# Small-Scale Web Archiving in an Age of Uncertainty

**2017 Federal Depository Library Conference**

**Kelly L. Smith**
**UC San Diego**

A US government website that used to warn about the risks of oil and gas drilling was changed to promote their economic benefits

The Census director resigns just as the $1.5 billion agency heads into its biggest test. Next in command may be a weather forecaster.

White House posts wrong versions of Trump's orders on its website

EPA website removes climate science site from public view after two decades

Feds fight suit over web takedown of animal abuse data

Trump taps anti-LGBTQ activist Roger Severino to lead HHS Civil Rights Office

FCC votes to advance net neutrality repeal

Here are the 66 programs eliminated in Trump's budget

The EPA’s Science Office Removed “Science” From Its Mission Statement

ICE ERO immigration arrests climb nearly 40%

Chaffetz Responds to FBI Letter Regarding Comey Memos

# Why archive?

- Gov websites have always changed/disappeared for variety of reasons, and every new administration changes web to reflect agenda
- What’s new: chaos and controversy
- Examples:
    - Change in agenda = significant policy changes = controversies
    - Executive Orders and regulatory changes
    - Page content and/or entire pages changed or deleted without notice
    - Defunding/elimination threats for offices
    - Lack of transparency
    - Greater blending of political/governmental (e.g. election commission)
    - Lack of governmental experience for top administrators
    - Business ties and other ethical considerations

# Large-scale archiving of federal info

- End of Term (EOT) project – between Fall 2016-Spring 2017, archived over 350 million federal URLs/files. UNT continuing effort.
- WebHarvest.gov – Congressional websites harvested at end of each Congress since 2006 by NARA
- FDLP Web Archive – crawls select sites within scope of FDLP (currently 140)
- Library of Congress collections – not specific to federal info, but some included (e.g. federal courts, congressional, Homeland Security). Not all are ongoing projects.
- GovernmentAttic and the Memory Hole - activist, not “official” sites; archive of documents obtain via FOIA and other means

# Other large-scale initiatives : new

- DataRefuge
  - focus on climate/environmental data
  - most content also available in Wayback; home interface preserves content difficult to harvest with web crawlers
- Environmental Data and Governance Initiative (EDGI)
  - focus on scientific data and web pages from EPA, DOE, NOAA, OSHA, NASA, USDA, DOI, and USGS
  - worked with EOT and sponsored several data rescue events; created public archiving tools
  - also conducts website monitoring

# Small-scale archiving

- Larger projects are great but not always immediate
- Opportunity for those with limitations (e.g. time/tech skills/networking) to participate
- Opportunity to embrace online collection development/maintenance
- Opportunity to enhance own current awareness & education
- Opportunity to contribute to preservation of historical record

# What should you consider for archiving?

- What you can easily incorporate into existing workflows
- What you care about – for yourself/your users
- What you stumble upon and think should be preserved for future research
- What you can track easily enough
- What you can get Administrative/IT support for
- What you can’t easily schedule for automated crawls

# What I’m doing - GovSpeak

- Tweaked my own understanding of the guide from just a list of acronyms to a directory of links for Wayback
- Checked 1000s of URLs submitted for EOT project against GovSpeak : 4200 links ----> 5500 links
- Tweaked processes:
    - every new entry immediately saved in Wayback to ensure at least one capture
    - each time I verify links, check questionable ones against Wayback and add as necessary
    - in February, began retaining links to old urls to help track

# GovSpeak: A Guide to U.S. Government Acronyms & Abbreviations: S

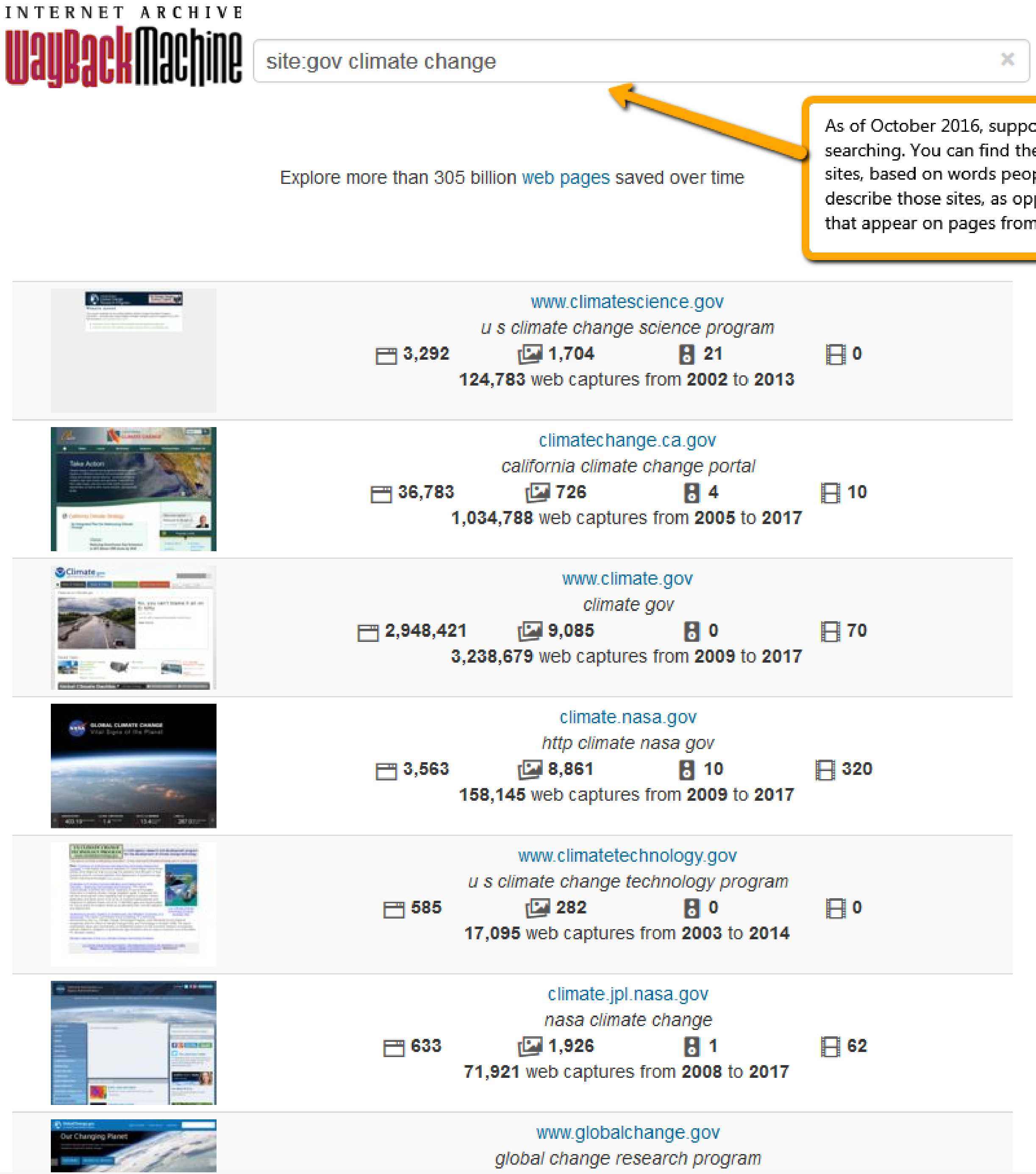
Search this Guide | Search

The most current, extensive and fully-linked listing of U.S. government agency, office, program and publication acronyms.

Home | A | B | C | D | E | F | G | H | I | J-K | L | M | N | O | P | Q | R | S | T | U | V | W | XYZ

S

| | |
|---|---|
| S&T | Science and Technology Directorate |
| S&TF | Science and Technology Facility |
| SAB | EPA Science Advisory Board |
| SABER | Statistical Analysis Battery for Epidemiological Research *(unable to verify)* |
| SABIT | Special American Business Internship Training |
| SABS | School Attendance Boundary Survey |
| SAC | Small Agency Council |
| SACC | Science Advisory Committee on Chemicals |
| SACGHS | Secretary's Advisory Committee on Genetics, Health and Society *(unable to verify 7/17)* |
| SACGT | Secretary's Advisory Committee on Genetic Testing *(absorbed by SACGHS)* |
| SACO | Subject Authority Cooperative Program |
| SACX | Secretary's Advisory Committee on Xenotransplantation *(defunct)* |
| SaDIP | Safety Data Improvement Program |
| SAE | State and Metro Area Employment, Hours, & Earnings |
| SAER | LLNL Site Annual Environmental Report |
| SAFECOM | Safety Communications |
| SAFER | Safety and Fitness Electronic Records \| Staffing for Adequate Fire & Emergency Response Grants Program |
| SAHIE | Small Area Health Insurance Estimates |
| SAIF | Saving Associate Insurance Fund *(merged into DIF)* |
| SAIPE | Small Area Income & Poverty Estimates |
| SAM | System Advisor Model \| System for Award Management *(older link archive)* |
| SAMHDA | Substance Abuse and Mental Health Data Archive |
| SAMHSA | Substance Abuse and Mental Health Services Administration |
| SAO | Smithsonian Astrophysical Observatory |
| SAOP | Senior Agency Officials for Privacy *(older link archive)* |
| SAP | Scientific Advisory Panel |
| SARA | Service and Advice for Research and Analysis \| Stock Assessment Results Archive |
| SARC | Stock Assessment Review Committee |
| SARD | Safety and Assurance Requirements Division |
| SARP | Sectoral Applications Research Program *(older link archive)* |
| SAS | Service Annual Survey \| Surveillance and Aviation Section |
| SASS | Schools and Staffing Survey |
| SAVE | Systematic Alien Verification for Entitlement |



INTERNET ARCHIVE WaybackMachine

site:gov climate change

Explore more than 305 billion web pages saved over time

As of October 2016, supports limited keyword searching. You can find the homepages of sites, based on words people have used to describe those sites, as opposed to words that appear on pages from sites.

www.climatescience.gov
*u s climate change science program*
3,292 1,704 21 0
**124,783** web captures from **2002** to **2013**

climatechange.ca.gov
*california climate change portal*
36,783 726 4 10
**1,034,788** web captures from **2005** to **2017**

www.climate.gov
*climate gov*
2,948,421 9,085 0 70
**3,238,679** web captures from **2009** to **2017**

climate.nasa.gov
*http climate nasa gov*
3,563 8,861 10 320
**158,145** web captures from **2009** to **2017**

www.climatetechnology.gov
*u s climate change technology program*
585 282 0 0
**17,095** web captures from **2003** to **2014**

climate.jpl.nasa.gov
*nasa climate change*
633 1,926 1 62
**71,921** web captures from **2008** to **2017**

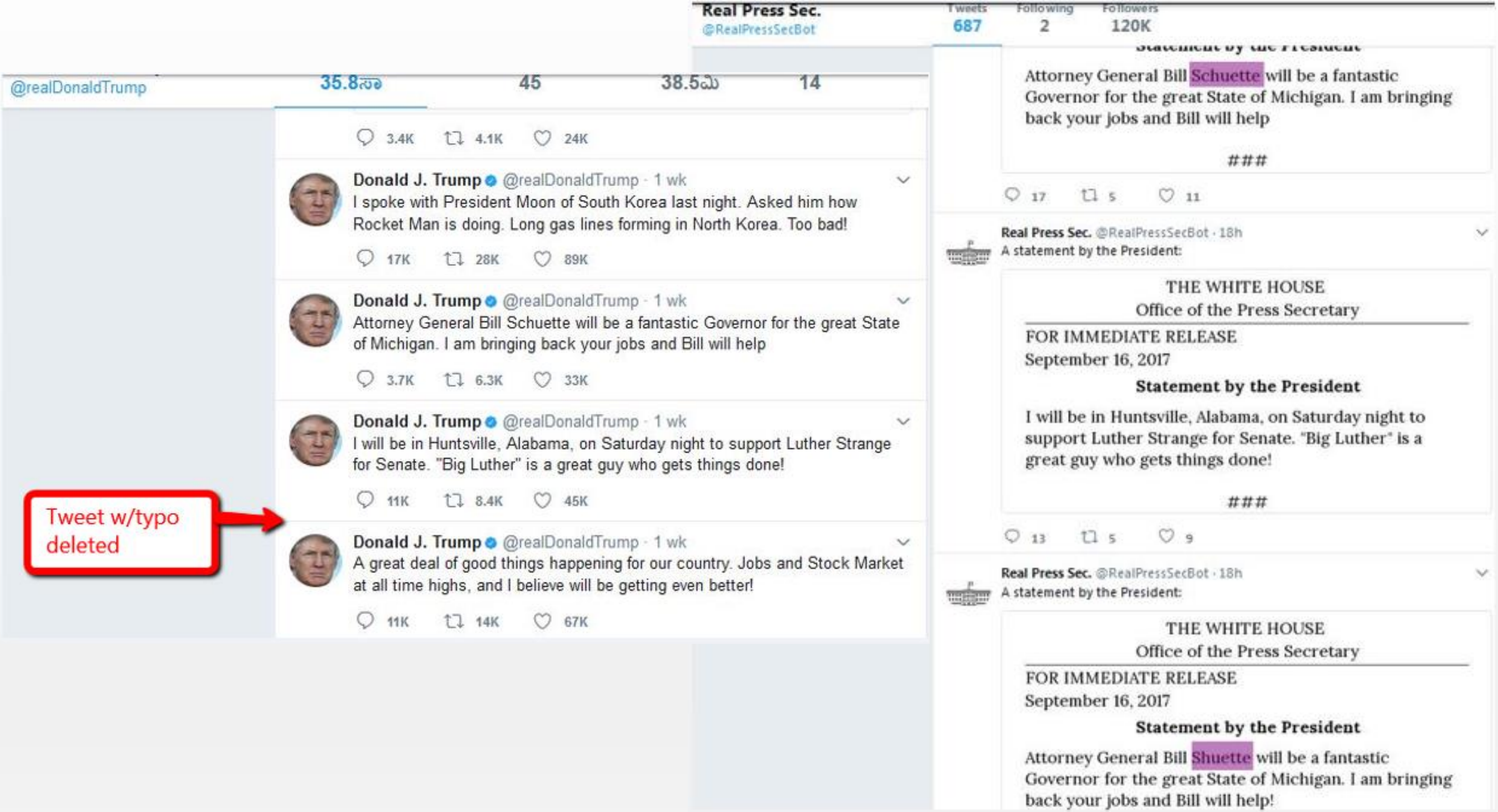
www.globalchange.gov
*global change research program*

# Additional archiving efforts

- 2016 Election Documents – official documents related to election controversies, including Clinton email investigations, Russian hacking, Trump ethics
- Discussed automated crawls – esp. for NOAA and ICE/USCIS - but wasn’t feasible
- Select news releases
  - All department-level (Commerce, Energy, Education, Homeland Security, etc.) Also BOC, CBP, CDC, CENTCOM, CFPB, CIA, DNI, EPA, FBI, FDA, ICE, NOAA, NSA, OGE, SBA, SEC, TREAS, USCIS, USGS, USTR
  - House & Senate committees & minority pages
  - Congressional leadership
  - Individual representatives & senators when important

# Additional archiving efforts

- Using page monitor Chrome extension:
  - Some OIG pages
  - Presidential advisory commission on election integrity
  - GAO reports & testimony
  - CBO reports & cost estimates
  - GSA electronic FOIA reading room (Trump hotel: Old Post Office building)
- Closely monitor news (esp. via Twitter) and do on-the-fly archiving as needed
- UCSD GovInfo Facebook page – posting daily gov link
- Twitter feeds – @RealPressSecBot and @RealDonaldTrump



@realDonaldTrump 35.8ಸಾ 45 38.5ಮಿ 14

3.4K 4.1K 24K

**Donald J. Trump** @realDonaldTrump · 1 wk
I spoke with President Moon of South Korea last night. Asked him how Rocket Man is doing. Long gas lines forming in North Korea. Too bad!

17K 28K 89K

**Donald J. Trump** @realDonaldTrump · 1 wk
Attorney General Bill Schuette will be a fantastic Governor for the great State of Michigan. I am bringing back your jobs and Bill will help

3.7K 6.3K 33K

**Donald J. Trump** @realDonaldTrump · 1 wk
I will be in Huntsville, Alabama, on Saturday night to support Luther Strange for Senate. "Big Luther" is a great guy who gets things done!

11K 8.4K 45K

**Donald J. Trump** @realDonaldTrump · 1 wk
A great deal of good things happening for our country. Jobs and Stock Market at all time highs, and I believe will be getting even better!

11K 14K 67K

Tweet w/typo deleted

**Real Press Sec.**
@RealPressSecBot
Tweets 687 Following 2 Followers 120K

Statement by the President

Attorney General Bill Schuette will be a fantastic Governor for the great State of Michigan. I am bringing back your jobs and Bill will help

###

17 5 11

**Real Press Sec.** @RealPressSecBot · 18h
A statement by the President:

THE WHITE HOUSE
Office of the Press Secretary

FOR IMMEDIATE RELEASE
September 16, 2017

**Statement by the President**

I will be in Huntsville, Alabama, on Saturday night to support Luther Strange for Senate. "Big Luther" is a great guy who gets things done!

###

13 5 9

**Real Press Sec.** @RealPressSecBot · 18h
A statement by the President:

THE WHITE HOUSE
Office of the Press Secretary

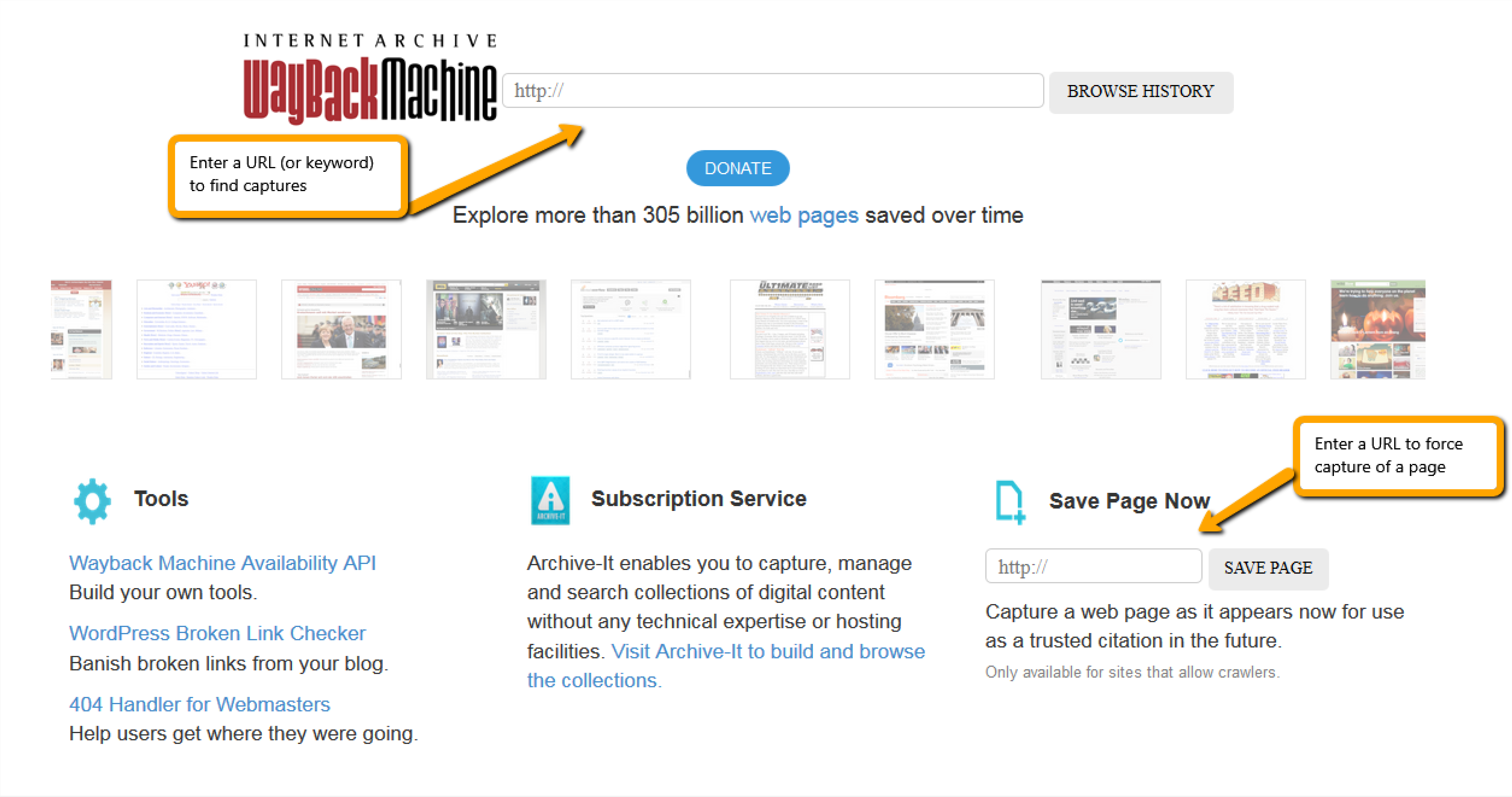
FOR IMMEDIATE RELEASE
September 16, 2017

**Statement by the President**

Attorney General Bill Shuette will be a fantastic Governor for the great State of Michigan. I am bringing back your jobs and Bill will help!

# Using Wayback Machine

- “save page now” is usually all you have to do, but…
- It won’t save pages with robots.txt files
- It won’t save media files
- It will only capture the exact page url you’ve entered – not the .pdf or other documents linked from that page
- Think about the path users will follow: home page > news > press releases > 2017
- There can be a time delay for saved pages to appear



Enter a URL (or keyword) to find captures

Explore more than 305 billion web pages saved over time

Wayback Machine Availability API
Build your own tools.

WordPress Broken Link Checker
Banish broken links from your blog.

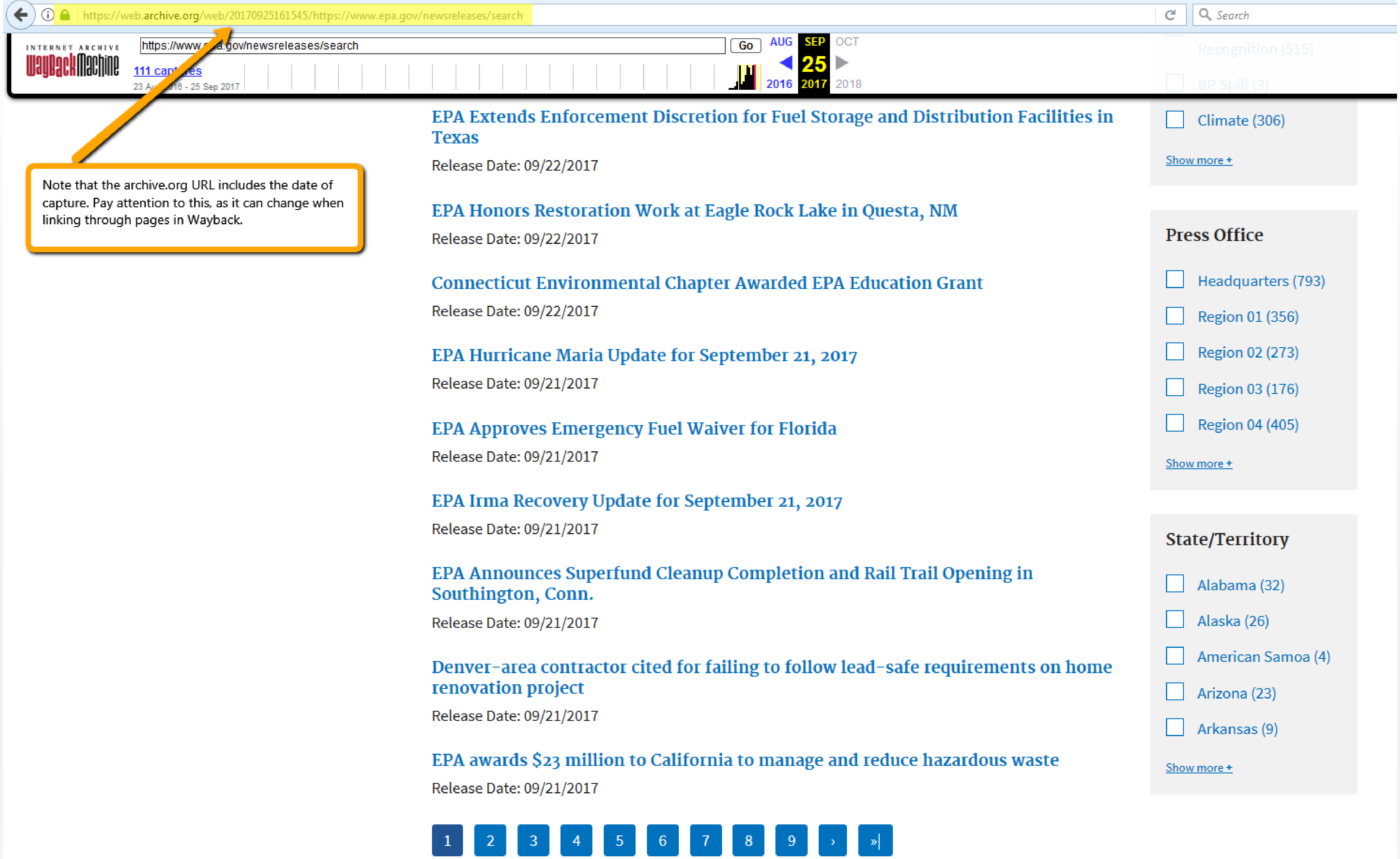
404 Handler for Webmasters
Help users get where they were going.

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit Archive-It to build and browse the collections.

Enter a URL to force capture of a page

Capture a web page as it appears now for use as a trusted citation in the future.

Only available for sites that allow crawlers.
]


https://web.archive.org/web/20170925161545/https://www.epa.gov/newsreleases/search
Search
INTERNET ARCHIVE
WaybackMachine
https://www.e a.gov/newsreleases/search
Go
111 capt es
23 Au 16 - 25 Sep 2017
AUG
SEP
OCT
2016
25
2017
2018

Note that the archive.org URL includes the date of capture. Pay attention to this, as it can change when linking through pages in Wayback.

EPA Extends Enforcement Discretion for Fuel Storage and Distribution Facilities in Texas

Release Date: 09/22/2017

EPA Honors Restoration Work at Eagle Rock Lake in Questa, NM

Release Date: 09/22/2017

Connecticut Environmental Chapter Awarded EPA Education Grant

Release Date: 09/22/2017

EPA Hurricane Maria Update for September 21, 2017

Release Date: 09/21/2017

EPA Approves Emergency Fuel Waiver for Florida

Release Date: 09/21/2017

EPA Irma Recovery Update for September 21, 2017

Release Date: 09/21/2017

EPA Announces Superfund Cleanup Completion and Rail Trail Opening in Southington, Conn.

Release Date: 09/21/2017

Denver-area contractor cited for failing to follow lead-safe requirements on home renovation project

Release Date: 09/21/2017

EPA awards $23 million to California to manage and reduce hazardous waste

Release Date: 09/21/2017

□ Recognition (515)

□ BP Spill (3)

□ Climate (306)

Show more +

### Press Office

□ Headquarters (793)

□ Region 01 (356)

□ Region 02 (273)

□ Region 03 (176)

□ Region 04 (405)

Show more +

### State/Territory

□ Alabama (32)

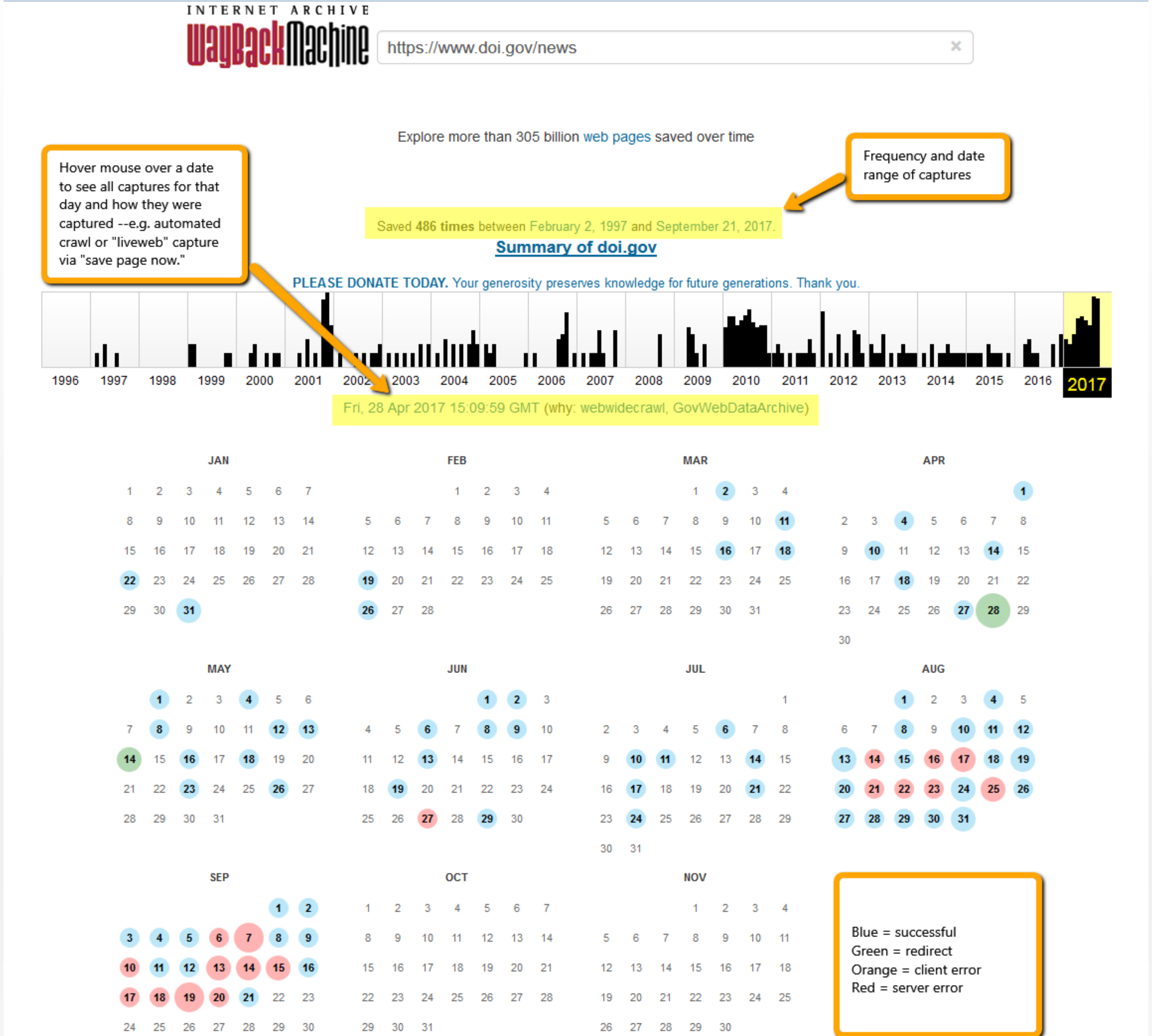
□ Alaska (26)

□ American Samoa (4)

□ Arizona (23)

□ Arkansas (9)

Show more +



Explore more than 305 billion web pages saved over time

Hover mouse over a date to see all captures for that day and how they were captured --e.g. automated crawl or "liveweb" capture via "save page now."

Frequency and date range of captures

Saved **486 times** between February 2, 1997 and September 21, 2017.

**Summary of doi.gov**

**PLEASE DONATE TODAY.** Your generosity preserves knowledge for future generations. Thank you.

Fri, 28 Apr 2017 15:09:59 GMT (why: webwidecrawl, GovWebDataArchive)

JAN

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 30 | 31 | | | | |

FEB

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | | | | |

MAR

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | 29 | 30 | 31 | |

APR

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | | | | | | |

MAY

| | | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 28 | 29 | 30 | 31 | | | |

JUN

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 25 | 26 | 27 | 28 | 29 | 30 | |

JUL

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | 31 | | | | | |

AUG

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 27 | 28 | 29 | 30 | 31 | | |

SEP

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 1 | 2 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 24 | 25 | 26 | 27 | 28 | 29 | 30 |

OCT

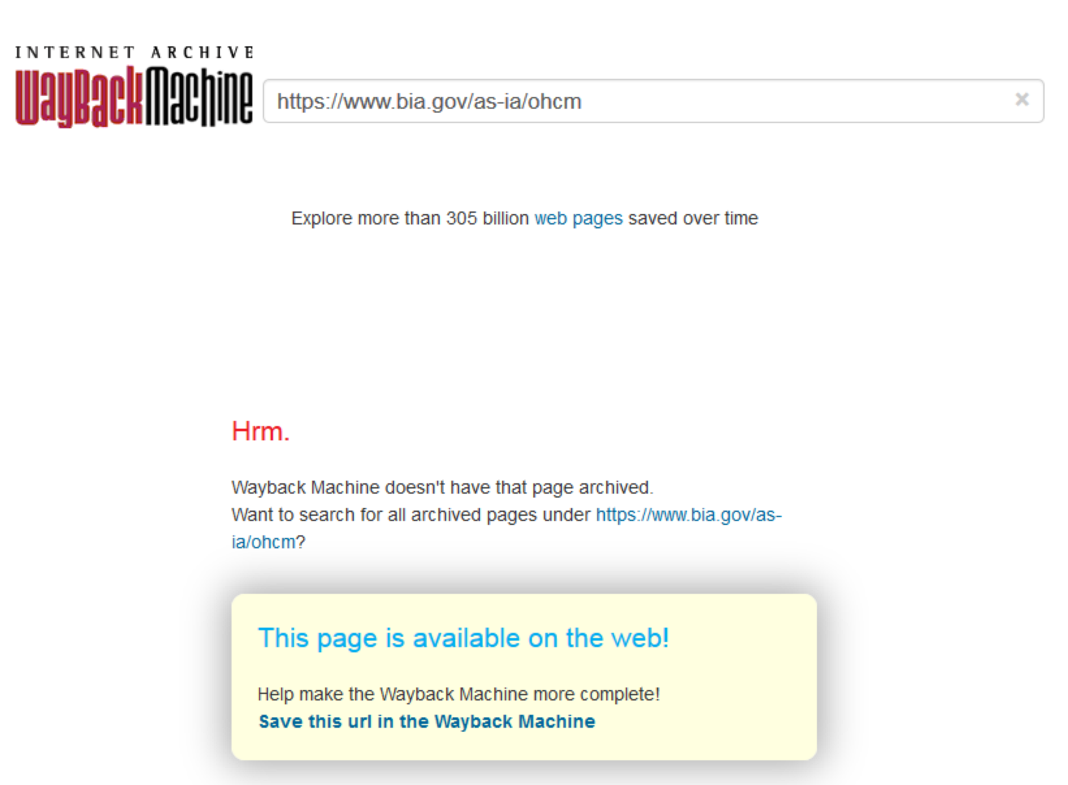
| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 30 | 31 | | | | |

NOV

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | 29 | 30 | | |

Blue = successful
Green = redirect
Orange = client error
Red = server error

# If no captures are found in Wayback…



https://www.bia.gov/as-ia/ohcm

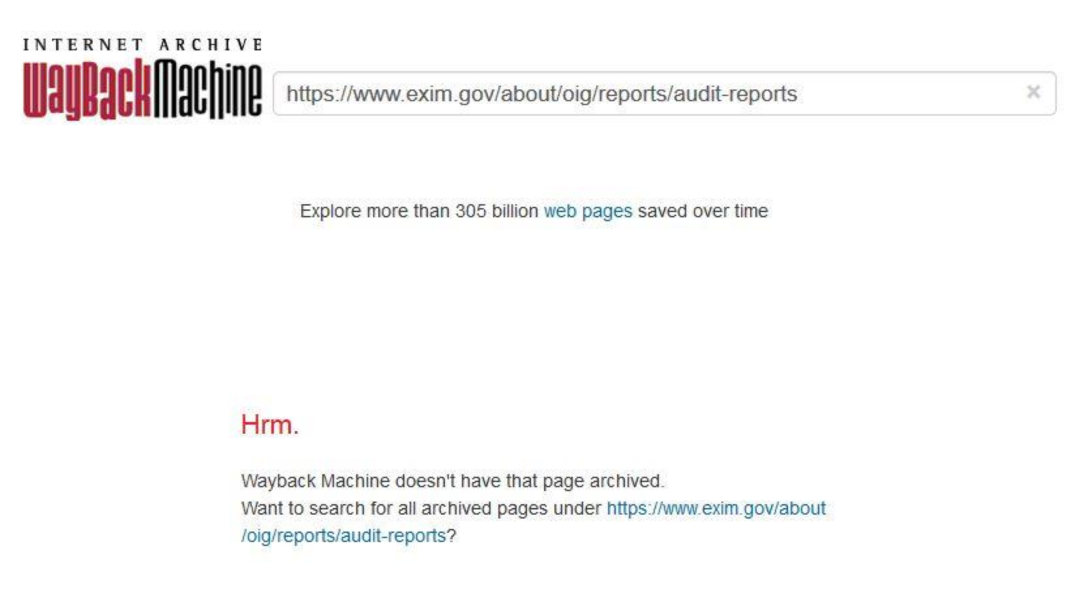
Explore more than 305 billion web pages saved over time

Hrm.

Wayback Machine doesn't have that page archived.
Want to search for all archived pages under https://www.bia.gov/as-ia/ohcm?

This page is available on the web!

Help make the Wayback Machine more complete!
**Save this url in the Wayback Machine**



https://www.exim.gov/about/oig/reports/audit-reports

Explore more than 305 billion web pages saved over time

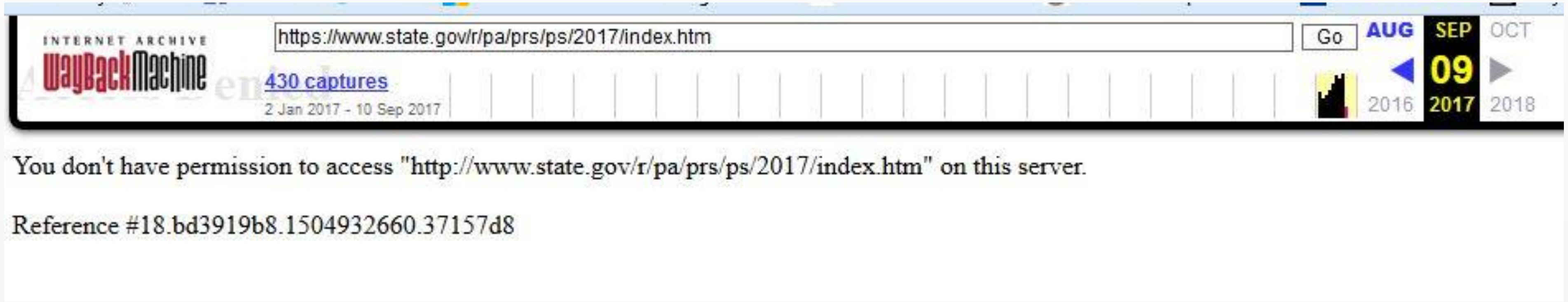
Hrm.

Wayback Machine doesn't have that page archived.
Want to search for all archived pages under https://www.exim.gov/about/oig/reports/audit-reports?

# Potential exceptions with Wayback Machine

- Dynamic pages (e.g. sequential pages, databases)
- Javascript
- Security certificates
- Permissions error
- Documents saved to Scribd or similar platforms
- Occasional “unable to archive” glitch – retry 2-3 times and it may save
- Automated crawls sometimes fail to capture successfully
- Excel files – not a problem, but…
- Rare but scary – pages saved one day but are seemingly not there the next
- Some things you just can’t save and you may not know why. Accept and move on.

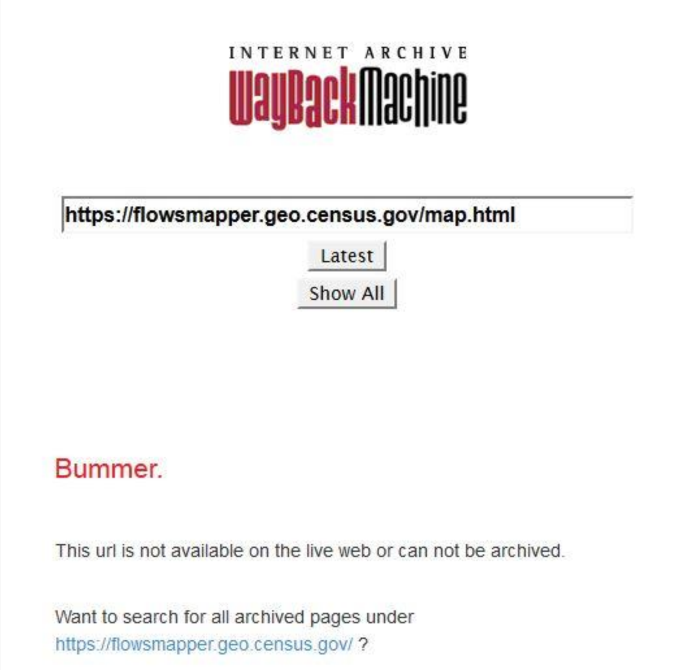
# If unable to archive in Wayback…



https://www.state.gov/r/pa/prs/ps/2017/index.htm

You don't have permission to access "http://www.state.gov/r/pa/prs/ps/2017/index.htm" on this server.

Reference #18.bd3919b8.1504932660.37157d8

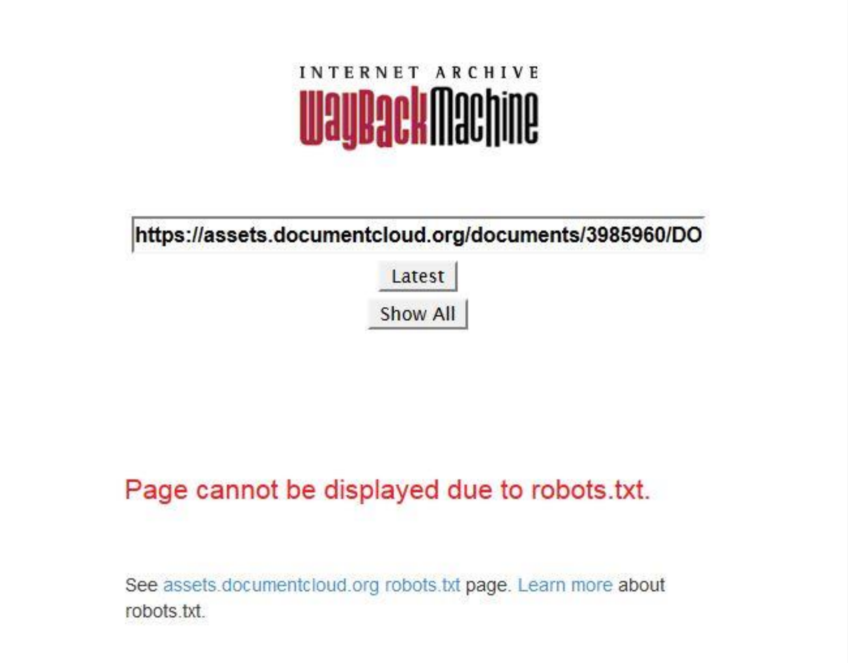


https://flowsmapper.geo.census.gov/map.html

Bummer.

This url is not available on the live web or can not be archived.

Want to search for all archived pages under https://flowsmapper.geo.census.gov/ ?



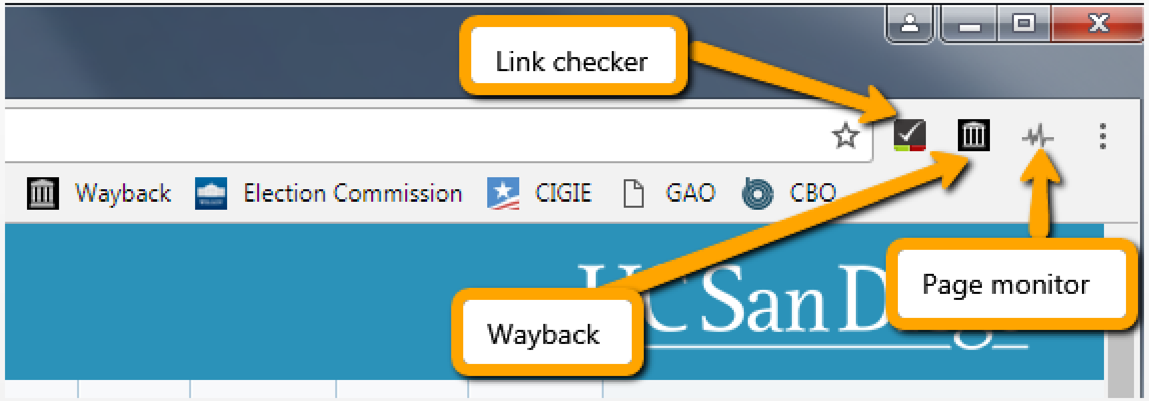
https://assets.documentcloud.org/documents/3985960/DO

Page cannot be displayed due to robots.txt.

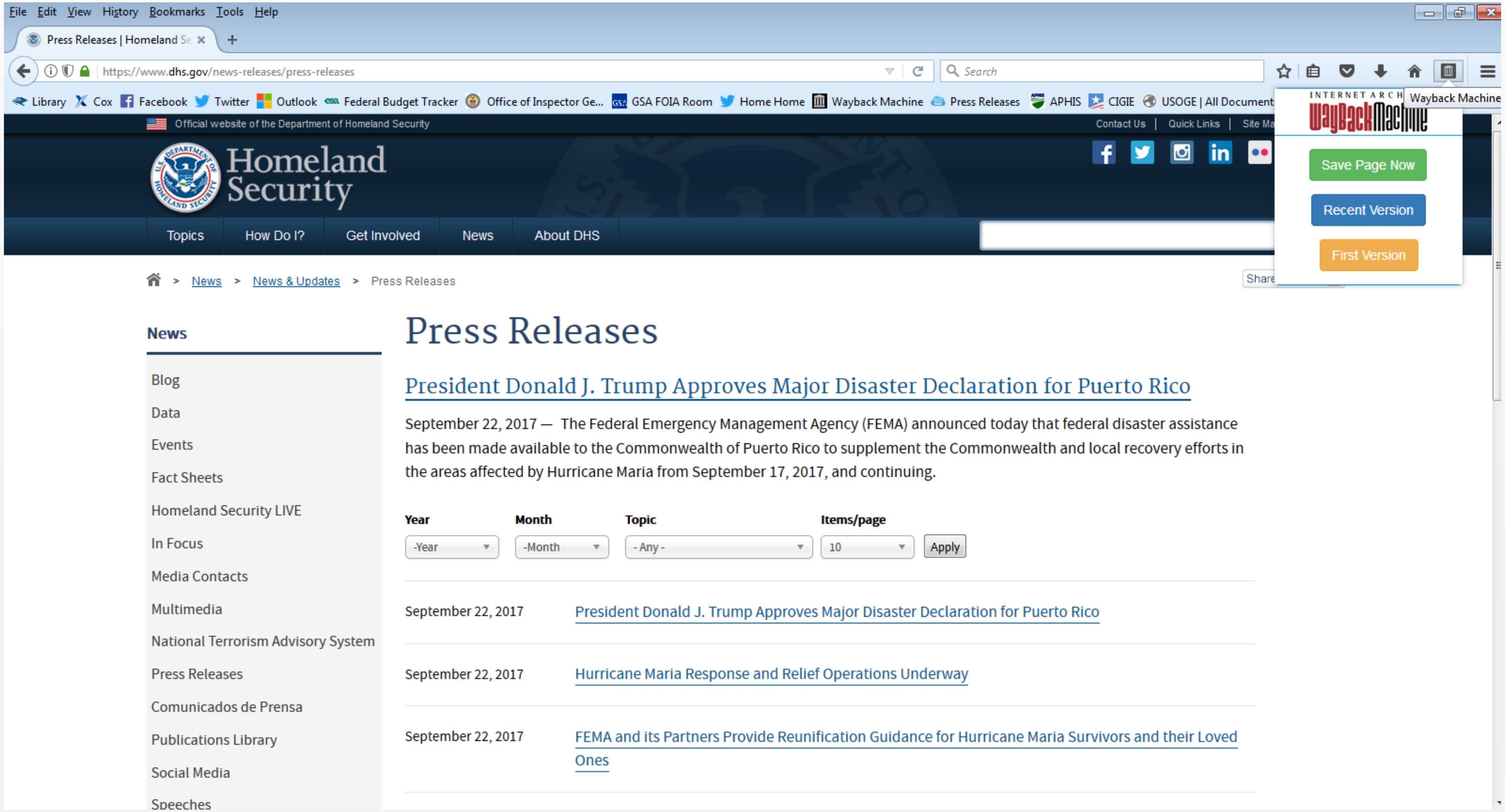
See assets.documentcloud.org robots.txt page. Learn more about robots.txt.

# Special tools

- Chrome Extensions
    - Wayback Machine
    - Check my links
    - Page monitor
- UNT nomination form
- Twitter bot - @LinkArchiver



# Wayback Machine extension



File Edit View History Bookmarks Tools Help
Press Releases | Homeland Se
https://www.dhs.gov/news-releases/press-releases
Search
Library Cox Facebook Twitter Outlook Federal Budget Tracker Office of Inspector Ge... GSA FOIA Room Home Home Wayback Machine Press Releases APHIS CIGIE USOGE | All Document
Wayback Machine
INTERNET ARCHIVE WayBackMachine
Save Page Now
Recent Version
First Version
Official website of the Department of Homeland Security
Contact Us | Quick Links | Site Ma
Homeland Security
Topics
How Do I?
Get Involved
News
About DHS
Share

News > News & Updates > Press Releases

## Press Releases

### President Donald J. Trump Approves Major Disaster Declaration for Puerto Rico

September 22, 2017 — The Federal Emergency Management Agency (FEMA) announced today that federal disaster assistance has been made available to the Commonwealth of Puerto Rico to supplement the Commonwealth and local recovery efforts in the areas affected by Hurricane Maria from September 17, 2017, and continuing.

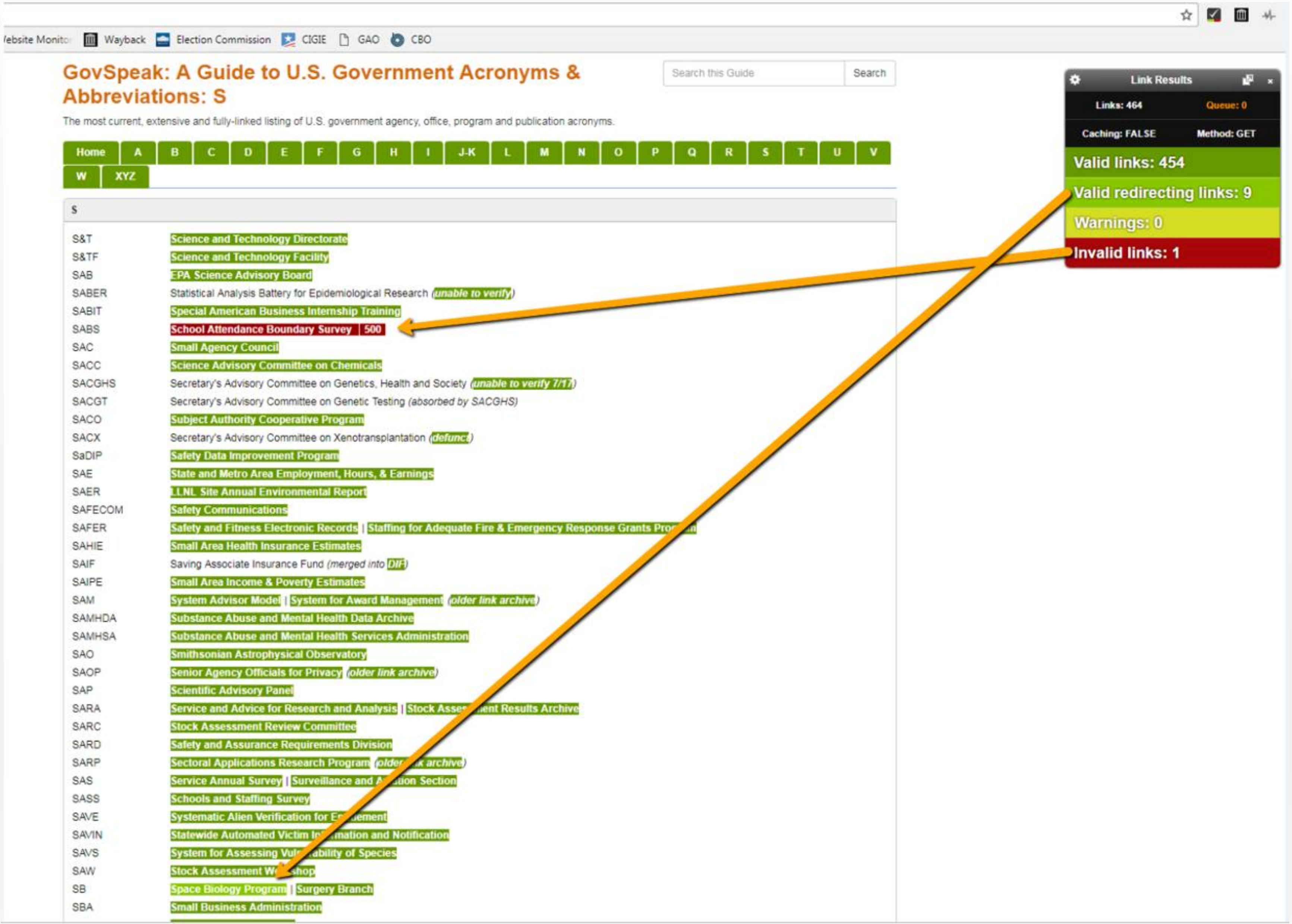
| Year | Month | Topic | Items/page | |
|---|---|---|---|---|
| -Year | -Month | - Any - | 10 | Apply |

| | |
|---|---|
| September 22, 2017 | President Donald J. Trump Approves Major Disaster Declaration for Puerto Rico |
| September 22, 2017 | Hurricane Maria Response and Relief Operations Underway |
| September 22, 2017 | FEMA and its Partners Provide Reunification Guidance for Hurricane Maria Survivors and their Loved Ones |

**News**

- Blog
- Data
- Events
- Fact Sheets
- Homeland Security LIVE
- In Focus
- Media Contacts
- Multimedia
- National Terrorism Advisory System
- Press Releases
- Comunicados de Prensa
- Publications Library
- Social Media
- Speeches

# Check my links extension



## GovSpeak: A Guide to U.S. Government Acronyms & Abbreviations: S

The most current, extensive and fully-linked listing of U.S. government agency, office, program and publication acronyms.

S

| | |
|---|---|
| S&T | Science and Technology Directorate |
| S&TF | Science and Technology Facility |
| SAB | EPA Science Advisory Board |
| SABER | Statistical Analysis Battery for Epidemiological Research *(unable to verify)* |
| SABIT | Special American Business Internship Training |
| SABS | School Attendance Boundary Survey \| 500 |
| SAC | Small Agency Council |
| SACC | Science Advisory Committee on Chemicals |
| SACGHS | Secretary's Advisory Committee on Genetics, Health and Society *(unable to verify 7/17)* |
| SACGT | Secretary's Advisory Committee on Genetic Testing *(absorbed by SACGHS)* |
| SACO | Subject Authority Cooperative Program |
| SACX | Secretary's Advisory Committee on Xenotransplantation *(defunct)* |
| SaDIP | Safety Data Improvement Program |
| SAE | State and Metro Area Employment, Hours, & Earnings |
| SAER | LLNL Site Annual Environmental Report |
| SAFECOM | Safety Communications |
| SAFER | Safety and Fitness Electronic Records \| Staffing for Adequate Fire & Emergency Response Grants Program |
| SAHIE | Small Area Health Insurance Estimates |
| SAIF | Saving Associate Insurance Fund *(merged into DIF)* |
| SAIPE | Small Area Income & Poverty Estimates |
| SAM | System Advisor Model \| System for Award Management *(older link archive)* |
| SAMHDA | Substance Abuse and Mental Health Data Archive |
| SAMHSA | Substance Abuse and Mental Health Services Administration |
| SAO | Smithsonian Astrophysical Observatory |
| SAOP | Senior Agency Officials for Privacy *(older link archive)* |
| SAP | Scientific Advisory Panel |
| SARA | Service and Advice for Research and Analysis \| Stock Assessment Results Archive |
| SARC | Stock Assessment Review Committee |
| SARD | Safety and Assurance Requirements Division |
| SARP | Sectoral Applications Research Program *(older link archive)* |
| SAS | Service Annual Survey \| Surveillance and [illegible] Section |
| SASS | Schools and Staffing Survey |
| SAVE | Systematic Alien Verification for Entitlement |
| SAVIN | Statewide Automated Victim Information and Notification |
| SAVS | System for Assessing Vulnerability of Species |
| SAW | Stock Assessment Workshop |
| SB | Space Biology Program \| Surgery Branch |
| SBA | Small Business Administration |

Link Results

Links: 464 Queue: 0

Caching: FALSE Method: GET

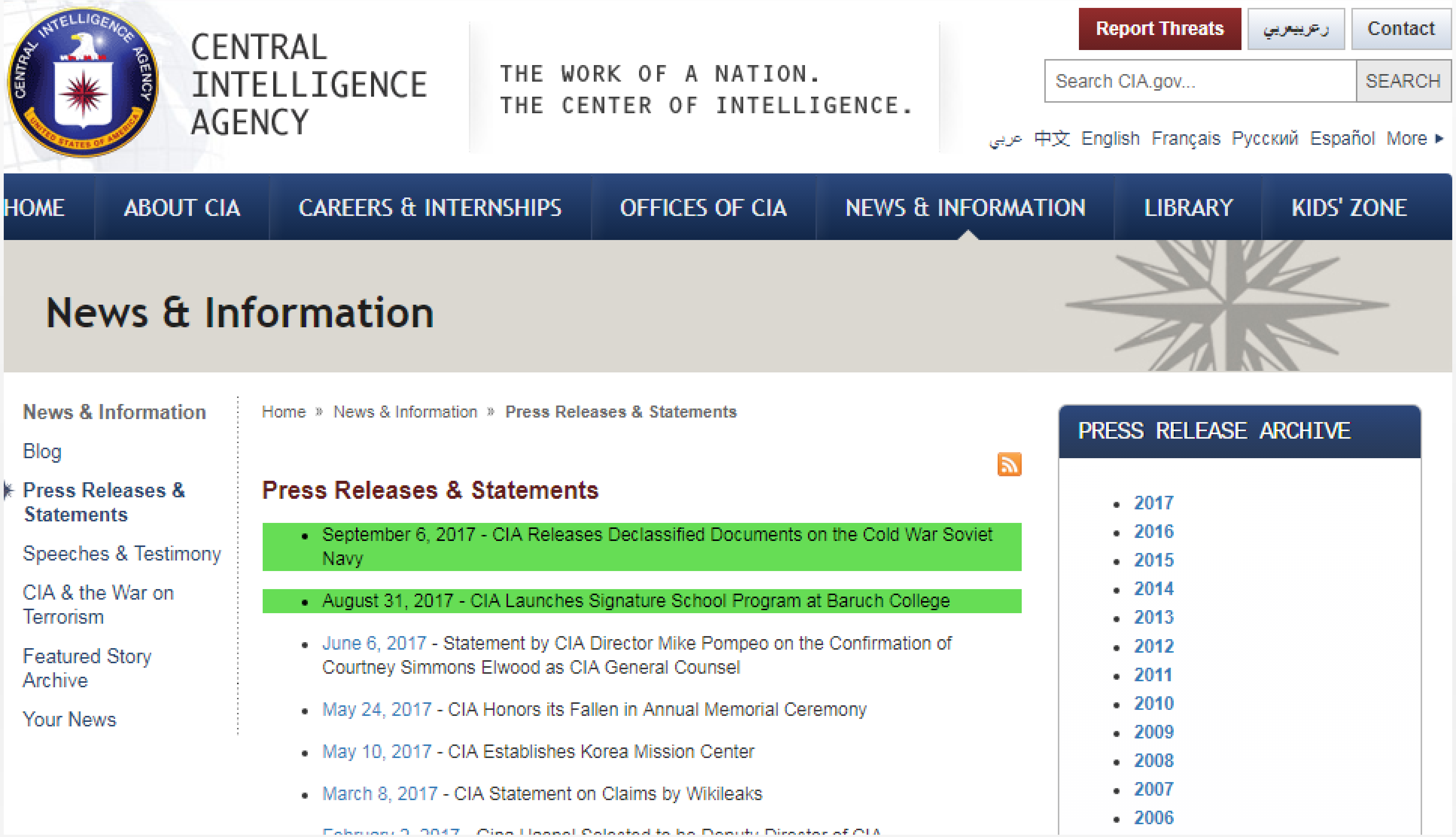
Valid links: 454

Valid redirecting links: 9

Warnings: 0

Invalid links: 1

# Page monitor extension



## Press Releases & Statements

- September 6, 2017 - CIA Releases Declassified Documents on the Cold War Soviet Navy
- August 31, 2017 - CIA Launches Signature School Program at Baruch College
- June 6, 2017 - Statement by CIA Director Mike Pompeo on the Confirmation of Courtney Simmons Elwood as CIA General Counsel
- May 24, 2017 - CIA Honors its Fallen in Annual Memorial Ceremony
- May 10, 2017 - CIA Establishes Korea Mission Center
- March 8, 2017 - CIA Statement on Claims by Wikileaks
- February 2, 2017 - Gina Haspel Selected to be Deputy Director of CIA

PRESS RELEASE ARCHIVE

- 2017
- 2016
- 2015
- 2014
- 2013
- 2012
- 2011
- 2010
- 2009
- 2008
- 2007
- 2006

# URLs for sites mentioned:

- EOT portal > http://eotarchive.cdlib.org/
- UNT nomination tool > http://digital2.library.unt.edu/nomination/GWDA/
- WebHarvest.gov > https://www.webharvest.gov/
- FDLP web archive > https://archive-it.org/home/FDLPwebarchive
- Library of Congress collections > http://www.loc.gov/webarchiving/
- Government Attic > http://governmentattic.org/
- Memory Hole > http://thememoryhole2.org/
- Data Refuge > https://www.datarefuge.org/
- EDGI > https://envirodatagov.org/
- UCSD LibGuide > http://ucsd.libguides.com/usgov
- Wayback Machine > http://archive.org/web/
- Twitter bot > https://parkerhiggins.net/2017/07/linkarchiver-a-new-bot-to-back-up-tweeted-links/
- Chrome extensions
  - Internet Archive > https://chrome.google.com/webstore/detail/wayback-machine/fpnmgdkabkmnadcjpehmlllkndpkmiak?hl=en-US
  - Check my links >  https://chrome.google.com/webstore/detail/check-my-links/ojkcdipcgfaekbeaelaapakgnjflfglf?hl=en-GB
  - Page Monitor > https://chrome.google.com/webstore/detail/page-monitor/ogeebjpdeabhncjpfhgdibjajcajepgg?hl=en