

# UC Office of the President

## UCOP Previously Published Works

### Title

Cobweb: A collaborative collection development platform for web archives

### Permalink

<https://escholarship.org/uc/item/8092g3d2>

### Journal

D-Lib Magazine, 22(11/12)

### ISSN

1082-9873

### Authors

Abrams, Stephen  
Goethals, Andrea  
Grappone, Todd

### Publication Date

2016-11-01

# Cobweb: A Collaborative Collection Development Platform for Web Archives

Contributed by

Stephen Abrams<sup>1</sup>, Andrea Goethals<sup>2</sup>, and Todd Grappone<sup>3</sup>

<sup>1</sup>California Digital Library, <sup>2</sup>Harvard Library, <sup>3</sup>UCLA Library

Stephen.Abrams [at] ucop.edu; andrea\_goethals [at] harvard.edu; grappone [at] library.ucla.edu

The California Digital Library ([CDL](#)), [Harvard Library](#), and [UCLA Library](#) have been awarded funding by IMLS to create [Cobweb](#), a collaborative collection development platform supporting the creation of comprehensive web archives by coordinating the independent activities of the web archiving community.

The demands of archiving the web in either comprehensive breadth or thematic depth easily exceed the technical and financial capacity of any single institution. In order to ensure that the limited resources of a given archiving program are deployed most effectively, it is important that its curators can know something about the collection development priorities and holdings of other, similarly-engaged institutions. Cobweb, <https://github.com/CobwebOrg/cobweb>, will meet this need by supporting three key functions: *nominating*, *claiming*, and *holdings*. The nomination function will let curators and stakeholders suggest web sites pertinent to specific thematic areas and provide seed-level descriptive metadata; the claiming function will allow archival programs to indicate an intention to capture some subset of nominated sites; and the holdings function will allow programs to document captured sites along with their collection-level description, structural and temporal scope, preservation policies, and terms of use.

How would Cobweb work? Imagine a fast-moving news event unfolding online via news reports, videos, blogs, and social media. Recognizing the importance of recording this event, a curator immediately creates a new Cobweb project and issues an open call for nominations of relevant web sites. Scholars, subject area specialists, interested members of the public, and event participants themselves quickly respond, contributing to a site list that is more comprehensive than could be created by any one curator or institution. Archiving institutions review the site list and publicly claim responsibility for capturing portions of it that are consistent with local collection development policies and technical capacities. After capture, the institutions' holdings information is updated in Cobweb to disclose the various collections containing newly available content. By distributing the responsibility, more content is captured more quickly with less overall effort than would otherwise be possible.

The International Internet Preservation Consortium ([IIPC](#)), of which all three project partners are members, has organized collaborative collecting relying on a nomination tool from the University of North Texas ([UNT](#)) and other ad hoc methods such as spreadsheets and email. While a valuable resource, the UNT tool supports nomination only and does not support other critical collecting activities; in particular, it has no mechanisms for indicating either an institution's collecting intentions or its actual holdings. Archive-It ([AIT](#)), the [Internet Archive](#)'s subscription service, has been used often for cross-institutional projects. However, ~~IA-AIT~~ does not have the legal, managerial, or technical infrastructure to

support large-scale, cross-institutional collecting, especially when the collaborating institutions do not already have formal AIT agreements. Nevertheless, Cobweb will leverage the [APIs](#) being developed for AIT for the potential retrieval of AIT holdings information, currently comprising over 3,500 collections from 350 institutions.

Cobweb will help libraries and archives make better informed decisions regarding the allocation of their programmatic resources, and promote more effective institutional collaboration and sharing.

This project was made possible in part by the [Institute of Museum and Library Services](#), #LG-70-16-0093-16. The views, findings, conclusions or recommendations expressed in this article do not necessarily represent those of the Institute of Museum and Library Services.