Open-source collaborative platform to collect content from over 350 institutions' archives

April 12 2016

With the technical and financial capacity of any currently existing single institution failing to answer the needs for a platform efficiently archiving the web, a team of American researchers have come up with an innovative solution, submitted to the U.S. Institute of Museum and Library Services (IMLS) and published in the open-access journal *Research Ideas and Outcomes* (*RIO*).

They propose a lightweight, open-source collaborative collection development platform, called Cobweb, to support the creation of comprehensive web archives by coordinating the independent activities of the web archiving community. Through sharing the responsibility with various institutions, the aggregator service is to provide a large amount of continuously updated content at greater speed with less effort.

In their proposal, the authors from the California Digital Library, the UCLA Library, and Harvard Library, give an example with the fast-developing news event of the Arab Spring, observed to unfold online simultaneously 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," explain the researchers. "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 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."

Unlike already existing tools supporting some level of collaborative collecting, the proposed Cobweb service will form a single integrated system.

"As a centralized catalog of aggregated collection and seed-level descriptive metadata, Cobweb will enable a range of desirable collaborative, coordinated, and complementary collecting activities," elaborate the authors. "Cobweb will leverage existing tools and sources of archival information, exploiting, for example, the APIs being developed for Archive-It to retrieve holdings information for over 3,500 collections from 350 institutions."

If funded, the platform will be hosted by the California Digital Library and initialized with collection metadata from the partners and other stakeholder groups. While the project is planned to take a year, halfway through the partners will share a release with the global web archiving community at the April 2017 IIPC General Assembly to gather feedback and discuss ongoing sustainability. They also plan to organize public webinars and workshops focused on creating an engaged user community.