UC Office of the President
CDL Staff Publications

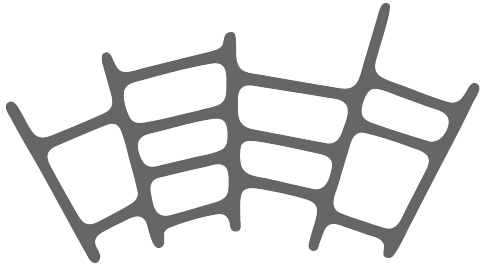
Title
Cobweb: Collaborative Collection Development for Web Archives

Permalink
https://escholarship.org/uc/item/050385r2

Authors
Stine, Kathryn
Abrams, Stephen

Publication Date
2018-02-08

eScholarship.org Powered by the California Digital Library
University of California



cobWeb

Collaborative Collection Development for Web Archives

#cobwebarchive

Stephen Abrams
Kathryn Stine
Jim Vanderveen

California Digital Library

Peter Broadwell
Andrew Wallace

UCLA

Janet Taylor
Ann Whiteside

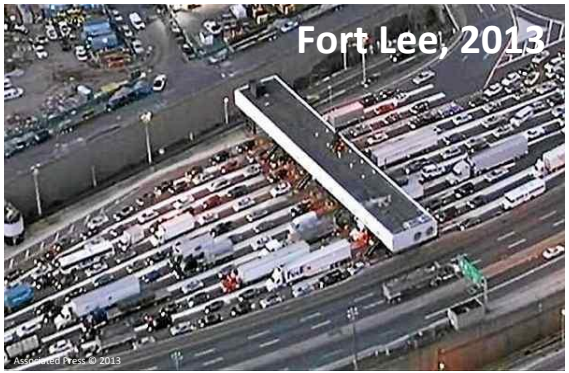
Harvard University

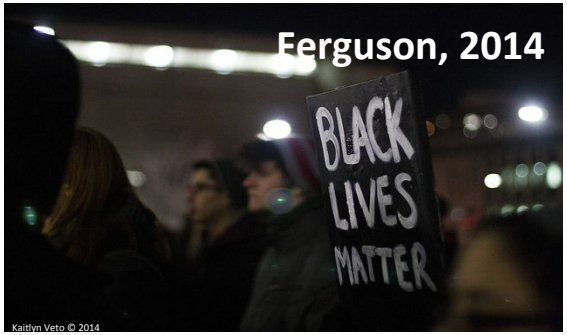
CC BY

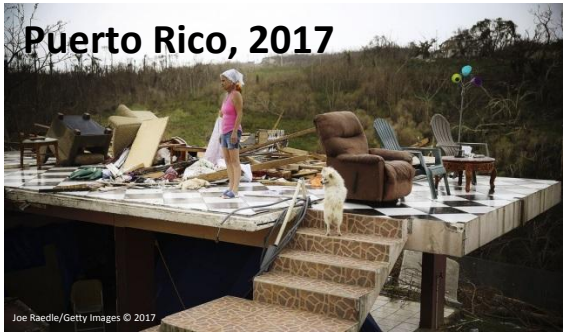
Imagine …

… 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*

cobWeb

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

University of California CDL California Digital Library

UCLA

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 ?

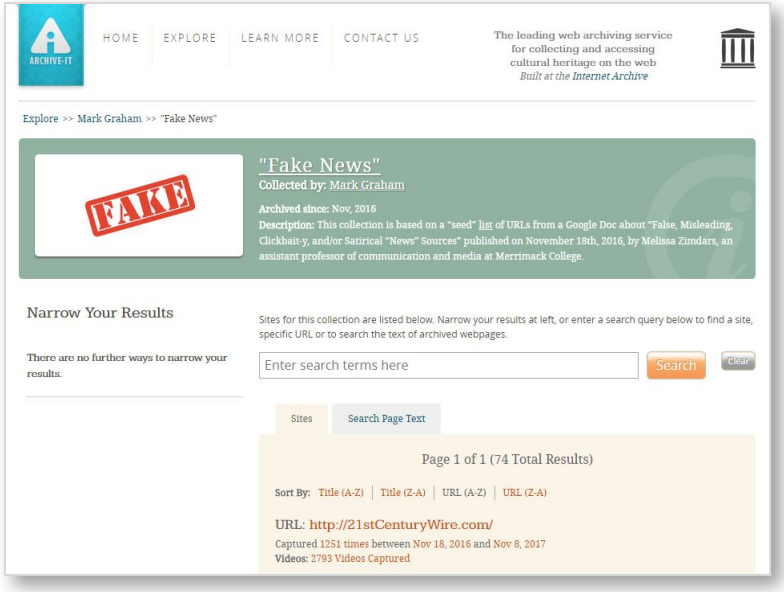
CA.gov archive of the California state government domain

- 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

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



University of California CDL California Digital Library

UCLA

Why cobWeb ?

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)

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

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

INSTITUTE of Museum and Library SERVICES

University of California CDL California Digital Library

UCLA

cobWeb roles and collaboration

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 outreach

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

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

cobWeb technical stack

Python/Django stack

Bootstrap framework

PostgreSQL DB

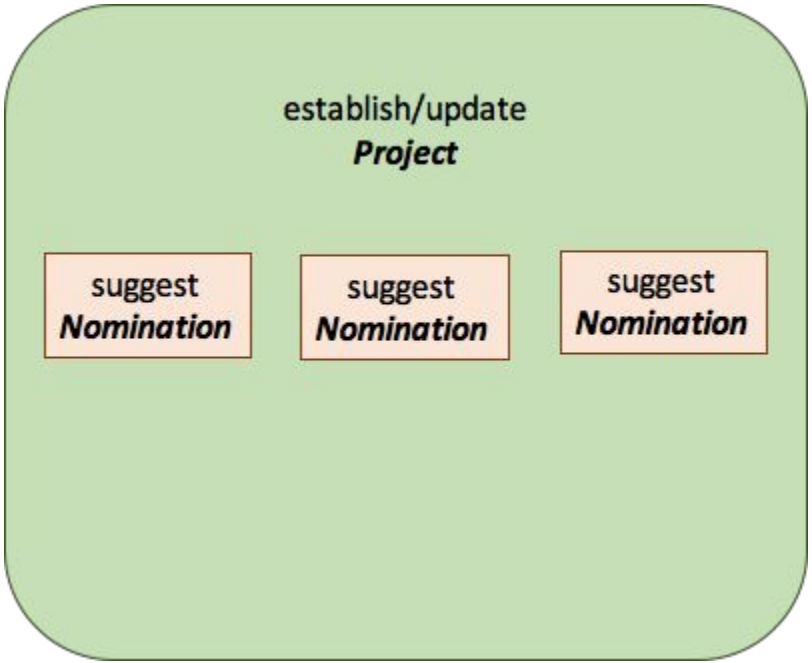
Solr search index

Agile, sprint-based

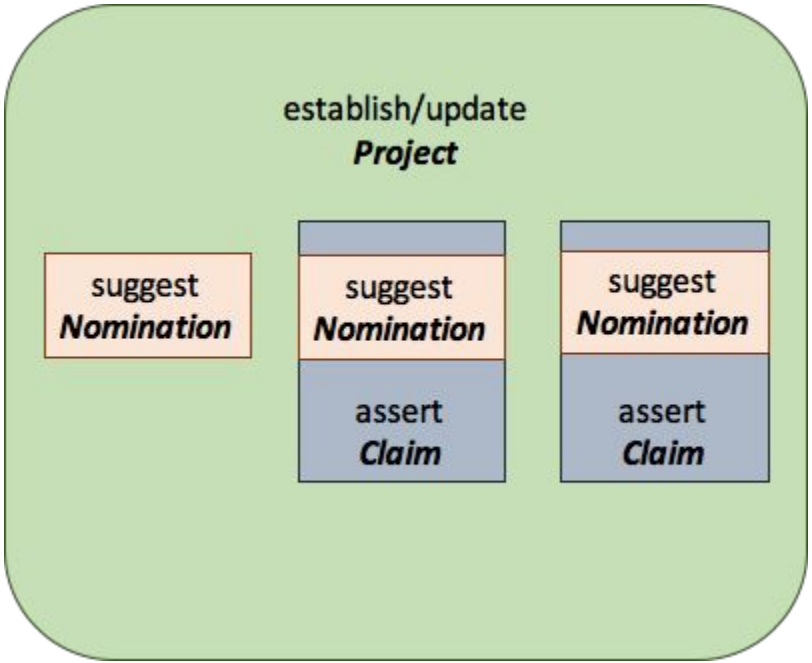
MIT license

https://github.com/CobwebOrg/cobweb

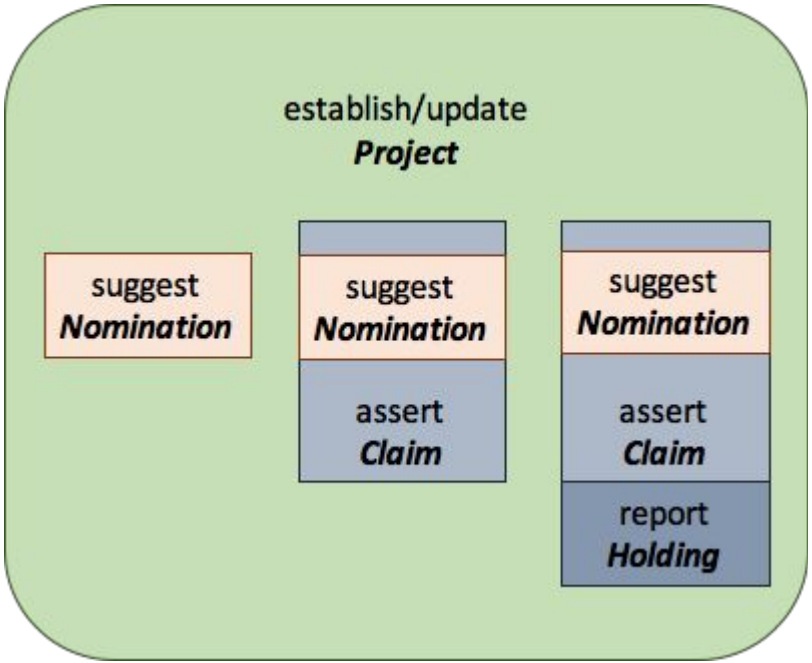
cobWeb functionality



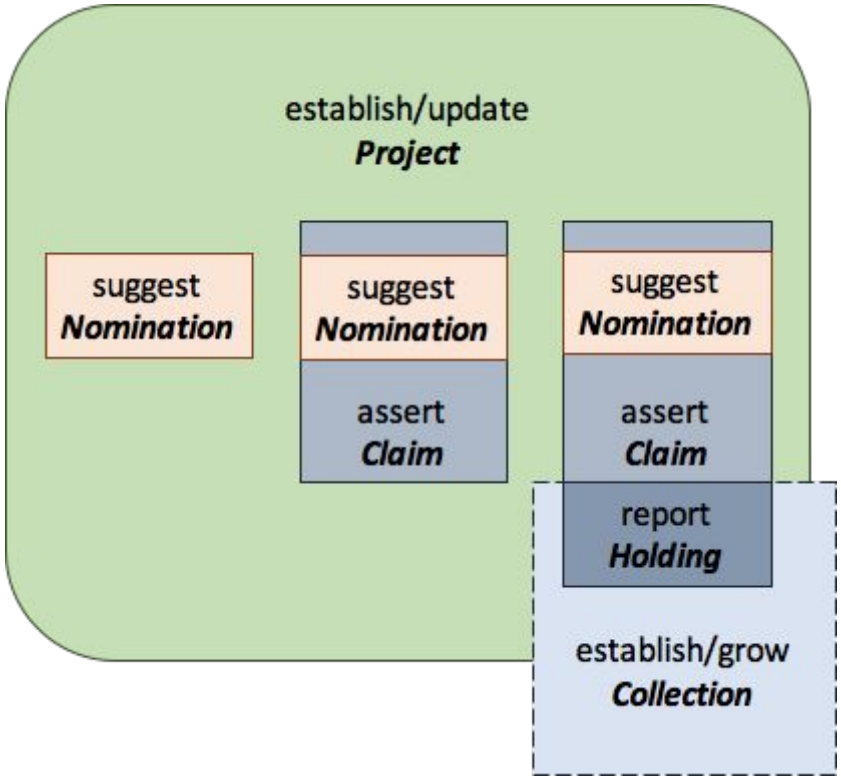
cobWeb functionality



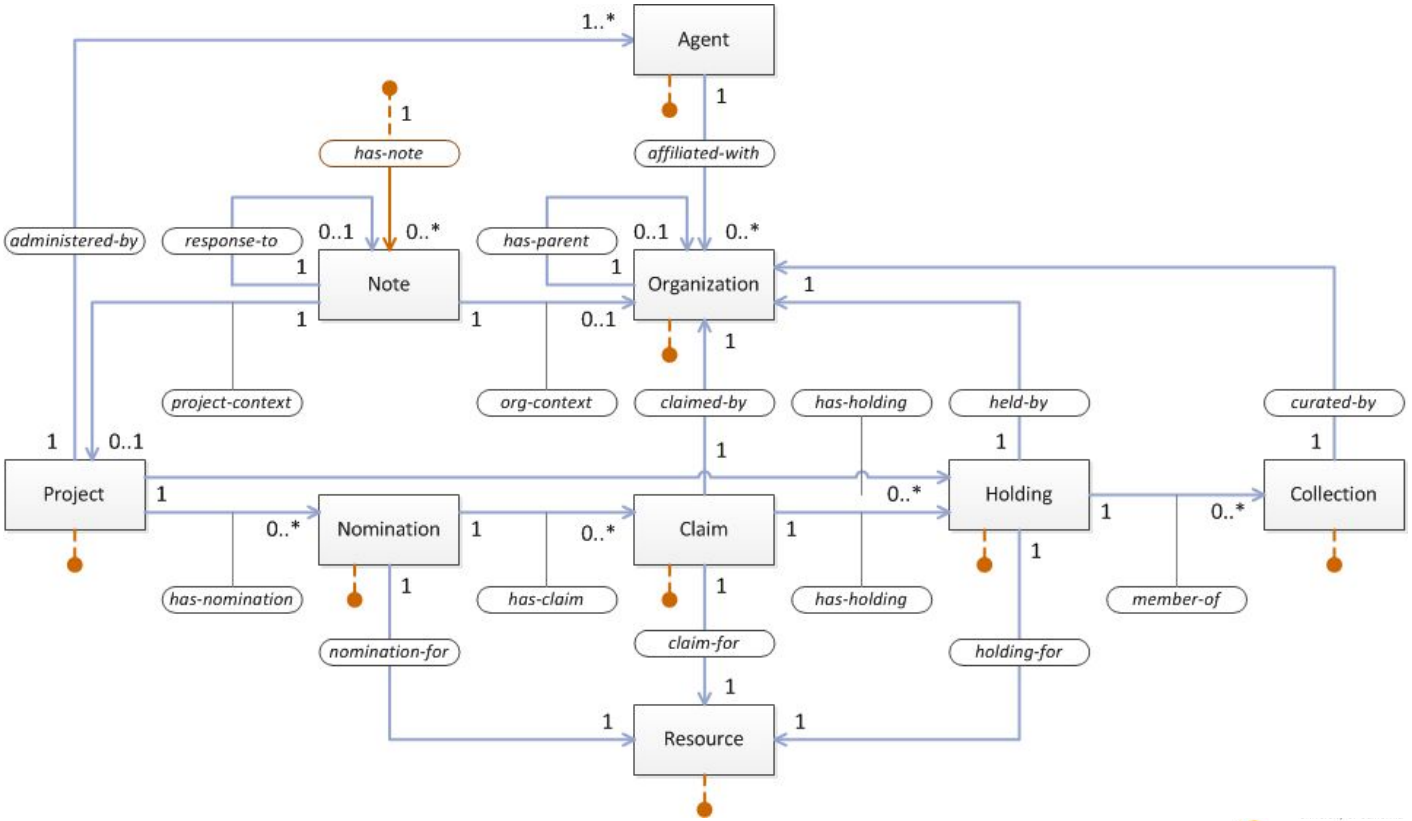
cobWeb functionality



cobWeb functionality

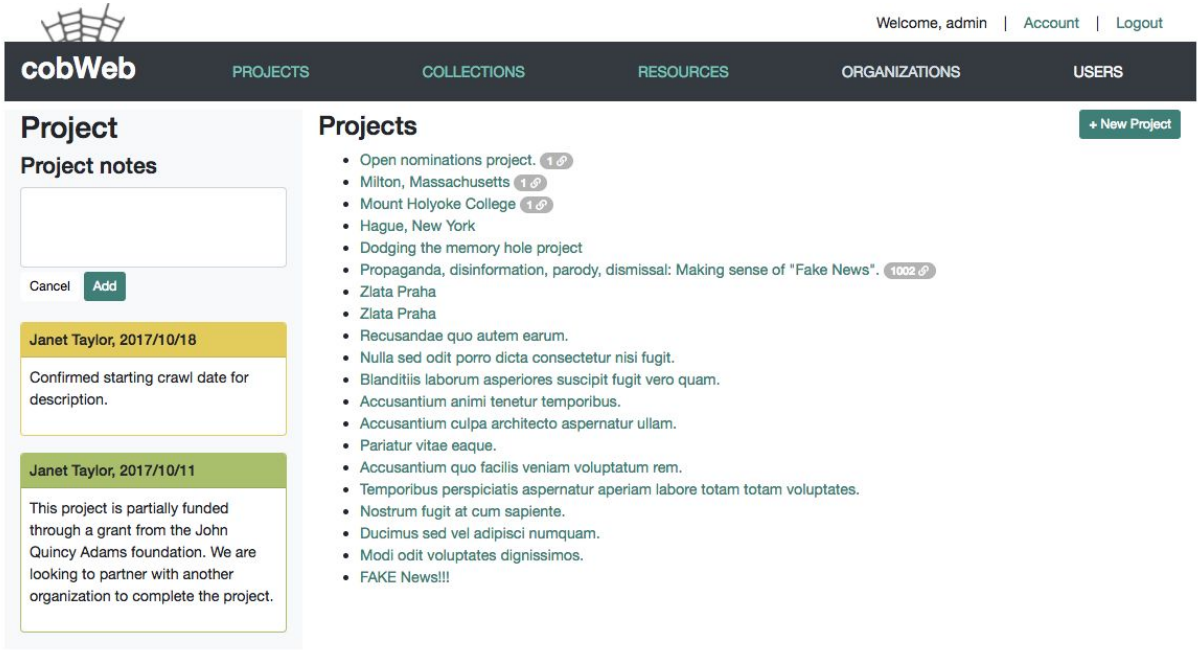


cobWeb data model

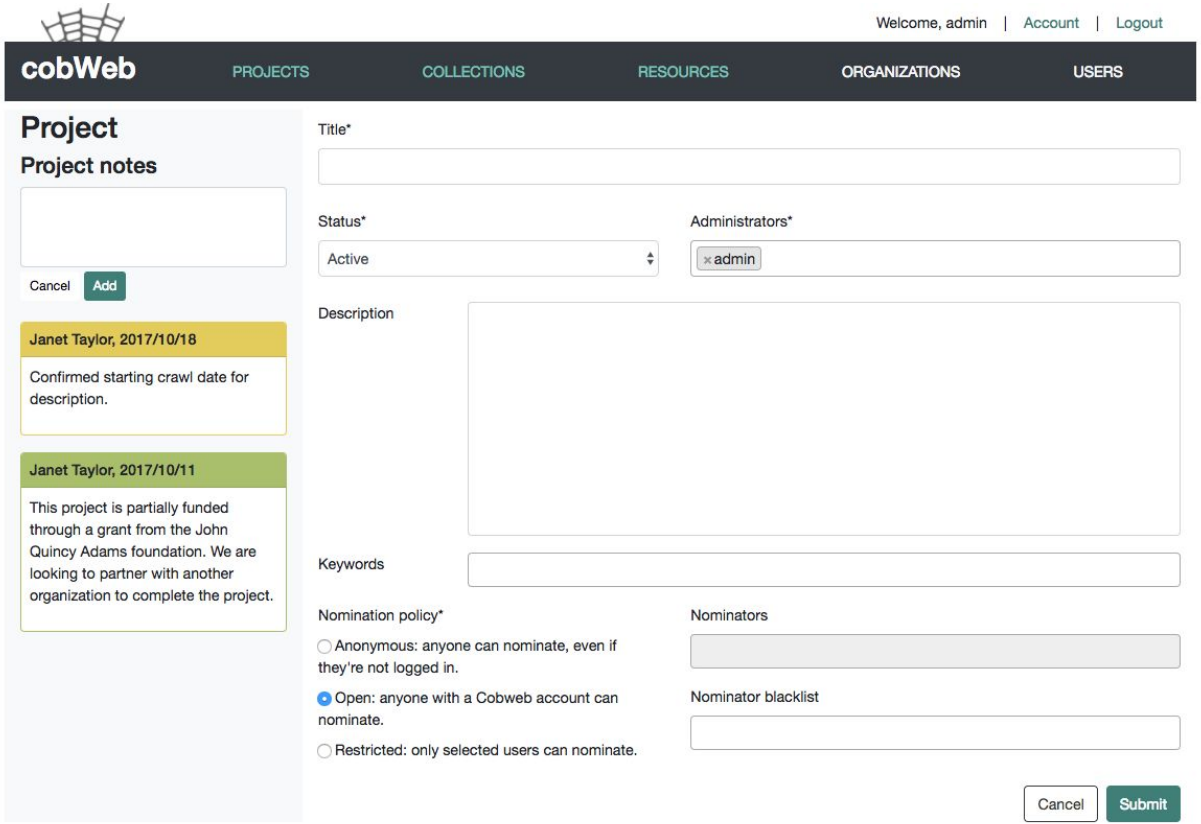


CDL California Digital Library
UCLA

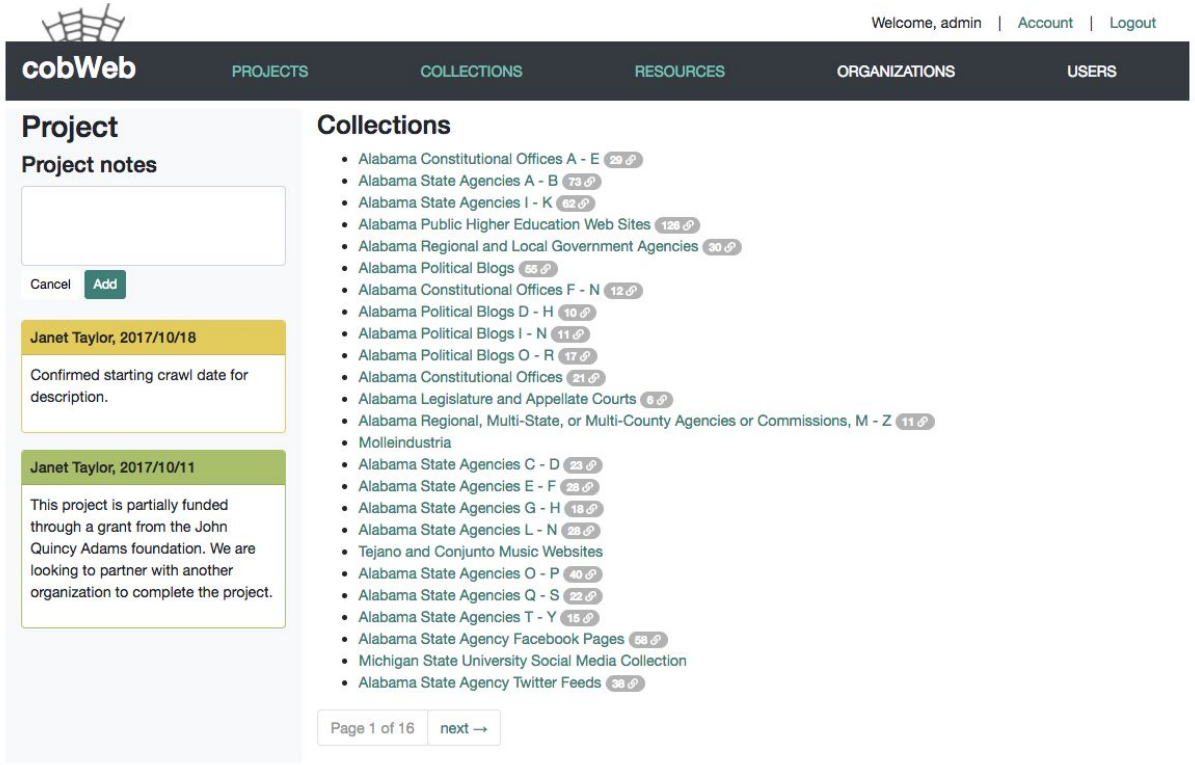
Prototype


Welcome, admin | Account | Logout
cobWeb
PROJECTS
COLLECTIONS
RESOURCES
ORGANIZATIONS
USERS
Project
Project notes
Cancel
Add
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
+ New Project
Open nominations project. 1
Milton, Massachusetts 1
Mount Holyoke College 1
Hague, New York
Dodging the memory hole project
Propaganda, disinformation, parody, dismissal: Making sense of "Fake News". 1002
Zlata Praha
Zlata Praha
Recusandae quo autem earum.
Nulla sed odit porro dicta consectetur nisi fugit.
Blanditiis laborum asperiores suscipit fugit vero quam.
Accusantium animi tenetur temporibus.
Accusantium culpa architecto aspernatur ullam.
Pariatur vitae eaque.
Accusantium quo facilis veniam voluptatum rem.
Temporibus perspiciatis aspernatur aperiam labore totam totam voluptates.
Nostrum fugit at cum sapiente.
Ducimus sed vel adipisci numquam.
Modi odit voluptates dignissimos.
FAKE News!!!

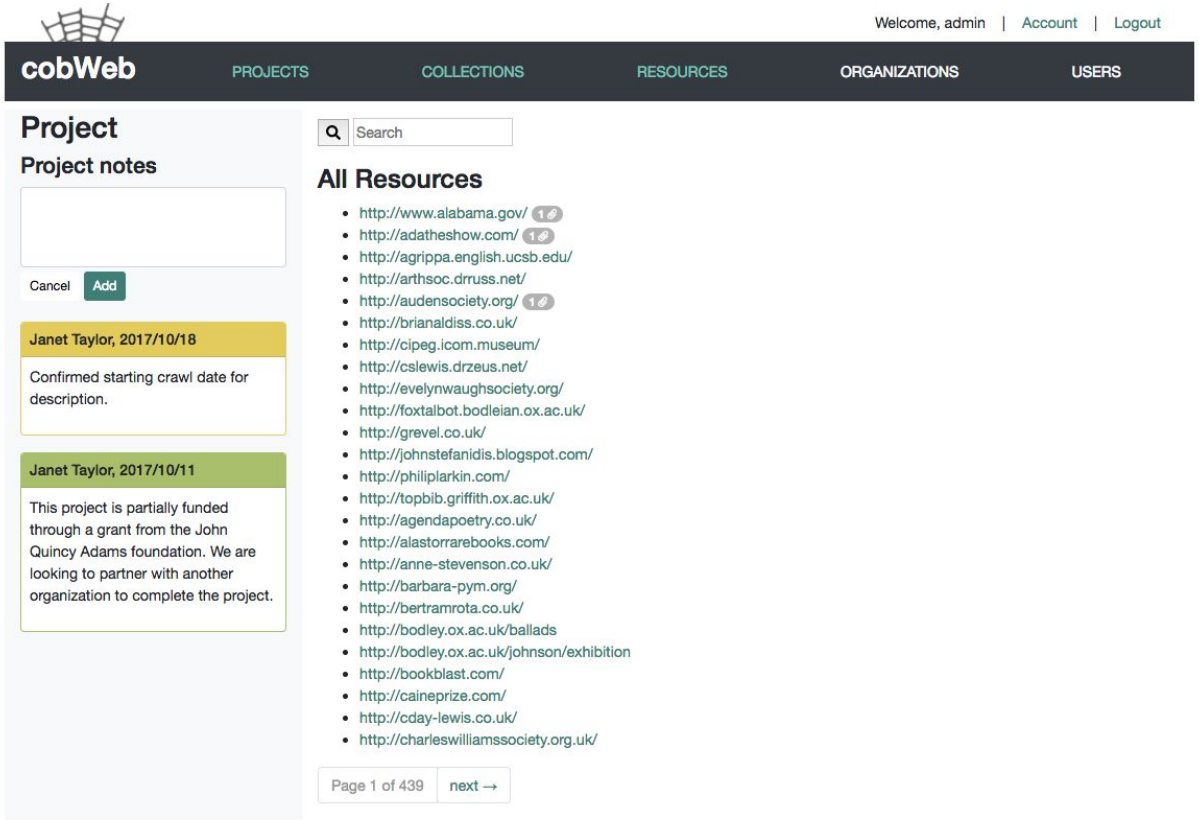
Prototype


Welcome, admin | Account | Logout
cobWeb
PROJECTS
COLLECTIONS
RESOURCES
ORGANIZATIONS
USERS
Project
Project notes
Cancel
Add
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.
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

Prototype


Welcome, admin | Account | Logout
cobWeb
PROJECTS
COLLECTIONS
RESOURCES
ORGANIZATIONS
USERS
Project
Project notes
Cancel
Add
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.
Collections
Alabama Constitutional Offices A - E 29
Alabama State Agencies A - B 73
Alabama State Agencies I - K 62
Alabama Public Higher Education Web Sites 126
Alabama Regional and Local Government Agencies 30
Alabama Political Blogs 55
Alabama Constitutional Offices F - N 12
Alabama Political Blogs D - H 10
Alabama Political Blogs I - N 11
Alabama Political Blogs O - R 17
Alabama Constitutional Offices 21
Alabama Legislature and Appellate Courts 6
Alabama Regional, Multi-State, or Multi-County Agencies or Commissions, M - Z 11
Molleindustria
Alabama State Agencies C - D 23
Alabama State Agencies E - F 28
Alabama State Agencies G - H 18
Alabama State Agencies L - N 28
Tejano and Conjunto Music Websites
Alabama State Agencies O - P 40
Alabama State Agencies Q - S 22
Alabama State Agencies T - Y 15
Alabama State Agency Facebook Pages 58
Michigan State University Social Media Collection
Alabama State Agency Twitter Feeds 38
Page 1 of 16
next →

Prototype


Welcome, admin | Account | Logout
cobWeb
PROJECTS
COLLECTIONS
RESOURCES
ORGANIZATIONS
USERS
Project
Project notes
Cancel
Add
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.
Search
All Resources
http://www.alabama.gov/
http://adatheshow.com/
http://agrippa.english.ucsb.edu/
http://arthsoc.drruss.net/
http://audensociety.org/
http://brianaldiss.co.uk/
http://cipeg.icom.museum/
http://cslewis.drzeus.net/
http://evelynwaughsociety.org/
http://foxtalbot.bodleian.ox.ac.uk/
http://grevel.co.uk/
http://johnstefanidis.blogspot.com/
http://philiplarkin.com/
http://topbib.griffith.ox.ac.uk/
http://agendapoetry.co.uk/
http://alastorrarebooks.com/
http://anne-stevenson.co.uk/
http://barbara-pym.org/
http://bertramrota.co.uk/
http://bodley.ox.ac.uk/ballads
http://bodley.ox.ac.uk/johnson/exhibition
http://bookblast.com/
http://caineprize.com/
http://cday-lewis.co.uk/
http://charleswilliamssociety.org.uk/
Page 1 of 439
next →

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...

University of California CDL California Digital Library

UCLA

Q & A

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

cobWeb

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

University of California CDL California Digital Library

UCLA