

UC Office of the President

CDL Staff Publications

Title

Cobweb: Collaborative Collection Development for Web Archives

Permalink

<https://escholarship.org/uc/item/89c466mn>

Authors

Stine, Kathryn
Abrams, Stephen
Wallace, Andrew

Publication Date

2017-11-16



Dodging the Memory Hole, Internet Archive, 15-16 November 2017

Cobweb

Collaborative Collection Development for Web Archives

#cobwebarchive

Stephen Abrams

Kathryn Stine

California Digital Library

Peter Broadwell

Andrew Wallace

UCLA

Janet Taylor

Ann Whiteside

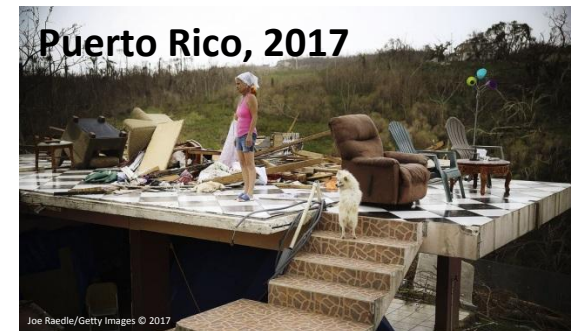
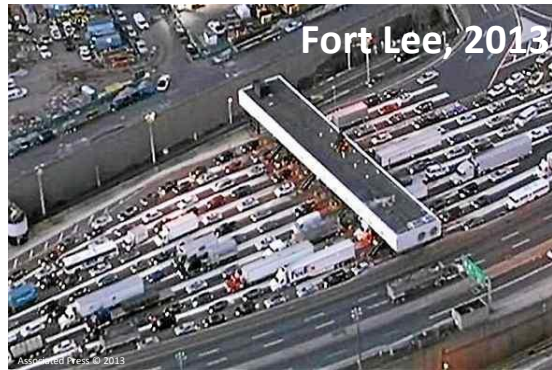
Harvard University



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

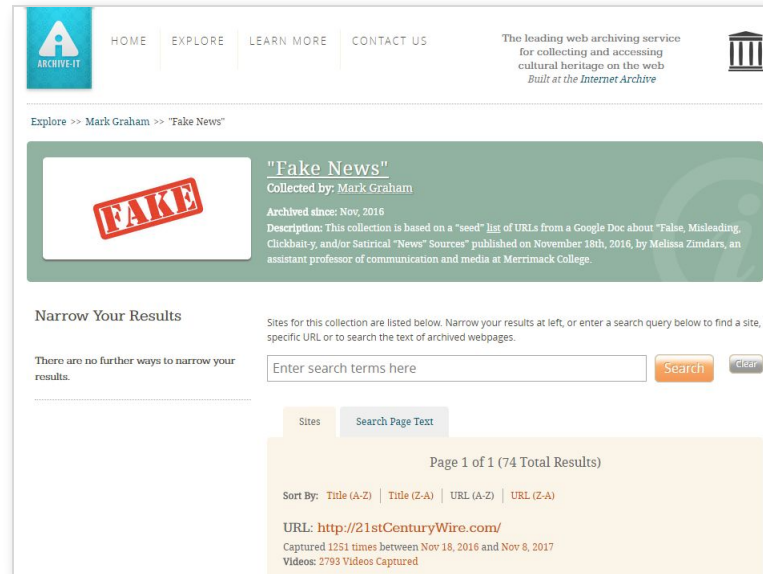
Relevant seed URLs can be meaningfully *contributed* by various stakeholders: curators, archivists, subject area specialists, scholars, journalists, event participants, and the public

Apportioning collection responsibility into granular pieces encourages *participation* by smaller institutions and programs

Why Cobweb?

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



The screenshot shows the Internet Archive website interface. At the top, there is a navigation bar with links for HOME, EXPLORE, LEARN MORE, and CONTACT US. The main heading is "Explore >> Mark Graham >> 'Fake News'". Below this, there is a green banner with a red "FAKE" stamp and the title "'Fake News'". The banner also states "Collected by: Mark Graham" and "Archived since: Nov, 2016". A description follows: "Description: This collection is based on a 'seed' list of URLs from a Google Doc about 'False, Misleading, Clickbait-y, and/or Satirical 'News' Sources' published on November 18th, 2016, by Melissa Zimdars, an assistant professor of communication and media at Merrimack College." Below the banner, there is a section titled "Narrow Your Results" with a search bar and a "Search" button. The search results show "Page 1 of 1 (74 Total Results)" and a list of results. The first result is for the URL "http://21stCenturyWire.com/", which was captured 1251 times between Nov 18, 2016 and Nov 8, 2017, and has 2793 videos captured.



Cobweb project

One-year collaborative project between CDL, Harvard University, and UCLA, funded by IMLS #LG-70-16-0093-16

Public online service hosted at CDL

Python/Django stack

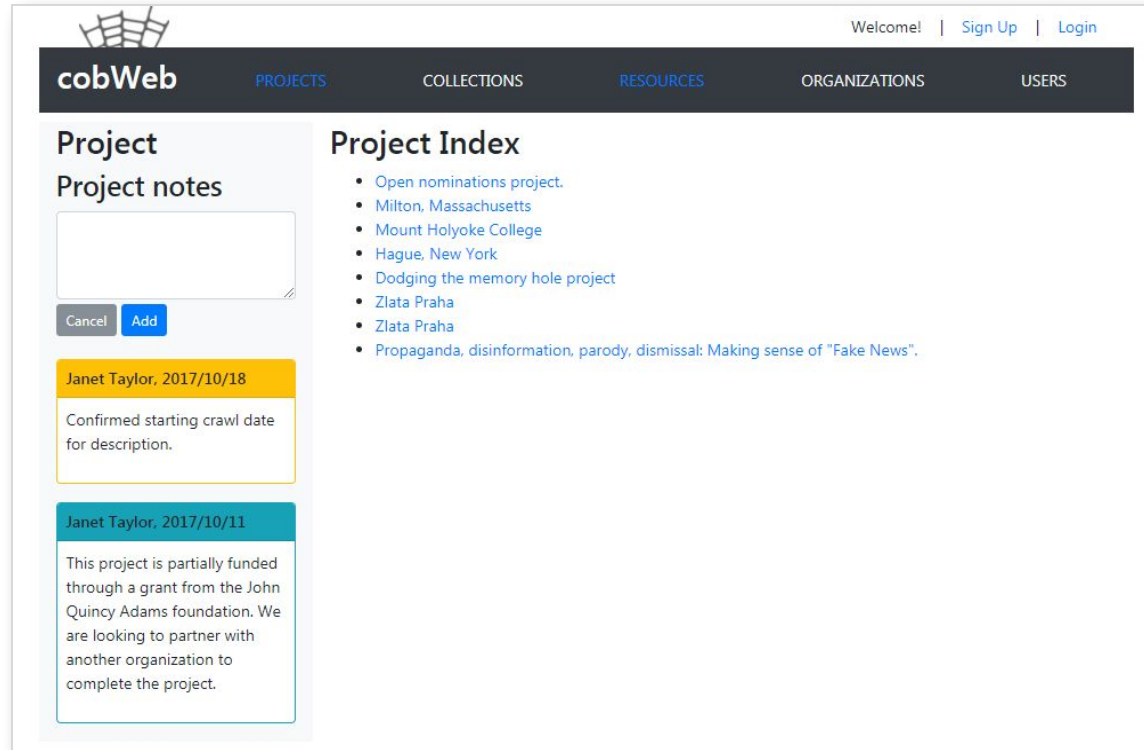
MIT license

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

[*https://cdlib.org/cobweb*](https://cdlib.org/cobweb)

[*https://github.com/CobwebOrg/cobweb*](https://github.com/CobwebOrg/cobweb)

Demo



The screenshot displays the cobWeb website interface. At the top right, there is a navigation bar with the text "Welcome!" followed by links for "Sign Up" and "Login". Below this is a dark navigation menu with the "cobWeb" logo on the left and menu items for "PROJECTS", "COLLECTIONS", "RESOURCES", "ORGANIZATIONS", and "USERS".

The main content area is divided into two columns. The left column is titled "Project" and contains a "Project notes" section. This section includes a text input field, "Cancel" and "Add" buttons, and two project entries:

- 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.

The right column is titled "Project Index" and contains a bulleted list of project links:

- [Open nominations project.](#)
- [Milton, Massachusetts](#)
- [Mount Holyoke College](#)
- [Hague, New York](#)
- [Dodging the memory hole project](#)
- [Zlata Praha](#)
- [Zlata Praha](#)
- [Propaganda, disinformation, parody, dismissal: Making sense of "Fake News".](#)

Next steps

Cobweb is a tool for collecting communities...
we need you!

Next steps

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

Q & A

Questions for us?
Questions *from* us!



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