

The University of Wisconsin–Madison Archives
Web Archiving Collection Policy

Fall 2015

Created by
Katlyn Griffin

Contents

I. Mission Statement.....	3
II. Selection	3
III. Born Digital Acquisition.....	5
IV. Web Acquisition	6
IV. Access and Use.....	6
Appendix 1: Information for Campus Webmasters or Potential Donors	7
References	8

I. Mission Statement

The University of Wisconsin-Madison University Archives preserves university records and materials of permanent and historical value. Designated an official state repository for records, the University Archives is charged by both the State of Wisconsin and the Board of Regents of the University of Wisconsin System to preserve records that have permanent administrative, legal, fiscal, or historical value. The University Archives is the only University agency that provides systematic assistance in managing non-current records. The University Archives will preserve the records appraised to have permanent or archival value created by the University of Wisconsin-Madison, University of Wisconsin System Administration, University of Wisconsin Extension, and University of Wisconsin Colleges. In accordance with this mission, the University Archives collects and preserves born-digital records pertaining to our collection areas.

Our web archiving efforts began in 2007 to further support this mission by capturing websites and born digital records developed and maintained by UW-Madison departments and organizations.

II. Selection

The University Archives actively collects digital records that document University of Wisconsin-Madison faculty and staff, administrators, organizations, departments and student life. The University Archives does not discriminate against format or time period. We also capture websites maintained by UW-System Administration, UW-Extension, and UW-Colleges. Digital material at risk of degradation or websites at risk of disappearing are a priority in our selection.

See attached guidelines if you feel your website or born-digital material meets our criteria (Appendix I).

Selection Areas

Administrators

The University Archives seeks out the digital materials of administrators that hold influential positions at the University of Wisconsin-Madison; for example,

- Chancellor, Provost, Vice Chancellor
- College and school deans
- Faculty and academic staff leadership
- UW System Administration, Extension, and Colleges Administrators
- Leadership within Governance, housing, campus recreation, campus police, health facilities, and other key entities

Faculty & Staff

The University Archives seeks out the digital materials of faculty and staff most likely to retire or leave the university in the near future, and/or those who have made a significant contribution to the University. The University Archives is interested in personal collections of faculty and staff as well as research associated with their time at the University of Wisconsin-Madison. The following is an example of kinds of digital materials the University Archives will collect:

- Personal websites that relate to their UW research/work
- Blogs that relate to their UW research/work
- Presentations and publications
- Photographs
- Audio and video recordings

Units

The University Archives seeks to collect a summary of units, including but not limited to, colleges and schools at the University of Wisconsin-Madison. The following are examples of digital records the University Archives will collect from units:

- Annual reports
- Faculty hires and retirements
- Research milestones
- Space and facilities
- Newsletters
- Publications
- Celebrations/key events

Organizations

The University Archives is interested in collecting digital records from organizations that are short lived, associated with the University, or have a high membership turn over or a long history at the University of Wisconsin-Madison. The following are examples of those organizations:

- Associated Students of Madison (ASM)
- Wisconsin Alumni Association (WAA)
- Center for Leadership and Involvement (CFLI)
- Wisconsin Union Directorate (WUD)
- Student Leadership Center (SLC)
- Software Training for Students (STS)

Student Life

The University Archives will collect both formal and informal digital records that document the leadership and extracurricular activities of students at the University of Wisconsin-Madison. Topics on diversity, recreational sports, religion, violence, and crime in relation to the student body are also of interest for the University Archives.

- Key events
- Social trends on campus
- Students' health
- Fashion
- Student employment
- Student lifestyles
- Traditions
- Changing demographic
- Diversity

III. Born Digital Acquisition

The University Archives actively collects born digital materials from campus organizations and departments, including university publications and reports, photos, and videos. The Archives also accept donations of born digital materials in many formats, including DVD/CDs, external hard drives, hard drives, flash drive/memory sticks, SD card, floppy disks, and shared webspace (Dropbox, Box, Google Drive). See attached guidelines if you are interested in making a donation (Appendix I).

The University Archives' Oral History Program is a large source of born-digital material. All interviews conducted by the OHP are digitally recorded and processed. The audio exists both in mp3 and wav formats. Indexes and transcripts are created and stored digitally with hardcopies existing in both our internal files and the public files.

As technology and stable preservation formats evolve, the University Archives will actively monitor trends and convert our current collection to more stable formats for future accessibility.

IV. Web Acquisition

Acquisition Method

Our web collections are harvested, stored, and accessed through Archive-It, a subscription service from the Internet Archive. The University Archives selects websites to be crawled through Heritrix, a web archiving tool developed by the Internet Archive. The crawler captures web domains or individual web pages, taking a snapshot of the page and storing a copy in the Internet Archive, which can be accessed through Archive-It and the WayBack Machine.

Frequency of Crawls

The frequency of a capture depends on the frequency of updates to a particular website and current relevance. Websites that change frequently—like news sites—are crawled more often. The University Archives has a plan that includes yearly, quarterly, semiannually, weekly crawls, or one-time. For example, we crawl <https://commencement.wisc.edu/> twice a year: in December and May.

Scope and Limitations of Crawl

The University Archives collects solely web content relating to the University of Wisconsin-Madison, the University of Wisconsin-System Administration, University of Wisconsin-Extension, and University of Wisconsin-Colleges. File formats and types captured are HTML, JavaScript, PDFs, images, and videos. Content that is not crawled includes database-driven or form-driven pages, forums, calendars, streaming media, password-protected sites, and sites protected by robots.txt files.

IV. Access and Use

Archive-It also provides access the content it collects. The content is searchable by keyword and facets, such as subjects, creator, date, type, etc. The sites are viewable through the WayBack Machine, the Internet Archive’s access tool.

The Internet Archive strives to replicate the look and feel of websites whenever possible by also collecting CSS files and displaying the websites as they existed.

While the University Archives and the Internet Archive strive to preserve the authentic experience of a website, it is not always possible. The limited scope of our web collections

may include external links from UW pages are not captured, resulting in a “Not Found in Archive” message.

Copyright & Permissions

The University Archives does not claim copyright for any materials in the collection. In many of the records in the collection, the rights holder is listed (when it was available). Because this is a university collection, the most common rights holder for this collection is the Board of Regents of the University of Wisconsin System.

If a rights holder or creator wishes for their content to no longer be crawled by the University Archives, place a robots.txt file on your web server that is set to disallow User-Agent: *ia_archiver*.

Appendix 1: Information for Campus Webmasters or Potential Donors

Web Archiving

The University Archives is not currently crawling every university webpage. If your page is not currently in our Archive-It collection (<https://www.archive-it.org/collections/2826>) but you think it fits our collection scope, please send an email to uwarchiv@library.wisc.edu with the subject line “Request to Archive Website.”

If your organization is planning to retire or redesign university-related website, please send an email to uwarchiv@library.wisc.edu with the subject line “Request to Archive Website.”

Please include the following information about your site:

- URL and brief description of scope of website, organization, or cause
- Relevance to our collecting areas (UW-Madison, UW-System Administration, and UW-Extension and Colleges)
- Frequency of web updates (static page / weekly news, etc)
- Date of retirement or redesign (if applicable)

Digital Materials

The University Archives also accepts donations of born-digital materials in a variety of formats, including DVD/CDs, external hard drives, hard drives, flash drive/memory sticks, SD card, floppy disks, and shared webspace (Dropbox, Box, Google Drive). While the

University Archives does not limit the file formats we accept, donors should be aware that it is possible we will be unable to recover or preserve some rare formats.

If you or your organization are interested in donating, please send an email to uwarchiv@library.wisc.edu with the following information:

- Brief description of contents and scope
- Relevance to our collecting areas (UW-Madison, UW-System Administration, and UW-Extension and Colleges)
- Quantity and format of material

References

The following policies were consulted in the development of this one:

Columbia University Libraries, Web Resources Collection Program

https://library.columbia.edu/bts/web_resources_collection.html

Michigan State University Archives & Historical Collections, Web Site Collection Plan,

http://archives.msu.edu/collections/documents/CollectionPlan_v3.pdf

Purdue University, Web Archiving Collection Policy

<https://www.lib.purdue.edu/sites/default/files/spcol/purdue-archives-web-archiving-policy.pdf>

Stanford University Libraries, Web Archiving Collection Development

<http://library.stanford.edu/projects/web-archiving/collection-development>

The University of Texas at San Antonio, Web Collection Development Guidelines:

<http://lib.utsa.edu/collections/web-collection-development-guidelines/>