

Human Rights Web Archive @ Columbia University Libraries

Why create a web archive for Human Rights?

- Growing reliance of researchers and activists on internet based communication tools
- Unprecedented amount of web-only content created by human rights activists, organizations and scholars
- Prevalence of websites and content that are at-risk for disappearance due to lack of resources for maintaining a web presence and/or existing political conditions
- Desire to expand collecting to include ephemeral online content of enduring research value that may be more significant than is recognized given the format (i.e. blogs)
- Commitment to develop primary source collections, especially to support the growing interest in the history of the human rights movement; complements and builds upon existing archival collections in human rights
- Potential of a web archive as big data that can enable new methods for inquiry and reveal new topics of research

Web Collecting Program

@Columbia University Libraries

Rationale for general web collecting program

Web archiving can...

- Extend collecting in areas of programmatic strength (e.g. Human Rights, Architecture & Urban Planning, New York City religious institutions)
- Aid in use and discovery of more born-digital materials of research value
- Allow us to adapt collecting methods to new publishing models and trends, especially content only published online
- Meet commitments to collect materials to preserve institutional memory (Columbia.edu and more) and complement current collections (i.e. websites to accompany manuscript and archives)
- Facilitate more responsive, timely and user focused collecting
- Provide opportunities to pursue collaborative work within and beyond Columbia University Libraries

The screenshot displays the Archive-IT interface for Columbia University Libraries. At the top, there is a navigation bar with links for HOME, EXPLORE, LEARN MORE, and CONTACT US. A logo for ARCHIVE-IT is on the left, and a tagline 'A web archiving service to harvest and preserve digital collections a service of the Internet Archive' is on the right. Below the navigation, the page title is 'Columbia University Libraries'. A search bar is present with the text 'Enter search terms here' and a 'Search' button. The main content area shows search results for 'Columbia University Libraries'. The results are sorted by 'Collection Name (A-Z)'. The first result is 'Avery Library Historic Preservation and Urban Planning', followed by 'Burke Library New York City Religions', 'General', 'Human Rights', 'Rare Book and Manuscript Library', and 'University Archives'. Each result includes a brief description, subject terms, and creator information. The page indicates 'Page 1 of 1 (6 Total Results)'.

HRWA: Engagement with Stakeholders

Questions:

How can we leverage the expertise and knowledge of human rights advocates and researchers in the selection of websites?

What rights do we have to capture and preserve freely available web content?

How do we design and develop an access environment for archived sites?

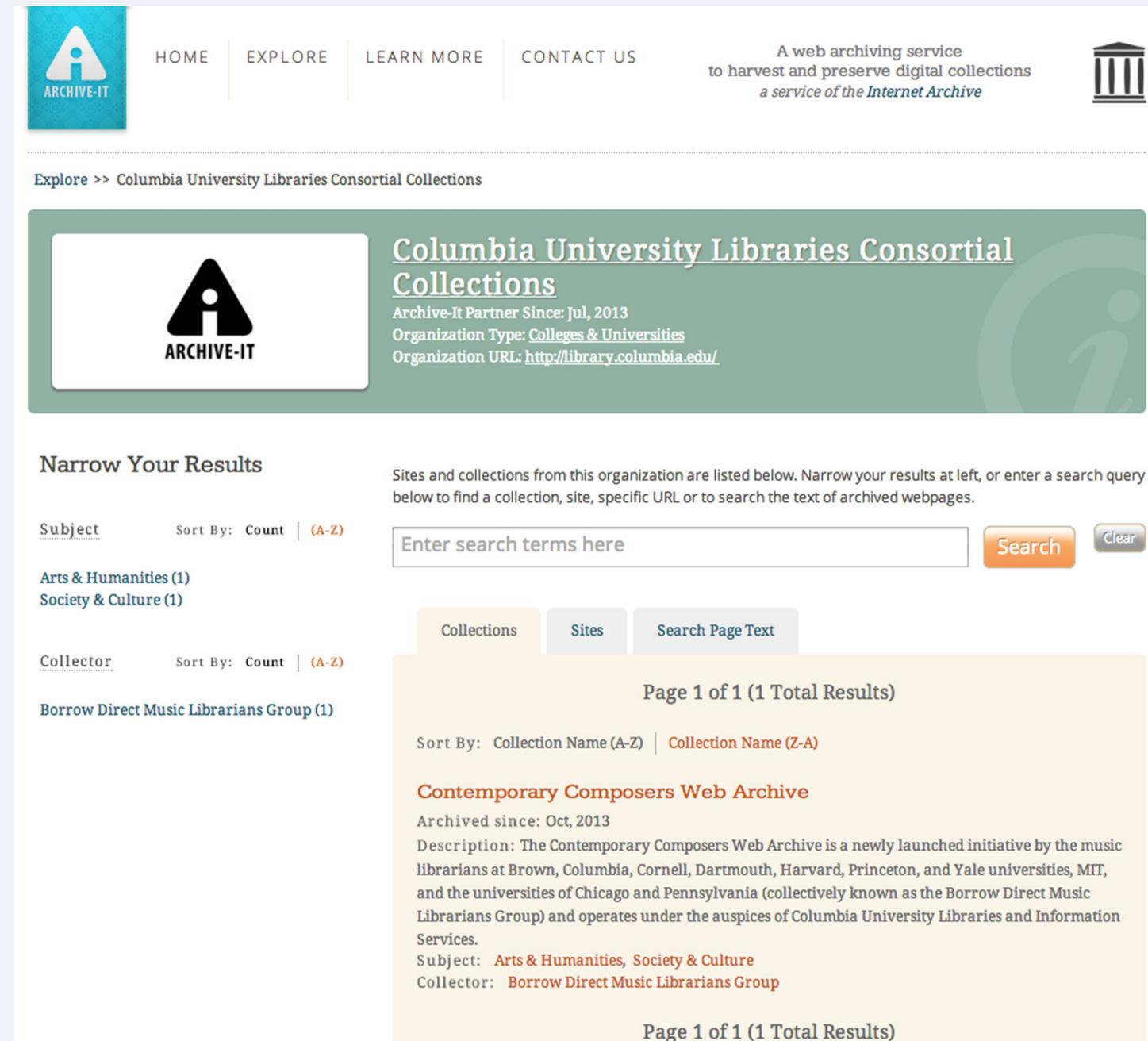
Approaches:

- Subject specialist librarians and curators created an initial list of websites, reflecting existing collection strengths and programmatic interests.
- We adapted an existing workflow for cataloging live websites to enable librarians and curators to submit archiving requests.
- Librarians were encouraged to select in response to emerging and current events, and to be attentive to at-risk websites. Web curators also identified new websites through their test crawls and quality assurance reviews of captured websites.
- We created an open nomination process to invite scholars and advocates to submit suggestions.
- Our permissions process notified organizations of our interest in archiving and requested consent. This process served to open a dialog with content creators, and encouraged suggestions for additional collecting.
- Research practices survey (2012) and usability tests (2012/ 2013) informed the design and functionality of the HRWA site.

Web Resources Archiving Collaboration

Building collaborations among

- The web archiving community
- Other research libraries, including CUL selectors and librarians
- Users and potential users of web archives
- Website creators



The screenshot displays the Archive-IT website interface. At the top, there is a navigation bar with links for HOME, EXPLORE, LEARN MORE, and CONTACT US. A tagline reads: "A web archiving service to harvest and preserve digital collections a service of the Internet Archive". The main content area features a header for "Columbia University Libraries Consortial Collections" with the Archive-IT logo and details: "Archive-It Partner Since: Jul, 2013", "Organization Type: Colleges & Universities", and "Organization URL: http://library.columbia.edu/". Below this is a "Narrow Your Results" section with filters for Subject (Arts & Humanities (1), Society & Culture (1)) and Collector (Borrow Direct Music Librarians Group (1)). A search bar is present with a "Search" button and a "Clear" button. The search results show "Page 1 of 1 (1 Total Results)" and a single result: "Contemporary Composers Web Archive", archived since Oct, 2013. The description states it is a newly launched initiative by music librarians at various universities, including Columbia, and operates under the auspices of Columbia University Libraries and Information Services. The subject is "Arts & Humanities, Society & Culture" and the collector is "Borrow Direct Music Librarians Group".

The current Web Resources Archiving Collaboration is partially funded by a grant from the Andrew W. Mellon Foundation.

The Human Rights Web Archive @ Columbia University

is a searchable collection of over 500 human rights websites created by non-governmental organizations, national human rights institutions, tribunals and individuals.

- Collecting began in 2008 and has been ongoing for active websites. New websites are added to the collection regularly.
- The HRWA is an initiative of the Center for Human Rights Documentation & Research and is a key focus of the Columbia University Libraries' Web Resources Collection Program. The HRWA was made possible by generous support from the Andrew W. Mellon Foundation.



Collaboration with librarians within and beyond Columbia University

- First collaborative project built with music librarians at Borrow Direct institutions to collect websites of contemporary composers
- A list for first round collecting was based on frequency of collection of composers' work among Borrow Direct partner institutions or via nomination by a participating music librarian
- Over 30 composers or their proxies have consented to have their websites archived in the newly formed Contemporary Composers Web Archive (CCWA)
- 69 composers are in the first round and 68 of their websites will be harvested soon, followed by cataloging and quality assurance of crawls

A screenshot of the Contemporary Composers Web Archive website. The top navigation bar includes links for HOME, EXPLORE, LEARN MORE, and CONTACT US. Below the navigation bar, there is a search bar and a "Search" button. The main content area features the "Contemporary Composers Web Archive" logo and a description of the project. The description states that the archive is a newly launched initiative by music librarians at various universities, including Brown, Columbia, Cornell, Dartmouth, Harvard, Princeton, and Yale, as well as MIT and the universities of Chicago and Pennsylvania. The archive is operated under the auspices of Columbia University Libraries and Information Services. The collector is identified as the Borrow Direct Music Librarians Group. The page also shows a "Narrow Your Results" section with a search bar and a "Search" button. Below the search bar, there are tabs for "Sites" and "Search Page Text". The results section shows "Page 1 of 1 (11 Total Results)" and lists two URLs: "http://conlonnancarrow.org/" and "http://dmitri.tymoczko.com/". Each URL is followed by the text "Captured 2 times between Oct 18, 2013 and Oct 19, 2013".

Human Rights Web Archive: Assessment and Usability

General Survey (2011)

- 200 respondents to online survey of research practices
- human rights researchers, activists, librarians, archivists

Usability Testing (2012-13)

- Internal Staff Survey
 - 24 respondents
- Usability Tests of HRWA
 - 10 subjects

FINDINGS:

- Lost access to web-based resources and documents is a consistent problem that has an impact on scholars' work
- Most users have not used a web archive and the web is not viewed as an "historical" entity
- Use cases for web archives are not well-established, known or defined
- Users are challenged by temporal dimensions of a web archive (e.g. What is a date of publication in this environment? How can one restrict searching or results by date?)

INITIAL USE CASES:

- Link or Reference Rot: a cited or known URL no longer exists or no longer points to desired content
- Analyze organizations' websites over time
- Known document/item search
- General publication search
- "Big Data" analysis of corpus of collected sites

RECOMMENDED STEPS:

- Clarify the concepts of searching "full-text" of captured pages versus searching website descriptions or metadata
- Provide more context and guidance for users, especially the difference between a search tool for the live web vs. how to search a web archive
- Make information about the size and scope of the HRWA more prominent; how and why we built this collection is highly relevant to users
- Continue to include scholars, website creators and other stakeholders in collection development for the web archive

Web Archiving: Ongoing Research and Opportunities for Growth

Web Archiving Incentive Program

- Awards to incentivize development of software to improve, enhance or innovate in the areas of web harvesting and use
- The process to select projects to fund is in progress

Journal Citation Study

- Collect citations for websites from scholarly articles published in human rights journals
- Then follow links to see if the cited web page/content is still online
- Assess comprehensiveness of the HRWA by reviewing cited websites
- Review cited websites not currently in the HRWA for possible inclusion in the web archive

Best Practices

- Collaborate with website creators and develop guidelines for web developers to produce websites more conducive to being archived
- Identify common roadblocks to effective harvest of websites and seek workarounds that will enable creators to make more easily archived sites without sacrificing certain types of functionality

Further Use Case Development

In-depth interviews with human rights scholars will explore . . .

- How can the HRWA support research and teaching?
- How does the researcher wish to work with the web archive data?
- What role should researchers play in collection development?

Presented by

Anna Perricci,
Web Archiving Project Librarian

Pamela Graham,
**Director, Center for Human Rights
Documentation & Research**

**Columbia University Libraries /
Information Services**