

**Our Vision** is to bring national recognition to the university for distinctive research collections documenting the diverse histories and development of San Antonio and South Texas.

**Our Mission** is to build, preserve, and provide access to distinctive archival, photographic, and printed materials, with a particular commitment to collections with significance to our region. We embrace the changing digital landscape by actively exploring new ways to enhance access to our collections. We support the university's ascent to Tier One status by building nationally recognized collections that inspire new knowledge, serving researchers at UTSA and from around the world.

### **Statement of Purpose**

In 2008 UTSA Special Collections began using the Internet Archive's web archiving service, Archive-It. Archive-It allows Special Collections staff to capture snapshots of relevant web content and ensure its long-term access through the Internet Archive's global web archive, the Wayback Machine. The archived web pages are then made publicly accessible on UTSA Special Collections' Archive-It partner page: <https://archive-it.org/organizations/318>

UTSA Special Collections is the official repository for records pertaining to the history of UTSA. As such, Special Collections uses Archive-It to preserve web content of enduring value to UTSA. This is the primary purpose of Special Collections' web archiving efforts, and UTSA-related crawls are given precedence as well as the majority of the data budget. Special Collections' secondary focus is to preserve web content of enduring value to the San Antonio and the South Texas region, in accordance with our other collecting priorities.

This Web Archives Policy was created to address the unique needs and circumstances of collecting web resources. This document is an extension of the University Archives Collecting Guidelines, the Manuscripts Collecting Policy, and the Rare Books Collection Development Policy. These broader policies should be consulted for more complete information on collecting areas, the history of UTSA Special Collections, the value of Special Collections to the University's Mission, the collection of government records, access and copyright restrictions, and user groups and use priorities.

This document supersedes the Web Archiving Methods and Collection Guidelines created in March 2016.

### **Selection and Acquisition**

The UTSA Special Collections Web Archiving Team will confer regarding selection of web resources and evaluate them based on the considerations outlined below.

#### **Inclusion Criteria**

- Materials must complement and/or supplement additional physical or digital archives held by UTSA Special Collections.

- The geographic focus should center on San Antonio and South Texas, but may expand to Texas statewide if deemed appropriate
- Materials must relate to the history, administration, or culture of UTSA (University Archives).
- Materials must align with the Primary Collecting Areas outlined in the Manuscripts Collection Development Policy.
- Materials must align with the Rare Books Collection Development Policy.
- Materials should support the research and instruction needs of the University.

### **Types of Web Content Generally Not Collected**

- Web resources created by private individuals for personal purposes;
- Password protected sites;
- Databases;
- Calendars;
- Non-UTSA web resources that have robots.txt exclusion requests; and
- Linked resources from utsa.edu that are not hosted by utsa.edu/not created by UTSA entities; the original URL of these will be captured in the crawl but Special Collections can currently only devote resources to UTSA-created content.

### **Resource Constraints**

The Web Archiving Team will consider the following resource constraints in its web archiving practice:

- The amount and cost of available storage space;
- Available staff time; and
- The department's ability to provide for the maintenance and preservation of, and access to, the materials.

Sampling may also be used to selectively acquire a manageable quantity of materials. These considerations will help ensure the sustainability of Special Collections' web archiving efforts.

### **Additional Considerations**

- The rapid extinction rate of digital content;
- Staff time involved in creating descriptive metadata for collections, groups and seeds;
- Usage statistics for related physical or digital collections;
- Ethics and privacy;
- The frequency of updates to the web resource;
- Web content that is currently captured by other institutions; and
- Content has been previously captured in crawls via links to existing seeds.

### **Reappraisal**

The Web Archiving Team may revisit earlier appraisal decisions and remove sites,\* discontinue future crawls, or reorganize/re-describe existing collections. The Web Archiving Team may perform any one or more of these actions when:

- The web resource/collection no longer aligns with the inclusion/exclusion criteria, resource constraints, and/or additional considerations outlined above;
- The value of the web resource/collection is ephemeral and limited to a specific time period (such as a one-time event, conference, or the work of a temporary committee);

- Usage statistics for the web resource/collection in Archive-It, and/or for related physical or digital collections show low usage (usage statistics will likely not factor into reappraisals of UTSA-related content);
- The targeted web resource has exhibited no or minimal change for one year (as verified by Archive-it curatorial tools and manual review by staff); and/or
- Archived versions of a web resource exhibit severe technical or display issues to the extent that patrons will not be able to use them for research. The Web Archiving Team may also continue to monitor the impacted web resource if it is associated with a significant organization, association, individual, or event.

\*Removal of sites from UTSA's Archive-It partner page does not remove the captured web resource from the Wayback Machine. UTSA Special Collections is unable to delete captured web resources from the Wayback Machine. For more information, please refer to the Takedown Policy section in this document.

## Methods

The Web Archiving Team:

- Identifies a topic/subject/theme for a collection (if creating a new collection);
- Selects relevant, specific web resources to crawl;
- Administers crawl mechanics and adds descriptions (metadata); and
- Determines the frequency of content changes/updates and sets crawl frequency accordingly.

Web Archiving Team discussion and approval is required for creation of new web collections/groups, seed selection, and for major changes to existing collections.

### Thematic Collecting

Generally, the Web Archiving Team creates web collections that address a specific theme or topic. Event-based crawls (related to a finite, time-sensitive event) may also be conducted at the seed or group level within a larger thematic collection.

### Capture Scope and Frequency

Crawls of web resources are based on lists of URLs (seed lists) developed by the Web Archiving Team. The Web Archiving Team will meet twice a year to determine crawls. University Archives web resources will generally be crawled at the conclusion of the fall and spring semesters. Non-UA resources will generally be crawled twice per year. Prior to crawls, the team will review and analyze collections/groups/seeds, previous crawl reports, and scoping rules. Data limits may be placed on crawls to stay within the relevancy of the collecting focus and data budget. Further considerations outlined in other sections of this document may also factor into capture decisions. Additional team meetings and event-based crawls may occur outside of this schedule, and will be conducted as the team determines necessary.

### Descriptive Metadata

The Web Archiving Team will create descriptive metadata using the DACS standard and an adaptation of Dublin Core's 15 standard fields. The adaptation was proposed by OCLC's Research Library Partnership Web Archiving Metadata Working Group. For more detail, see <https://www.oclc.org/content/dam/research/publications/2018/oclcresearch-wam-recommendations.pdf>

Metadata creators must be conscious of contextualization and biased language. To assist in metadata creation and in an effort to avoid bias, the Web Archiving Team will create guidelines and generate a list of subject headings which will be provided to metadata creators. Newly created metadata must be approved by a member of the Web Archiving Team prior to publication.

### **Ongoing Maintenance Activities**

As part of ongoing collection development activities, the Web Archiving Team will:

- Perform quality assurance (QA) on all captures to determine the success of capture and verify the accuracy of target URLs (this will be done immediately following the completed capture);
- Update and enhance site metadata to ensure the accuracy of descriptive information and facilitate the search and retrieval of content (this may be done in the weeks and months following the completed capture);
- Remove seeds from captures that have significant technical issues, as necessary.

### **Authenticity and Limitations**

All archived web resources will appear with a yellow header at the top of the page so as not to be confused with the current, live version of the resource. This header displays the following text:

“You are viewing an archived web page, collected at the request of University of Texas, San Antonio using Archive-It. This page was captured on [time/date], and is part of the University of Texas at San Antonio [collection name] collection. The information on this web page may be out of date. See all versions of this archived page.”

As a disclaimer, some archived websites contain internal links that are not captured by the Archive-It web crawler, and will divert to the live web. Similarly, other links may not have been included in the crawl, and will direct users to a web page with the message “Not in Archive.”

The Web Archiving Team performs quality control reviews of captures following the completion of crawls and strives to ensure that the crawled web resources closely resemble the original. However, the team may not review every seed captured due to the exceedingly time-consuming nature of seed-level quality control. Additionally, social media websites, calendars, JavaScript-based content, audio/video, and items that are protected by web authors using robots.txt files may not be adequately captured and therefore may not render exactly as they did in their original form on the live web. Therefore, Special Collections cannot guarantee fully functional or comprehensively complete web captures.

### **Intellectual Property**

Special Collections strives to respect the rights of content owners and to follow professional best practices for intellectual property rights management in web resource preservation. The Web Archiving Team seeks to follow the Section 108 Study Group’s recommendations for changes to the Copyright Act for web resource 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 passwords, entry forms, or subscriptions) and that all governmental web resources should be freely accessible to web crawlers.

The Web Archiving Team does not determine the copyright status of web resources. All intellectual property rights are retained by the legal copyright holders. If UTSA does not clearly hold the copyright to a web resource, we cannot grant or deny permission to use the material. Researchers are responsible for obtaining formal permission from the copyright holder to reuse or republish captured content.

Please visit the UTSA Special Collections website for more [information regarding copyright compliance](#).

## **Takedown Policy**

Special Collections acknowledges that organizations and individuals as content creators of web resources have agency over their born-digital content. If you believe we may have harvested your web content in error, or that maintaining your content in our web archive does not adequately reflect your organization, please contact us. Also, if you are the copyright owner of material found in our web archive and believe UTSA has used the work beyond fair use and without permission, we want to hear from you. Please email us at [specialcollections@utsa.edu](mailto:specialcollections@utsa.edu).

While we are able to remove captured web resources from UTSA's Archive-It partner page, we cannot remove these resources from the Wayback Machine. Removing a web resource from our partner page will diminish its discoverability, but we do not have the power to fully delete a captured website.

As a preemptive measure, the Web Archiving Team strives to capture content which was created for public consumption (e.g., a public official's Twitter feed, the website of a business, etc.). Content within official UTSA web resources (any content on utsa.edu or content created by UTSA for University purposes) is predominantly considered to be public record. In the rare event that we select web resources created by private individuals for personal purposes for capture, an opt-out notification will be distributed to groups or individuals prior to crawls.

## **Preservation**

Special Collections recognizes the research value of curated web collections and is committed to preserving select web archive crawl files (WARC files) and descriptive metadata for long term access.

## **Access**

Special Collections provides access to captured web resources via our partner page in Archive-It in order to facilitate private study, scholarship, or research. We reserve the right to restrict access to captured content, and may not make captured web resources available immediately after capture.

## **Related Documents**

Web Archiving LibGuide

University Archives Collecting Guidelines and UTSA Records Retention Schedule

[Manuscripts Collection Development Policy](#)

Rare Books Collection Development Policy

Photographs Collecting Policy

Digital Curation and Preservation Policy (forthcoming)

[UTSA Libraries Strategic Plan](#)

## **References**

The following documents were referenced during preparation of these guidelines:

[Montana State University Special Collections & Archives' Web Archiving Procedures](#)

[Purdue University Virginia Kelly Karnes Archives and Special Collections' Web Archive Collecting Policy](#)

[Seton Hall University Libraries Web Archive Collecting Policy](#)

[Texas A&M University Libraries Web Archiving Methods and Collection Guidelines](#)

[University of Michigan Bentley Historical Library Web Archives Collection Development Policy](#)