Establishing and growing a multi-institutional web archiving collaboration for the Collaborative Architecture, Urbanism and Sustainability Web Archive (CAUSEWAY)

Anna Perricci
Columbia University Libraries
ARLIS/NA 2015
3/21/15
Web Resources Archiving Collaboration

Many thanks to the Mellon Foundation

Building collaborations among

• Web archiving communities
• Other research libraries
• Users and potential users of web archives
• Website creators
What is the Collaborative Architecture, Urbanism and Sustainability Web Archive (CAUSEWAY)?

The Collaborative Architecture, Urbanism and Sustainability Web Archive (CAUSEWAY) is a newly launched pilot project to archive websites devoted to the related topics of architecture, urban fabric, community development activism, public space and sustainability. Made possible by a grant from the Andrew W. Mellon Foundation, CAUSEWAY is being curated by art and architecture librarians at Brown, Columbia, Cornell, Dartmouth, Harvard, Johns Hopkins, Princeton, and Yale universities, MIT, and the universities of Chicago and Pennsylvania (collectively known as the Irvies Plus Art and Architecture Group) and operates under the auspices of Columbia University Libraries and Information Services.

What is the purpose of CAUSEWAY?

The overarching goal of this project is to preserve (and document the evolution over time of) the selected websites in a secure digital archive to guarantee the continuing availability of these important but potentially ephemeral resources for researchers and scholars.

How are websites chosen for CAUSEWAY?

Participating librarians are choosing websites that fit into the themes of CAUSEWAY: Urban Fabric (e.g. historic preservation, urban renewal, urban preservation), Public Space (e.g. parklands, community gardens), or Community Activism (e.g. historic preservation initiatives, associations). Each librarian is making nominations focused on the geographic region in which her or his institution is located.

How can I view the websites that have been archived? Will access always be free?

Archived websites will remain freely accessible to the public. Websites included in CAUSEWAY will be viewable by date of capture via our Internet Archive partner pages beginning in summer 2014. Additional means of viewing archived websites will be explored by program staff.

I want more information about web archiving

Please see our Website Owner FAQ
Contact us at causeway@library.columbia.edu
Progress on CAUSEWAY

By the numbers:

• Curators from 9 Ivies Plus institutions (up to 20 seeds per institution)

• 144 seed URLs active (over 100 harvested and being released as sites are tested, cataloged and assigned metadata in Archive-It)

• 51 GB of content archived (over 1 million URLs so far)

• Over 80 sites available in Archive-It (over 60 of these sites have MARC records and Dublin Core metadata to facilitate access via Archive-It)
Feedback and learning from Avery collecting

Avery Library Historic Preservation and Urban Planning

Collected by: Columbia University Libraries

Archived since: Jan, 2010

Description: A collection of websites chosen by subject specialists from the Avery Architectural and Fine Arts Library at Columbia University. The collection’s principal thematic focus is documenting the evolution of the built environment and public spaces through the interaction of historic preservation efforts and new development projects within urban planning debates, particularly in and around New York City.

Subject: Arts & Humanities, Government - Cities, Architecture, Historic preservation, City planning, Parks, Public spaces

Creator: Avery Architectural and Fine Arts Library, Columbia University Libraries
Collector: Avery Architectural and Fine Arts Library, Columbia University Libraries
Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY)

Collected by: Columbia University Libraries Consortial Collections

Archived since: Jun, 2014

Description: The Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY) is a newly launched initiative by the art and architecture librarians at Brown, Columbia, Cornell, Dartmouth, Johns Hopkins, Harvard, Princeton, and Yale universities, MIT, and the universities of Chicago and Pennsylvania, under the auspices of Columbia University Libraries and Information Services. This project aims to preserve websites devoted to the related topics of architecture, urban fabric, community development activism, public space and sustainability in order to guarantee the continuing availability of these important but potentially ephemeral documents for researchers and scholars.

Subject: Arts & Humanities, Society & Culture, Universities & Libraries

Narrow Your Results

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.
### BAC/A+P, Barbara A. Campagna/Architecture + Planning, True green cities

**URL:** [http://barbaracampagna.com/](http://barbaracampagna.com/)

**Description:** "BAC/A+P is an architectural firm rooted in the belief that historic preservation values equal the best of green building practices. Our firm’s work demonstrates that the artistic, scientific and cultural aspects of remaking and greening historic and existing buildings are crucial to a sustainable future.” Website includes blog: True green cities.

**Captured on Aug 28, 2014**

**Subject:** Architectural firms, Sustainable architecture, Historic preservation

**Creator:** Barbara A. Campagna/Architecture + Planning (Firm)

**Language:** English

**Coverage:** Buffalo (N.Y.)

**Collector:** Cornell University Library

### Buffalo Central Terminal

**URL:** [http://buffalocentralterminal.org/](http://buffalocentralterminal.org/)

**Description:** "Buffalo Central Terminal, besides being an instantly recognizable element of the Buffalo skyline and a cultural landmark, could once again serve as a gateway to the city by utilizing the opportunities presented by its location and the current system of networks which pass through or near the structure."

**Captured on Aug 28, 2014**

**Subject:** Buffalo Central Station (Buffalo, N.Y.), Railroad stations, Historic buildings—Conservation and restoration, City planning

**Creator:** Central Terminal Restoration Corporation (Buffalo, N.Y.)

**Language:** English

**Coverage:** Buffalo (N.Y.)

**Collector:** Cornell University Library
Brown nomination
University of Chicago nomination
Cornell nomination
Harvard nomination

Lenny Zakim Fund Challenge Grant

The Lenny Zakim Fund has issued Discover Roxbury with a $10,000 Challenge Grant. If Discover Roxbury raises $10,000 in new or increased funds, the Lenny Zakim Fund will match it dollar for dollar. If you’ve thought about donating - or if you’ve lapsed for a year - this is a great time to double your donating power. If you’ve donated recently, an increase over your previous donation will count toward the match. Please click here to donate or visit this page to learn more.

Seasonal Highlights
Johns Hopkins nomination
Join & make a stand!
Penn nomination

You are viewing an archived web page, collected at the request of Columbia University Libraries Consortial Collections using Archive-It. This page was captured on 16:56:07 Sep 24, 2014, and is part of the Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY) collection. The information on this web page may be out of date. See All versions of this archived page. Metadata

Philadelphia Trolley Tracks
Interested in a talk on Newark's history or architecture?

Welcome to Newarkology, a page dedicated to the exploration of New Jersey's greatest city. From Woodside to Weequahic and from Ivy Hill to the Ironbound, come see what a fascinating place Newark is. Click on sections of the map below to be taken on a tour of any city ward you want.

**Complete Tour List**

**Cemeteries**

**Streets**

**Parks**

**Churches**

**Statues**

**About Us**

**Recent Features**

9/24/2010 - St. Casimir's Church

8/4/2008 - Clinton Avenue

12/10/2007 - St. Lucy's Church

10/29 - Colleoni Statue

9/29 - Ivy Hill Park

8/23 - Immaculate Heart of Mary
Yale nomination
Program Components

• Core Services

• Archive-It Subscription

• Supplemental Quality Review

• MARC Metadata

• Local preservation storage
CAUSEWAY Permissions

137 Site owners

- Yes (74)
- No (3)
- Later (2)
- No contact info (2)
- Did not respond (56)
Cataloging expertise

- Alex Thurman’s expertise in cataloging architecture and urban planning sites (built through work on the Avery collection of web archives) enables him to make more specific MARC records.

- Alex is working with our Bibliographic Assistant, Naeema Akter to put appropriate metadata for better browsing in the Archive-It interface.
Facets in Archive-It
Suggested additions requested

<table>
<thead>
<tr>
<th>Subject</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historic buildings--Conservation and restoration</td>
<td>17</td>
</tr>
<tr>
<td>Historic preservation</td>
<td>16</td>
</tr>
<tr>
<td>Community development</td>
<td>14</td>
</tr>
<tr>
<td>Historic districts--Conservation and restoration</td>
<td>10</td>
</tr>
<tr>
<td>Neighborhood planning</td>
<td>10</td>
</tr>
<tr>
<td>Parks</td>
<td>9</td>
</tr>
<tr>
<td>Historic buildings</td>
<td>7</td>
</tr>
<tr>
<td>City planning</td>
<td>6</td>
</tr>
<tr>
<td>Historic sites--Conservation and restoration</td>
<td>6</td>
</tr>
<tr>
<td>Historic districts</td>
<td>5</td>
</tr>
<tr>
<td>Housing</td>
<td>4</td>
</tr>
<tr>
<td>Neighborhoods</td>
<td>2</td>
</tr>
<tr>
<td>Regional planning</td>
<td>2</td>
</tr>
<tr>
<td>Sustainable architecture</td>
<td>2</td>
</tr>
<tr>
<td>Sustainable development</td>
<td>2</td>
</tr>
</tbody>
</table>

Coverage

<table>
<thead>
<tr>
<th>Title</th>
<th>URL</th>
<th>Description</th>
<th>Subject</th>
<th>Created</th>
<th>Captured once on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffalo skyline and the opportunities near the structures captured once on</td>
<td><a href="http://b">http://b</a></td>
<td>Buffalo skyline and the opportunities near the structures captured once on</td>
<td>Buffalo and restoration, C</td>
<td>Creator: Central</td>
<td>Language: Eng</td>
</tr>
<tr>
<td>Buffalo (N.Y.)</td>
<td>1</td>
<td>Buffa</td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago (Ill.)</td>
<td>8</td>
<td>High rise buildings</td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providence (R.I.)</td>
<td>8</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston (Mass.)</td>
<td>6</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buffalo (N.Y.)</td>
<td>6</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>3</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Haven (Conn.)</td>
<td>3</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philadelphia (Pa.)</td>
<td>3</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>3</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newark (N.J.)</td>
<td>2</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarasota Springs (N.Y.)</td>
<td>2</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albany (N.Y.)</td>
<td>1</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ashbury Park (N.J.)</td>
<td>1</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blackstone Park Conservation District (Providence, R.I.)</td>
<td>1</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston Harbor (Mass.)</td>
<td>1</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charles River Esplanade (Boston, Mass.)</td>
<td>1</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title</th>
<th>URL</th>
<th>Description</th>
<th>Subject</th>
<th>Created</th>
<th>Captured once on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago skyline and neighborhood infrastructure captured once on</td>
<td><a href="http://c">http://c</a></td>
<td>Chicago skyline and neighborhood infrastructure captured once on</td>
<td>Neighborhoods</td>
<td>Morris: Rogers, Eric Ally</td>
<td>Language: Eng</td>
</tr>
<tr>
<td>Collector: Morris</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown University Library</td>
<td>16</td>
<td></td>
<td>Collector: Brown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cornell University Library</td>
<td>11</td>
<td></td>
<td>Collector: Cornell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Chicago Library</td>
<td>9</td>
<td></td>
<td>Collector: University of Chicago</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harvard Library</td>
<td>7</td>
<td></td>
<td>Collector: Harvard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIT Libraries</td>
<td>6</td>
<td></td>
<td>Collector: MIT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Princeton University Library</td>
<td>5</td>
<td></td>
<td>Collector: Princeton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yale University Library</td>
<td>5</td>
<td></td>
<td>Collector: Yale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penn Libraries</td>
<td>3</td>
<td></td>
<td>Collector: Penn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheridan Libraries at Johns Hopkins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Getting the CAUSEWAY records in your OPAC via OCLC

- Records have been released to WorldCat
- A query can be built for OCLC WorldShare to obtain the MARC records for CAUSEWAY. The records can be delivered in a batch one time or periodically on an ongoing basis
Contemporary Composers Web Archive (CCWA)

By the numbers:

• 11 curators participating
• 56 sites currently available in Archive-It all with MARC records in WorldCat
  – Russell Merritt (music cataloger) collaboratively developed MARC records for composers websites; further cataloging of sites might happen in 2CUL
  – 268,828 URLs and 27 GB archived

Outreach

• SAA presentation on MARC records for CCWA
  http://www.slideshare.net/annaperricci/lightning-talk-for-session-703-of-society-of-american-archivists
• Over 30 sites tested for quality by five music librarians; bibliographic assistant on the grant tested all sites in collection
Climate change pilot & lessons learned so far

- 156 seeds nominated by at least 27 selectors from 6 institutions

Selectors from a great range of fields:
- Wide variety of area studies
- Social science
- Science and environmental science
- Medical, Law, Special Collections, Preservation
- Collection Development Associate University Librarians

- A lot of enthusiasm for topic, potential recognized
What we’ve learned about workflows and scale

• Collaborative effort builds the project and new tasks promote professional growth

• Quality Assurance and cataloging integral to process of creating high quality collections of web archives

• Distributing work does not reduce costs
Goals for duration of remainder of grant

- Collect all nominated sites in scope, test for quality, create a MARC record for each archived website (ongoing in 2015)

- Evaluate quality and solicit feedback (ongoing)

- Establish ongoing workflows and goals (2015 and onward)

- Plan a webinar with more details on what kinds of sites tend to be most successfully archived or problematic (using examples from CAUSEWAY)

- End of pilot phase: December 2015

- Hopefully to be continued 2016 onward
Future: moving beyond the pilot phase

• Discussions with Collections Development Associate University Librarians (good progress, funding news expected in April 2015)

• Coordination
  – Determining if point person(s) will lead progress (e.g. with pilot collections) and work within shared governance structure being formed

• Policy questions
  – Scope
  – Governance and sustainability
  – Place in Ivy Plus landscape
Engaging with web archiving initiatives focusing on art resources

An initiative designed to address the “urgent need to document the dynamic web-based versions of auction catalogues, catalogues raisonnés, and scholarly research projects, as well as artist, gallery, and museum websites” (http://www.nyarc.org/content/web-archiving)

Artists Files Special Interest Group
Thanks to some of many collaborators

- Bob Wolven (Primary Investigator), Alex Thurman, Naeema Akter
- Pamela Graham, Kate Harcourt, Christina Harlow, Stephen Davis
- Ivies Plus Art & Architecture Group (special thanks to Carole Ann Fabian, Chris Sala, Paula Gabbard and Karen Bouchard)
- Borrow Direct Associate University Librarians for Collection Development group
- Elizabeth Davis, Russell Merritt & Borrow Direct music librarians
- Climate change selectors at Borrow Direct/Ivy Plus institutions
- Archive-It staff
- Community for discussion and participation
  Including: NYARC, METRO, International Internet Preservation Consortium (IIPC), SAA Web Archiving Roundtable, ARLIS/NA Artist Files SIG
Again many thanks to Ivies Plus partners!
Growing web archives
Thanks!

Anna Perricci
alp2198@columbia.edu
@AnnaPerricci
Columbia University Libraries