Web Archiving Expertise at Yale
(July 2016)

This document identifies the scope of Yale University’s web archiving initiatives and the tools used by its staff. To date, these initiatives have included projects undertaken as Ivy Plus initiatives, unit-level use of the Archive-It service, work of the University Web Archiving Working Group, and various collaborations with partners both within and beyond Yale University.

Yale University Library Ivy Plus Initiatives

1. **Contemporary Composers Web Archive (2013)**
   The Contemporary Composers Web Archive was begun in 2013 to capture and preserve the websites of contemporary composers. The project uses Archive-It to capture and provide access to crawled sites. The project is administered by Columbia University Libraries; Yale’s current involvement is as a selector through the Yale University Library’s Irving S. Gilmore Music Library.

   Key contact: Suzanne Lovejoy

   The Collaborative Architecture, Urbanism, and Sustainability Web Archive was begun in 2014 to capture and preserve websites devoted to the topics of architecture, urban fabric, community development activism, public space and sustainability.

   Key contact: Molly Dotson

   Ivy Plus institutions have proposed to collaboratively collect and preserve web resources of scholarly interest. They have also sought to provide collective financial support for this purpose, including funding for staff resources and an Archive-It subscription for Ivy Plus collections.

   Key contact: Daniel Dollar

Yale University Archive-It Contract

In spring 2014, after evaluating web archiving services and software, the Yale Center for British Art initiated a test of Archive-It, a web harvesting service provided by the Internet Archive that uses a Heritrix crawler, to capture content from Yale websites and social media. The trial involved web content identified by the Yale Center for British Art, Beinecke Library, Manuscripts and Archives, the Office of the President, and the Yale Office of Public Affairs and Communications and utilized the feedback of staff from these units plus the Office of General Counsel and Yale ITS.
After a successful test period, the Beinecke Library, Yale ITS, Manuscripts and Archives, and the Yale Center for British Art entered into a one-year contract with Archive-It in 2015. The Archive-It service enables account holders to collect portions of the web and provide full-text searching of captured URLs and metadata. Among the sites crawled during the 2015-2016 year were Yale University websites, select social media sites, and websites related to Beinecke’s and Manuscripts and Archives’ collecting areas.

Yale’s Archive-It contract was recently renewed for a second year by Beinecke Library, Manuscripts and Archives, and the Yale Center for British Art. The Yale Center for British Art currently administers the Archive-It service agreement but will be transferring management to Collection Development at Yale University Library beginning in July 2016.

Key contacts:
- **Archive-It users:** Rachel Chatalbash, Mike Lotstein, Gabby Redwine, and Julie McEvoy (FY16)
- **Contract administration:** Rachel Chatalbash (FY15 and FY16), Daniel Dollar (FY17), and Ray Frohlich (Security Design Review)

**Web Archiving Working Group:**
The University-wide Web Archiving Working Group (WAWG) was established in 2015 to support web archiving activities at Yale. The Working Group’s charge is to develop a web archiving strategy for Yale University that includes website harvesting, description of the archived web content, development of access methods, and investigation and management of rights issues. The Working Group has also committed to consulting with other staff members as needed about digital preservation, description, and discovery. It will also maintain necessary web archiving contracts.

During its first year of work, the Web Archiving Working Group produced a three-year strategic plan for web archiving at Yale and began work to execute that plan.

**Web Archiving Working Group members:**
- Rachel Chatalbash (Yale Center for British Art, co-chair)
- Melissa Fournier (Yale Center for British Art)
- Gabriela Redwine (Beinecke Library, co-chair)
- Michael Lotstein (Manuscripts and Archives)
- Maureen Callahan (Manuscripts and Archives, Sept 2015-Feb 2016)
- Daniel Dollar (Yale University Library, sponsor)
- Haruko Nakamura (Department of Area Studies and Humanities Research Support)
- Tang Li (Department of Area Studies and Humanities Research Support)
- Kim Pasko (Information Technology Services, Sept 2015-Nov 2016)
- Pam Patterson (Center for Teaching and Learning)
- Louis King (Information Technology Services)
Jeffrey Campbell (Information Technology Services, March 2016- )
Suzanne Lovejoy (Music Library)
Steve Wieda (Yale University Library IT)
Andrea Belair (President’s Office)

Manuscripts and Archives Capture University Sites
Manuscripts and Archives used HTTrack (https://www.httrack.com/), an open source website copier utility, from 2004 to 2008 to capture University sites that were being removed.

Key contact: Kevin Glick

Web Archiving the 2015 Nepal Earthquake
Assisted by Archive-It, Columbia University Library, Yale University Library, and the Bodleian Library worked to select web and social media sites for archiving in the wake of the 2015 earthquake in Nepal: https://archive-it.org/collections/5764. The project focused on the capture of non-US sites in the days immediately following the natural disaster.

Key contact: Richie Richard

Lillian Goldman Law Library Perma.cc Partnership
The Yale Lillian Goldman Law Library promotes and administers Perma.cc: https://perma.cc/about for the Yale Law School. Overseen by the Harvard Law Library, this service allows users to identify linked websites found in scholarly journals for digital preservation, thus preventing “link rot” for legal citations.

Key contact: Jason Eiseman

Yale Community Voices Archive
The Yale Community Voices Archive captures digital materials, including web and social media sites, related to student discussions of race, ethnicity, and social justice at Yale. The Omeka-based project was developed in 2016 by the Yale University Library Digital Humanities Lab in partnership with Manuscripts and Archives.

Key contacts: Carol Chiodo and Mike Lotstein

Web Archiving and Digital Preservation
Preservica, Yale University Library’s digital preservation system, has the capability to capture, preserve, and provide access to archived web content.

Key contact: Euan Cochrane