



CASE 13

On the Development of the University of Michigan Web Archives: Archival Principles and Strategies

AUTHOR:

MICHAEL SHALLCROSS

Assistant Archivist, University Archives and Record Program, Bentley Historical Library University of Michigan

PAPER DATE:

April 2011

CASE STUDY DATE:

July 2010

ISSUE:

The University Archives and Records Program (UARP) at the Bentley Historical Library initiated a large-scale website preservation project as part of a broader effort to proactively capture and maintain select electronic records of the University of Michigan.

KEYWORDS:

Preservation of Electronic Records, Web Archives, Website Preservation

Copyright by Michael Shallcross.

Institutional Context

The University of Michigan was established in Detroit in 1817 and relocated to Ann Arbor in 1837. In addition to the 19 schools and colleges of the main campus, the university has regional campuses in Flint (which opened in 1956) and Dearborn (which followed in 1958). As of February 2011, the three branches employ 7,092 faculty and 34,592 regular and supplemental staff.¹ In 2010, 58,947 undergraduate, graduate, and professional students were enrolled in the three campuses and a grand total of 14,690 degrees were awarded.² A leader in higher education, research and athletics, the University of Michigan's research expenditures topped \$1.14 billion in 2010 and were among the largest of any American university.³

The University of Michigan is home to 29 libraries, including the Bentley Historical Library. Established in 1935 by the University of Michigan Regents, the Bentley Historical Library serves as the official archives of the university and documents the history of the state of Michigan and the activities of its people, organizations and voluntary associations. The Bentley is comprised of three divisions: the Michigan Historical Collections (MHC), the University Archives and Records Program (UARP), and Access and Reference Services. The Bentley Historical Library reports to the Provost and Executive Vice President for Academic Affairs.

UARP has successfully managed and preserved large collections of electronic records since 1997, when former President James J. Duderstadt donated the contents of his personal computer to the archives. 2010 marked the beginning of a two-year grant from the Mellon Foundation to fund the Bentley Historical Library's MeMail Project, the goal of which is to develop a robust and flexible system with appropriate tools, resources, and procedures so that UARP may be more proactive in the identification and management of select electronic records. Much of the MeMail Project's work has focused on 'internal' records: email and administrative records stored in shared drives, content management systems, and desktop computers. The large-scale preservation of 'external' digital content came about with the creation of the University of Michigan Web Archives in July 2010

¹ Office of Budget and Planning. "All Campus Faculty and Staff Headcount" (February 11, 2011), accessed on 14 February 2011, http://sitemaker.umich.edu/obpinfo/files/um_system_faculty_10.pdf.

² Ibid. "All Campus Enrollment Data" (November 4, 2010), accessed on 14 February 2011, http://sitemaker.umich.edu/obpinfo/files/um-system_enrll_2010.pdf. "All Campus Degree Data" (November 4, 2010), accessed on 14 February 2011, http://sitemaker.umich.edu/obpinfo/files/um-system_degr.pdf.

³ Office of the Vice President for Research. "Overview of U-M Research and Scholarship" (September 2010), accessed on 14 February 2011, <http://research.umich.edu/quick-facts/overview-of-u-m-research-and-scholarship/>.

(see Figure 1).⁴ This case study details the strategies and procedures UARP followed to develop its collection of archived websites.



Figure 1

Overview of the University of Michigan Web Domain

The University of Michigan “domain” refers to websites administered by the university and denoted by the *umich.edu* domain name. Through its official websites, the University of Michigan promotes the cutting-edge research, innovative teaching, and significant creative work of students and faculty that have made it one of the world’s leading public universities. This highly diverse web presence represents the university’s goals, achievements, and initiatives to a global audience. The University of Michigan Gateway (<http://umich.edu/>) alone averages 276,000 daily hits, with visitors hailing from Indiana to Indonesia.⁵

At a more local level, U of M websites serve as information and resource clearinghouses for faculty, students, staff, and administrators. Many important documents, such as course catalogs, degree requirements, and departmental newsletters, are now available

⁴ The University of Michigan Web Archives became publicly available on February 28, 2011 and may be accessed at <http://webarchives.cdlib.org/a/universityofmichigan>. Introductory information on this collection and UARP’s associated responsibilities may be found at <http://bentley.umich.edu/uarp/home/webarchives/guidelines.php>.

⁵ Michigan Marketing & Design. “U-M Gateway” (2011), accessed on 14 February 2011, <http://mmd.umich.edu/portfolio.php?pid=39>

exclusively online. With interactive features and robust opportunities for collaboration, websites are vital to the university's daily operations. Rich multimedia content, image galleries, and podcasts document the intellectual and social life of the campus; without active preservation, these resources will be irrevocably lost.

Website Preservation at the Bentley Historical Library

UARP has long recognized the significance of the University of Michigan's online resources. In addition to the inherent evidential and informational value of such content, the websites are official university records. According to the University of Michigan's *Standard Practice Guide*, "'University records' are defined as all records, regardless of their form, prepared, owned, used, in the possession of, or retained by administrators, faculty acting in administrative capacities, and staff of University units in the performance of an official function."⁶ Academic and administrative units are also aware of their websites' transience and have requested support in preserving important content. In light of these factors, the Bentley Historical Library has a strong mandate to archive University of Michigan online resources of unique, essential, and enduring value.

UARP previously has been active in website preservation by offering guidance on web design and maintenance⁷ and by capturing select university websites. Starting in 2000, archivists crawled the homepages of major academic and administrative units with Teleport Pro and HTTrack and stored the captured content on CD-ROM. The growing volume and complexity of university websites in recent years has led UARP to find a more efficient and cost-effective model for large-scale web archiving.

Inauguration of the University of Michigan Web Archives: a Shared Responsibility Approach

As UARP looked for a scalable solution for website preservation, it became apparent that an in-house program would be impractical. The technical expertise and infrastructure required for such endeavors are costly and difficult to develop.⁸ Even though open source resources such as the Heritrix web crawler, the Wayback Machine access tool, and the NutchWAX search engine are freely available, the knowledge and time required to configure and support these products render them unrealistic for many archivists. In addition to software implementations, an institution must allot sufficient resources to store an ever-growing corpus of content and associated metadata and then ensure that these materials retain their integrity and accessibility over the long term. Access poses an

⁶ University of Michigan. "Identification, Maintenance, and Preservation of Digital Records created by the University of Michigan," *Standard Practice Guide* (2009). Accessed on 14 February 2011, <http://spg.umich.edu/pdf/601.08-1.pdf>.

⁷ UARP has posted guidelines for the design and maintenance of websites at <http://bentley.umich.edu/uarp/home/manual/special/digital/index.php>.

⁸ For an introduction to web archiving, please see: Masanès, Julian. *Web Archiving*. Springer: Berlin, 2006 and Brown, Adrian. *Archiving Websites: A Practical Guide for Information Management Professionals*. Facet Publishing: London, 2006.

additional challenge since it requires the archives to maintain web servers and configure a search and retrieval mechanism.

Rather than create and administer such an extensive infrastructure, UARP proposed that the Bentley Historical Library use an external partner to manage the project's technical aspects. By outsourcing these functions, archivists would be able to focus on the appraisal, selection, and description of historically significant websites. After reviewing available options, UARP subscribed to the California Digital Library's Web Archiving Service (WAS) in July 2010.⁹ Under the service agreement, UARP may create an unlimited number of projects (WAS terminology for a collection of archived websites), each of which may contain an unlimited number of archived sites. The yearly contract includes one terabyte of storage in the University of California Curation Center's Merritt Repository and stipulates basic digital preservation activities such as fixity checks and data redundancy.

Definition of a Collecting Policy and the Initial Appraisal of Content

In anticipation of working with the California Digital Library (CDL), UARP articulated its collecting policy for website preservation. Archivists recognized the distinct relationship between online resources and other materials at the Bentley Historical Library. Although the formats and access methods of websites are unique in comparison with its other holdings, UARP views them as integral parts of larger collections and record groups. Selecting content for the web archives would therefore require an adaptation and extension of UARP's basic collecting principles and practices. As a result, archivists did not draft a separate collection policy for website preservation but relied upon existing guidelines, previously conducted analyses, a consideration of archival theory, and surveys of the University of Michigan web domain.

UARP's emphasis on continuity across record formats allowed the Bentley Library's *Records Policy and Procedures Manual* to be the basis for the web archives' collection development.¹⁰ The manual limited the archives' scope to include the university and its community and designated the major academic and administrative units as primary archival priorities. Secondary priorities included centers and institutes, museums and libraries, athletic teams, student organizations, and individual faculty members. The manual also ensured that archival principles—and provenance in particular—were preeminent in the planning process. Organizational charts were especially helpful in establishing the provenance of specific websites and identifying content that was essential for the collections.

The definition of this collecting policy was aided by analyses of content in the University Archives. A review of archived University of Michigan publications found that many departmental materials, such as newsletters, course catalogs, and degree requirements,

⁹ For more information on WAS, please visit <http://webarchives.cdlib.org/>.

¹⁰ Available at <http://bentley.umich.edu/uarp/home/manual/index.php>.

were no longer accessioned on a regular basis once they began to be published online. Additional reviews showed that fine arts faculty members were under-documented, especially in regards to their classroom instruction and professional creative work. UARP's familiarity with its collections led to a greater emphasis on the preservation of online academic publications and the websites of faculty from the School of Art and Design and the School of Music, Theatre, and Dance.

A consideration of core university functions, as set forth by Helen Samuels in *Varsity Letters* (1992), suggested additional criteria with which to select potential websites for preservation. While reviewing Michigan's online resources, archivists were keenly aware of the extent to which websites help confer credentials (from the recruitment of students through their graduation), convey knowledge, foster socialization, conduct research, sustain the institution, provide public services, and promote a distinctive culture.¹¹ Furthermore, many of the functions identified by Samuels as being difficult to acquire in traditional accessions (such as socialization or the promotion of a distinct culture) were readily apparent in websites. In this respect, social media sites such as Facebook, Twitter, and YouTube initially seemed to be promising candidates for preservation, especially since many departments have official accounts with these services. After conducting an extensive review of social media use at the university, archivists discovered that these sites largely repeat news and information posted to other university web pages. The structure and design of social media sites also posed significant challenges for accurate website preservation. UARP decided to exclude such content from the web archives but remains mindful of social media's significance and the potential for such content to be preserved in the future. Archivists continue to monitor the professional community's progress on this front and will reassess the university's use of social networking services in 2012.

A close acquaintance with the University of Michigan's web presence was also important in establishing a collecting policy. Periodic surveys of the university's web domain over the past decade helped archivists identify many of the most important and influential sites related to business, research, academics, and creative work. With this knowledge—and an awareness of the above-mentioned points and criteria—UARP created a 55-page spreadsheet that eventually became the foundation for the actual appraisal of sites. This spreadsheet organized content of interest according to provenance and collecting priority and also identified related subdomains that would need to be captured as separate seed URLs.¹² Although additional resources were discovered during the appraisal and creation of archived sites, this initial list provided archivists with a good starting point for the methodical selection of content for preservation.

¹¹ Samuels, Helen. *Varsity Letters: Documenting Modern Colleges and Universities*. Chicago: Society of American Archivists, 1992.

¹² For instance, the Gerald R. Ford School of Public Policy (<http://www.fordschool.umich.edu/>) hosts information on its Science, Technology, and Public Policy (STPP) Program at <http://www.stpp.fordschool.umich.edu/>. To completely capture content related to STPP, this latter URL needed to be entered as a separate seed for the WAS robot to crawl.

Although UARP did not have a formal collection development policy at the outset of its website preservation project, archivists were guided in the identification and appraisal of content by well-established criteria and principles. After the public release of the collection February 28, 2011, UARP recorded its decision-making process and selection guidelines in the “University of Michigan Web Archives Collection Development Policy and Methodology.”¹³ This document is intended to provide transparency to library patrons and clients and serve as a model for other institutions engaged in website preservation. It also reflects UARP’s belief that a collection development policy needs flexibility so that archivists may document breaking news, respond to special requests, and preserve online materials for people, organizations, and events associated with (but not necessarily part of) the University of Michigan.

Technologies for Website Preservation

When UARP’s subscription to WAS began in July 2010, archivists needed to familiarize themselves with the service’s functionality and better understand how “web crawlers” operate. Professional literature and user guides, informational videos, and webinars hosted by the California Digital Library proved highly informative in this respect.¹⁴

Before the nature and use of web archiving technologies are discussed, some basic definitions may be helpful. A “website” is a collection of plain text documents formatted with the Hypertext Markup Language (HTML) and typically accompanied by other digital assets such as images, audio and/or video files, embedded media players, and style sheets. In a ‘static’ website, these files are stored within folders on a web server where each folder corresponds to a section of the website. When a user types a URL into a web browser or clicks on a link (i.e. <http://bentley.umich.edu/exhibits/index.php>), the browser retrieves the HTML document (i.e. “index.php”) and any embedded files (such as an image) from the location (i.e. the folder “/exhibits/”) in which it is stored on the web server. ‘Dynamic’ web pages constructed with PHP, JavaScript, or Flash and those based on database-backed content management systems (such as Drupal) store and display content in a different manner and pose additional challenges for website preservation (as discussed later in this paper).

A “web crawler” (also referred to as a “spider” or “robot”) is a computer program that methodically copies a website from a web server and then saves the content (and its directory hierarchy) to a local file server. Given the progressive and thorough nature of this copying process, the application may be said to “crawl” through the website.¹⁵ The crawler may also be instructed to follow any hypertext links on the pages of the target

¹³ The web archives’ collection development policy and methodology is available at http://bentley.umich.edu/uarphome/webarchives/UM_WebArchives_Policy_20110324.pdf.

¹⁴ Web curator training materials are available at: <http://webarchives.cdlib.org/p/curators>.

¹⁵ The term “crawl” may be used as a verb to describe the act of capturing (i.e. copying) a website for preservation and also as a noun to refer to a copying session.

site. To initiate a crawl, the archivist provides a “seed” URL as a target for the application and specifies whether the robot should copy:

- a. Everything on the website (i.e. <http://bentley.umich.edu/>).
- b. All the files in a single folder on the web server (i.e. <http://bentley.umich.edu/exhibits/>).
- c. A single page (and embedded files) identified by a URL (i.e. a letter written by Abbie Hoffman to John Sinclair, located at <http://bentley.umich.edu/exhibits/sinclair/ahletter.php>).

Content is saved in the International Standards Organization (ISO)-approved WARC (Web ARChive) file format, a container (or “wrapper”) that combines “multiple digital resources into an aggregate archival file together with related information.”¹⁶ Once the preserved websites have been copied and stored locally, indexing software (WAS uses the NutchWAX web archive search engine) allows the archivist to conduct keyword searches. A specialized browser (the Wayback Machine) is then used to access and render the archived content.¹⁷

Archivists conducted a number of test crawls in early July 2010 to more fully comprehend the features and performance of the WAS implementation of the Heritrix web crawler.¹⁸ In its “native” state, Heritrix is a highly configurable command-line tool; the WAS curatorial interface greatly simplifies these options by focusing on some of the most essential settings:

Scope: the breadth of a crawl. The archivist may elect to capture the entire host site, a specific directory, or a single page.

Linked pages: the depth of a crawl. The archivist may crawl only within the seed URL or have the crawler follow hypertext links one ‘hop’ from the source page.¹⁹

¹⁶ National Digital Information Infrastructure & Preservation Program. “WARC, Web ARChive file format,” *Sustainability of Digital Formats: Planning for Library of Congress Collections* (September 2, 2009); accessed 14 February 2011,

<http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml>

¹⁷ The Internet Archive has supported the development of the WARC (Web ARChive) archival format, NutchWAX, and the Wayback Machine. For WARC format specifications please see <http://archive-access.sourceforge.net/warc/>; for NutchWAX (Nutch + Web Archive eXtensions) please see <http://archive-access.sourceforge.net/projects/nutch/>; and for the Wayback Machine, please see <http://www.archive.org/web/web.php>. WAS plans to replace NutchWAX with the Apache Solr indexing engine (<http://lucene.apache.org/solr/>) in 2011.

¹⁸ For more detailed information on Heritrix, an open source web crawler developed by the Internet Archive, please see <http://crawler.archive.org/>.

¹⁹ In other words, the crawler will capture and preserve only the web page or resource that is immediately linked to the original page. If the archivist would like to preserve additional content from the ‘linked’ site, s/he must enter that site’s URL as a separate seed.

Maximum time: the duration of a crawl. The archivist may select ‘brief’ (1 hour) or ‘full’ (36 hours) and the crawl will continue until all content has been captured or the allotted time period has ended (in which case all available content may not have been captured).

Capture frequency: how often a crawl will be repeated. The archivist may elect to crawl the site once or configure the robot to perform daily, weekly, monthly, or custom captures.

These tests crawls allowed archivists to see how different settings—as well as the design and components of individual websites—influenced the results of web captures. Archivists achieved an even greater proficiency after initiating hundreds of crawls in the succeeding months.

Strategies for the Organization of Archived Websites

Another important consideration at the project’s advent involved how the archived websites would be organized and structured. Archivists were split between two main options: (1) a single ‘University of Michigan’ collection that would encompass the entire institutional web domain or (2) numerous stand-alone projects that would parallel existing University Archives collections and record groups. Due to the nature of WAS and the intended scope of the web archives, each possibility had its benefits and drawbacks.

1. Having one large collection would permit users to navigate across a diverse range of websites and thereby approximate the experience of browsing the ‘live’ University of Michigan web domain. At the same time, WAS (as of July 2010) lacked features that would optimize search and retrieval in a single collection. The most obvious shortfall was that archivists could not group related sites according to provenance or develop a hierarchy to represent the university’s organization. As a result, users might lack important information related to context or provenance while searching sites.
2. Splitting the collections into discrete units (so that separate collections would exist for individual academic and administrative units) would respect the principle of provenance and allow archivists to closely manage and describe specific subsets of content. Because each project would mirror an existing record group, UARP could insert a standard description into relevant finding aids to denote the archived websites as separate series. A significant problem with this approach, however, would be the need to capture duplicate content in multiple collections. The University of Michigan hosts many multi-disciplinary initiatives such as the Nonprofit and Public Management Center (<http://nonprofit.umich.edu/>), a joint undertaking of the Stephen M. Ross School of Business, the Gerald R. Ford School of Public Policy, and the School of Social Work. To ensure that the

Center's archived website would be accessible from the collection of each school, UARP would have to capture it three times.

In the end, the ability to aggregate related sites according to provenance (i.e., all content produced by departments in the College of Engineering) proved to be more important than the drawbacks posed by redundant captures and limited cross-collection searches. UARP thus decided that the University of Michigan Web Archives would be comprised of multiple collections. UARP would revisit this decision in December 2010 after WAS announced new upgrades and archivists had more experience working with a robust array of archived websites.

Workflows for the Creation and Description of Archived Websites

Based upon the collecting priorities identified in the *Records Policy and Procedures Manual*, archivists first preserved the websites of the university's nineteen schools and colleges and central administrative units. Attention was then turned to the athletic department, centers and institutes, and libraries and museums. A second phase initiated in February 2011 is devoted to the preservation of sites related to prominent faculty members and student organizations. Additional phases may follow based upon the collecting needs of the University Archives.

Creation of Archived Websites

After conducting a series of trial captures, archivists developed and implemented a standard workflow for the configuration and initiation of web crawls. These procedures have been followed in both collecting phases of the University of Michigan Web Archives but may be revised as WAS introduces new features and/or archivists develop new techniques.

1. *The archivist creates a project based upon a specific theme or collecting focus.*

A "project" (to use WAS terminology) is a discrete collection of archived websites that is organized around a specific theme or collecting focus. To create a new project, the archivist opens the Institutional Administrator screen of the WAS interface, selects the "Create New Project" option, and enters the project name (see Figure 2).

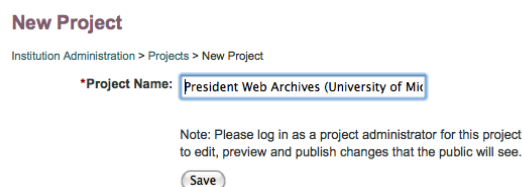


Figure 2

This step only needs to be completed once; thereafter, the archivist merely selects the appropriate project when s/he wishes to create a new site.

2. *The archivist identifies the “seed” URL for a website selected for preservation.*

While often straightforward, this task requires the archivist to verify if the site’s content is hosted from more than one domain or subdomain (i.e. from different root URLs). For example, the Athletic Department’s website on the history of Michigan and Ohio State’s football rivalry is hosted at both <http://bentley.umich.edu/athdept/football/> and <http://library.osu.edu/sites/archives/OSUvsMichigan/>. The complete preservation of this site required archivists to specify both domains as seed URLs in the WAS interface.

At the same time, the different domains within a site may merit preservation as separate websites. For example, the Office of the Vice President of Research (<http://research.umich.edu/>) maintains a large body of information related to research administration (<http://www.drda.umich.edu/>) and human research compliance (<http://www.ohrcr.umich.edu/>). Although these latter sites could be included as secondary seeds for the Vice President of Research’s site, their scope and informational value led archivists to preserve them separately.

3. *With the seed URL(s) identified, the archivists creates a new archival site manually or via the WAS browser button.*

In the manual process, the archivist first opens the appropriately themed project from the “Choose Role” screen of the WAS interface (see Figure 3).

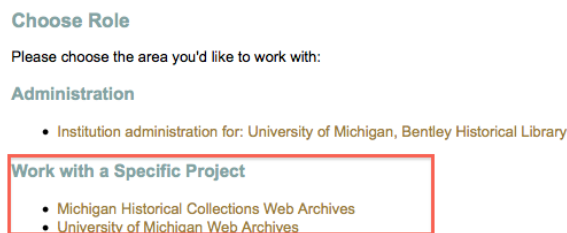


Figure 3

After a project has been selected the archivist opens the “Create Site” screen and enters the site name, seed URL(s) and relevant crawl settings (scope, linked pages, and maximum time; see Figure 4).

Create Site

Capture Settings | **Scheduling** | **Descriptive Data**

*** Required field**

***Site Name:** President Web Archives (University)

***Seed URLs:**
Ex.: http://www.example.com

Scope: Directory

Capture Linked Pages: ☐ No ☒ Yes

Max. Time: Full Capture (36 hours) | Brief Capture (1 hour) | Full Capture (36 hours)

Figure 4

UARP standardized the names of its archived sites by using the title found at the top of the website or, in the absence of a formal title, the name of the creating unit. UARP follows the best practices for collection titles as established by Describing Archives: a Content Standard (DACS). To ensure that the provenance and nature of the collections are clear, archivists supply “Web Archives” and “(University of Michigan)” to the final title. Complete names for archived websites thus follow the pattern “President Web Archives (University of Michigan).”

The “browser button” method automates several steps in the process and is therefore convenient when creating a large number of archival sites. The archivist first installs the “Add to WAS” button in the web browser’s toolbar bookmark menu (see figure 5).²⁰

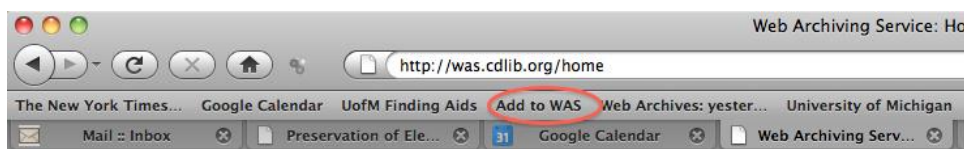


Figure 5

The archivist then selects the appropriate web archiving project in one browser tab, navigates to the target website in a separate tab, and clicks the “Add to WAS” button. A confirmation page then opens with information from the target site (see figure 6); the archivist merely clicks the “Add” button to include the site in the project.

²⁰ The WAS browser button is supported for Internet Explorer, Firefox, Chrome, and Safari web browsers.

Sites	Captures	Administration	Public Access
-------	----------	----------------	---------------

Adding: Home | President Mary Sue Coleman

The browser button allows you to quickly add name and URL information for a site.

The information shown below was automatically detected for the page you're adding. You'll be able to modify this information as you continue.

Site Name: Home | President Mary Sue Coleman

Seed URL: <http://www.umich.edu/pres/>

Where should this seed be added?

☒ To a newly created site

☐ To the site selected below

Academic Support Services Web Archives (University of Mic...

Add

Figure 6

The “Create Site” screen will then open with the site name and seed URL fields automatically completed with information drawn from the target page’s HTML header. The ‘site name’ generated in this step often requires revision because the title metadata is nonexistent, imprecise (the Department of Nuclear Engineering’s home page bore the title “Michigan Engineering | Letter from the Chair”), or unsuitable for archival description (i.e. “Home”). After supplying necessary title information, the archivist follows the same steps as in the manual process.

4. *The archivist sets the frequency of future crawls or creates a customized capture schedule.*

The WAS “Scheduling” tab (see Figure 7) provides various options to arrange for future crawls depending on how frequently the site in question needs to be captured.

Capture Settings | **Scheduling** | Descriptive Data

Capture Frequency: ☐ Off
☐ Daily End Date: December 24, 2010
☐ Weekly
☐ Monthly
☒ Custom

Day of the month: 1

Months to run:

- ☐ January
- ☐ February
- ☐ March
- ☒ April
- ☐ May
- ☐ June
- ☐ July
- ☐ August
- ☒ September
- ☐ October
- ☐ November
- ☐ December

Figure 7

Sites with rapidly changing content or related to time-sensitive events may require daily or weekly captures while other content may only need monthly (or even less frequent) captures.

After some experimentation, UARP decided that the majority of the university's sites would only be captured once a year with the exception of the President and Provost (quarterly), the athletic department (monthly), and course schedules (every semester). Should events or content require additional captures, archivists will adjust schedules accordingly.

5. *The archivist enters information related to the site's creation and subject matter to provide a descriptive context for end-users.*

The "Descriptive Data" tab allows archivists to manually enter information related to "Site Description," "Creator," "Publisher," "Subjects," and "Geographic coverage" (see Figure 8).

The screenshot shows a web archiving interface with three tabs: 'Capture Settings', 'Scheduling', and 'Descriptive Data'. The 'Descriptive Data' tab is active. It contains several text input fields:

- Site Description:** A large text area containing two paragraphs. The first paragraph states: "Website for the Office of the President. From the current site, it states "Mary Sue Coleman has led the University of Michigan since being appointed its 13th president in August 2002. As president, she has unveiled several major initiatives that will have an impact on future generations of students, the intellectual".
- Creator:** A text field containing "Office of the President".
- Publisher:** A text field containing "The Regents of the University of Michigan".
- Subjects:** A text field containing "University of Michigan--Administration".
- Geographic coverage:** A text field containing "Michigan--Ann Arbor".

At the bottom right of the form are two buttons: "Cancel" and "Save (all tabs)".

Figure 8

Although these metadata fields mirror elements in the Dublin Core Metadata Set, UARP needed to establish local definitions and conventions for their use. A series of internal discussions led archivists to adopt the following practices: “Site Description” would generally be taken from content found on the websites (such as an “About Us” page); the “Creator” would refer to the specific unit responsible for the website’s content; and “Geographic coverage” would identify where the activities described in the site took place (most often, Ann Arbor). Ongoing uncertainty about the nature and use of the “Subjects” element led archivists to refrain from entering information in this field during the initial phase of the project.

This workflow step was revised when WAS added the “Publisher” field in December 2010; UARP elected to use the Regents of the University of Michigan as the body ultimately responsible for the production and presentation of content. That same month, UARP learned that the “Subjects” element could be used in a manner analogous to MARC subject access fields; archivists thus returned to the sites in December 2010 and January 2011 to enter a single subject authority heading for each.²¹

6. *The archivist clicks the “Save (all tabs)” button to preserve the settings and metadata and then initiates the capture (see Figure 9).*

²¹ This large-scale editing of site metadata was conducted in concert with the review of sites after the consolidation of the University of Michigan Web Archives (to be discussed below). UARP only entered a single term because it was unclear if delimiters could be used to separate successive items. Likewise, MARC numbered fields, indicators, and subfield codes were not used because it was unclear if WAS would develop a system for the search and/or display of these terms.

Site Summary: President Web Archives (University of Michigan)

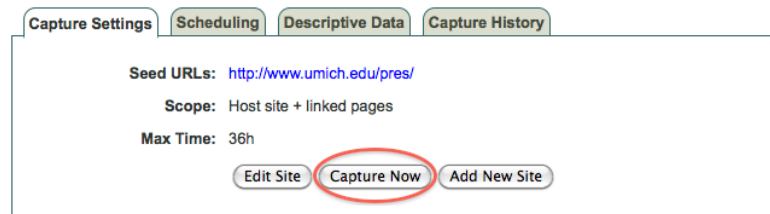


Figure 9

At this point, the web crawl has been initiated and its status is displayed in the “Manage Sites” screen of the curatorial interface (see Figure 10) so that archivists may track its progress or be notified of technical difficulties.



Figure 10

Over three months, UARP devoted the equivalent of two FTE archivists to the creation and description of archived websites. By early October 2010, Associate Archivist Nancy Deromedi and Assistant Archivist Michael Shallcross had created a total of 123 projects that contained 665 sites and amounted to 730 GB of preserved content.

Description of Archived Website Collections

While the identification and selection of seed URLs was time consuming, description proved to be a particularly labor intensive aspect of web archiving. In addition to the site descriptions, archivists created additional contextual information at the collection level via the WAS “Project Administration” interface. The primary means for describing a collection is the “Project Description” (see Figure 11), which UARP used to provide a general overview of the content creator and salient features of the archived websites.

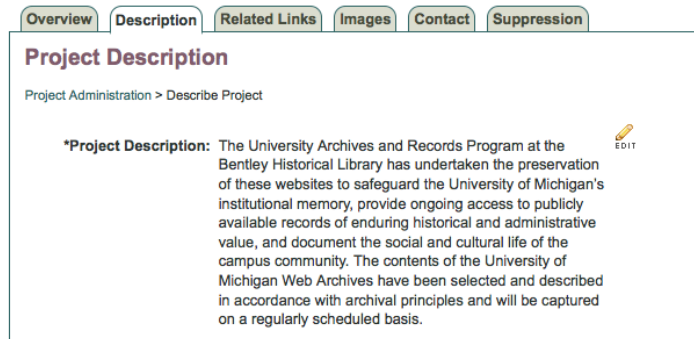


Figure 11

This information is then prominently displayed when users visit the collection's main landing page (see Figure 12).

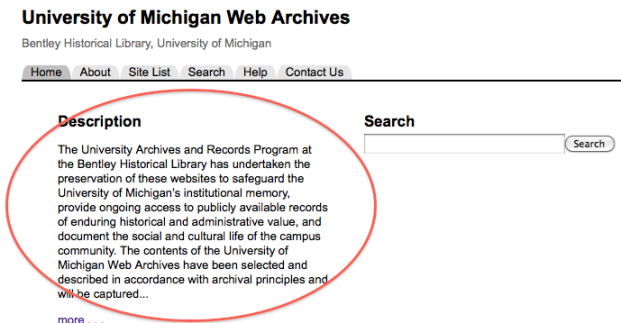


Figure 12

WAS also allows archivists to create links to related resources for each collection. To better integrate the web archives with existing record groups and manuscript collections, UARP inserted links to relevant UARP EAD finding aids into the collections' metadata (see Figure 13).

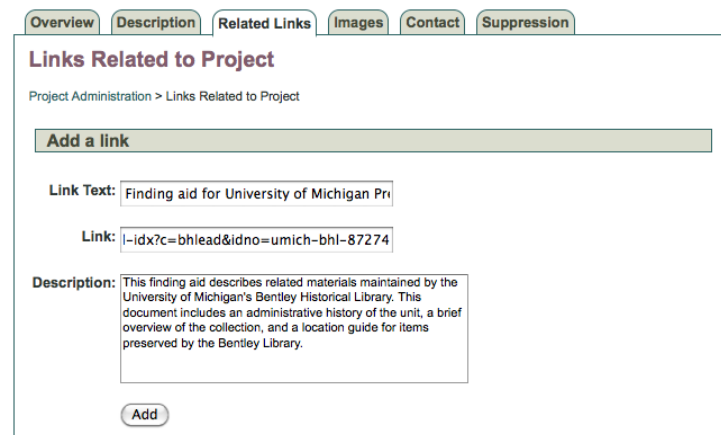


Figure 13

These "Related Resources" are viewable by users who click the "About" tab on the collection's main landing page (see Figure 14).

University of Michigan Web Archives

University of Michigan, Bentley Historical Library

[Home](#) [About](#) [Site List](#) [Search](#) [Help](#) [Contact Us](#)

Description

The University Archives and Records Program at the Bentley Historical Library has undertaken the preservation of these websites to safeguard the University of Michigan's institutional memory, provide ongoing access to publicly available records of enduring historical and administrative value, and document the social and cultural life of the campus community. The contents of the University of Michigan Web Archives have been selected and described in accordance with archival principles and will be captured on a regularly scheduled basis.

Related Resources

[An Introduction to the University of Michigan Web Archives](#)

Information on website preservation by the University Archives and Records Program (UARP) at the University of Michigan's Bentley Historical Library.

Figure 14

Archivists also planned to promote access to the web archives by inserting standardized language into UARP finding aids. Before this step could be taken, WAS introduced new features (due in part to the size and complexity of the University of Michigan Web Archives) that led UARP to revise its strategies for collection-level description.

Influence of the University of Michigan Web Archives on the California Digital Library's Web Archiving Service

By the fall of 2010, UARP had become one of the most prolific users of WAS outside of the University of California system. The University of Michigan Web Archives was among the largest bodies of content devoted to a single subject within WAS and was likely the most extensive collection—in any repository—of archived websites related to a single university.²² Although UARP had anticipated challenges as a result of creating and maintaining hundreds of individual projects, these issues had become more pronounced by December 2010. The large number of collections made it difficult to update and manage individual sites and archivists had to manage multiple spreadsheets to track blocked crawls, broken hypertext links, and technical problems within WAS.

The sheer bulk of the University of Michigan Web Archives, and the attendant challenges this size imposed upon UARP and WAS, hastened the California Digital Library's introduction of significant service upgrades. New features included improvements in the layout of the curatorial interface, expanded content management functionality (i.e. the ability to batch-reschedule crawl frequencies), and the introduction of tags to facilitate user navigation.²³ In preparing for the December 2010 release of the upgrades, WAS technicians used University of Michigan collections to test the new features, sought feedback from archivists during the beta-phase of development, and responded to various requests from UARP.

²² This statement is based upon a review of university-related collections in WAS, Archive-It, IIPC member archives, and general Internet searches. The author offers his sincerest apologies if a collection has been overlooked.

²³ As utilized by WAS (as well as countless blogs, Facebook, Flickr, etc.), a 'tag' is "a non-hierarchical keyword or term assigned to a piece of information [that] helps describe the item and allows it to be found again by browsing or searching." (Wikipedia. "Tag (metadata)" (February 13, 2011), accessed on 15 February 2011, http://en.wikipedia.org/wiki/Tag_%28metadata%29.)

The biggest change for UARP, however, came with the suggestion that the 123 projects in the University of Michigan Web Archives should be consolidated into a single collection. While the upgrade (and tagging in particular) obviated many of UARP's concerns about presenting the web captures in a single collection, the proposed consolidation would force archivists to reassess a number of key decisions and abandon work on collection descriptions and links to related resources. After extensive discussions among archivists and several teleconferences with WAS staff, UARP decided to proceed with the consolidation. WAS technicians automated the process and in one day (December 14, 2010) the University of Michigan Web Archives was reduced to a single collection of 531 sites. While a significant amount of duplicate content was deleted as part of the consolidation, archivists conducted a systematic reassessment of the collection to weed out inferior or redundant captures (i.e. captures that occurred only a month apart so that few if any changes were present). These actions further concentrated the size of the web archives to 426 GB—a reduction of nearly 300 GB. Archivists took advantage of this extensive review process to revise and complete the metadata of individual sites (especially in regards to the “Subjects” fields).

Future of the University of Michigan Web Archives

UARP continues to select sites for preservation and, as of February 14, 2011, the University of Michigan Web Archives is comprised of 607 archived websites that total 531 GB of data. The first installment of the collection was formally accessioned by UARP on February 1, 2011 and subsequent content will be accessioned annually on this date. The accession record included the overall size of the web archives and was accompanied by a spreadsheet listing all archived sites, the number of times each was captured, and the amount of data collected for each. UARP unveiled the web archives to the general public on February 28, 2011.

After the excitement of creating sites and implementing the WAS upgrades, the University of Michigan Web Archives has coalesced into a more stable form. In addition to collection development activities, archivists are now handling various ongoing issues that include description and access, content management, and intellectual property rights.

Description and Access

Given the broad range of the collection (which, two months after consolidation has grown by approximately 75 sites and 100 GB of data), UARP's strategies for description and access will be important to help patrons find relevant information. Although the collections will be indexed by Internet search engines, UARP recognizes that preserved websites will benefit from archival mediation and description as do other holdings at the Bentley Historical Library. After initial plans were revised due to the collection consolidation, UARP focused on three main strategies to describe the University of Michigan Web Archives: tags, finding aids, and catalog records.

- a. Tags were an important factor in UARP's content consolidation because they accomplish the original goal of multiple collection approach by identifying and aggregating related content. Tags will not bring users into the web archives in the

same way as ‘external’ access points such as finding aids or electronic catalog records. They will, however, be of the utmost importance as end-users navigate and retrieve content from the University of Michigan Web Archives.

WAS technicians made tagging an even more attractive option by automating the initial application of tags. Because tagging is only practical if a significant number of sites (i.e. 5 or more) are associated with a tag, UARP identified the most prominent groups of related content and devised tags of approximately 30 characters for each. As of February 2011, tags include the abbreviated names of the university’s 19 schools and colleges, “Administration,” “Athletics,” “Faculty,” “News & Events,” and “Museums & Cultural Attractions.” Archivists then entered the tags in a separate column on spreadsheet of the University of Michigan’s archived websites (many of which received multiple tags). During the consolidation process, WAS technicians simply imported the spreadsheet and the tags were automatically associated with the appropriate sites.

As a result of the procedure, all the sites that had previously been in the College of Engineering project were now assigned a “College of Engineering” tag. With the tags in place, end-users may refine the site list of the University of Michigan Web Archives to only browse content related to a specific topic (see Figure 15).



Figure 15

The introduction of tagging prompted UARP to add an additional step to the workflow for the creation and description of sites. After an archivist has saved the settings and metadata for a site, s/he has the opportunity to apply tags (see Figure 16).



Figure 16

Many sites are exempt from tagging because they do not fit into the established categories; for those that do, however, the archivist merely selects the appropriate tag(s) from a drop-down menu and clicks the “add” button. If the archivist captures a site that merits a new tag, s/he simply enters a 30-character name or phrase in the “Site Tags” field and clicks add. Given the ease with which tags may be created, the process must be carefully controlled to ensure that tags are accurate and succeed in helping patrons navigate the web archives. Even the slightest difference in a tag name (entering, for instance, “News and Events” instead of “News & Events”) will result in two content categories instead of one.

In introducing tagging functionality, WAS also rolled out several features to help administer tags. For instance, a batch-process option allows multiple sites to be tagged from the “Manage Sites” screen of the curatorial interface. The archivist simply marks a checkbox next to each site’s name, selects the appropriate tag, and clicks the “Add tag to selected” button (see Figure 17).



Figure 17

Archivists may also edit tag names or completely remove a tag from the collection via the “Manage Tags” screen (see Figure 18). All sites that are associated with the tag in question will automatically inherit the changes.

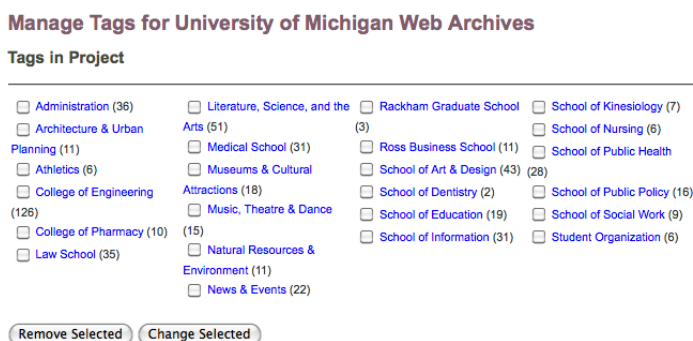


Figure 18

The true test of tagging’s efficacy will occur when they are used by patrons to navigate through the collection. Archivists look forward to working with the Bentley Library’s Access and Reference Services division to gather feedback from users and optimize the use of the tags.

- b. As mentioned above, UARP's strategies for the use of finding aids have evolved in the aftermath of the web archives' consolidation. As of February 2011, archivists have resolved the following issues:
1. Although the University of Michigan Web Archives is essentially a stand-alone collection, UARP will not create a paper-based or EAD finding aid for it in its entirety. Such an endeavor would be both time-consuming and impractical since the finding aid could only approximate the search and retrieval functionality built into the web archives. Users who access the public web archives are able to browse alphabetized lists of site names, perform keyword searches (across site content and URLs), and filter sites according to tags and content types. As such, it seems doubtful that a finding aid could provide additional functionality or utility.
 2. UARP will not update the finding aids for all record groups and manuscript collections that have related content in the University of Michigan Web Archives. The broad inclusion of this information might improve access, but the necessary time and resources required to emend the finding aids would likely outweigh any benefits.
 3. UARP will include standardized language in the finding aids of its most prominent record groups, those of the Board of Regents, President, Provost, and the 19 schools and colleges. The archived websites will be included as a new series (or, in some cases, as a continuation of an existing "Archived Website" series) and the date range will indicate that the captures commenced in 2010 and are ongoing. The EAD version of the finding aid will furthermore have a direct link to the persistent URL of the archived site. A high level description will then be included in the series' scope and content note to indicate the overall purpose and function of the archived site and note that captures will continue on a regular basis.²⁴

UARP formed a working group in January 2011 to examine the use of finding aids with digital records in general and this work may yield additional recommendations for the description of archived websites.

- c. MARC catalog records present another important means to describe and provide access to the web archives. UARP aims to create records for each site in the University of Michigan Library's online catalog (Mirlyn) that will include direct links to archived content. Because they are intended to give end-users a toehold to identify and access relevant University of Michigan content, the MARC records will only

²⁴ The Board of Regents EAD finding aid illustrates how the archived website series was added to an existing collection (<http://quod.lib.umich.edu/cgi/f/findaid/findaid-idx?c=bhlead&idno=umich-bhl-8722>).

contain the metadata entered by archivists (site name, creator, description, publisher, subject) as well as fixed-length data fields that remain to be determined. The record creation process will need to be automated due to the large number of preserved websites and so archivists are working closely with WAS technicians and administrators of the University of Michigan Library's Aleph ILS to accomplish this goal. Although details are still being resolved (as of February 2011), UARP expects that catalog record creation will be accomplished in the following steps:

1. WAS will provide a spreadsheet that lists all archived websites and important metadata (site name, archival URL, creator, subject, date of first capture, etc.).
2. UARP will convert the information on this spreadsheet to MARC format using MarcEdit, a freely available MARC editing application.²⁵
3. MLibrary will ingest and publish the MARC records in Mirlyn, the University of Michigan's online catalog.

Once the process is firmly established, archivists will receive a spreadsheet of uncataloged sites every six months to convert to MARC format and submit to the Michigan Library. UARP will also devise procedural safeguards to prevent the creation of duplicate catalog records.

Archivists have also constructed an access page on the Bentley Library homepage to orient users to the web archives and provide additional contextual information.²⁶ Although WAS allows its subscribers to include custom text and links to related resources on the web archives' "About" page²⁷, UARP wanted to develop a resource for the large number of patrons likely to discover the collection through the Bentley website. Archivists have therefore included a description of the project, various access methods (including a direct link to the collection, a full-text search box, and links to different archival subjects based upon tags), a preferred citation style, and a collection development policy. Because WAS indexes all archived web pages, users may conduct full-text and keyword searches on URLs, site names, and HTML, PDF, Word, Excel, Flash, and text files within the websites. Once patrons have gained access to the web archives itself, they may conduct full-text searches, browse a list of sites, view a specific subset of sites (based upon tags), and peruse descriptions and other metadata for individual sites (see Figure 19).

²⁵ MarcEdit is developed and supported by Terry Reese. Information and downloads are available at <http://people.oregonstate.edu/~reese/marcedit/html/index.php>.

²⁶ Please see <http://bentley.umich.edu/uarp/home/webarchives/index.php> to view UARP's web archives access page.

²⁷ See <http://webarchives.cdlib.org/a/universityofmichigan/about> for descriptive and contextual information on the University of Michigan Web Archives provided by UARP archivists.

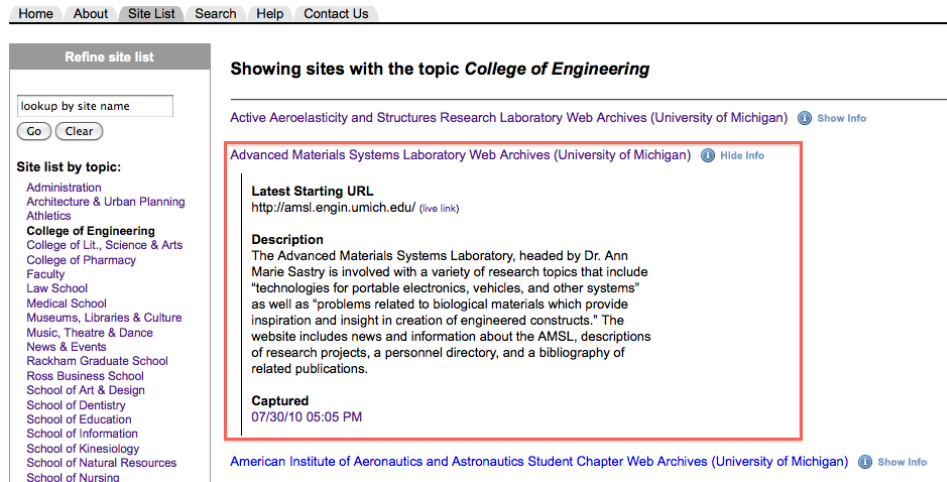


Figure 19

UARP is also interested in utilizing the unique permanent URLs associated with all content in the web archives (web pages as well as embedded PDFs, images, audio files, etc.). Archivists may be able to prepare subject guides for specific themes or content types (such as course catalogs, degree requirements, newsletters, etc.) and then link directly to relevant materials.

Content Management

By virtue of UARP's contract with the California Digital Library, WAS technicians are responsible for the secure storage of content and the performance of various digital preservation activities (such as integrity checks, data replication, and disaster recovery planning). UARP's immediate content management responsibilities will therefore relate to the arrangement and description of content (already discussed above), quality assurance, and the administration of new and previously scheduled crawls.

Quality assurance has proven to be difficult due to the complex nature of websites and the time it requires of archivists. One of the major problems involves the captures themselves: due to the intricacies of advanced web design and the limitations of web crawler technology, it is sometimes impossible for archivists to preserve the exact form, functionality, and content of websites as they are experienced on the 'live' web. Furthermore, even if the Heritrix web crawler has successfully captured a site, the Wayback Machine may be unable to properly render or display aspects of it for end-users. The following types of content are known to be particularly difficult to capture and/or display:

- Dynamic scripts or applications such as JavaScript and Adobe Flash
- Streaming media and embedded players with video or audio content
- Database-driven content


- Content that requires user interaction with the site (such as forms, dropdown menus, radio dials, password entry, Captcha authorization, etc.)
- Exclusions specified in robots.txt files

To verify if particular items have been preserved, archivists may review the Heritrix crawl log, a detailed report that records every item that the web crawler encounters in its operation. By consulting this log, archivists can see if specific objects were preserved, whether errors prevented particular captures, or if the crawler completely missed the objects in question. Even with this resource, UARP has had to contact WAS technicians to check whether or not important materials have been captured and preserved in the digital repository.

As indicated above, quality assurance for web captures is a highly labor-intensive process, especially with a collection as large as the University of Michigan Web Archives. As of February 2011, WAS is developing more detailed reports and tools to assess crawl results. In the absence of such resources, archivists have been forced to review sites one at a time to discover blocked crawls, dead or erroneous seed URLs, technical issues, and similar problems. When the number of archived sites was relatively small, archivists could devote time to these personal inspections. Now, with 607 sites and more to come, UARP reserves in-depth audits for high profile captures such as the Office of the President homepage. In the future, if UARP determines that additional quality assurance is necessary, the unit might explore having graduate student workers conduct more thorough reviews.

In the absence of more advanced tools and a larger workforce, UARP has established basic guidelines for the review of capture results. As an initial step, archivists check the “View Captures” screen of the WAS interface to see if the crawl was completed and if any problems may have arisen (see Figure 20).

View Captures

Click  to view the captures for a site.

1-1 of 1

display: 25 | 50 | 100




SITE NAME / CAPTURE DATE	STATUS	FILES	DURATION	ACTIONS
 President Web Archives (University of Michigan) (2)				Compare
01/06/11 02:58 PM Settings: Host site + linked pages, 36h Captured by: Michael Shallcross	Preserved	13460	8h 18m 47s	View Results 
08/12/10 10:35 AM Settings: Host site + linked pages, 36h Captured by: Nancy Deromedi	Preserved	13167	8h 14m 7s	View Results 

Figure 20

The “Status” column may indicate a variety of conditions: the crawl may be (a) ongoing, (b) paused or subject to technical difficulties, (c) completed and in the midst of processing, or (d) successfully preserved. If the capture has been preserved, archivists

check the “Files” and “Duration” columns to identify captures with relatively low or high figures for crawl duration or the number of files captured. An extremely short crawl or a scant number of files may indicate an error with the seed URL or the presence of robots.txt exclusions. Subsequent procedures are as follows:

- a. The archivist clicks the “View Results” link and verifies that the seed URL is accurate (see Figure 21).

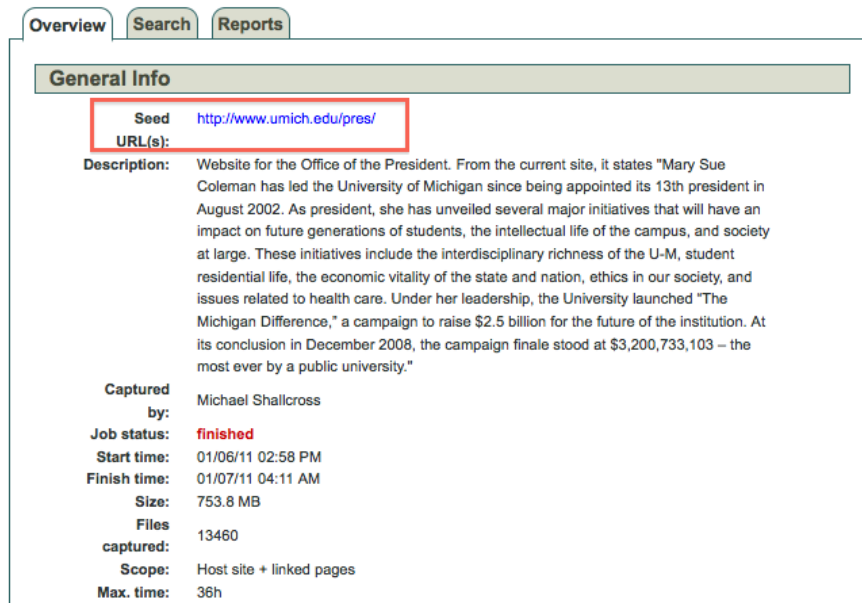


Figure 21

- b. The archivist checks the WAS crawl overview to see if robots.txt exclusions may have halted the Heritrix robot (see Figure 22).²⁸

²⁸ A “robots.txt” file is an Internet convention used by webmasters to prevent all or certain sections of websites from being crawled by a robot. The robots.txt must reside in the root of the site’s domain and its presence may be verified by typing ‘/robots.txt’ after the root URL (i.e. <http://umich.edu/robots.txt>). By convention, a web crawler or robot will read the robots.txt file of a target site before doing anything else. This text file will specify what sections of a site the robot is forbidden to crawl. A typical robots.txt exclusion statement is as follows:

User-agent: *

Disallow: /

‘User-agent’ refers to the crawler; ‘*’ (a wildcard symbol) indicates that the exclusion applies to *all* robots; and ‘/’ applies the exclusion to all pages on the site. Alternatively, a webmaster might exclude only certain directories (entering each one on a separate line) or open the whole site to a robot by leaving the field blank after “Disallow.”

Other Statistics

Robot Exclusions

The following robots.txt files were discovered on servers specified in your seed URL list. The robots.txt files are archived in order to document the host server policies on the date this job was run. Click the link to view the archived version of the robots.txt file.

- <http://www.umich.edu/robots.txt>

See [Robots Help](#) for more information.

Figure 22

- c. If the robots.txt file has blocked the crawler, the archivist contacts the webmaster to explain the purpose of the University of Michigan Web Archives and request an exception to the exclusion.
- d. Excessively long captures (in which the crawler runs the full 36 hours) may result in an incomplete site capture or indicate the presence of a ‘crawler trap.’²⁹ To resolve these issues, the archivist will:
 1. Check crawl reports to see how many of URLs of the target host were not captured.
 2. Verify the seed URL(s) and check the live site for content such as online calendars that could result in crawler traps.
 3. Enter only specific directories as seed URL(s).
 4. Prevent the crawler from following associated links.
 5. Limit the time allotted for the capture to “Brief (1 hour).”

Ongoing experience with WAS has allowed UARP to troubleshoot many of the problems that arise with crawls; at the same time, archivists occasionally must rely upon the technical expertise of WAS to resolve particularly vexing issues.

Moving forward, UARP will need to be aware of events, initiatives, or content that could impact the frequency of crawls. Archivists check university news releases on a daily basis to learn about new or previously unknown sites and to see if any news or developments require immediate preservation. UARP is also building relationships with webmasters and administrative assistants across campus so that archivists can be alerted when existing websites are changed or new ones are launched. UARP has yet to discover a strategy to identify broken or dead seed URLs for which crawls are scheduled, but archivists hope that new reporting tools under development at WAS may be of service.

²⁹ Such a trap is essentially an infinite loop from which a robot is unable to escape; online calendars are among the most common examples. The crawler will start with the present date and capture page after page of the calendar until the crawl expires without preserving more meaningful site content.

Intellectual Property Rights

Generally speaking, the success and legitimacy of any web archives depends on the archivists' ability to legally acquire content and respect the rights of content owners. Although "no provision of the Copyright Act expressly allows libraries and archives to capture publicly disseminated online content and create a permanent copy of it for their collections,"³⁰ UARP adheres to the Section 108 Study Group's recommendations for changes to the Copyright Act for website preservation.³¹ This group of copyright experts asserts that archives and libraries should have the right to capture "publicly available" content (i.e. materials that do not require a password, entry forms, or subscriptions) as well as websites related to federal, state, and local governments. UARP further acknowledges the rights of content owners (on its own and as a subscriber to WAS) with the following steps:

- a. The WAS web crawler respects all exclusions in robots.txt files and will not capture content designated as off-limits by a webmaster. (See note 23 for more information on robots.txt files.)
- b. WAS will stop a capture if it detects any degradation of service or negative impact on the host's web server.
- c. All archived materials will be prominently labeled as an "archived copy for study and research" to avoid confusion with the live websites (see Figure 23).



Figure 23

- d. Content owners may request that portions of their site be suppressed from public view and can choose to opt out entirely from captures.
- e. While the vast majority of UARP's archived websites are created and published by university units, archivists have captured the personal pages³² of faculty

³⁰ Section 108 Study Group. *The Section 108 Study Group Report* (March 2008) accessed on 14 February 2011, <http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>

³¹ The Study Group is named for the section of the Copyright Act that permits libraries and archives to reproduce and make use of copyrighted material to serve the public. A copy of its report may be found at <http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>.

³² These "personal" pages include privately developed and hosted websites (i.e. <http://johndoe.com>) as well as sites hosted by the university (i.e. <http://www-personal.umich.edu/~faculty/member>).

members in the School of Art & Design and future efforts will be made to preserve content related to other important professors. UARP will distribute communications to these faculty to explain the purpose of the University of Michigan Web Archives, inform them of their right to opt out or suppress content, and invite questions or concerns. UARP will also notify major academic and administrative units when their content is going to be publicly available.

Final Thoughts

The Bentley Historical Library's MeMail Project provides a four-pronged approach for the capture and preservation of select electronic records as they are produced and stored in various offices at the University of Michigan. This content includes email, records maintained in structured environments (such as SharePoint or other policy-driven content management systems), records maintained in unstructured environments (such as shared drives, desktops, and removable media), and publicly available material on websites. The University of Michigan Web Archives thus form an important part of UARP's strategy to document academic and administrative units in the present and for the foreseeable future.

Website preservation provides an exciting opportunity to manage a dynamic facet of the university's historical record and at the same time to unite researchers with important information. The development of the University of Michigan Web Archives has also helped UARP engage with other archives and establish itself as a resource in this growing field. Archivists recently provided training in website preservation to staff in the Bentley Library's Michigan Historical Collections (the state-wide collecting division), presented to students in the University of Michigan's School of Information, and provided consultation to the University of Michigan-Dearborn Archives. Requests for information and training materials for website preservation have been initiated by several peer institutions and archivists have been asked to participate in CDL webinars on best practices and procedures. While important challenges lie ahead—especially in terms of content description, resource tracking and management, and the promotion of user access—UARP welcomes the opportunity to contribute to the professional dialogue and share its experiences with others.