DIY WEB ARCHIVING

PEOPLE'S MEDIA CAMP

OCTOBER 5, 2025

RACHEL APPEL

CYNTHIA HEIDER

INTROS

- Rachel Appel, she/her, Digital Preservation Librarian

- Cynthia Heider, she/her, Public Digital Scholarship Librarian

What's your favorite thing to come in a can? 🥫

SESSION ROADMAP

- Introduction to web archiving: What is it, and why would you want to do it?

- How does web archiving work?

- Overview of tools

- Live demo of Archive Web.page by Webrecorder and hands-on session

- Live demo of Conifer and hands-on session

- Live demo of "Save Page Now" by Internet Archive Wayback Machine if we have time

- Next steps: Services, resources, projects, and upcoming events
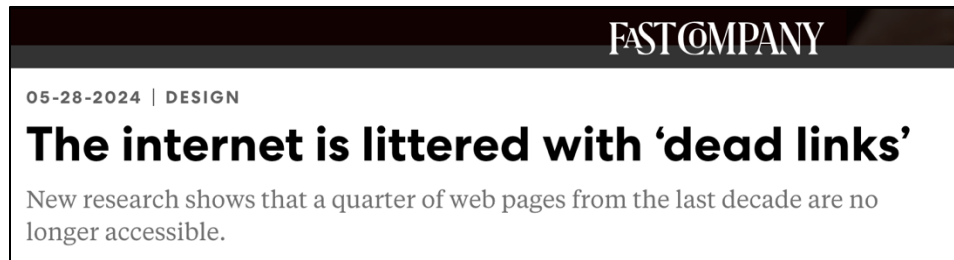
- Q&A

WHY WEB ARCHIVING?
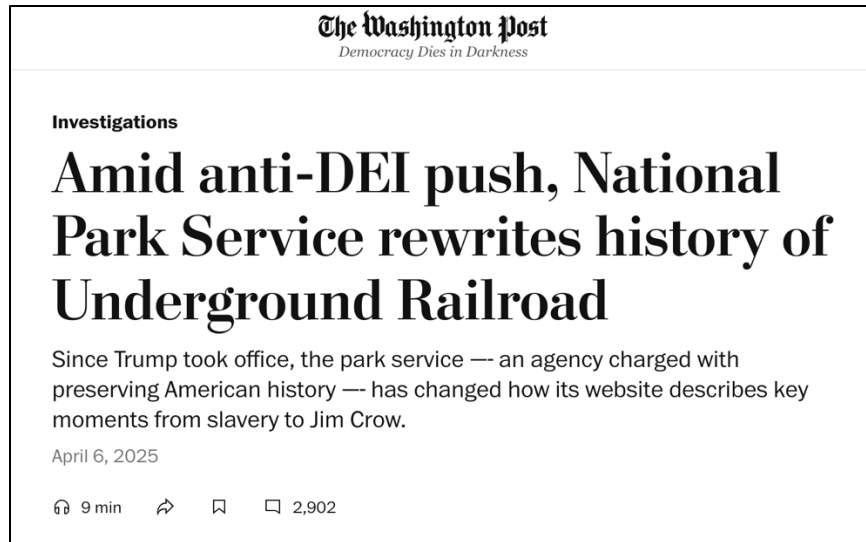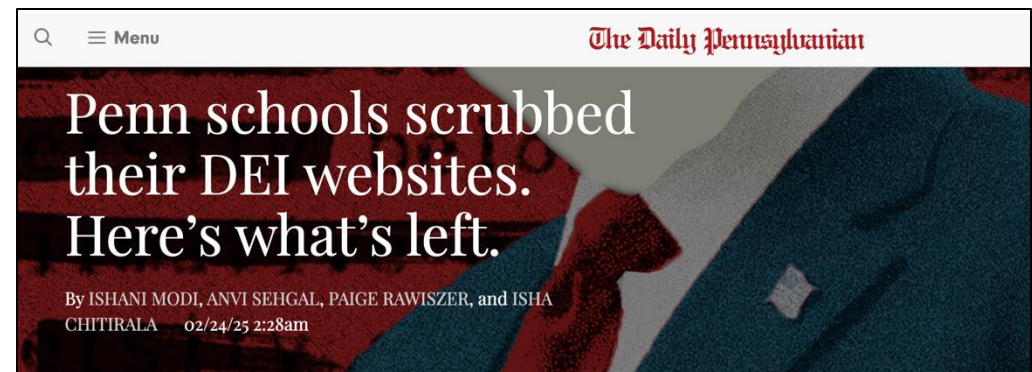
Here today...
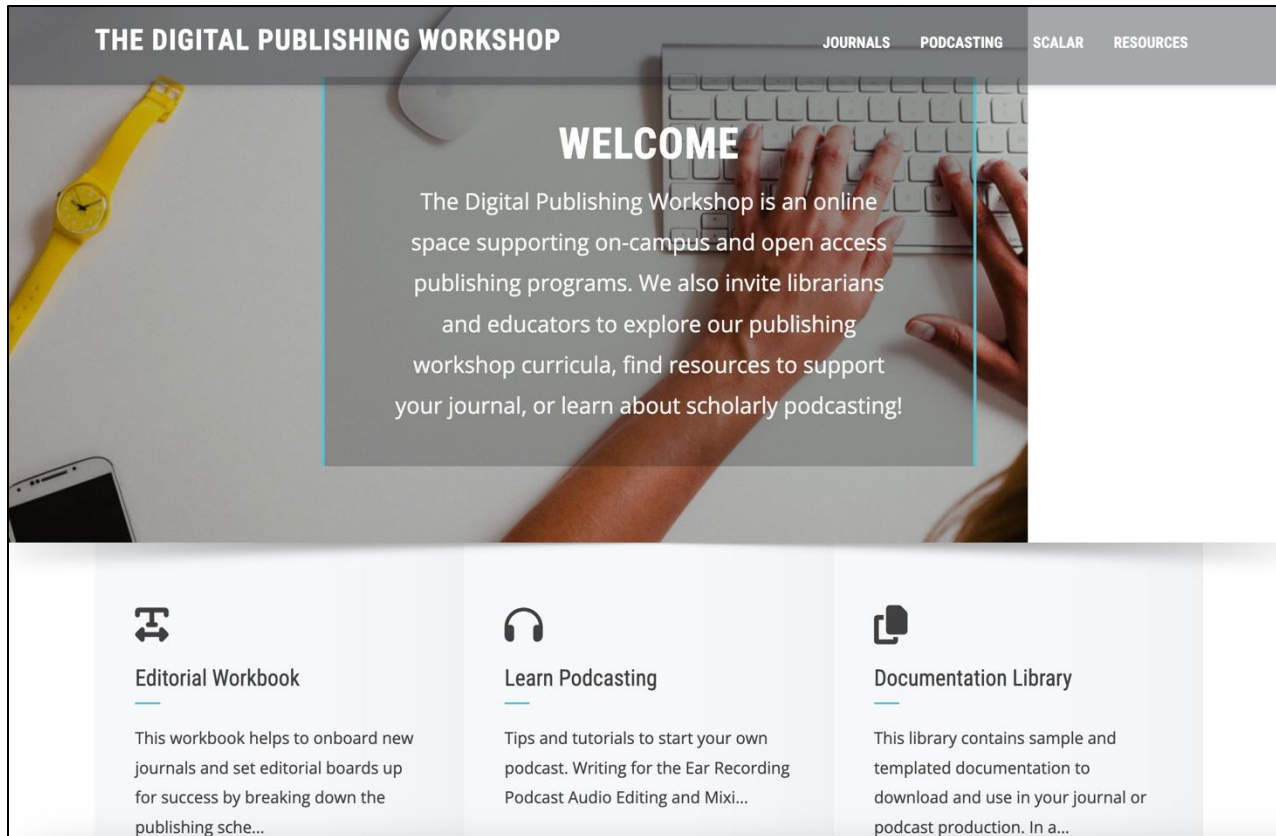
... Gone tomorrow
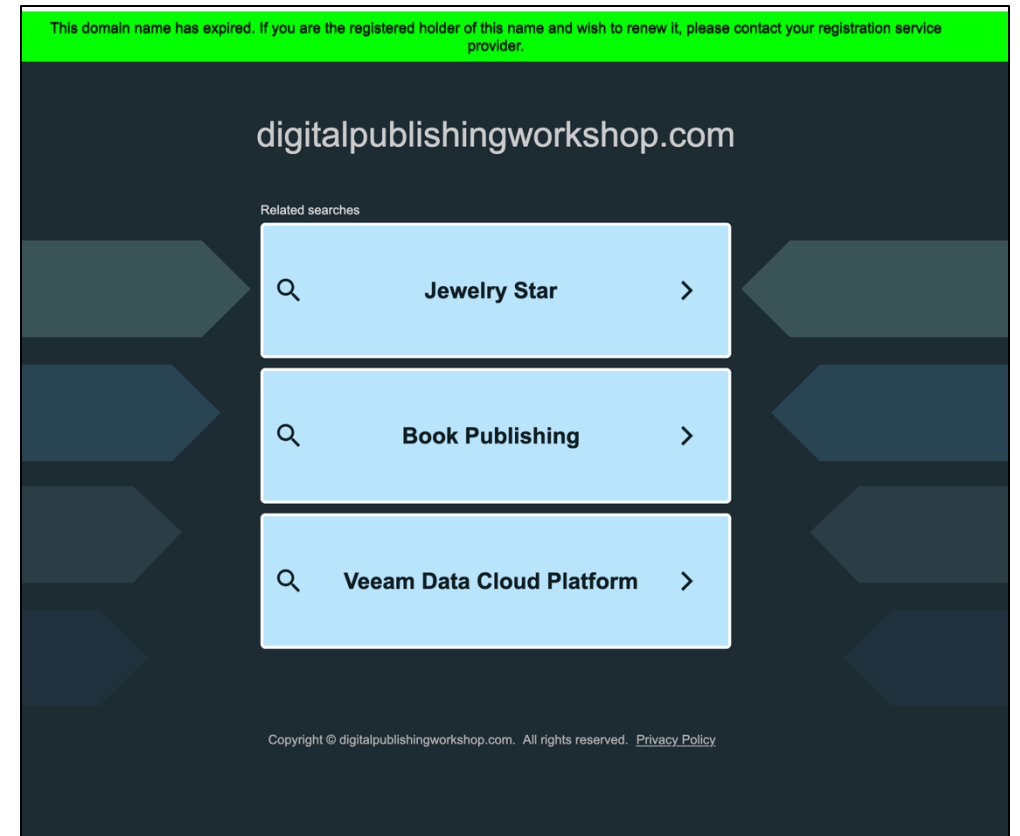
WHY WEB ARCHIVING?

"The web is not a storage service – it's a publication platform." - Sara Day Thomson

05-28-2024 | DESIGN

The internet is littered with 'dead links'

New research shows that a quarter of web pages from the last decade are no longer accessible.

Fast Company

Penn schools scrubbed their DEI websites. Here's what's left.

By ISHANI MODI, ANVI SEHGAL, PAIGE RAWISZER, and ISHA CHITIRALA   02/24/25 2:28am

The Daily Pennsylvanian

Investigations

Amid anti-DEI push, National Park Service rewrites history of Underground Railroad

Since Trump took office, the park service —- an agency charged with preserving American history —- has changed how its website describes key moments from slavery to Jim Crow.

April 6, 2025

9 min   2,902

Washington Post

home • artnews • news

British Library Slowly Recovers Following Cyberattack

BY FRANCESCA ATON   October 14, 2024 3:48pm

ARTnews

WHAT IS WEB ARCHIVING TECHNICALLY?

- Collecting & preserving online content

- Web crawlers capture sites

- Stored in WARC format

- Replay tools (e.g., Wayback Machine)

- Challenges: dynamic, multimedia, storage

LEGAL AND COPYRIGHT CONSIDERATIONS

- Copyright applies to online content

- Fair use/fair dealing

- Website terms may restrict crawling

- Hosted platforms (Squarespace, WordPress, Wix) add complexity

- Platform terms = contract law risk

- Institutions set policies, permissions, takedown options

SO MANY TOOLS, SO LITTLE TIME

Paid

- [Archive-It](#) by the Internet Archive

- [CivicPlus](#)

- [Webrecorder](#)

- [Perma.cc](#)

Free

- * [Archive Webpage](#) by Webrecorder

- * [Conifer](#) by Rhizome

- [WARCreate](#)

- [Web Curator Tool](#)

- [WAIL](#)

- [Heritrix](#)

*we'll be testing these today!
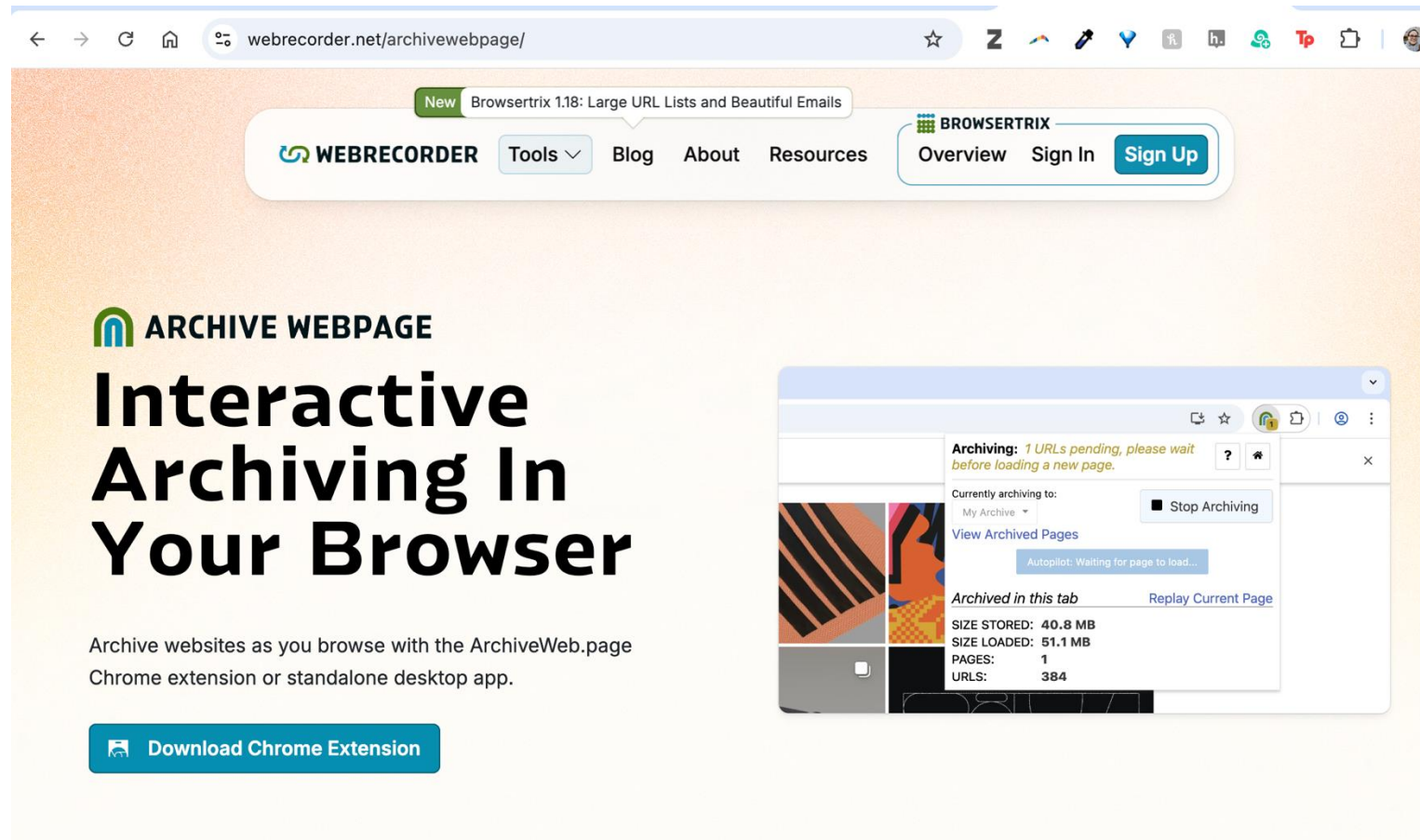
ARCHIVE WEB.PAGE BY WEBRECORDER

- Free, open-source tool

- Use Chrome browser extension or download the desktop app

- Stores in WARC and WACZ file formats

- All data is saved on your computer

- Peer-to-peer sharing feature in development

- Not for large-scale or automated crawling

https://webrecorder.net/archivewebpage/

ARCHIVE WEBPAGE

HTTPS://WEBRECORDER.NET/ARCHIVEWEBPAGE/

CONIFER

- Free, open-source tool

- User-driven captures

- Saves interactive, dynamic sites

- Stores in WARC format

- Great for small, detailed projects

- Not for large-scale crawling

https://conifer.rhizome.org/

Conifer
by RHIZOME.org

HTTPS://GUIDE.CONIFER.RHIZOME.ORG

Quick Start: Capture a Webpage

SAVE PAGE NOW BY THE WAYBACK MACHINE

- Free, tool provided by the Internet Archive's Wayback Machine

- Archives a live web page and create a permanent link to its archived version

https://web.archive.org/save

INTERNET ARCHIVE

WayBack Machine

HANDS-ON ACTIVITY

SERVICES AND RESOURCES

- Community Webs from the Internet Archive

  - Community Webs Web Archiving 101 training with videos and reference materials

- Wayback Machine version of NC State University Libraries Social Media Archives Toolkit

- Personal digital archiving of social media info (University of Missouri)

- Council of State Archivists video on Social Media & Web Archiving

- Zine Bakery DIY web archiving zine

- Webrecorder teaching materials (formats, presentations, glossary)

- Tools and software list from International Internet Preservation Consortium

COOL PROJECTS AND INITIATIVES

- Human Rights Web Archive

- End of Term Web Archive

- Saving Ukrainian Cultural Heritage Online (SUCHO)

- Invisible Histories Project

- Archiving the Black Web

UPCOMING EVENTS OF INTEREST

- October 8, 2025 - Archiving the Black Web webinar series: "Digital Storage, Backups, and Repositories"

- October 9, 2025 – International Internet Preservation Consortium: Social Media Archiving Q&A Webinar

- October 16, 2025 - #AskAnArchivist Day on social media

- October 22, 2025: Internet Archive 30 year celebration: "The Web We've Built"

Q & A

THANK YOU!

Rachel Appel (rappel@upenn.edu)
Cynthia Heider (heider@upenn.edu)