Collaborative Collection Development for Web Archives

#cobwebarchive

Stephen Abrams  
Kathryn Stine  
Jim Vanderveen  

Peter Broadwell  
Andrew Wallace  

Janet Taylor  
Ann Whiteside  

California Digital Library  
UCLA  

Harvard University
Imagine …

… a fast-moving event unfolding online as much as on the ground

How can we respond as a community appropriately and responsibly?
Imagine …

Harnessing the domain knowledge and technical capabilities of the entire community

Enabling local collection development decisions based on global information

Complementary, cooperative, and collaborative collecting

Institutional participation at a level commensurate with local expertise and capacity

Increasing scholarly and public awareness
Centralized **catalog** of collection- and seed-level metadata

Establishment of thematic collecting **projects**

Open **nomination** of topical seed URLs by interested stakeholders

**Claiming** of seeds by archival institutions intending to harvest

**Holdings** records for seeds actually harvested

Thematic **discovery** of web archives of interest
The demands of archiving the web in comprehensive breadth or thematic depth *exceeds* the technical and financial capacity of any one institution. Curators cannot make rational collection development *decisions* without knowledge of what others have collected or intent to collect. Relevant seed URLs can be meaningfully *contributed* by various stakeholders: curators, archivists, subject area specialists, scholars, journalists, event participants, and the public. Apportioning collection responsibility into granular pieces encourages *participation* by smaller institutions and programs.
CA.gov archive of the California state government domain

- Collaborative effort of government information librarians at UC and Stanford and the CA State Library and CA State Archive
- More than 700 seeds (URLs) captured total
- Collecting 2007-present, major content refresh 2016-2017
- Coordinating workflows, approaches to QA and metadata
Peter Broadwell at UCLA was well into collecting “fake” news sites before it occurred to him to wonder if anyone else was doing something similar.

There was; Mark Graham at IA.
Collaborative Web Archiving Projects

- UC Libraries, Stanford, CA State Library, CA State Archives
  - CA.gov
- Ivy Plus Libraries Web Resources Collection Program
  - CAUSEWAY, Contemporary Composers
- IIPC Content Development Group
  - Olympics, Global News
- Internet Archive, LoC, Stanford, CDL, et al
  - End of Term
- Canadian Web Archiving Coalition (CWAC)
- Preservation of Electronic Government Information (PEGI)
One-year collaborative project between CDL, Harvard University, and UCLA, funded by IMLS #LG-70-16-0093-16

Public online service hosted at CDL

Targeting initial production release in conjunction with the November 2018 IIPC General Assembly and Web Archiving Conference

https://cdlib.org/cobweb
https://github.com/CobwebOrg/cobweb
**The Team**

- Project and Outreach Management: CDL
- Development and Technical Oversight: UCLA
- UI/UX Development and Testing: Harvard

**Our Process**

Agile approach (two-week sprints), iterative and collaborative approaches:
- collecting user data
- translating user data -> workflows -> functionality/interfaces
- confirming functional reqs
- sprint prep/retros
Conferences/Professional Associations:
- SAA (Web Archiving Section)
- AoIR
- Dodging the Memory Hole
- UC-DLFx
- CNI
- IIPC

Platforms/Communities:
- Internet Archive’s Archive-It
- Rhizome’s webrecorder.io
## cobWeb Timeline

<table>
<thead>
<tr>
<th>Activity</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop user personas and user stories, user data gathering</td>
<td>August – October 2017</td>
</tr>
<tr>
<td>Establish functional requirements, refine metadata specifications</td>
<td>August 2017 – January 2018</td>
</tr>
<tr>
<td>Develop core functionality, import sample holdings metadata</td>
<td>September – November 2017</td>
</tr>
<tr>
<td>Draft workflows and wireframes</td>
<td>October – November 2017</td>
</tr>
<tr>
<td>First prototype rollout</td>
<td>November 2017</td>
</tr>
<tr>
<td>User testing, data gathering</td>
<td>February - March 2018</td>
</tr>
<tr>
<td>Iterative development on prototype refinements, continue importing holdings metadata</td>
<td>December 2017 – March 2018</td>
</tr>
<tr>
<td>Second prototype rollout</td>
<td>March/April 2018</td>
</tr>
<tr>
<td>Beta testing</td>
<td>April - September 2018</td>
</tr>
<tr>
<td>Production system rollout</td>
<td>September/October 2018</td>
</tr>
<tr>
<td>Sustainability planning</td>
<td>July – October 2018</td>
</tr>
</tbody>
</table>
cobWeb technical stack

Python/Django stack
Bootstrap framework
PostgreSQL DB
Solr search index
Agile, sprint-based
MIT license

https://github.com/CobwebOrg/cobweb
establish/update
Project

suggest Nomination
suggest Nomination
suggest Nomination
functionality

establish/update
Project

suggest Nomination
assert Claim

suggest Nomination

suggest Nomination
assert Claim
cobWeb functionality

establish/update Project

suggest Nomination

assert Claim

suggest Nomination

assert Claim

report Holding

establish/grow Collection
Prototype

Project
Project notes

Janet Taylor, 2017/10/18
Confirmed starting crawl date for description.

Janet Taylor, 2017/10/11
This project is partially funded through a grant from the John Quincy Adams foundation. We are looking to partner with another organization to complete the project.

Projects
- Open nominations project
- Milton, Massachusetts
- Mount Holyoke College
- Hague, New York
- Dodging the memory hole project
- Propaganda, disinformation, parody, dismissal: Making sense of "Fake News"
- Zlata Praha
- Zlata Praha
- Recusandae quo autem earum.
- Nulla sed odio porro dicta consectetur nisi fugit.
- Blanditiis laborum asperiores suscipit fugit vero quam.
- Accusantium animi tenetur temporibus.
- Accusantium culpa architecto aspernatur ullam.
- Paritur vitae esque.
- Accusantium quo facilis veniam voluptatum rem.
- Temporibus perspiciatis aspernatur aperiam labore totam totam voluptates.
- Nostrum fugit at cum sapiere.
- Ducimus sed vel adipisci nuncum.
- Modi odio voluptates dignissimos.
- FAKE News!!!
Prototype
Prototype

Project

Project notes

Janet Taylor, 2017/10/18
Confirmed start date for description.

Janet Taylor, 2017/10/11
This project is partially funded through a grant from the John Quincy Adams Foundation. We are looking to partner with another organization to complete the project.

Collections

- Alabama Constitutional Offices A - E
- Alabama State Agencies A - B
- Alabama State Agencies I - K
- Alabama Public Higher Education Web Sites
- Alabama Regional and Local Government Agencies
- Alabama Political Blogs
- Alabama Constitutional Offices F - H
- Alabama Political Blogs D - H
- Alabama Political Blogs I - N
- Alabama Political Blogs O - R
- Alabama Constitutional Offices C - H
- Alabama Legislature and Appellate Courts
- Alabama Regional, Multi-State, or Multi-County Agencies or Commissions, M - Z
- McLeodstralia
- Alabama State Agencies C - D
- Alabama State Agencies E - F
- Alabama State Agencies G - H
- Alabama State Agencies L - N
- Tejano and Conjunto Music Websites
- Alabama State Agencies O - P
- Alabama State Agencies Q - S
- Alabama State Agencies T - Y
- Alabama State Agency Facebook Pages
- Michigan State University Social Media Collection
- Alabama State Agency Twitter Feeds
Prototype
Next steps

Cobweb is a tool for collecting communities...
Next steps

- Keep learning about your collaborative collecting projects and workflows
- Confirming functional requirements
- Refining and validating use cases
- Engaging testers to walk through prototypes (early spring 2018)
- Iterative development, responding to user testing findings
Next, next steps

Cobweb is a tool for collecting communities and perhaps not just those collecting web resources...
Questions for us?

Questions *from* us!
Discussion

Cobweb affordances for:

- Establishing collecting (collection management?) projects
- Nominating resources for...
- Claiming intentions for nominated resources
- Reporting holdings
Collaborative Collection Development for Web Archives
#cobwebarchive

California Digital Library   https://www.cdlib.org/
Harvard University Library   http://library.harvard.edu/
UCLA Libraries   http://www.library.ucla.edu/

https://cdlib.org/cobweb
https://github.com/CobwebOrg/cobweb

Kathryn Stine, Cobweb Outreach Manager
Kathryn.Stine@ucop.edu