

GPO Captures Moments in Time Through Web Archiving

“The web evolved faster than the Government and guidance could keep up with,” says GPO Archives Specialist Dory Bower.

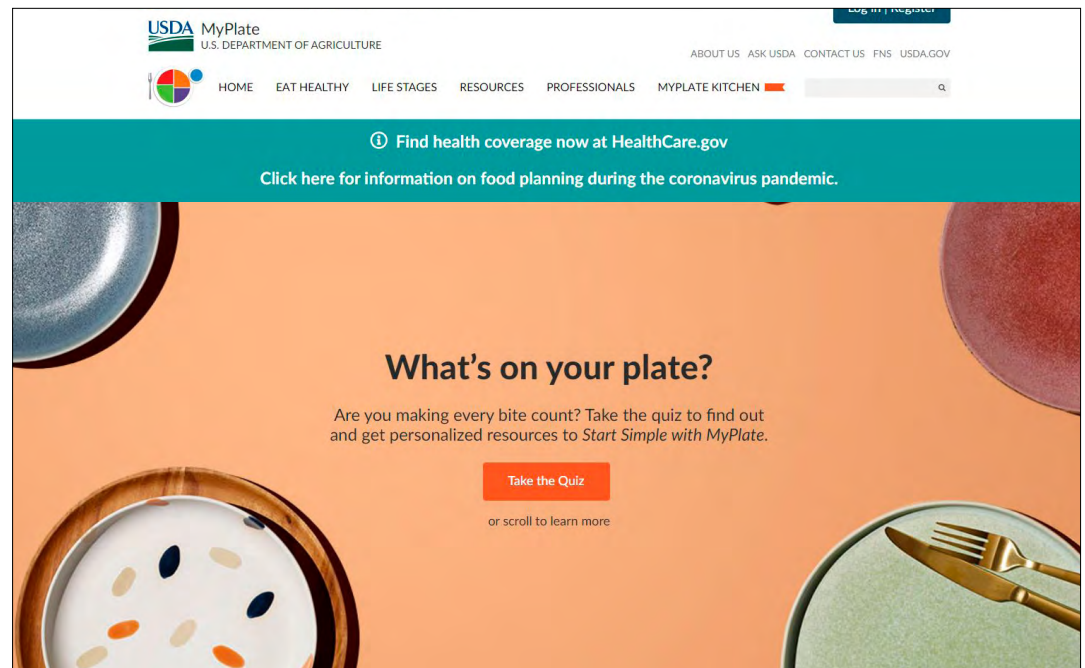
She explains that it was due to this dynamic nature of the web that the End of Term (EOT) project began in 2008.

The EOT project was founded to execute a comprehensive harvest of the Federal Government domains in the final months of the Bush administration, and to document changes in Federal Government websites as agencies transitioned to the Obama administration. GPO works with the Library of Congress, the Internet Archive, Stanford University Libraries and the University of Texas.

“It is common for web content and sites to be removed at the time of transition from one administration to the next,” said Bower. “We’re trying to capture a moment in time before some content is taken off the web forever.”



GPO works with the Library of Congress on the EOT project to determine what will be crawled.



GPO helps archive Federal agency websites, such as USDA's MyPlate. Above is an archived version of the site, previously called ChooseMyPlate. Below is the site today.



GPO Archives Specialist Dory Bower assists with the End of Term (EOT) project which helps ensure web content is not lost during transitions to new administrations.



Federal Depository Library Program Web Archive

Archive-It Partner Since: Nov, 2011

Organization Type: [National Institutions](#)

Organization URL: <http://www.fdlp.gov>

Description: Harvested by: Library Services and Content Management, U.S. Government Publishing Office. The Federal Depository Library Program, FDLP, was established by Congress in 1895 to provide free, permanent public access to Federal government information. The US Government Publishing Office administers the FDLP on behalf of the participating libraries and the American people. More information about the FDLP can be found at www.fdlp.gov. To provide permanent public access to Federal Agency Web content, the Federal Depository Library Program harvests selected U.S. Government Web sites in their entirety. Access to these sites is made available through links in our online public access catalog, the Catalog of U. S. Government Publications <http://catalog.gpo.gov>. For more questions about our web archiving program contact us at FDLPwebarchiving@gpo.gov. Note: Works of the U.S. Government are not subject to copyright protection pursuant to 17 U.S.C. § 105. Content in this web archive may contain copyrighted material used with permission of the copyright owner.

Members of the public can visit Archive-It to view Federal agency web content that has been archived.

Bower says the project speaks to GPO's mission.

"We never erase or alter anything," says Bower. "It's part of our mission *Keeping America Informed* to maintain public access to this Government information."

The goals of the project have pretty much stayed the same throughout the years but the scope has expanded. The size of the web has greatly increased from what it was in 2008. There has been increased effort to ensure there is capture of non .gov domains; government websites that may have .org, and ensuring full capture of .mil sites. The aim is to capture all content that is publicly available on a website, which includes all social media.

Members of GPO's Library Services and Content Management team, such as Bower, don't do the archiving themselves, but rather help get the word out about the project. She has been helping with the project since 2012.

"GPO primarily does the outreach through our libraries

in the Federal Depository Library Program," says Bower of GPO's role. "We provide information about how the public can get involved. The public can nominate sites to make sure those websites become part of the crawl and get archived."

GPO also provides seed lists for the EOT project. Seed lists include thousands of starting URLs of websites that should be captured and archived. These lists help to identify Government websites for crawlers.

In addition to the EOT project, GPO has been crawling websites for about nine years and currently has 211 collections of Government websites archived on the FDLP Web Archive. Some examples of those are www.myplate.gov, a United States Department of Agriculture (USDA) website about nutrition, the Centers for Disease Control and Prevention website, and the National Highway Traffic Safety Administration website.

"We are archiving

Government websites all the time," says Bower. "Agency websites have become an important way for agencies to communicate their message to the public. We figured web archiving was a good way to fulfill our mission *Keeping America Informed*."

Bower says one of the primary challenges of archiving is that the size of the Government web has greatly increased over the years.

"This past year we've seen an even greater increase in data and video on agency websites," says Bower. "We think the coronavirus pandemic is one reason we've seen that increase. There have been a lot more online video meetings and messages."

Bower explains that crawl sizes are getting larger and larger with increased amounts of video data, and with a wider variety of hosting platforms. All agencies are using most major social media services and many agencies have dozens, even hundreds, of social media accounts. Bower explains that capturing and playing back social media is a

constant challenge. Once one challenge is resolved with one social media service, a new one develops.

GPO typically crawls Federal agency websites on a bi-annual basis and does quality assurance, troubleshooting, then more crawling, to ensure the best capture and playback possible.

"Sometimes the crawlers have to change their scope and manipulate the crawl to make sure we capture as much of the site as possible," said Bower.

GPO participates in a Federal Web Archiving working group that communicates regularly. Also part of the group are the National Archives and Records Administration (NARA), the Library of Congress (LOC), the National Library of Medicine (NLM) and the Smithsonian.

Archived sites are also available through GPO's Catalog of U.S. Government Publications (CGP).