

## **End of Term Web Archive: Harvesting and Preserving the Federal Government Web Domain – Chat Log**

from James Jacobs to everyone: 2:02 PM  
Hi Kelly I'm online and ready to go

from Chris Fox to everyone: 2:13 PM  
I'm having trouble getting access to the slides/handouts for each session. When I click on the slides link by a session I'm interested in, it just brings me to the conference home page. Help?

from Kelly Seifert to everyone: 2:13 PM  
We have seen a few issues with Handouts from the Portal - they are also accessible on FDLP.gov at <https://www.fdlp.gov/file-repository/1044/handouts-and-slides>

from James Jacobs to everyone: 2:13 PM  
FYI there was an article published today in Forbes on the End of Term crawl.  
<https://www.forbes.com/sites/lesliekatz/2024/10/23/inside-the-race-to-capture-government-websites-before-they-vanish-forever/>

from Chris Fox to everyone: 2:14 PM  
Also, is there going to be a list of attendees with their emails? I've attended conferences where they do that.

from Chris Fox to everyone: 2:14 PM  
Thanks @Kelly! I appreciate it.

from Kelly Seifert to everyone: 2:16 PM  
@Chris - in the Portal, you can view the Attendee Directory and you should be able to see email addresses in there.

from Chris Fox to everyone: 2:17 PM  
Portal?

from Chris Fox to everyone: 2:18 PM  
(Sorry. I'm a dolt!)

from Ashley Dahlen to everyone: 2:18 PM  
The DLC event site that you logged into. There is a menu option on the left side for Attendee Directory.

from Scott Matheson to everyone: 2:19 PM  
<https://secure.iCohere.com/FDLP-Fall2024>

from Kelly Seifert to everyone: 2:19 PM  
Sorry Chris, that conference access point where you log in is known as the Conference Portal

from Chris Fox to everyone: 2:19 PM  
Duh! Thanks @Ashley and @Kelly.

from Becki White to everyone: 2:20 PM

@Kelly - wow, missed that in the portal, and was wondering how to contact people. Thank you!!!

from Ron Nakao to everyone: 2:22 PM

<https://www.forbes.com/sites/lesliekatz/2024/10/23/inside-the-race-to-capture-government-websites-before-they-vanish-forever/>

from Daniel Cornwall to everyone: 2:25 PM

And so many sites use third parties for their storage. For example

<https://www.fs.usda.gov/projects/tongass/landmanagement/projects>, a federal sites where the actual projects are stored on box.com

from Scott Matheson to everyone: 2:38 PM

<https://digital2.library.unt.edu/nomination/eth2024/>

from David Walls to everyone: 2:41 PM

Other than the IA's Wayback machine, is there a complete mirror site of the content on any repository?

from Katherine Marshall to everyone: 2:43 PM

You mentioned that multiple seeds per site is recommended because larger sites may get 'lost' as they skip through external links on subpages, etc. Can you explain what types of subpages/sub-seeds you'd recommend?

from ben amata to everyone: 2:43 PM

should or is NARA capturing databases on the web where one needs the content & the retrieval software? i believe the Canadian national archives does.

from Katherine Marshall to everyone: 2:44 PM

Thanks!

from Beth Downing to everyone: 2:46 PM

What about the changes in Congress at the midterms? Is there a similar crawl to capture the sites of members who are leaving or are new?

from Marija Gudauskas to everyone: 2:46 PM

As you increase the number of seed urls for top level domains, how do you effectively manage the crawl reports to ensure that you're capturing everything you want to capture without falling into traps?

from Malea Walker to everyone: 2:49 PM

LC has both a Congressional Web Archive and Elections Web Archive:

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

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

from ben amata to everyone: 2:49 PM

perhaps we want to revive an agency volunteer program to monitor. many hands light work.

from Kelly Smith to everyone: 2:49 PM

Some websites, such as the State Dept. OIG, seem to block individual captures in the Wayback Machine. Are there work-arounds for the EOT crawls that allow you to capture that content?

from Daniel Cornwall to everyone: 2:51 PM

US Forest Service is a good example where regional offices have a lot of unreported publications.

from Megan Minta to everyone: 2:51 PM

Information about reporting publications to GPO: <https://fdlp.gov/collection-tools/reporting-pubs-to-gpo>

from ben amata to everyone: 2:52 PM

i'm finding more webpages that are programs where the sites have multiple publications (pdfs & html) organized in groups.

from ben amata to everyone: 2:55 PM

can't unmute

from Ashley Dahlen to everyone: 2:55 PM

you can now!

from Kelly Seifert to everyone: 2:55 PM

you can now, Ben

from Daniel Cornwall to everyone: 2:56 PM

Ben is amazing for current awareness and advocating for metadata.

from James Jacobs to everyone: 2:56 PM

Many thanks to Kelly Smith for submitting a bulk list of seeds!

from Danielle Gensch to everyone: 2:58 PM

Perhaps Greenhouse Gas Reporting Program? <https://www.epa.gov/ghgreporting/find-and-use-ghgrp-data>

from Becki White to everyone: 2:58 PM

Just curious if the New Electronic Titles list is used as a seed suggestion?

from Danielle Gensch to everyone: 2:58 PM

This was from Kelly's roundup and Ben's discussion, but I will nominate it, too!

from Bruce Sarjeant to everyone: 2:59 PM

Ben, the Forest Service has that kind of thing going on.

<https://usfs.maps.arcgis.com/home/item.html?id=06e6dc77363b4ddab2b8531fa56b1208>

from Ben Petersen to everyone: 2:59 PM

<https://www.archives.gov/about/plans-reports/strategic-plan/strategic-plan-2022-2026#toc-strategic-goal-3-maximize-nara-s-value-to-the-nation>

from Ben Petersen to everyone: 2:59 PM

See 3.1 for NARA strategic goal to issue guidance on websites

from Dory Bower to everyone: 3:01 PM

GPO provides a seed list of all seeds in FDLP Web Archive and what is on the PURL server

from David Walls to everyone: 3:01 PM

Not strictly end of term, but there was a discussion of a collaborative approach to harvesting links found in the courts documents on Gov Info.

from Mark Ames to everyone: 3:02 PM

Thanks!

from Andrea Stelljes to everyone: 3:02 PM

Thank you!

from Charlie she-her Amiot to everyone: 3:02 PM

Thank you!

from Christopher Brown to everyone: 3:02 PM

Thanks James!

from Rebecca Hyde to everyone: 3:02 PM

Thanks so much, James! That was very informative!

from Carol Spector to everyone: 3:02 PM

Thanks, James!!

from Jacqui Celecia to everyone: 3:02 PM

Thank you

from Carolyn Frenger to everyone: 3:02 PM

Thank you, James for such an informative session!

from Joshua Finnell to everyone: 3:02 PM

Thanks, James!

from Angela Bonnell to everyone: 3:03 PM

Thank you!

from Roberta Boyd to everyone: 3:03 PM

Thank you!

from Katherine Marshall to everyone: 3:03 PM

Thanks!

from Carolyn Frenger to everyone: 3:03 PM

Thanks too to everyone who shared urls with us! : )

from James Jacobs to everyone: 3:03 PM  
thanks everyone for your attention today!

from Fang Gao to everyone: 3:03 PM  
Thanks, James!

from Cynthia Etkin (privately): 3:16 PM  
Thanks!