



BRINGING DATA BACK FROM THE DEAD

**DATA PRESERVATION WORKSHOP | SCARY DATA WEEK 2025
JESSICA BREEN, PHD | OLIVIA IVEY, MLIS, MSSW**

WELCOME & WORKSHOP GOALS



Learn why online information disappears



Practice using the **Internet Archive's Wayback Machine**



Understand archiving as both a **research skill** and **civic act**



Leave with practical ways to preserve your digital sources

THE LOST DATA PROBLEM

Reports, news
interactives, and
datasets disappear
or change every day

URLs in academic
citations often break
within a few years

Even reputable
organizations update
pages without saving
prior versions



**WHAT ONLINE RESOURCE
WOULD YOU MISS IF IT
VANISHED?**

WHY DATA AND INFORMATION VANISH



Institutional change or project retirement



Website redesigns or domain changes



File formats that become unreadable



Policy shifts or broken links



Assumption that “someone else is preserving it”



***PRESERVATION
IS NOT AUTOMATIC.***

***IT'S SOMETHING
WE CHOOSE TO DO.***

THE INTERNET ARCHIVE



A **nonprofit digital library** that preserves websites, books, media, and data portals



Houses the **Wayback Machine**, a searchable archive of web pages since 1996



Anyone can add pages, no login or special software required



All saved pages are public and citable

WAYBACK MACHINE AS A RESEARCH TOOL

Archiving helps you:

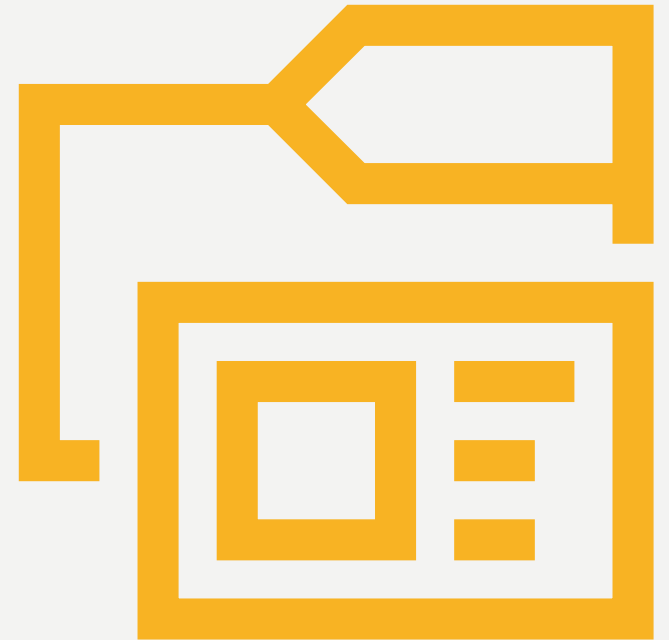
- Capture the version of a source you actually used
- Cite it confidently even if it changes later
- Compare older and newer versions of a page
- Document the evolution of ideas or narratives



WAYBACK MACHINE AS A CIVIC TOOL

Archiving also helps:

- Preserve public information and accountability
- Keep government and research data accessible to all
- Empower individuals—not just libraries or agencies—to steward digital knowledge



FINDING WEBSITES IN THE INTERNET ARCHIVE

1. Go to web.archive.org
2. Enter the **URL** of a webpage you want to look up into the Wayback Machine
 - a) Example: `https://brookings.edu/` or `https://epa.gov/climate-change`
3. Click **Browse History**
4. Explore the **timeline** to see how often it's been saved
5. Click a **blue date dot** to view that version of the page

ARCHIVING A WEBSITE

1. Go to web.archive.org
2. Find “**Save Page Now.**”
3. Paste the URL of any live page → click **Save Page**
4. Once the archiving is done you can use the provided link to see what was saved

💡 Try saving one of these:

- A class reading or resource from a faculty website
- A digital news story or visualization you’ve used in a project

HANDS-ON: DATA RESCUE TIME

[20 MIN]

01

Choose a website relevant to your studies or interests

02

Use the Wayback Machine to:

- View earlier snapshots
- Save the current version
- Note what content is missing or incomplete

03

Share something surprising you found or saved

WHAT DOES THE INTERNET ARCHIVE ACTUALLY SAVE?

- ✓ Page text and HTML structure
- ✓ Most static images (JPEG, PNG, GIF)
- ✓ Many PDFs and downloadable files linked directly on the page
- ✓ Basic CSS for layout and some simple scripts
- ✗ Password-protected or database-driven content
- ✗ Interactive elements (maps, embedded videos, forms)
- ✗ Large datasets or files stored on external servers
- ✗ Content blocked by site owners with `robots.txt` or paywalls

BEYOND THE INTERNET ARCHIVE

- [DataLumos](#) for depositing civic and social science data
- [Internet Archive Scholar](#) preserved research outputs
- [AU Government Information Data Rescue LibGuide](#) helps locate missing U.S. government data



DISCUSSION: ETHICS & IMPACT



Who gets to decide what's worth preserving?



Are there materials that shouldn't be archived?



How do individual actions connect to larger preservation efforts?



What roles do libraries, students, and researchers share in this work?

The web is fragile.
Archiving makes it
durable


The Wayback
Machine supports
research
integrity and
civic
participation

Anyone can help
preserve digital
knowledge

Open repositories
like **DataLumos**
ensure datasets
stay available long-
term

KEY TAKEAWAYS

RESOURCES

- web.archive.org – Wayback Machine
 - datalumos.icpsr.umich.edu – Open data repository
 - scholar.archive.org – Archived scholarly outputs
 - [AU Government Information Data Rescue LibGuide](#) - Finding guide for missing governmental data
- 



**THANK
YOU!**