
Bentley Historical Library Web Archives: Collection Development Policy

Version 4.0 (September 11, 2014)

Table of Contents

Introduction	2
Mission of the Bentley Historical Library	3
Scope of the Bentley Historical Library Web Archives	3
Designated Community of Users	4
Access to the Collections	4
Roles and Responsibilities for Website Preservation	5
Bentley Historical Library:	5
California Digital Library:	5
Content owners:	5
Intellectual Property Rights	6
Collection Development Policies	7
University Archives and Records Program	7
Michigan Historical Collections.....	8
Social Media.....	Error! Bookmark not defined.
Ongoing Activities	10
Maintenance	10
Reappraisal	10
Version History	11
Appendix A: Sample Communication to University of Michigan faculty	12

Introduction

This collection development policy for the Bentley Historical Library Web Archives will guide archivists' appraisal and selection of materials and serve as a reference for researchers, content owners, and potential donors. In addition to identifying the scope and focus of the Bentley Historical Library Web Archives, this document articulates the library's underlying criteria and strategies for collection development. Given the rapid pace of change in the content, features, and functions of websites and the continuous development of website preservation technology, this collection development policy will be reviewed on a regular basis and revised accordingly.

For a glossary of common terms associated with website preservation, please see:
<https://webarchive.jira.com/wiki/display/ARIH/Glossary+of+Web+Archiving+Terms>

Mission of the Bentley Historical Library

The Bentley Historical Library was established in 1935 by the University of Michigan Regents to carry out two functions: to serve as the official archives of the university and to document the history of the state of Michigan and the activities of its people, organizations and voluntary associations. The library is currently comprised of two departments: Collections and Curation.

The Bylaws of the University of Michigan Board of Regents ([Sec. 12.04](#)) establish the mission of the Bentley Historical Library:

The Michigan Historical Collections will be maintained for the purpose of collecting, preserving, and making available to students manuscripts and other materials pertaining to the state, its institutions, and its social, economic, and intellectual development. The University Archives and Records Program shall be maintained to collect, preserve, and make available the records generated by the university in the conduct of its business.

In accordance with this mandate, the library has preserved select websites and online resources since 2000.

Scope of the Bentley Historical Library Web Archives

The Bentley Historical Library administers two collections of archived websites: the Michigan Historical Collections Web Archives and the University of Michigan Web Archives (referred to hereafter as the MHC Web Archives and U-M Web Archives, respectively). Each collection reflects the mission and priorities of the Bentley Historical Library as it seeks to preserve online resources of unique, essential, and enduring value. As such, the MHC Web Archives focus on the activities of individuals and organizations in the state of Michigan, with particular emphasis on religious groups, social justice, ethnic communities, commerce and industry, and politics. The U-M Web Archives, on the other hand, document the university's administration, academics, research, athletic competitions, student life, and cultural activities.

Archivists actively seek to identify new, modified, or decommissioned websites; given the dynamic nature of these materials, the Bentley Historical Library Web Archives will continue to evolve over time.

Designated Community of Users

The Bentley Historical Library Web Archives will be of interest and value to researchers with varied interests related to the state of Michigan and the University of Michigan. This designated community of users will include citizens of the state, historians, genealogists, scholars, and the faculty, students, staff, administrators, and alumni of the university.

The Bentley Historical Library welcomes comments, questions, and suggestions in regards to these collections and their use. Please feel free to contact curators at bhlwebarchive@umich.edu.

Access to the Collections

Content stored in the Bentley Historical Library Web Archives may be browsed, searched, and accessed via the public collection pages of the U-M Web Archives (<http://webarchives.cdlib.org/a/universityofmichigan>) and the MHC Web Archives (<http://webarchives.cdlib.org/a/MichHistColl>).

In addition to these resources, the Bentley Historical Library has developed the following access points:

- **BHL Web Archives Portal:** A single access page for both the MHC and U-M Web Archives; provides full-text search for each collection along with the ability to browse archived content by key subject terms; also includes citation guidelines. See <http://bentley.umich.edu/dchome/webarchives/index.php>
- **Finding Aids:** Archived websites will be included as a distinct archival series in finding aids as they are periodically updated by staff members. Online EAD finding aids will include direct links to archived content stored in the University of Michigan's Deep Blue institutional repository as well as the California Digital Library's Web Archiving Service (WAS).¹ Scope and content notes indicate the overall purpose and function of the archived site and explain that captures will continue on a regular basis.
- **MARC Catalog Records:** The Bentley Historical Library has created catalog records for content in the MHC Web Archives with metadata provided by the Web Archiving Service. Archivists will begin the process of creating catalog records for the U-M Web Archives in the fall of 2014.

¹ See <http://deepblue.lib.umich.edu/handle/2027.42/65133> for the Bentley Historical Library's collections in the Deep Blue institutional repository and <http://webarchives.cdlib.org/> for information on the Web Archiving Service.

Roles and Responsibilities for Website Preservation

On July 1, 2010 the Bentley Historical Library initiated its subscription to the California Digital Library's Web Archiving Service (WAS)² in an effort to preserve online resources in a more efficient and cost-effective manner. This arrangement permits archivists to focus on the identification, appraisal, and description of content while the California Digital Library (CDL) supports the requisite hardware and software infrastructure. While no active participation is required of content owners, they may take several steps to ensure that websites are preserved as completely as possible. The following identifies specific roles in the process of website preservation as well as associated responsibilities:

Bentley Historical Library:

- Identify, appraise, and select websites that reflect the library's mission and collecting priorities.
- Provide descriptions and contextual information for materials.
- Facilitate access to content (via metadata, catalog records, and web interface) and promote its use.
- Respect the intellectual property rights of content owners.
- Reach out to webmasters or content owners when site design or configurations pose issues for the accurate capture of content.
- Monitor the development of tools, relevant technical issues, and larger trends in web archiving related to access, description, rights management, etc.

California Digital Library:

- Maintain and configure web archiving software; support curatorial interface.
- Provide secure and reliable storage of archival content and web hosting for access copies of material.
- Introduce new features and functionalities within WAS to improve the ability to capture, manage, and view content.
- Offer general technical assistance and customer support.

Content owners:

- Follow best practices for website design and accessibility.
- Amend site robots.txt files³ so the [Heritrix](http://www.heritrix.org/) web crawling software may access historically-significant, non-sensitive content.⁴ The Bentley's web crawler may be referenced in robots.txt files as *cdlwas_bot*.
- Inform the library if a website is scheduled to be launched, decommissioned, or undergo significant changes.

² For more information on the Web Archiving Service, see <http://webarchives.cdlib.org/>.

³ For information on robots.txt files and exclusions, please see <http://www.robotstxt.org/>.

⁴ The Heritrix web crawler cannot access materials that are password protected, require user authentication, or are excluded by robots.txt files.

Intellectual Property Rights

The Bentley Historical Library strives to respect the rights of content owners and to follow professional best practices for intellectual property rights management in website preservation. As part of its subscription to WAS, the library follows the Section 108 Study Group's [recommendations](#) for changes to the Copyright Act for website preservation. This group of copyright experts asserts that archives and libraries have the right to capture "publicly available" content (i.e. materials that do not require a password, entry forms, or subscriptions) and that all governmental websites should be freely accessible to web crawlers.

To address the rights of content owners, the Bentley Historical Library and its service provider have taken the following steps:

- The capture of a site will cease if WAS detects any degradation of service or negative impact on the host's web server.
- All preserved materials will be prominently labeled as an "archived copy for study and research" to avoid confusion with the live websites.
- By default, all websites are embargoed for six months so they will not be mistaken for 'live' sites or divert viewers from more current content.
- Content owners may request that portions of their site be suppressed from public view and can choose to opt out entirely from captures.
- When the Bentley Historical Library preserves the websites of private individuals, organizations, or associations (including U-M faculty and student organizations), every effort will be made to inform the content owners of our activity and to inform them of their right to opt out or suppress content (see Appendix A for a sample letter).

The Bentley Historical Library will generally respect all robots.txt exclusions except in cases where donor materials are held in third-party sites (including social media platforms such as Twitter, Facebook, Flickr, Instagram, etc.). In these cases, the Library will not honor the robots.txt exclusions so that it may capture content solely for the purposes of private study and research and in accordance with Fair Use exceptions to copyright law. The Bentley Historical Library will otherwise respect and adhere to robots.txt exclusions, but may contact content owners to request changes to rules or seek permission to override them.

Content owners are encouraged to contact the collection curators by email (bhlwebarchive@umich.edu) with questions, concerns, or requests.

Collection Development Policies

The Bentley Historical Library has articulated collection development policies that govern the identification, appraisal, and selection of content for the respective web archives of each division. These policies are informed by the library's main collecting priorities, archival principles, professional best practices, and analyses of manuscript collections and record groups.

The library makes every effort to produce exact captures of targeted websites, but acknowledges that some features of contemporary sites and the technical limitations of tools will result in the incomplete preservation of a given site's content, appearance, and/or functionality.

For more information on the Bentley Historical Library's methodology for the acquisition of content, please see <http://deepblue.lib.umich.edu/handle/2027.42/94164>.

University of Michigan Web Archives

The collection development policy for the U-M Web Archives is based upon the University of Michigan Standard Practice Guide [601.08](#) and the mandate set forth in Section [12.04](#) of the Board of Regents Bylaws. The Bentley Historical Library seeks the following types of content for inclusion in the U-M Web Archives:

- Websites created, owned, or used by university units, faculty, or students in carrying out university-related business, functions, or activities.
- Websites that reflect basic functions or activities associated with the university's administration, teaching, research, service, student life, and athletic competitions.
- Websites that complement existing record groups and manuscript collections.
- Websites that fill gaps in collections.
- Websites with unique and meaningful content that is periodically updated.

To ensure that its policy remains flexible, the Bentley Historical Library has identified several exceptions to the above criteria. On a case-by-case basis, archivists may consider websites related to alumni or organizations, individuals, and events affiliated with (but not part of) the university. Archivists may also select a wider range of content in case of important events, breaking news, or upon special request by university units.

In identifying and selecting content for inclusion in the U-M Web Archives, the Bentley has focused on websites and online resources related to:

- Central administrative offices
- Academic units (with special emphasis on course catalogs, degree requirements, newsletters and publications)
- Research, instruction, and creative work within the schools and colleges (particularly among the School of Art + Design and the School of Music, Theatre & Dance)
- Centers and institutes
- Museums and libraries
- Athletic department

The library is also interested in documenting the careers of prominent faculty as well as student organizations.

Michigan Historical Collections Web Archives

The collections development policy for the MHC Web Archives is based upon the mandate set forth in Section [12.04](#) of the Board of Regents Bylaws⁵ and MHC's existing collecting priorities. In selecting material for permanent preservation in the MHC Web Archives, the Bentley Historical Library seeks:

- Websites from an organization or person whose archives we are committed to preserve.
- Websites that fill gaps in our existing collections.
- Websites with unique content that documents the work and/or thought of a person or organization.
- Websites that periodically incorporate new content or feature regular updates.
- Websites with content that is not likely to be duplicated in an individual or organization's paper records.

In light of the library's mission to document "the state, its institutions, and its social, economic, and intellectual development," the Bentley Historical Library collects content in the following topical areas:

- Agriculture
- Commerce and Industry,
- Communications
- Creative Expression
- Education
- Ethnicity
- Family
- Gender and Sexuality (including lesbian, gay, bisexual, transgender, and queer communities)
- Labor
- Leisure
- Military
- Natural Resources
- Pioneer Michigan
- Politics and Public Policy
- Professionals
- Recreation
- Religion
- Science and Technology
- Transportation

To ensure documentation of the entire state, priorities within these 19 areas have been developed and are periodically reviewed and adjusted. The process of setting collecting priorities is described by Christine Weideman's "A New Map for Field Work: Impact of

⁵ See <http://regents.umich.edu/bylaws/bylaws12.html>

Collections Analysis on the Bentley Historical Library”⁶ and Judith E. Endelman's “Looking Backward to Plan for the Future: Collection Analysis for Manuscript Repositories.”⁷

As of 2013, the library will continue to collect in the above-mentioned areas, but will concentrate efforts on documenting the transformation of Michigan's economy (especially the auto industry and emerging industries), the revitalization of Detroit, and ethnic communities (with a focus on Arab American, African American, Asian American, and Latino communities). In addition to these priorities, the Bentley Historical Library remains open to the consideration of material with great research potential in any area of Michigan's history.

⁶ *American Archivist*, Winter 1991, Vol. 54, Issue 1, pp. 54-60.

⁷ *American Archivist*, Summer 1987, Vol. 50, Issue 3, pp. 340-355.

Ongoing Activities

Given the dynamic and ephemeral nature of online resources, the Bentley Historical Library Web Archives may be described as a work in progress. As such, the collections require regular maintenance and upkeep in addition to the reappraisal of content on an as-needed basis.

Maintenance

As part of ongoing collection development activities, archivists will:

- Perform quality assurance (QA) on all captures to determine the success of captures and verify the accuracy of target URLs. More information on QA procedures may be found at <http://deepblue.lib.umich.edu/handle/2027.42/94162>.
- Update and enhance site metadata to ensure the accuracy of descriptive information and facilitate the search and retrieval of content.
- Delete individual captures that have significant technical issues.
- Identify, appraise, and select new websites in accordance with the above-mentioned criteria and priorities.
- Monitor news reports and (when possible) maintain relationships with content owners and webmasters to learn of significant changes to sites.

If an organization changes its name but retains its original functions, the Bentley Historical Library will follow archival best practice by amending the title of the archived website and noting the change in the associated descriptive metadata.

Reappraisal

Depending on the accuracy or quality of captures as well as the status of the original website, archivists may revisit earlier appraisal decisions and either remove sites or discontinue future crawls.

If archived versions of a website exhibit severe technical issues to the extent that patrons will not be able to use them for research, the captures may be deleted and the site removed from the Web Archives. Archivists may continue to monitor the site if it is associated with a significant organization, association, individual, or event.

There are two main scenarios in which an archivist may discontinue captures of a site:

- The value of the site is limited to a specific time period (such as a conference, one-time event, or the work of a temporary committee).
- The targeted website has exhibited no or minimal change for three consecutive years (as verified by WAS curatorial tools and manual review by staff).

Version History

The Bentley Historical Library will review this collection development policy on an annual basis and make updates to reflect changes to the Web Archiving Service, archived websites, archival best practices, or other relevant issues.

Version No.	Date
4.0	September 11, 2014
3.0	August 30, 2013
2.0	August 2, 2011
1.0	March 23, 2011

Appendix A: Sample Communication to University of Michigan faculty

The University Archives and Records Program (UARP) at the Bentley Historical Library would like to preserve your personal website (at)as a representation of your academic career at the University of Michigan. Since your website is likely to change over time, UARP intends to capture it on an annual basis to document the evolution of your work and accomplishments.

Since its inception in 1935, the Bentley Historical Library has served as the official archives of the University of Michigan. While web pages have long been recognized as valuable university records, UARP launched the University of Michigan Web Archives in July 2010 to capture and preserve a greater number of historically significant sites. As part of this collection, your website would be an important record of your intellectual contributions to the U of M. Additional information on this initiative may be found at <http://bentley.umich.edu/uarp/home/webarchives/index.php>.

This letter is intended to share information about our project and respond to any questions or concerns you might have. UARP will preserve your website with a 'web crawler,' a computer application that methodically copies the site's content and structure. This archived version will then be made available to researchers through the University of Michigan Web Archives at <http://webarchives.cdlib.org/a/universityofmichigan>. UARP will only capture and preserve publicly available materials and will never copy content that is password protected or requires registration or data entry. In addition, all preserved content will be embargoed for six months before being made public and will then be prominently labeled as an "archived copy for study and research" to avoid confusion with your live website. This process involves no special preparation of the website and is designed to have no negative effects on your web server's performance.

As the content owner, you have the right to opt out of allowing UARP to preserve your website and you may also request that specific sections of your site be suppressed from public view in the Web Archives. If you would like your site to become part of the University of Michigan's official archives, no further actions are required on your part.

Please contact Michael Shallcross or Nancy Deromedi by phone (734.764.3482) or by email (bhlwebarchive@umich.edu) should you have any questions or concerns about the Bentley Historical Library's University of Michigan Web Archives or if you are interested in establishing an archive of additional content from your academic career. For more information about faculty collections in the University Archives, please see <http://www.bentley.umich.edu/research/um/facpapers.php>.

Thank you for your time and consideration.