cobWeb

Collaborative Collection Development for Web Archives

#cobwebarchive

Stephen Abrams
Kathryn Stine
Jim Vanderveen

California Digital Library

Peter Broadwell
Andrew Wallace

UCLA

Lindsay Deal
Janet Taylor
Ann Whiteside

Harvard University

CC BY

CDL and Web Archiving

CDL has been involved in web archiving since 2005

- NDIIPP Web-at-Risk
  http://www.cdlib.org/services/uc3/partners/webatrisk.html
- Web Archiving Service (WAS)
- End-of-Term Archive
  http://eotarchive.cdlib.org/
- CA.gov Archive
  https://archive-it.org/collections/5763

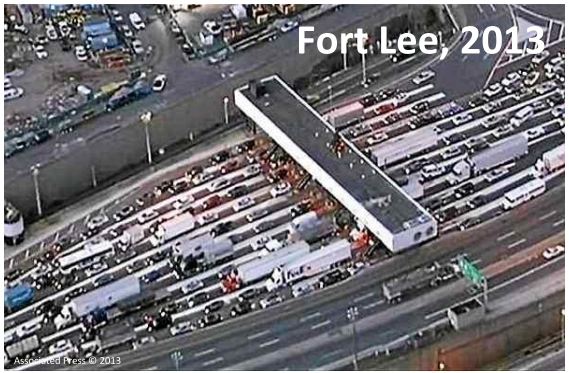
The 2015 decommissioning of WAS in favor of Archive-It decentralized web archiving activities at the University of California, so we are particularly interested in ways of promoting collaboration

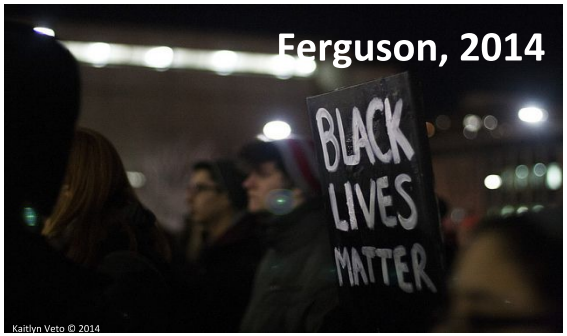
So …

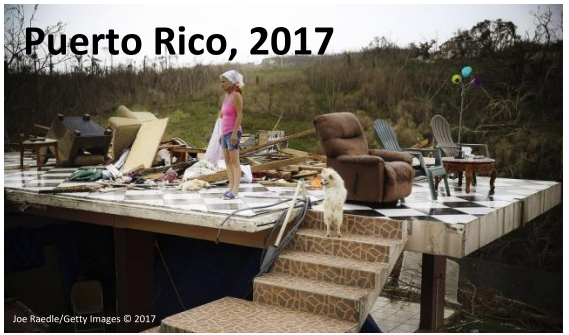
Imagine …

… 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

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

Centralized *catalog* of collection- and seed-level metadata

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

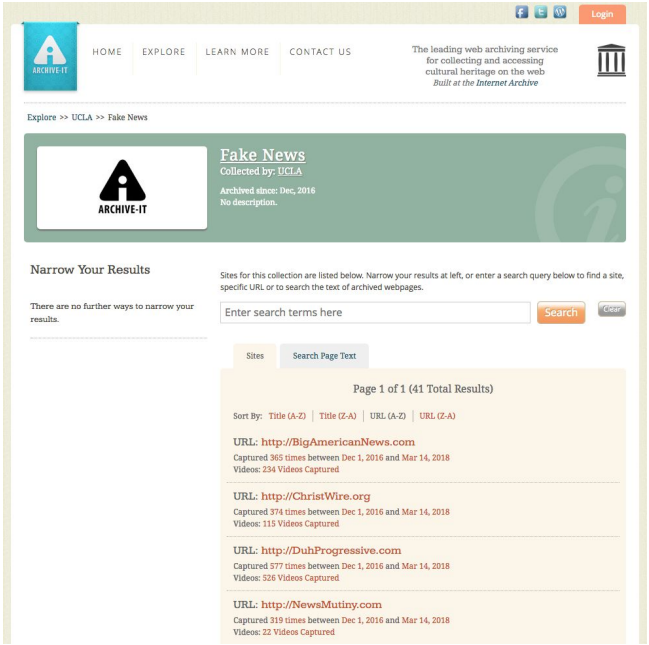
Why cobWeb ?

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



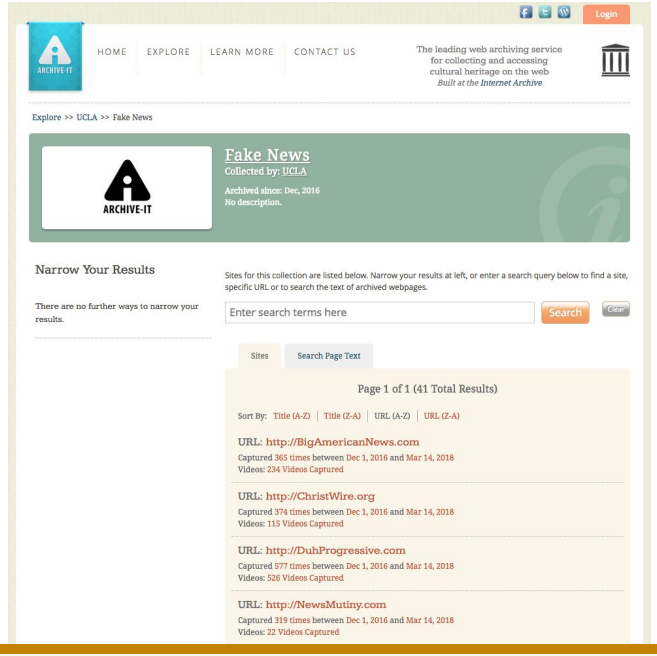
University of California CDL California Digital Library

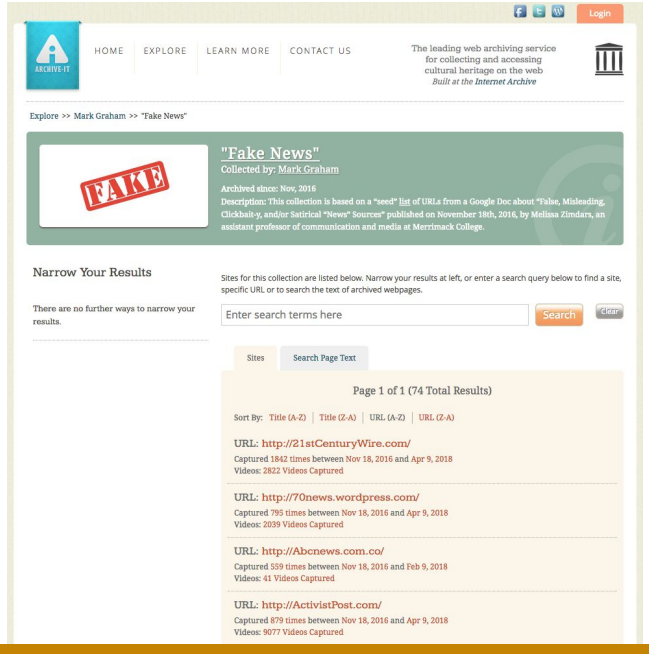
UCLA

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
- New York Art Resources Consortium (NYARC)
- 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/services/cobweb/
https://github.com/CobwebOrg/cobweb

INSTITUTE of Museum and Library SERVICES

University of California CDL California Digital Library

UCLA

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	December 2017 – July 2018
Iterative development on prototype refinements, continue importing holdings metadata	December 2017 – July 2018
Prototype rollouts	May 2018, July 2018
Beta testing	April - September 2018
Production system rollout	September/October 2018
Sustainability planning	July – October 2018

cobWeb outreach

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 we've been talking to:

- UC Libraries, Stanford, CA State Library, CA State Archives
  - CA.gov
- Ivy Plus Libraries Web Resources Collection Program
  - CAUSEWAY, Contemporary Composers
- New York Art Resources Consortium (NYARC)

cobWeb roles and collaboration

The Team

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

Team Intersections

- Data model
- Functional requirements
- Metadata flows, requirements, specs
- User data, feedback
- Wireframe review
- Prototype review and walkthroughs

cobWeb roles and collaboration

Our Process

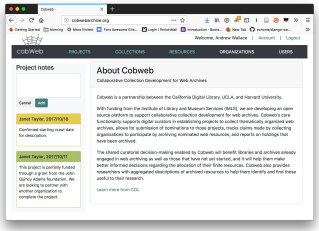
Agile approach (two-week sprints), iterative and collaborative approaches:

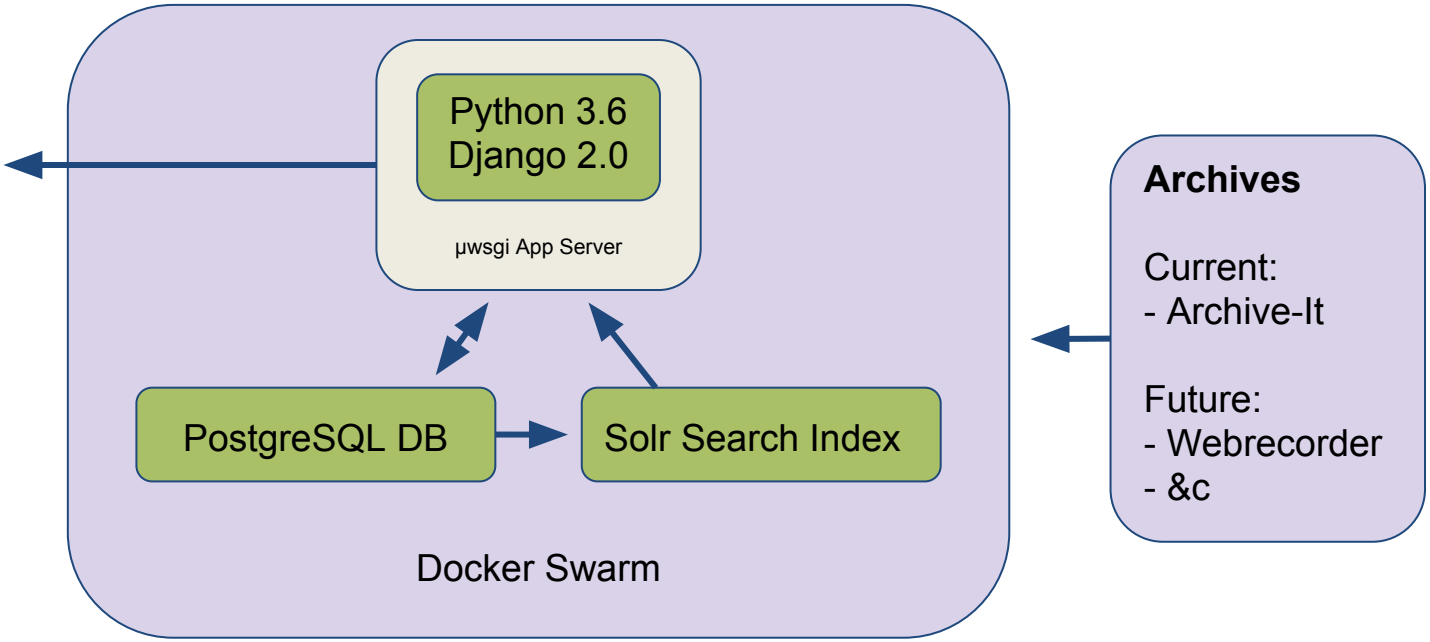
- collecting user data
- translating user data -> workflows -> functionality/interfaces
- confirming functional reqs
- sprint prep/retros

Collaboration Tools

- bi-weekly Zoom calls
- Slack for day-to-day check ins
- Google suite for doc drafting and sharing
- GitHub for source control, planning boards, and issue tracking

cobWeb technical stack



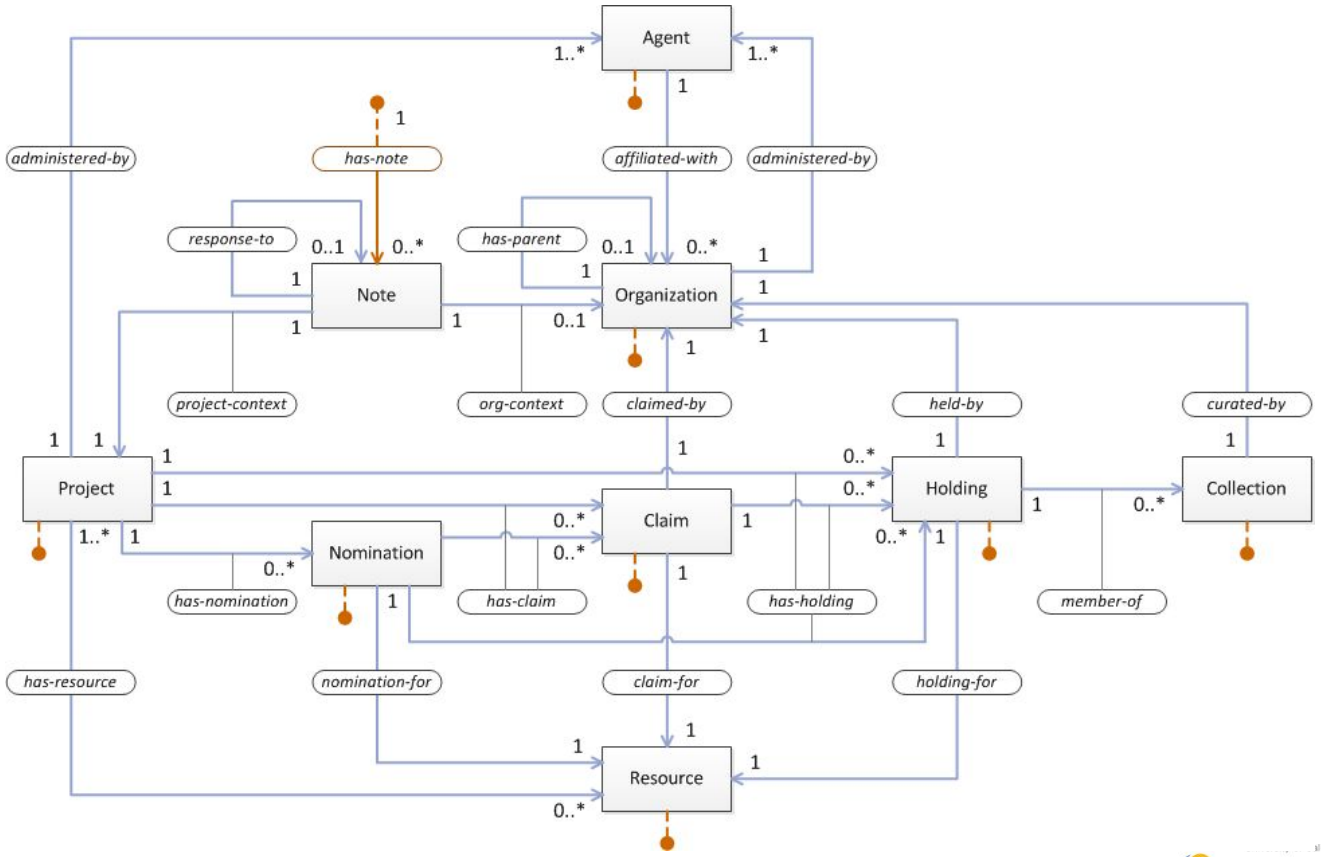


Agile, sprint-based development

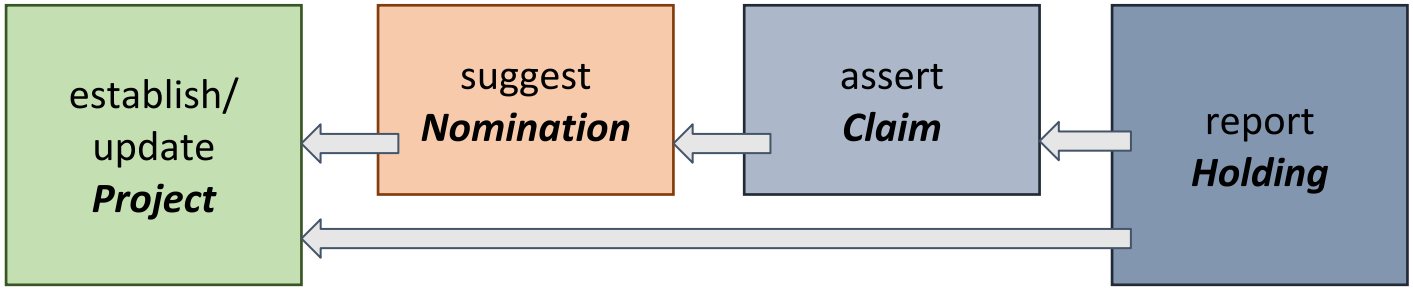
MIT software license

https://github.com/CobwebOrg/cobweb

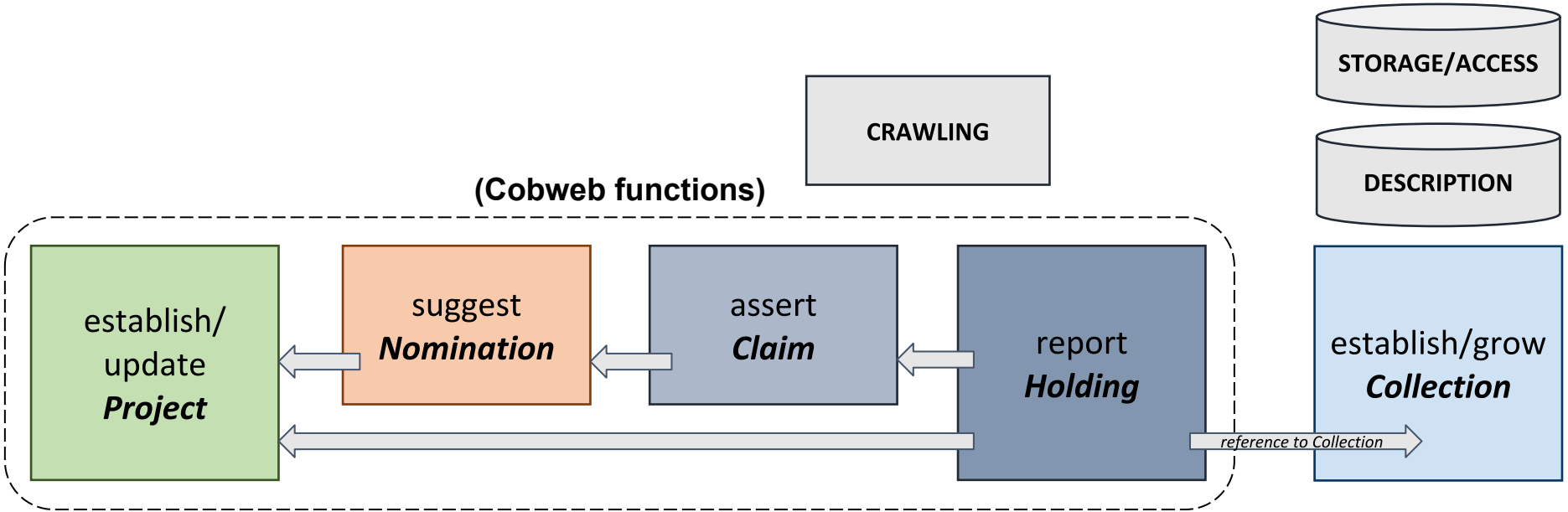
cobWeb data model



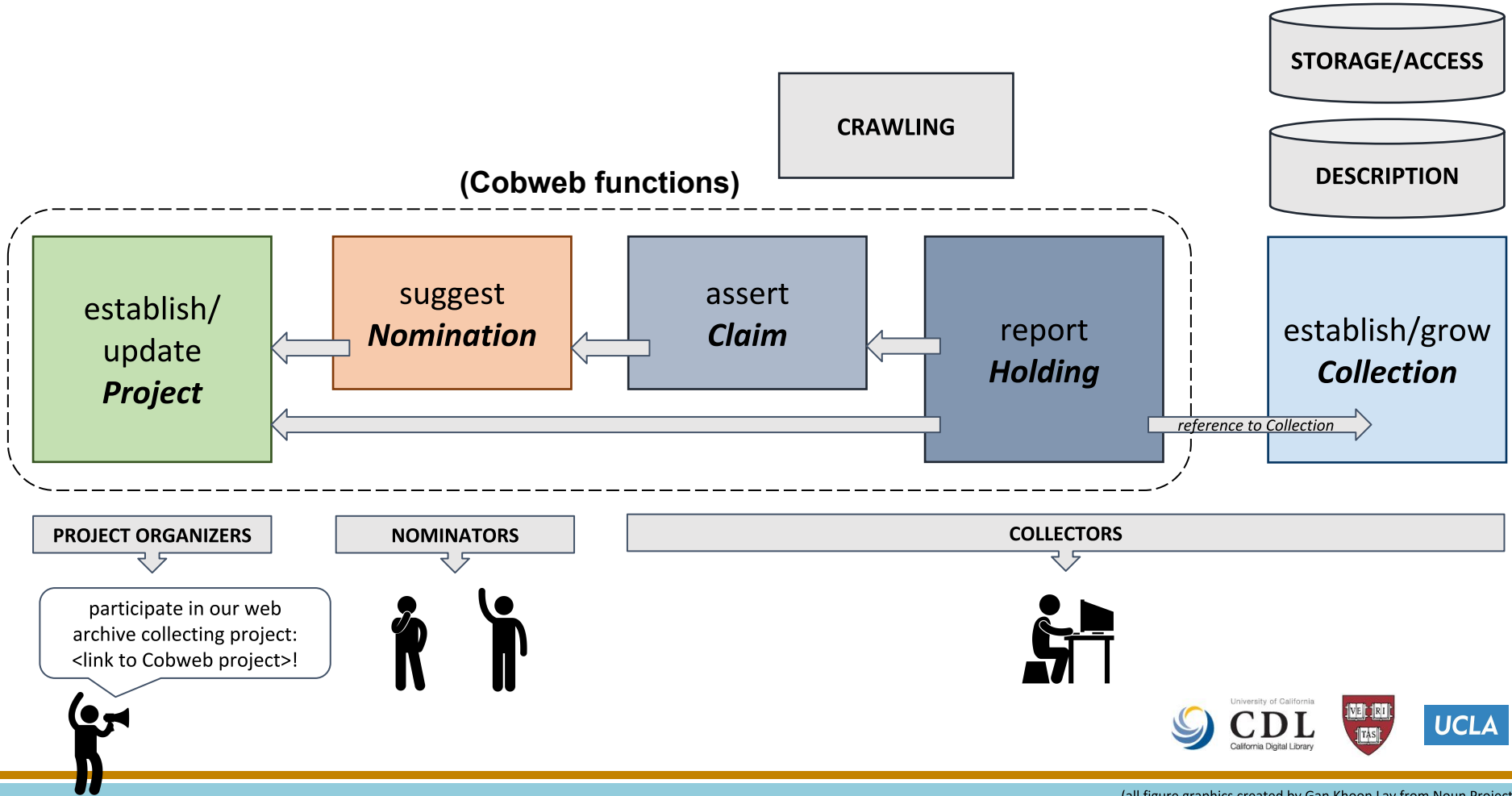
cobWeb functionality - core activities



cobWeb functionality - related external activities

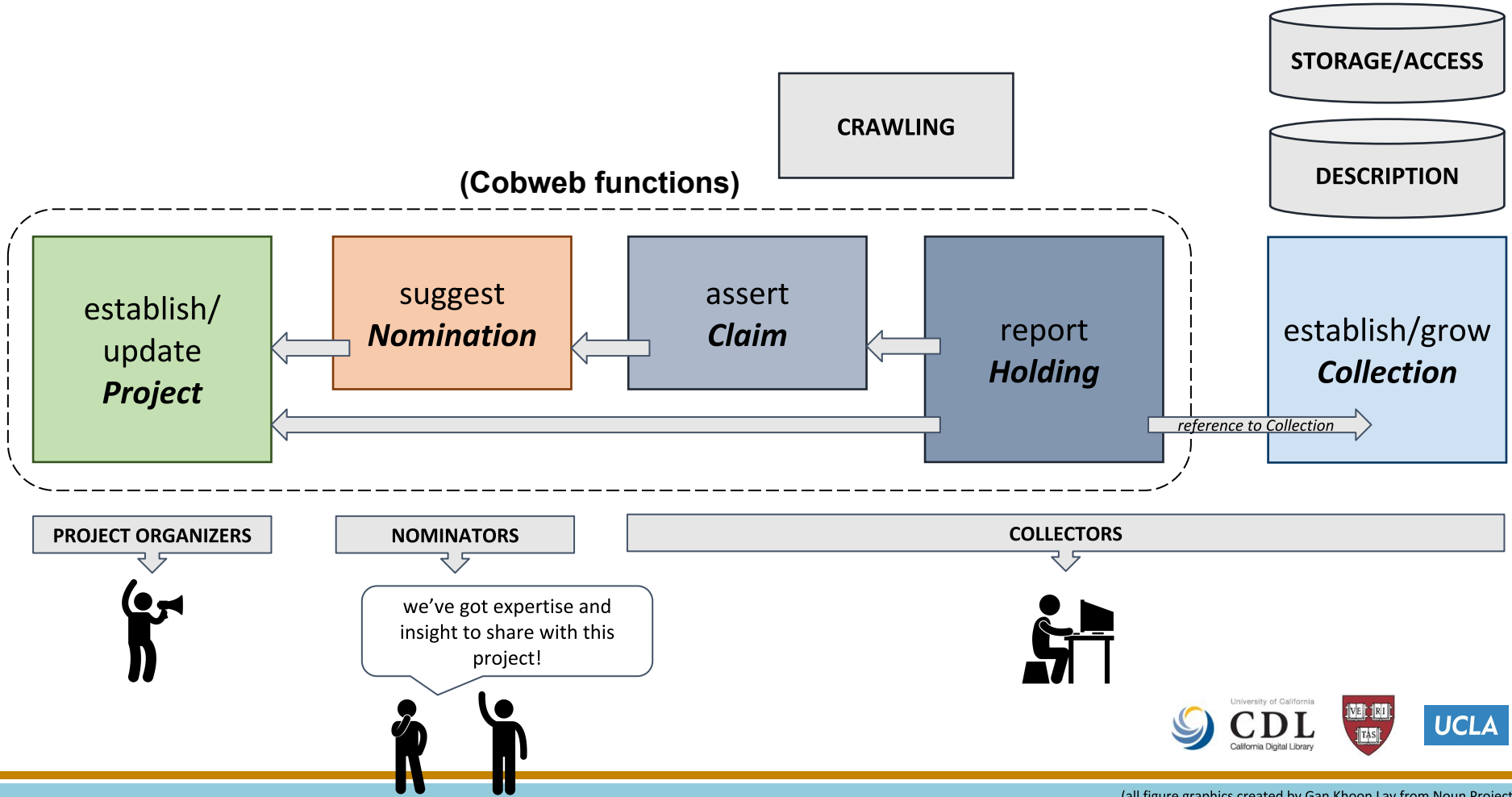


cobWeb functional user roles



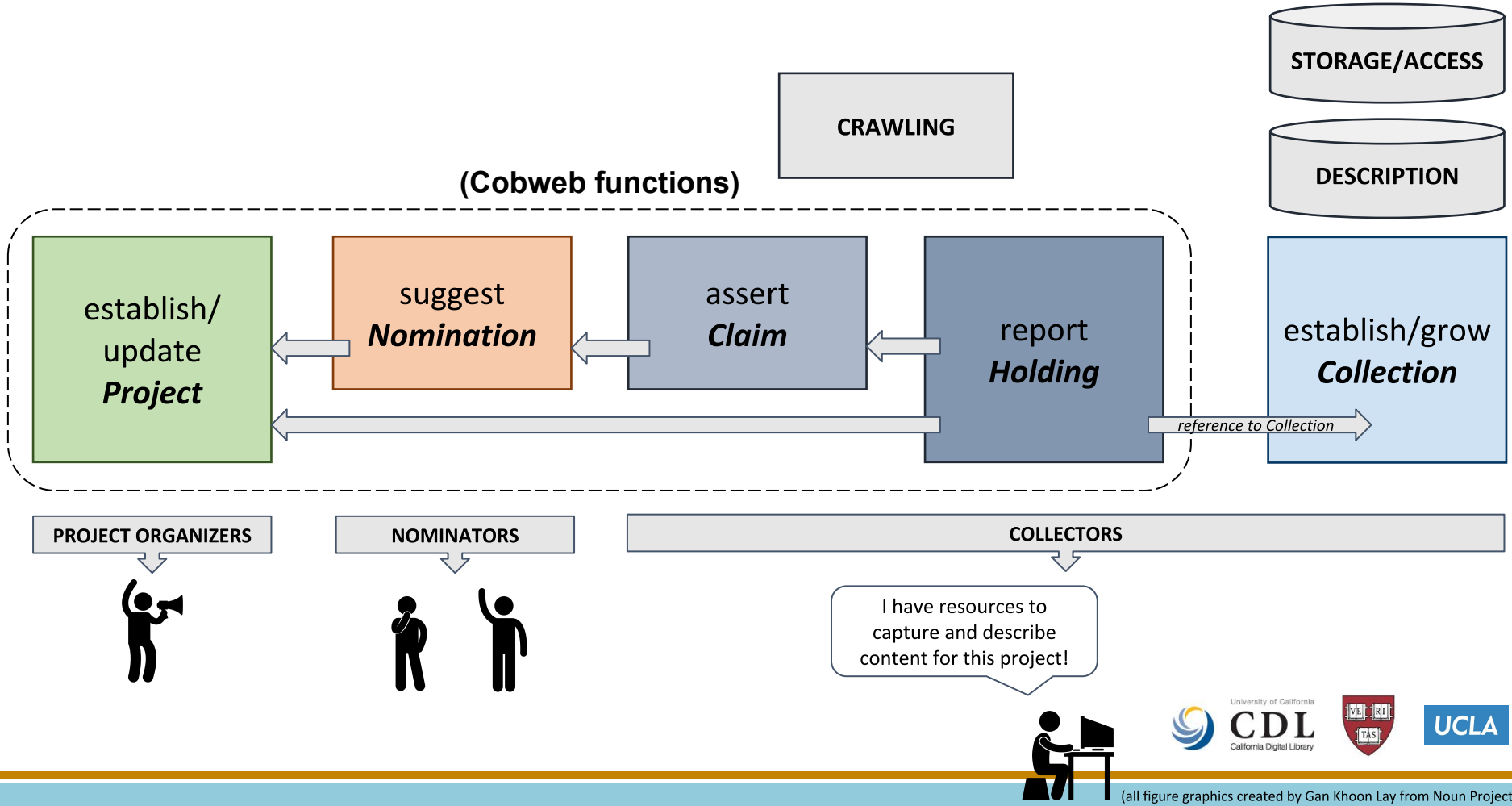
(all figure graphics created by Gan Khoon Lay from Noun Project)

cobWeb functional user roles



(all figure graphics created by Gan Khoon Lay from Noun Project)

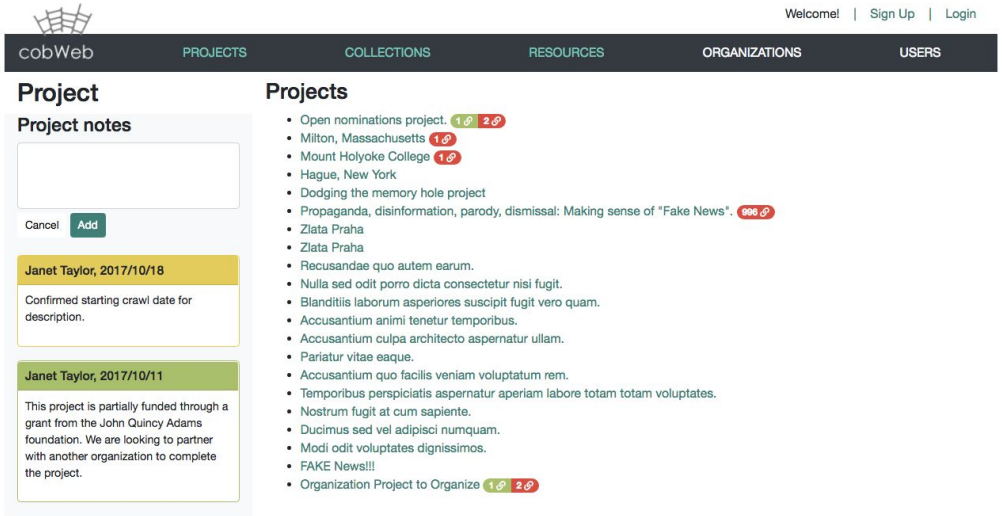
cobWeb functional user roles

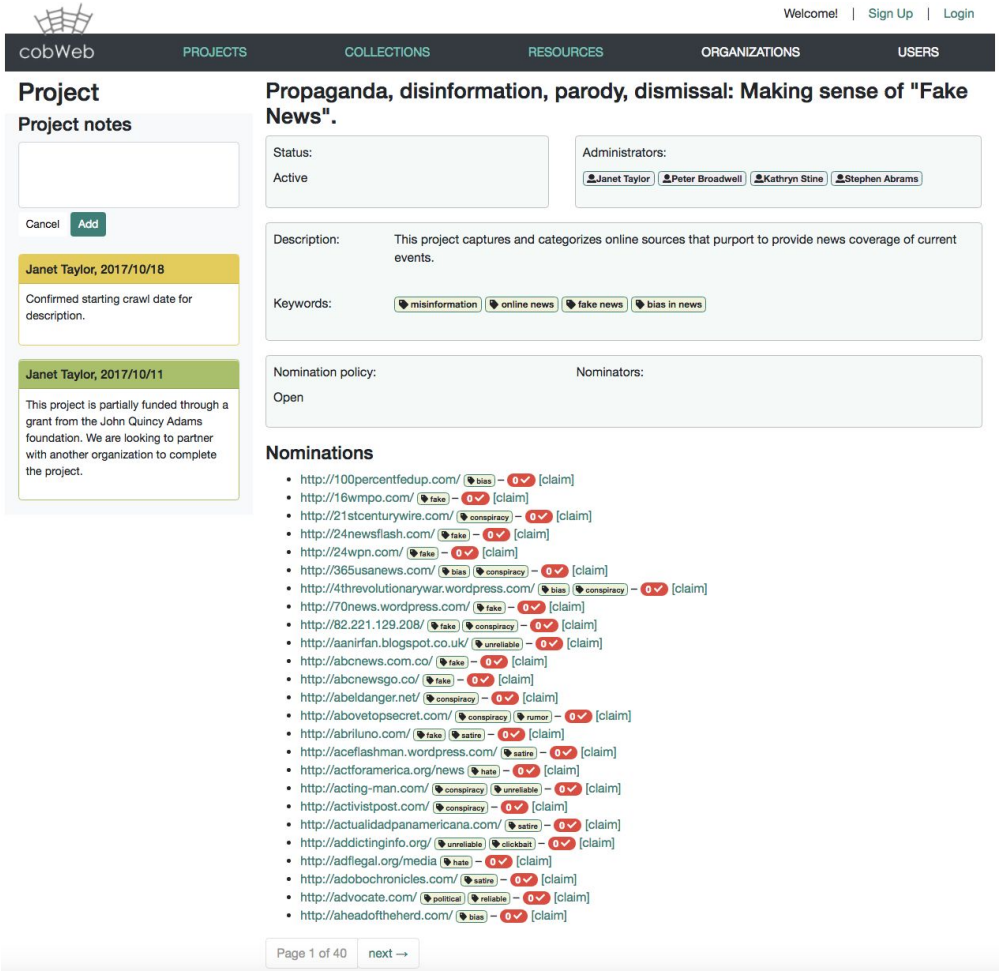


(all figure graphics created by Gan Khoon Lay from Noun Project)

cobWeb functionality - addressing current workflows

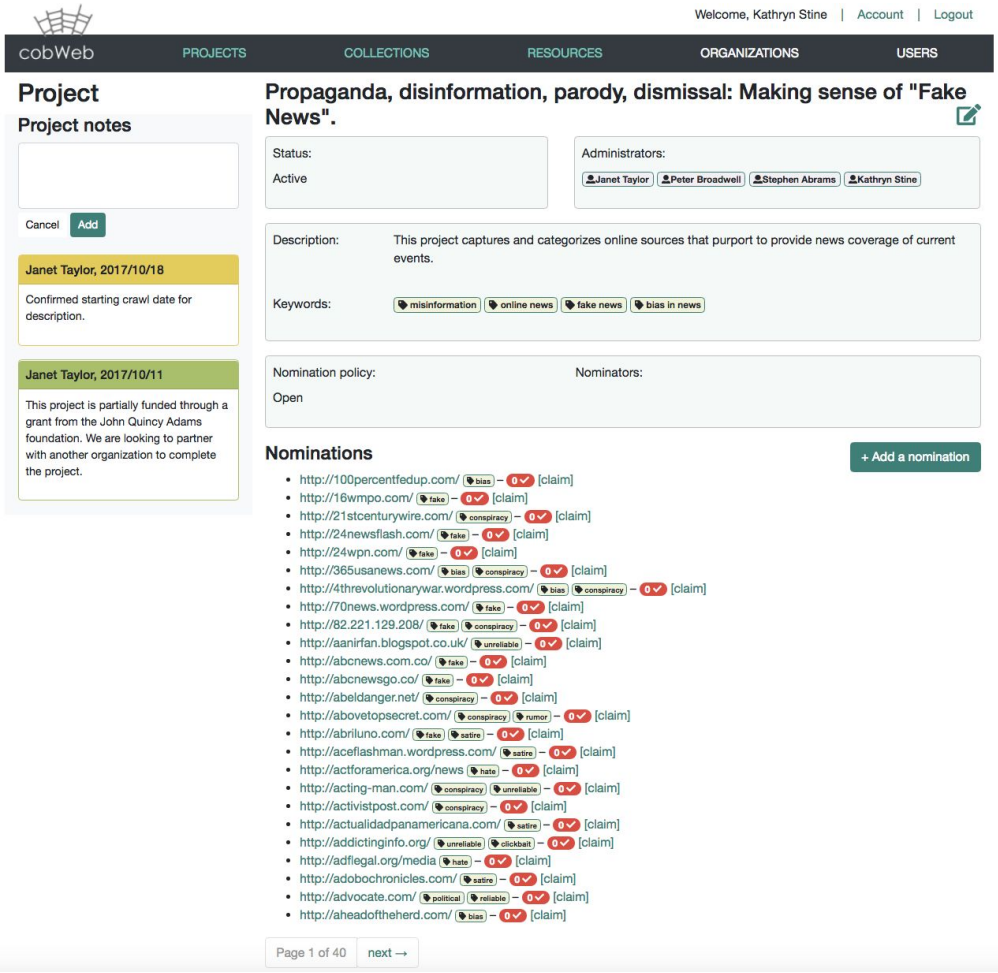
pre-Cobweb collaboration approaches			
Collecting Scope	Nomination Management	Crawling Distribution	Holdings Description
- negotiated through ad-hoc communication channels - established without good visibility into others' collecting activities - shared publicly on independent websites or nomination recruitment forms	- input via shared spreadsheets or forms (e.g., Google products) or UNT nomination tool - seed list grooming via shared spreadsheets - need for addressing time-sensitive nominations	- not well supported by any existing tools - many collaborative web archiving projects centralize crawling activities, though some distribute labor within a team (e.g., multi-agent access to a shared AIT account)	- typically part of thematic collections, established within crawling/storage/access platforms - typically described by DC or MARC metadata at the collection and seed level within one or more discovery environments


Welcome! | Sign Up | Login
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
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 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!!!
Organization Project to Organize



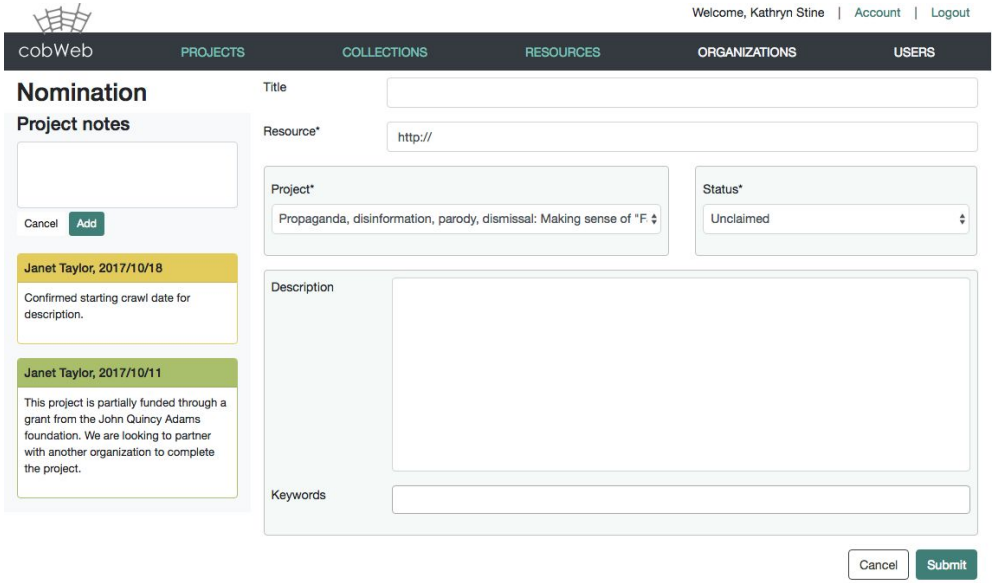
University of California CDL California Digital Library

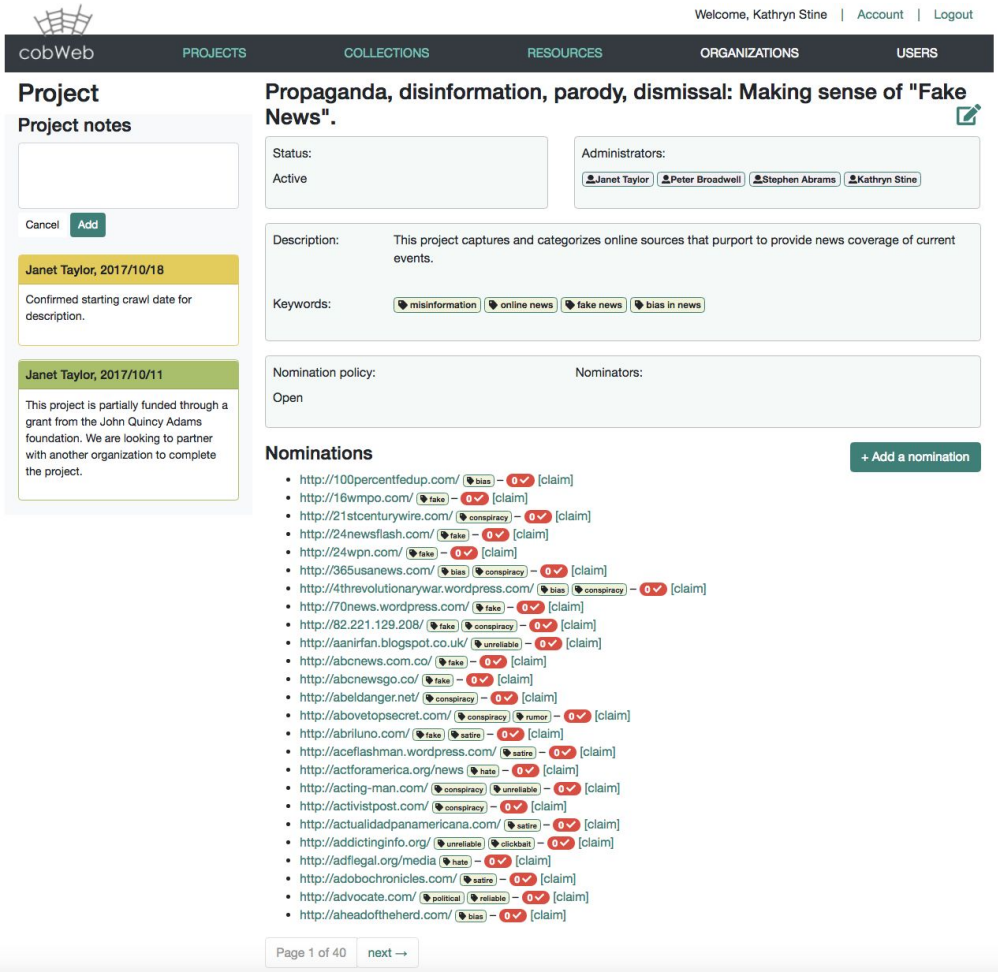
UCLA



University of California CDL California Digital Library

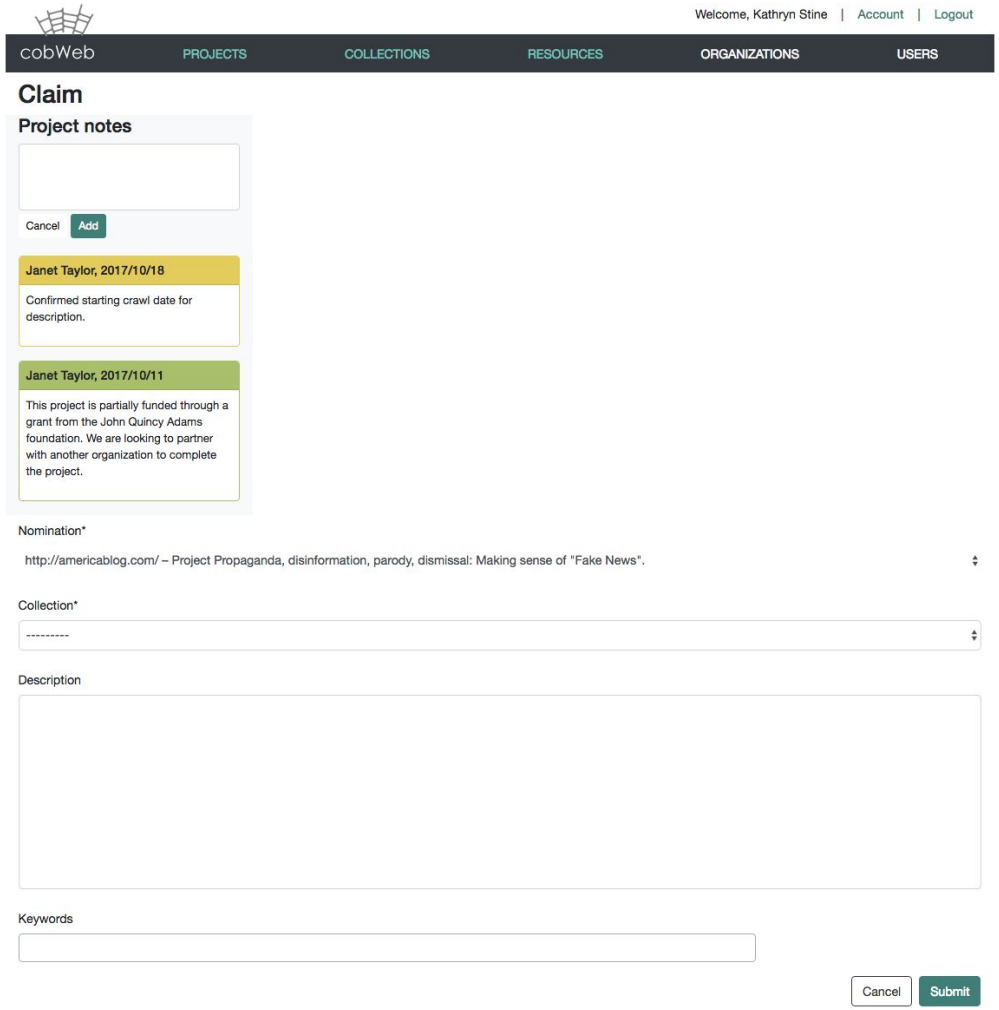
UCLA


Welcome, Kathryn Stine | Account | Logout
cobWeb
PROJECTS
COLLECTIONS
RESOURCES
ORGANIZATIONS
USERS
Nomination
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
Resource*
http://
Project*
Propaganda, disinformation, parody, dismissal: Making sense of "F
Status*
Unclaimed
Description
Keywords
Cancel
Submit



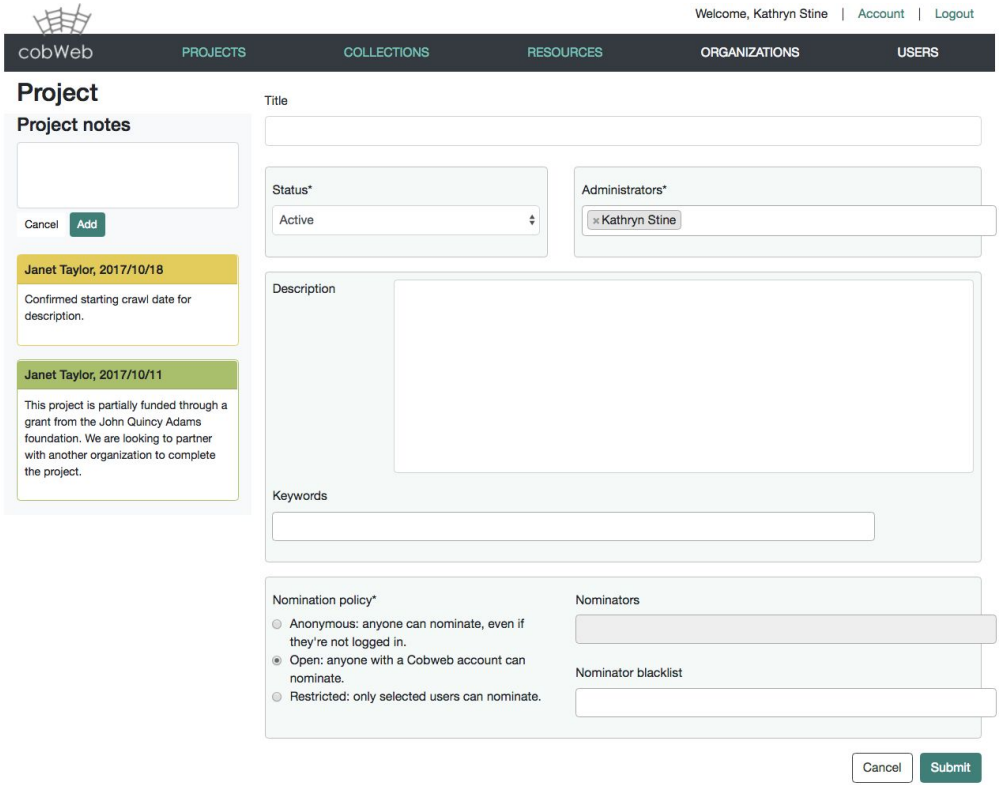
University of California CDL California Digital Library

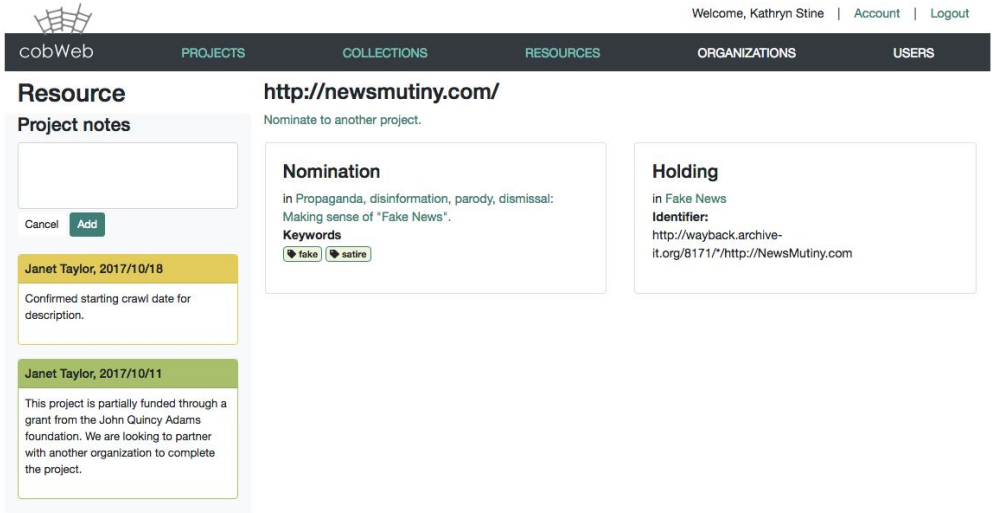
UCLA

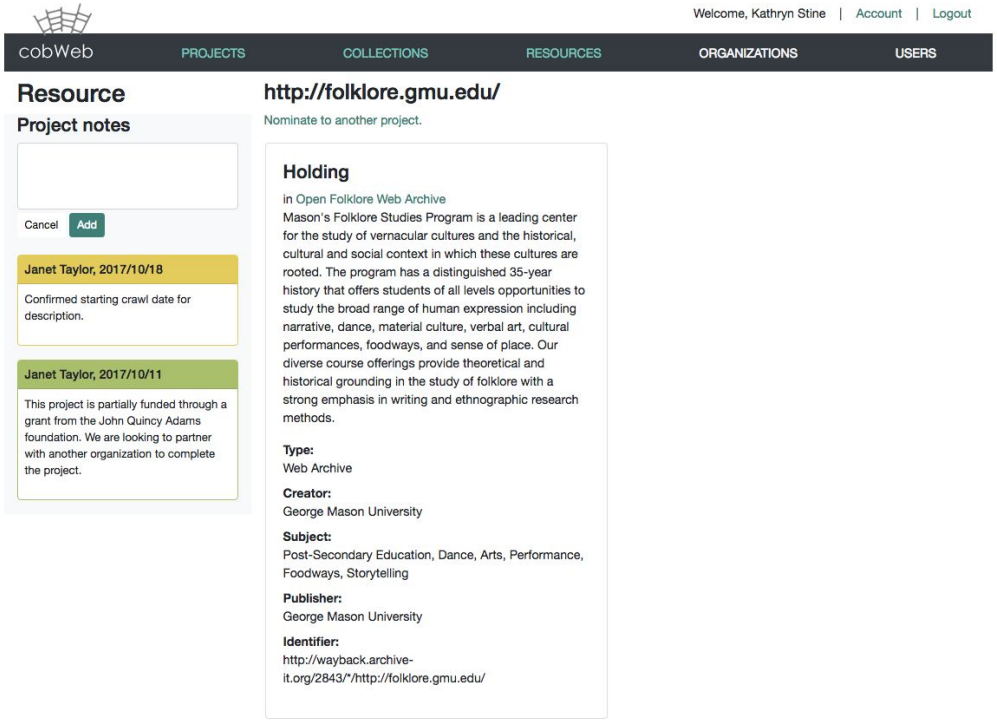


University of California CDL California Digital Library

UCLA


Welcome, Kathryn Stine | 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*
Kathryn Stine
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


Welcome, Kathryn Stine | Account | Logout
cobWeb
PROJECTS
COLLECTIONS
RESOURCES
ORGANIZATIONS
USERS
Resource
http://newsmutiny.com/
Project notes
Nominate to another project.
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.
Nomination
in Propaganda, disinformation, parody, dismissal: Making sense of "Fake News".
Keywords
fake
satire
Holding
in Fake News
Identifier:
http://wayback.archive-it.org/8171/*/http://NewsMutiny.com



University of California CDL California Digital Library

UCLA

Cobweb is a tool for collecting communities...

cobWeb *our next steps*

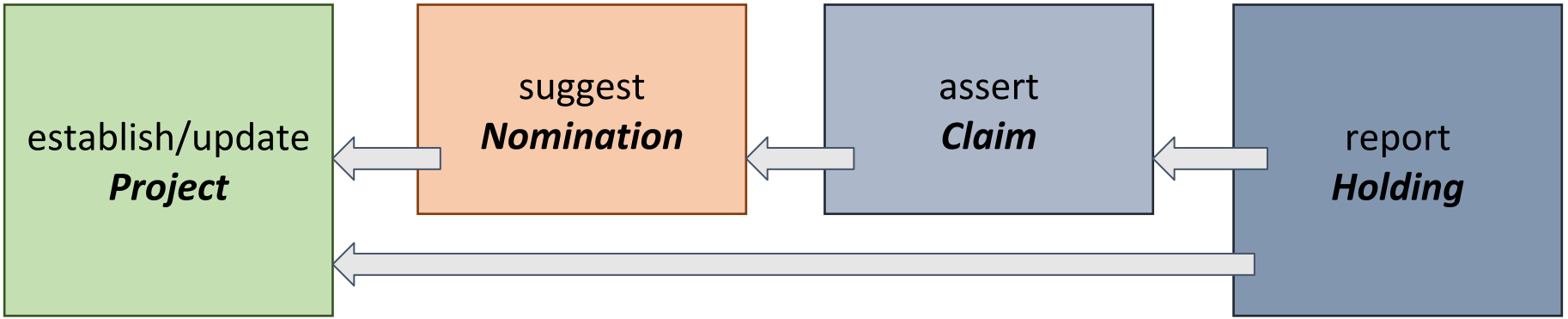
- Keep learning about collaborative web archive collecting projects and workflows
  - April: Interviews about workflows (focus on roles and collaboration)
  - May: Task-based prototype testing
- Functional requirements -> wireframing -> next prototype development
- Responding to user testing findings
- Considering how Cobweb fits with other tools and workflow steps
- Focusing on metadata inputs, outputs

next, next steps

Cobweb is a tool for collecting communities

and perhaps not just those collecting web resources...

cobWeb functionality - core activities



discussion

Cobweb affordances for:

- Establishing collecting projects
  - collection development?
  - collection management?
- Nominating resources for...
- Claiming intentions for nominated resources
- Reporting holdings

Q & A

Questions?

cobWeb acknowledgements

Institute for Museum and Library Services (IMLS) #LG-70-16-0093-16

Cobweb Advisory Board:

- Jefferson Bailey, Internet Archive
- Andrea Goethals, National Library of New Zealand
- Michael Nelson, Old Dominion University
- Mark Phillips, University of North Texas
- Alex Thurman, Columbia University
- Herbert Van de Sompel, Los Alamos National Lab

Martin Klein, Los Alamos National Lab

Rosalie Lack, UCLA

cobWeb

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/*

University of California CDL California Digital Library

UCLA