Tools for Online Research
When Content Goes Missing

ORALL, October 19, 2017
Stephanie Ziegler & Ingrid Mattson

• Quick Primer
• The Internet Archive’s Wayback Machine
• Additional Tools for Finding “Missing” Content
• Specialized Digital Preservation Projects & Collections
A Couple of Internet Basics

• **URLs – Uniform Resource Locators**
  • Domain name – “website”
  • top-level domain - .com, .edu, .gov
  • File name – individual web pages on the website; usually follows after a forward slash

http://moritzlaw.osu.edu/faculty/scholarship-digest/
http://moritzlaw.osu.edu/faculty/professor/ingrid-mattson/

http://sentencing.typepad.com/sentencing_law_and_policy/

The Wayback Machine

• M.I.T. graduate Brewster Kahle started the Internet Archive in 1996 in his attic, aiming to build “the Library of Alexandria Two.”

• The Wayback Machine was created by the Internet Archive in 2001.

• Named after Mr. Peabody’s WABAC Machine, it archives cached pages of websites on a biweekly to monthly basis.
The Wayback Machine

• You can search the Wayback Machine by URL or website name.
  • Also, keyword searching is in beta as of October 2016!

• Emails, chats, and internet forums are not archived.
  • Nor are sites where the owners have requested they not be.

• Uses for the Wayback Machine:
  • Finding pages that have been “lost” or changed.
  • Researching the history of a website.
  • Making a permanent record of a page you want to use.

Additional Tools for Finding “Missing” Content
Cached Pages

- Many search engines take a snapshot of each page they crawl and cache these as a back-up.

- If you click on the cache link, you will see the web page as it looked the previous time it was indexed.

- There will be no "Cached" link for
  - sites that have not been indexed or
  - sites whose owners have requested their content not be cached.

- Yahoo and Bing offer a similar service.

404. That's an error.

The requested URL /cse?q=00203415331144001350%3Atsq%3&ie=utf-8&sa=X&ved=0ahUKEwjB8NS37N2FAhV2woMKHd50B8oQ_BwklAE
Search was not found on this server. That's all we know.
Perma.cc Links

- Westlaw, Lexis, HeinOnline, and Google Scholar all include articles with perma.cc citations in footnotes.

- Perma.cc launched in 2013, so the usefulness of this type of search is still growing.

- Search for all or part the url a website within the same sentence as “perma.cc.”
Perma.cc Links

• Example: Dallas DA Craig Watkins to Push for Law Allowing Appeals Based on Racial Factors


• Result: Perma.cc Screenshot

Facebook Posts
Politwoops

- **Politwoops** tracks deleted tweets by public officials, including people currently in office and candidates for office.

- Search by word or politician
Specialized Digital Preservation Projects & Collections

Specialized Digital Preservation Projects & Collections
General

• Archive-It

• WebCite

• Library of Congress Digital Collections
Specialized Digital Preservation Projects & Collections

- Court URLs
  - Supreme Court of Canada
  - U.S. Supreme Court Web Citations
- Campaign Pages & Ads
  - Political TV Ad Archive
- Page Vault
- PageFreezer

Thank You.