cobWeb

Collaborative Collection Development for Web Archives
#cobwebarchive

Stephen Abrams  Peter Broadwell  Janet Taylor
Kathryn Stine  Andrew Wallace  Ann Whiteside
Jim Vanderveen

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
Why cobWeb?

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
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/services/cobweb/
https://github.com/CobwebOrg/cobweb
Conferences/Professional Associations:
  Archive-It Partner Meeting, AoIR, Dodging the Memory Hole, ARLIS/NA, UC-DLFx, CNI,
  SAA (Web Archiving Section), DLF, IIPC

Platforms/Communities:
  Internet Archive’s Archive-It, Rhizome’s webrecorder.io

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, UNT, Stanford, CDL, et al
    - End of Term
  - Canadian Web Archiving Coalition (CWAC)
  - New York Art Resources Consortium (NYARC)
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
cobWeb technical stack

- Docker Swarm
- Python 3.6
- Django 2.0
- µwsgi App Server
- PostgreSQL DB
- Solr Search Index
- Archives
  - Current - Archive-It
  - Future - Webrecorder - &c

Agile, sprint-based
MIT license

https://github.com/CobwebOrg/cobweb
cobWeb data model
establish/update 
*Project*

- suggest *Nomination*
- suggest *Nomination*
- suggest *Nomination*
cobWeb functional user roles

- Project Organizer
- Nominator
- Collector
- Researcher
### Projects

- Open nominations project
- Milton, Massachusetts
- Mount Holyoke College
- Hauge, New York
- Dodging the memory hole project
- Propaganda, disinformation, parody, dismissal: Making sense of "Fake News"
- Zita Prana
- Zita Prana
- Recusande quo autem earum
- Nulla sed odio portio dicta consequuntur nisi fugit.
- Blanditiis laborum aspernatur ut
- Accusantium nisi
- Accusantium cumque voluptas et
- Paratur vitae qui
- Accusantium quo facilis veniam voluptatum
- Temporibus perspicilis voluptatum
- Nostrum fugit aut cum sapientia.
- Ducimus sed vel adipsi
d
- Moli odit voluptates
dignissimos.
- FAKE News!!

### Project Notes

- Janet Taylor, 2017/1/0/18
  
  Confirmed starting crawl date for description.

- Janet Taylor, 2017/1/0/11
  
  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.
cobWeb prototype

Project

Project notes

Cancel  Add

Janet Taylor, 2017/10/18
Confirmed starting crawl data 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 organisation to complete the project.

Title

Status*

Active

Administrators*

= admin

Description

Keywords

Nomination policy*

- Anonymous: anyone can nominate, even if they’re not logged in.
- Open: anyone with a CobWeb account can nominate.
- Restricted: only selected users can nominate.

Nominators

Nominator blacklist

Cancel  Submit
### Project

**Propaganda, disinformation, parody, dismissal: Making sense of "Fake News".**

#### Status:
Active

#### Administrators:
- Janet Taylor
- Peter Broaddus
- Kathryn Stone
- Stephen Abrams

#### Description:
This project captures and categorizes online sources that purport to provide news coverage of current events.

#### Keywords:
- disinformation
- fake news
- satire news
- bias in news

#### Nomination policy:
Open

#### Nominations:
- [http://100percentlieurop.com/](http://100percentlieurop.com/) (claim)
- [http://16mpo.com/](http://16mpo.com/) (claim)
- [http://266ixanews.com/](http://266ixanews.com/) (claim)
- [http://4threvo.blogspot.com/](http://4threvo.blogspot.com/) (claim)
- [http://7news.wordpress.com/](http://7news.wordpress.com/) (claim)
- [http://82.221.129.208/](http://82.221.129.208/) (claim)
- [http://aaronfer.blogspot.co.uk/](http://aaronfer.blogspot.co.uk/) (claim)
- [http://abonestowards.com/](http://abonestowards.com/) (claim)
- [http://abotwagenssecret.com/](http://abotwagenssecret.com/) (claim)
- [http://abriluno.com/](http://abriluno.com/) (claim)
- [http://aofahammar.wordpress.com/](http://aofahammar.wordpress.com/) (claim)
- [http://activistpost.com/](http://activistpost.com/) (claim)
- [http://actualjokepanamanian.com/](http://actualjokepanamanian.com/) (claim)
- [http://addictinginfo.org/](http://addictinginfo.org/) (claim)
- [http://affsoft.org/media/](http://affsoft.org/media/) (claim)
- [http://adwordtechnicalities.com/](http://adwordtechnicalities.com/) (claim)
- [http://advocate.com/](http://advocate.com/) (claim)
- [http://ahechothelord.com/](http://ahechothelord.com/) (claim)
Nomination

Project notes

Janet Taylor, 2017/10/18
Confirmed starting crawl data 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.

Title

Resource
http://

Project
Propaganda, disinformation, parody, dismiss: Making sense of "F1;"

Status
Unclaimed

Description

Keywords

Cancel   Submit
Cobweb is a tool for collecting communities...
- 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
Join us at lunch!

We’re hosting a Web Archiving Birds of a Feather

- consider collaborative web archiving collecting use cases
- learn about your web archiving projects and workflows, challenges and successes
- lay groundwork for ongoing systemwide communication about web archiving
Questions for us?
Collaborative Collection Development for Web Archives
#cobwebarchive

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

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

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