gov,dot,fla)
gov,dot,fla.transit-safety)
gov,dot,isdcd)
gov,dot,its)
gov,dot,its.benefitcost)
gov,dot,its.pcb)
gov,dot,its.standards)
gov,dot,marad)
gov,dot,nhtsa)
gov,dot,nhtsa.www-esv)
gov,dot,nhtsa.www-fars)
gov,dot,nhtsa.www-nrd)
gov,dot,nhtsa.www-odi)
gov,dot,oig)
gov,dot,ost.airconsumer)
gov,dot,ost.dotcr)
gov,dot,ost.dothr)
gov,dot,ost.testimony)
gov,dot,phmsa)
gov,dot,phmsa.npms)
gov,dot,phmsa.opsweb)
gov,dot,phmsa.primis)

gov,ems)
gov,energy)
gov,energy.afdc)
gov,energy.betterbuildingsolutioncenter)
gov,energy.buildingdata)
gov,energy.catalyst)
gov,energy.eere)
gov,energy.eere,apps1)
gov,energy.eere,apps2)
gov,energy.etc)
gov,energy.fossil)
gov,energy.genomicscience)
gov,energy.hss)
gov,energy.hydrogen)
gov,energy.nnsa)
gov,energy.pi)
gov,energy.science)
gov,energy.ssi)
gov,energy.codes)
gov,energy.savers)
gov,energy.star)
gov,enfield-ct)
gov,ennistx)
gov,enterpriseal)
gov,eop)
gov,epa)
gov,epa.archive)
gov,epa.blog)
gov,epa.cfpub)
gov,epa.cumulis)
gov,epa.developer)
gov,epa.gispub4)
gov,epa.iaspub)
gov,epa.nepis)
gov,epa.ofmpub)
gov,epa.semspub)
gov,epa.water)
gov,epa.yosemite)
gov,epa.yosemite1)
gov,erie)
gov,erie.gis1)
gov,erie.gis2)
gov,erieco)
gov,erieco.engage)
gov,eriecountypa)
gov,essexct)
gov,eugene-or)
gov,eugene-or.ceapps)
gov,eugene-or.pdd)
gov,eulesstx)
gov,exeternh)

gov,house.bobbyscott)
gov,house.brown)
gov,house.castor)
gov,house.chris-smith)
gov,house.chu)
gov,house.clerk)
gov,house.cole)
gov,house.cummings)
gov,house.delbene)
gov,house.denham)
gov,house.desjarlais)
gov,house.docs)
gov,house.donovan)
gov,house.duckworth)
gov,house.edworkforce)
gov,house.energycommerce)
gov,house.farr)
gov,house.flores,rs)
gov,house.foreignaffairs)
gov,house.foreignaffairs,dc)
gov,house.fosteryouthcaucus-karenbass)
gov,house.gabbard)
gov,house.gosar)
gov,house.grothmanforms)
gov,house.gutierrez)
gov,house.heck)
gov,house.history)
gov,house.homeland)
gov,house.issa)
gov,house.jones)
gov,house.jordan)
gov,house.jee)
gov,house.lgbt-polis)
gov,house.messer)
gov,house.mulvaney)
gov,house.naturalresources)
gov,house.norton)
gov,house.oversight)
gov,house.oversight,dc)
gov,house.paulgossar)
gov,house.perry)
gov,house.peteking)
gov,house.quigley)
gov,house.resourcecommittee)
gov,house.roses)
gov,house.scalise)
gov,house.scalise,rs)
gov,house.schiff)
gov,house.science)
gov,house.sensenbrenner)
gov,house.smallbusiness)
gov,house.timryan)

gov,fcc)
gov,fcc,apps)
gov,fcc,appsdemo)
gov,fcc,consumercomplaints)
gov,fcc.esupport)
gov,fcc.fallfoss)
gov,fcc.hraunfoss)
gov,fcc.licensing)
gov,fcc.reboot)
gov,fcc.stations)
gov,fcc.transition)
gov,fcc.wireless2)
gov,fcc.wireless2)
gov,fda)
gov,fda.accessdata)
gov,fda.blogs)
gov,fdic)
gov,fdicig)
gov,fdip)
gov,fdip.puril)
gov,fdisy)
gov,fec)
gov,fec.docquery)
gov,fec.eqs)
gov,federalregister)
gov,federalreserve)
gov,federalreserve,oig)
gov,federalreserveconsumerhelp)
gov,fedshiresets)
gov,fedthefuture)
gov,fema)
gov,fema.asd)
gov,fema.beta)
gov,fema.careers)
gov,fema.citizenscorps)
gov,fema.community)
gov,fema.emilms)
gov,fema.gis)
gov,fema.hazards)
gov,fema.m)
gov,fema.msc)
gov,fema.ndms)
gov,fema.training)
gov,fema.usfa)
gov,fema.usfa,apps)
gov,ferc)
gov,ferndalemi)
gov,ffiec)
gov,ffiec,thandbook)
gov,fgdc)
gov,fhfa)
gov,fido.xml)

EOT 2016 Press ++

- **Press**

- Dozens of articles and interviews

- **Collaborations**

- Data Refuge
- EDGI
- GSA / 18F
- data.gov

The screenshot shows the Internet Archive website interface. At the top, there's a navigation bar with links for HOME, EXPLORE, LEARN MORE, and CONTACT US. A social media bar includes Facebook, Twitter, and YouTube icons, along with a Login button. The main header features the Internet Archive logo and a tagline: "The leading web archiving service for collecting and accessing cultural heritage on the web. Built at the Internet Archive".

The breadcrumb trail reads: Explore >> Internet Archive Global Events >> End of Term 2016 Press.

The main content area has a green background. On the left is the Internet Archive logo. To the right, the title "End of Term 2016 Press" is displayed, followed by "Collected by: Internet Archive Global Events". Below this, it states "Archived since: Dec, 2016". The description reads: "Description: This is a collection of press articles about the End of Term 2016 project. End of Term is a collaborative project to preserve public United States Government web sites at the end of the current presidential administration ending January 20, 2017." The subject is "Government, Politics & Elections, Computers & Technology, Web archiving, Government" and the collector is "Internet Archive".

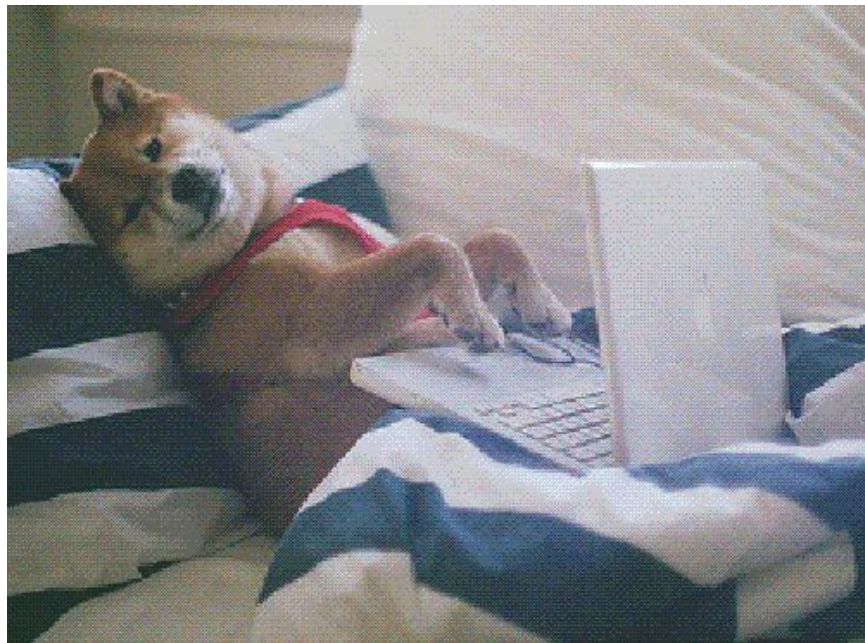
Below the main content, there's a section titled "Narrow Your Results". It includes a "Creator" filter with a list of sources: The New York Times (2), BBC News (1), Democracy Now! (1), Detroit News (1), Faculty of Information (iSchool) | University of Toronto (1), and a "More" button. There's also a "Sort By: Count" option and a "(A-Z)" link.

To the right of the filters is a search bar with the placeholder text "Enter search terms here", a "Search" button, and a "Clear" button. Below the search bar are two tabs: "Sites" and "Search Page Text".

At the bottom, it shows "Page 1 of 1 (74 Total Results)". The "Sort By" options are "Title (A-Z)", "Title (Z-A)", "URL (A-Z)", and "URL (Z-A)". The URL displayed is "http://arlington.wickedlocal.com/news/20161229/arlington-public-library-news/" and it notes it was "Captured once on Jan 12, 2017".

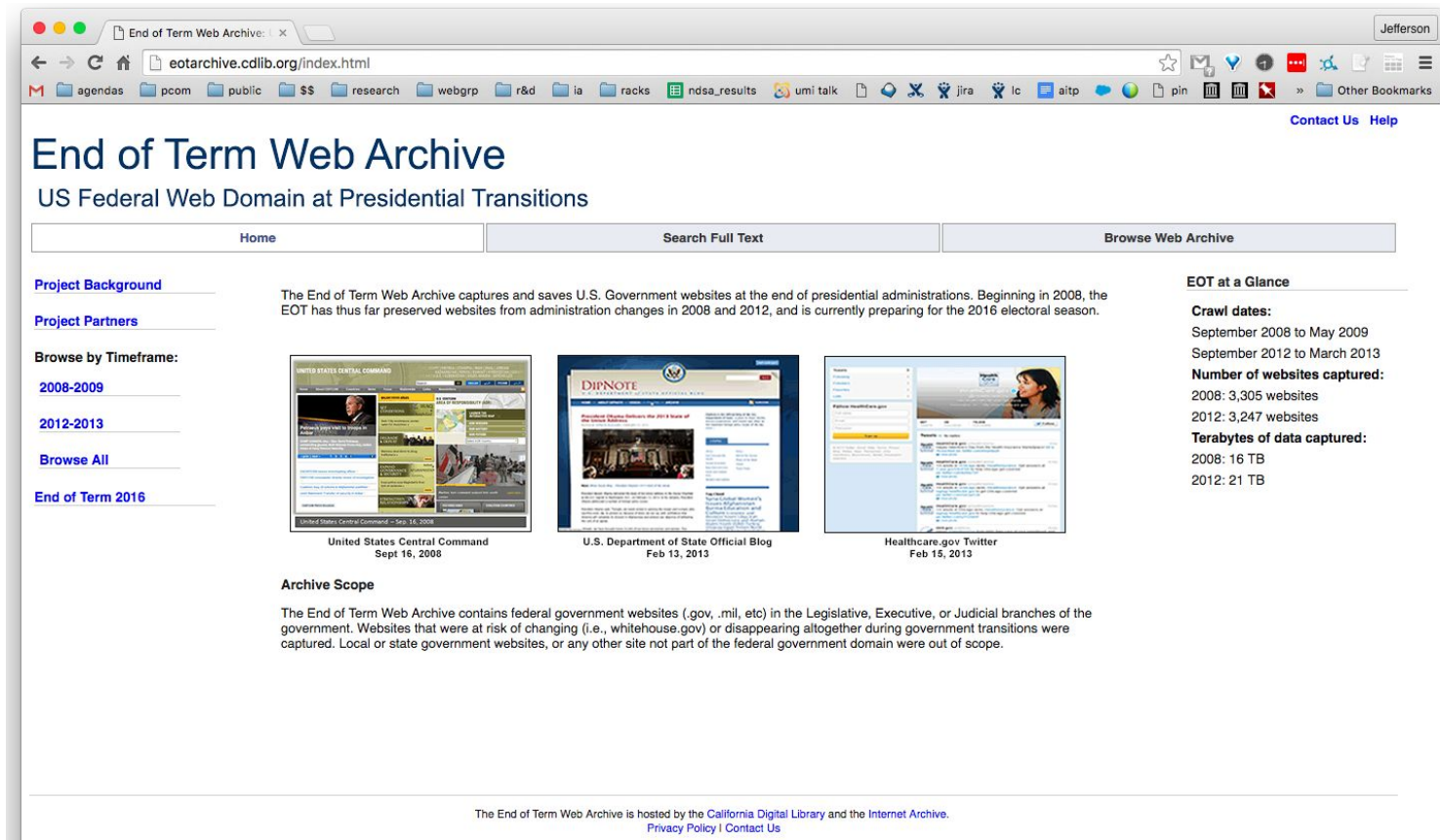
EOT Challenges

- **Typical web archiving challenges**
 - complexity of content
 - volume & proliferation
 - “you get what you get” w/ little cataloging or QA
- **Distribution of work**
 - more partners = more project/partner mgmt
 - contributed seed lists
- **Resource constraints**
 - the “it isn’t anyone’s actual job” problem
 - tech, time limitations & scale of data
 - funding = (there is none)



Using the EOT Archive

End of Term Web Archive



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

End of Term Web Archive

US Federal Web Domain at Presidential Transitions

[Home](#)[Search Full Text](#)[Browse Web Archive](#)

Results: 67 Items

Sorted by: [relevance](#) [Go](#)

Browse Web Archive

Page: [1](#) [2](#) [3](#) [4](#) [Next](#)

Site Lookup:

[Go](#)[Clear](#)

Look up a site by keywords in the title, description or URL.

Administration

- [2008](#) (63)
- [2012](#) (4)

Government Agency Domains

Select domain(s) that contain a specific agency name(s) in the URL

- [whitehouse](#) (67)
- [senate](#) (2)
- [petitions](#) (1)

Title: [Senator Sheldon Whitehouse | Senator for Rhode Island: Home](#)Archival URL: <http://eot.us.archive.org/eot/20080915214855/whitehouse.senate.gov/>Live URL: <http://whitehouse.senate.gov/>

Coverage: September 15, 2008 - August 14, 2009

Description: Sheldon Whitehouse - United States Senator for Rhode Island and Providence Plantations

Title: [Home | United States Senator Sheldon Whitehouse](#)Archival URL: <http://eot.us.archive.org/eot/20130321005322/http://whitehouse.senate.gov/>Live URL: <http://whitehouse.senate.gov/>

Coverage: September 13, 2012 - March 21, 2013

Title: [White House Radio](#)Archival URL: <http://eot.us.archive.org/eot/20080915223400/whitehouse.gov/radio/>Live URL: <http://www.whitehouse.gov/radio/>

Coverage: September 15, 2008 - August 25, 2009

Title: [Ask the White House](#)

Site Lookup:

[Go](#)[Clear](#)

Look up a site by keywords in the title, description or URL.

Administration

- [2008](#) (3305)
- [2012](#) (3247)

Government Agency Domains

Select domain(s) that contain a specific agency name(s) in the URL

- [house](#) (1056)
- [nasa](#) (310)
- [senate](#) (271)
- [facebook](#) (257)
- [uscourts](#) (257)
- [supremecourt](#) (156)
- [twitter](#) (155)
- [noaa](#) (151)
- [usgs](#) (139)
- [nih](#) (123)
- [gsfc](#) (98)
- [youtube](#) (94)
- [usda](#) (82)
- [army](#) (76)
- [whitehouse](#) (67)

[more](#)

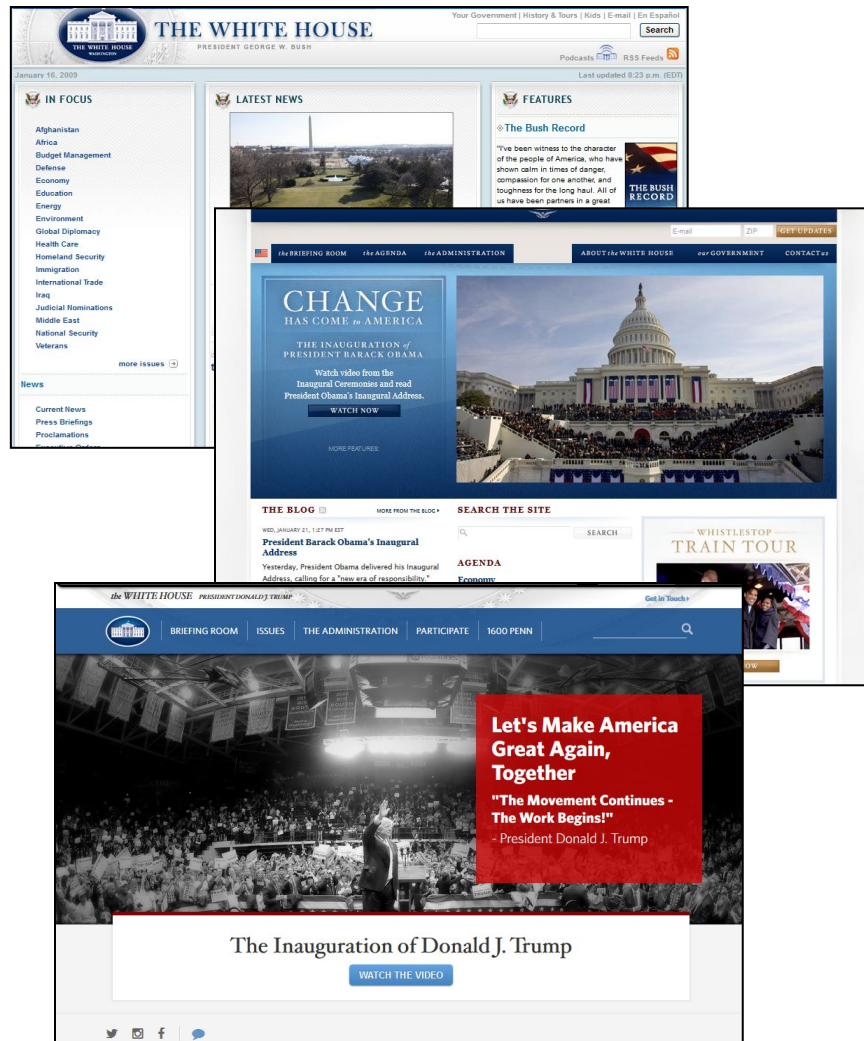
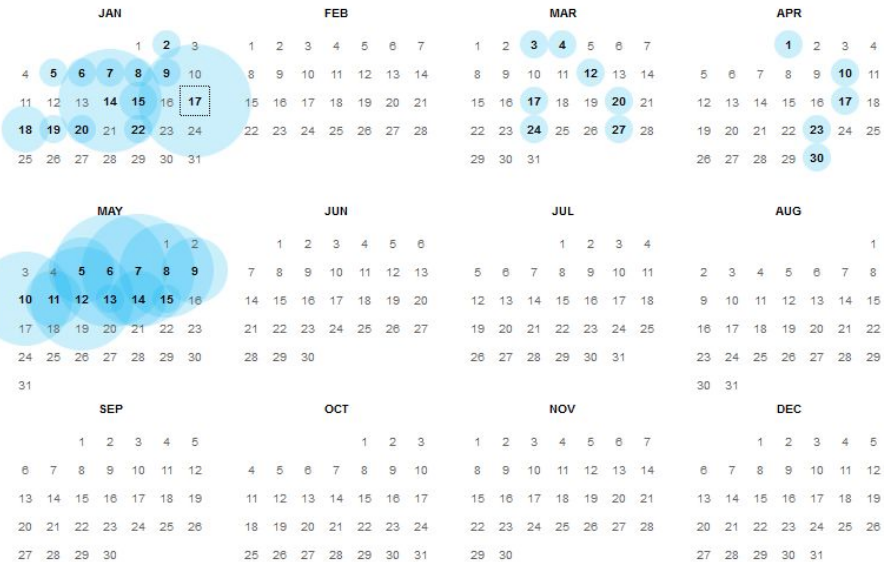
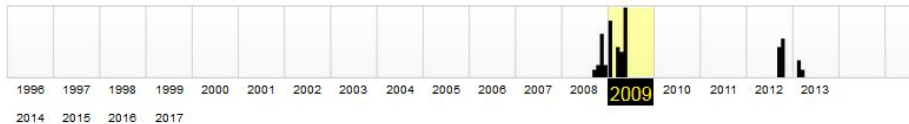
http://eot.us.archive.org/eot/*/www.whitehouse.gov



<http://www.whitehouse.gov>

Go Wayback!

<http://www.whitehouse.gov> has been crawled 154 times going all the way back to September 15, 2008. A crawl can be a duplicate of the last one. It happens about 25% of the time across 420,000,000 websites. [FAQ](#)



Plans for release of 2016 – 2017

- All web crawl data from IA, LC, UNT has been ingested at IA.
- Derivative datasets for all the data (WATs, WANEs, extracted page text) have been generated.
- Components to integrate new content into portal are being worked on (metadata, search, thumbnails, Wayback indexes). Once finalized, CDL will begin process to update the portal.
- We have a number of researchers interested in the data (IA working with them)

analysis by project team members

mark e. phillips journal

[HOME](#) [ABOUT](#)

SEARCH RESULTS FOR: EOT

eot

Search

Comparing Web Archives: EOT2008 and EOT2012 – Curator Intent

This is another post in a series that I've been doing to compare the End of Term Web Archives from 2008 and 2012. If you look back a few posts in this blog you will see some other analysis that I've done with the datasets so far. One thing that I am interested in understanding [...]

This entry was posted in [thinking outloud](#) on [June 20, 2016](#).

RECENT POSTS

[Metadata Quality Interfaces: Cluster Dashboard \(OpenRefine Clustering Baked Right In\)](#)

[Metadata Interfaces: Search Dashboard](#)
[Metadata Quality Interfaces: Element Count Dashboard](#)

[Metadata Quality Interfaces: Facet Dashboard](#)

[Updating Metadata Interfaces: Item Views](#)

Comparing Web Archives: EOT2008 and EOT2012 – What disappeared

ARCHIVES

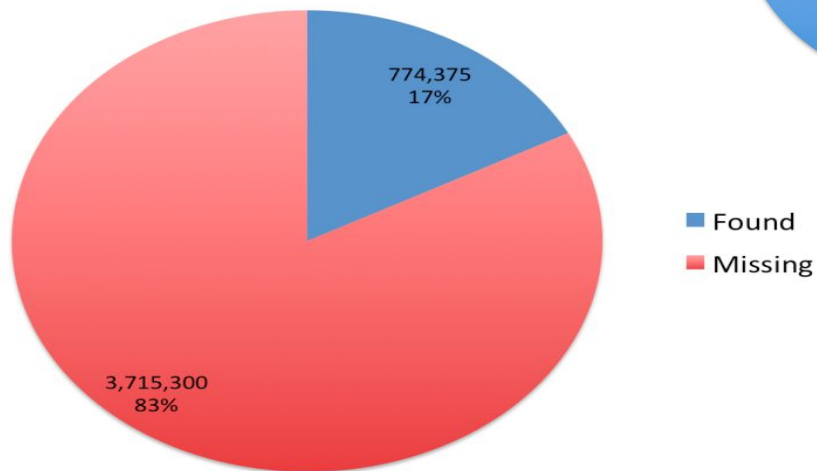
This is the fourth post in a series that looks at the End of Term Web Archives captured in 2008

[September 2017](#)

<http://vphill.com/journal/?s=eot>

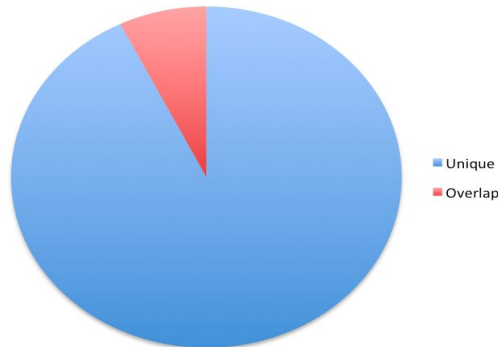
Comparing PDFs in EOT from 2008 to 2012

EOT2008 PDFs in EOT2012 Web Archive

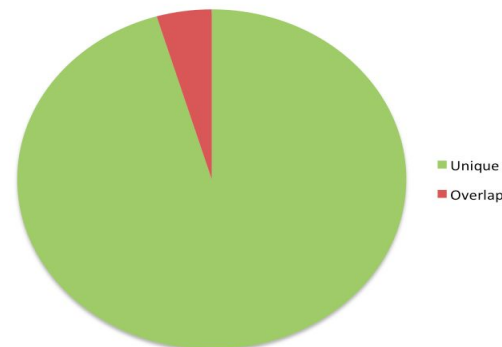


<http://vphill.com/journal/post/5872/>

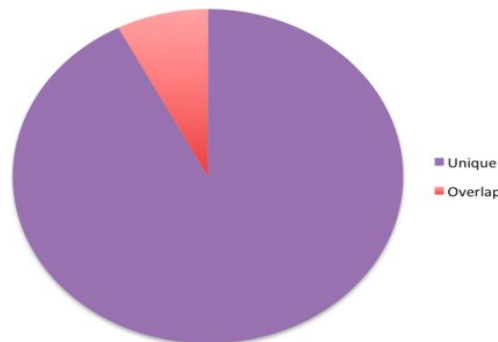
IA Unique



LOC Unique



UNT Unique



<http://vphill.com/journal/post/5861/>

.gov & .mil biggest change

Domain	EOT2008	EOT2012	Change	% Change
house.gov	13,694,187	35,894,356	22,200,169	162%
senate.gov	5,043,974	9,924,917	4,880,943	97%
gpo.gov	8,705,511	3,888,645	-4,816,866	-55%
nih.gov	5,276,262	1,267,764	-4,008,498	-76%
nasa.gov	6,693,542	3,063,382	-3,630,160	-54%
navy.mil	94,081	3,611,722	3,517,641	3,739%

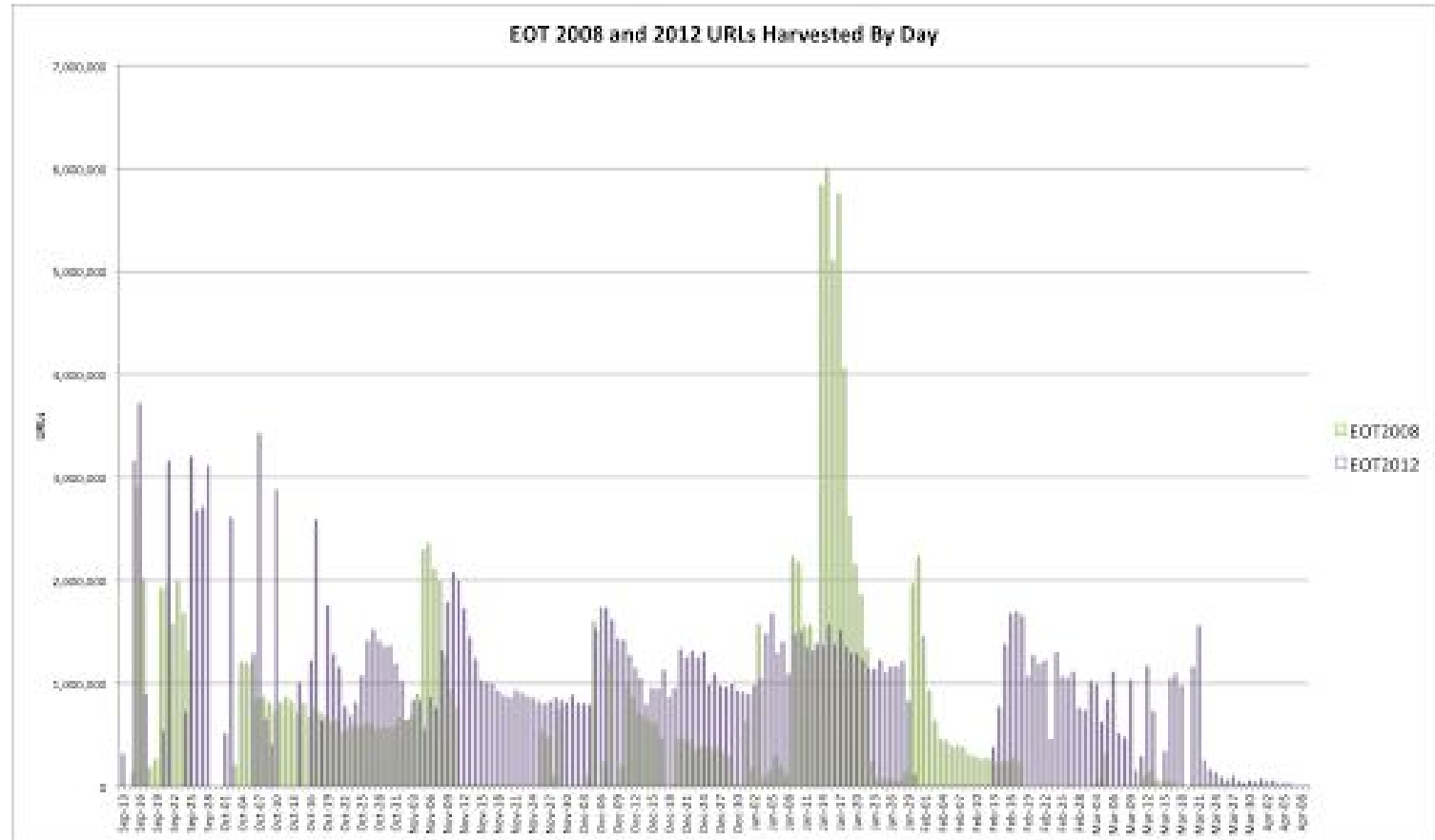
Top 15 .gov & .mil domains present in 2008 but missing in 2012

Domain	URLS
geodata.gov	812,524
nifl.gov	504,910
stat-usa.gov	398,961
tradestatsexpress.gov	243,729
arnet.gov	174,057
acqnet.gov	171,493
dccourts.gov	161,289
web-services.gov	137,202
metrokc.gov	132,210
sdi.gov	91,887
davie-fl.gov	88,123
belmont.gov	87,332
aftac.gov	84,507
careervoyages.gov	57,192
women-21.gov	56,255

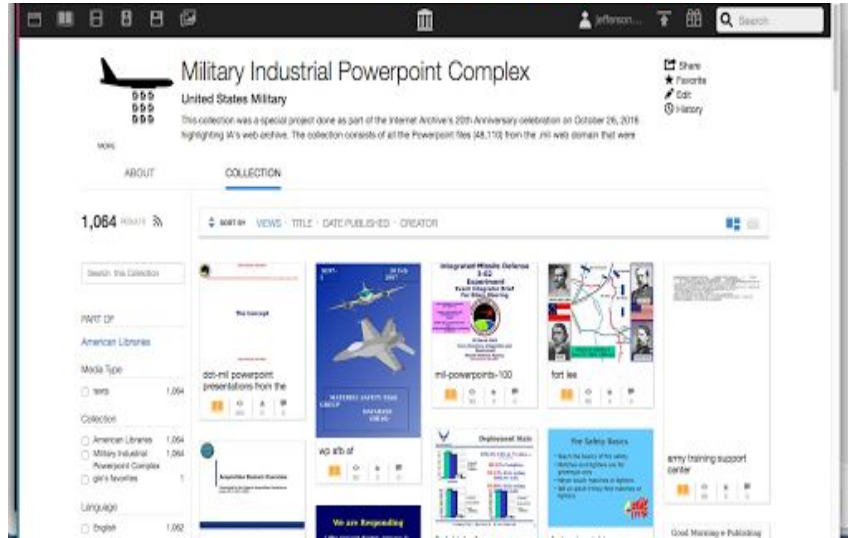
Top 15 .gov & .mil domains new in 2012

Domain	URLS
militaryonesource.mil	859,843
consumerfinance.gov	237,361
nrd.gov	194,215
wh.gov	179,233
pnnl.gov	132,994
eia.gov	112,034
transparency.gov	109,039
nationalguard.mil	108,854
acus.gov	93,810
404.gov	82,409
savingsbondwizard.gov	76,867
treasuryhunt.gov	76,394
fedshirevets.gov	75,529
onrr.gov	75,484
veterans.gov	75,350

EOT2008 and EOT2012 Crawling Schedule



Extracted Special Web Collections



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

<http://archive.org/~vinay/20th-century-gov-headshots.html>

<http://archive.org/~vinay/20th-century-gov-groupshots.html>



eot 2016 “raw” content

The screenshot displays the Internet Archive interface for the 'End of Term 2016 Web Crawls' collection. At the top, a dark navigation bar includes icons for audio, video, and text, followed by a search bar and a user profile 'jefferson...'. Below this, a secondary navigation bar lists links: ABOUT, CONTACT, BLOG, PROJECTS, HELP, DONATE, JOBS, VOLUNTEER, and PEOPLE.

The main header features the collection title 'End of Term 2016 Web Crawls' next to an American flag icon. A descriptive paragraph states: 'This collection contains web crawls performed as part of the End of Term Web Archive, a collaborative project that aims to preserve the U.S. federal government web presence at each change of administration. Content includes publicly-accessible government websites'. To the right of the title are links for 'Share', 'Favorite', 'Edit', 'History', and 'Play All'.

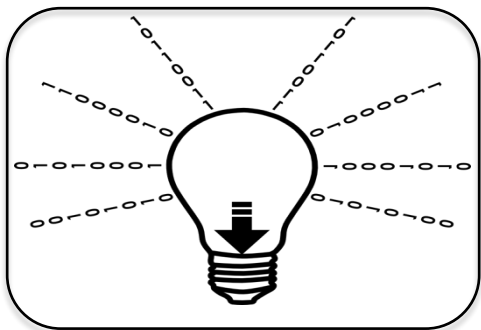
Below the header, two tabs are visible: 'ABOUT' and 'COLLECTION', with 'COLLECTION' being the active tab. On the left side, a sidebar shows '19,084 RESULTS' and a search box labeled 'Search this Collection'. Below this, the 'PART OF' section lists 'End of Term Web Crawls' and 'Internet Archive Web Crawls'. The 'Media Type' section shows 'collections' (6) and 'web' (19,078). The 'Topics & Subjects' section lists various categories with counts: 'crawldata' (15,168), 'congress' (5), '2016' (4), 'end of term' (4), 'federal government' (4), and 'president' (4). A 'More' link is also present. The 'Collection' section shows 'End of Term 2016' with 19,084 items.

The main content area displays a grid of 10 items. Each item has a thumbnail image, a title, and a count of items. The items are:

- End Of Term 2016 Pre-Inauguration Crawls (4,693 items)
- Government Web & Data Archive (6,268 items)
- End of Term 2016 Post-Inauguration Crawls (5,308 items)
- End Of Term 2016 Library of Congress Crawls (3,892 items)
- End of Term 2016 UNT Crawls (1,275 items)
- EndofTerm2016WebCrawls Crawldata 2016-11- (21,648 items)
- EndOfTerm2016PreinaugurationCrawls Crawldata (0 items)
- EndofTerm2016WebCrawls Crawldata 2016-11- (0 items)
- Archive Team Contributed GOV FTP Grabs

<https://archive.org/details/EndOfTerm2016WebCrawls>

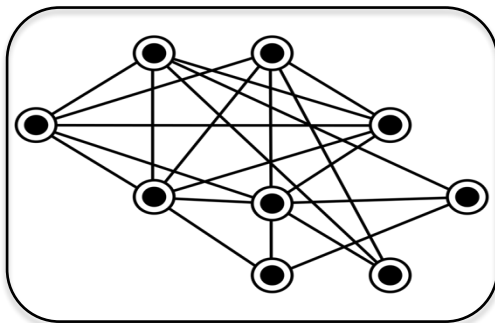
researcher access to .gov



WAT Datasets

(Web Archive
Transformation)

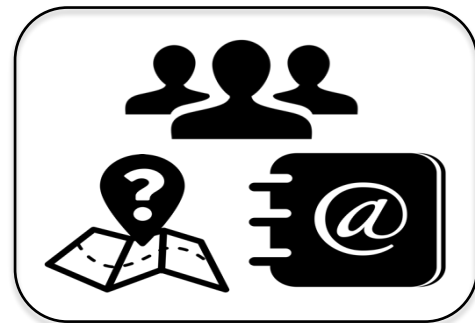
Key Metadata from Every
Resource



LGA Datasets

(Longitudinal
Graph Analysis)

What Links to What
over Time



WANE Datasets

(Web Archive
Named Entities)

Names of People, Places,
Organizations

Web Archive Datasets
(via platform, disk, APIs, etc.)

<http://webarchives.ca/>

web preservation for content creators

What can website creators do to help?

- Create preservation-friendly websites
- Understand the nature of your organization's website to determine the extent needed to preserve it:
 - What are the domains, subdomains?
 - What content is being hosted by third-party companies, for instance social media?
 - Is your content changing frequently?
- Archive your own content, using open source or subscription-based tools. Or at least make sure it is being preserved by someone else (check web.archive.org)
- Participate in our next End of Term 2020!

Creating Preservable Websites

- | | |
|--|---|
| ✓ Follow web standards and accessibility guidelines | ✓ Consider using a Creative Commons license |
| ✓ Be careful with robots.txt exclusions | ✓ Use sustainable data formats |
| ✓ Use a site map, transparent links, and contiguous navigation | ✓ Embed metadata, especially the character encoding |
| ✓ Maintain stable URIs and redirect when necessary | ✓ Use archiving-friendly platform providers and content management systems. |

For more details see Library of Congress Guide to Creating Preservable Websites

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

More Guidance for Website Creators

- **International Internet Preservation Consortium Tools and Software**

<https://netpreserve.org/web-archiving/tools-and-software/>

- **Recommended Formats Statement for Websites**

<https://www.loc.gov/preservation/resources/rfs/websites.html>

- **Stanford University's Archivability Guidelines**

<http://library.stanford.edu/projects/web-archiving/archivability>

- **Columbia University's Guidelines for Preservable Websites**

https://library.columbia.edu/bts/web_resources_collection/guidelines_for_preservable_websites.html

- **Princeton University's Guidelines for Designing Preservation-Friendly Websites**

<https://rbcs.princeton.edu/policies/guidelines-designing-preservation-friendly-websites>

- **Archive Ready, a free website archivability evaluation tool** <http://www.archiveready.com/>

EOT Going Forward

- **Processing the 2016 archive:** Full-text search, build index, make derivatives, simple extractions like domains, comparisons against 2012/2008
- **Providing access:** Updating the EOT portal, exploring media search, new access interfaces, sharing derivative data sets with researchers
- **Preservation copies:** Transferring to Library of Congress this year.
- And of course ...**End of Term 2020!**

Follow us on Twitter for updates: <https://twitter.com/eotarchive>

Questions?

abgr@loc.gov

eot-info@archive.org