



The Pedagogical and Scholarly Prospects of Web Archiving

California Academic & Research Libraries 2018 Conference
April 13-15, 2018, Redwood City, California

Dana Ospina, Digital Initiatives Librarian, California State University, Dominguez Hills
Mark Bilby, Scholarly Communications Librarian, California State University, Fullerton
Zach Vowell, Digital Archivist, California Polytechnic State University, San Luis Obispo
Jillian Lohndorf, Web Archivist, Internet Archive

Introduction



- Who, what, how of web archiving
- Uses and pedagogical considerations
- Collaboration
- Challenges and Considerations

Web Archiving: What is it?

Preserving web-based data

- Capture: Web crawlers
 - Both open source, and proprietary tools exist
 - Heritrix, HTTrack, Wget
 - Internet Archive: Heritrix, Umbra, Brozzler
- Storage: Web Archive File (WARC)
 - Best practice is redundant storage
- Playback: Mechanism is necessary
 - Wayback
- Search
 - NutchWax, Elastic Search
- Metadata
 - Solr



Web Archiving: Curation



- Create a thematic/topical web archive on a specific subject or event
 - Capture web-based primary sources
 - Document current events
 - Document different perspectives, social commentaries, online communities
 - Supplement traditional collecting activity
- Fulfill a mandate to capture/preserve evolving web history
 - Support an electronic records system to meet records retention requirements
 - Collect publications/documents that are no longer in print form
 - Construct a historical record of an institution or individual's web/social media presence
 - Fulfill FOIA requests

Pedagogy: Library Reference and Instruction

- Reference Desk: just in time use of Wayback Plugin
- Research Consultations: Archive-It searching
- Library Instruction Sessions
 - Wayback Machine permalink tutorial
 - Historicizing digital primary sources
- LibGuides Content Backup



Scaling Up Web-Archiving Pedagogy

- Instructional Designers: best practices for course readers and course shells
- Faculty Development Workshops
 - Link-rot-proofing course materials
 - Webliography assignments
 - Course research focus: top ten seeds
- Writing Center Collaboration
- University Writing Requirements



Pedagogy: In the Classroom



- Persistent resources
 - Web-based readings in a curriculum
 - Loss of context (and often “look and feel”) if you just save a web page as a PDF
 - Is it just the text that you need? Or the information as it lives as a web document? What do you want the student to take away from the web resource?
 - Journal article VS. General audience article (typically not as a PDF) VS. “Digital” journal VS. News sites VS. Anything with dynamic/interactive/audio-visual content
- A Tool to facilitate critical thinking
 - Is it really “out there forever?”
 - But it’s not a “scholarly article!”
 - Two examples
 - Tufts University environmental justice project
 - Middlebury College Web Archives

Scholarly Collaboration within Universities

- University Websites to Prioritize
 - Web-based student work and faculty scholarship
 - Student newspapers/media
 - University Policies
 - Course catalogs (web and PDFs)
 - Special projects (initiatives, renovations)
 - Student life

The screenshot shows a web browser window with the URL http://wayback.archive-it.org/7738/20160923201821/http://harlotofhearts.org/issues/issue_15/helmbrechtLove/index.html. A yellow banner at the top states: "You are viewing an archived web page, collected at the request of California Polytechnic State University, San Luis Obispo using Archive-It. This page was captured on 20:18:21 Sep 23, 2016, and is part of the Cal Poly Scholarship collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#) [Enable QA](#)".

The website header includes "TheAgora Prime Politics" and a search bar. Navigation links include Home, Binders, Shoes, Thoughts, Sources, and YourReviews.

The main content area features the title "From Product Reviews to Political Commentary: Performances in Amazon.com Reviews" by Meredith A. Love | Brenda M. Helmbrecht, with a web developer credit to Erin Kathleen Bahl. To the right is a collage of images including the US Capitol, Barack Obama, and various political symbols.

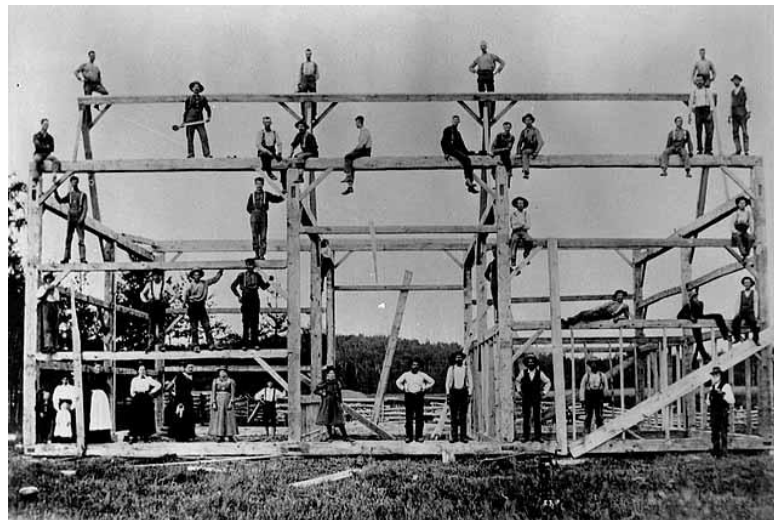
The "Featured Reviews" section displays two reviews:

- Durable Binder:** "Reviewed: Amy's Durable View Binder with Silver Wire Ring, 10.5 x 11.5". 4199 of 5917 people found this review helpful. Rating: 4.5 stars. "One Missing Bit of Information You Might Want to Know". October 18, 2012. Reviewer: -PenName220 (Click review to expand).
- Mizuno Women's Wave Rider 16:** "Reviewed: Mizuno Women's Wave Rider 16 Running Shoe (Blue/Red/White)". 3 of 5 people found this review helpful. Rating: 5 stars. "Perfect!". July 29, 2013. Reviewer: -kimberly piccolo (Click review to expand).

On the right side of the featured reviews, there is a book cover for "CHIMAMANDA NGOZI ADICHIE WE SHOULD ALL BE FEMINISTS".

Scholarly Collaboration across Institutions

- Cobweb
- Professional Organizations
 - International Internet Preservation Consortium
 - OCUL/COPPUL
- Research projects
 - WASAPI
 - Archives Unleashed
- National web archives
- Local (university, regional, etc)
 - KAIC
 - NYARC
 - Stanford
 - Ivy Plus



Minnesota Historical Society

Scaling Up Web-Archiving Scholarly Collaboration

- Instructional and Graphic Design Talent Needed!
 - Web archiving model LibGuide and content blocks
 - Instructional video/tutorial on creating and embedding permalinks with Wayback Machine
- Repository and Discovery Systems Integration
 - IR and Discovery integrations of locally curated Archive-It
 - IR and Discovery integrations of ALL Archive-It
- CARL/ACRL Resolutions and/or Recommendations to Consider
 - Creating and citing permalinks as a core InfoLit competency
 - Permanent, public URL requirements for (Future) Manuals of Style
 - WB permalink as standard web reference in Wikipedia
 - Automated permalink checking in Citation Management Software



Challenges and Considerations



- Allocation of Time and Division of Labor
 - Who does the work and how do we prepare them for it?
 - Faculty participation
 - Scalability
- Messaging
 - How should we present these collections to the public?
- Discoverability
 - Cataloging? Finding Aids?
 - OCLC Recommendations for metadata (<https://bit.ly/2GMVFhO>)
- Privacy Concerns
 - Student content



Thank you!