

University of Winnipeg Archives
Web Archiving Policy

Web Archiving Policy	
Approved By	Librarians Committee
Responsible Executive Officer	University of Winnipeg Archivist and Digital Curator
Effective Date	12 December 2019
Review Frequency	Every 5 years, or as needed
Review Date	12 December 2024
Applies To	All staff, students, volunteers, and external parties

Purpose

- 1.1 The purpose of this document is to articulate the Web Archiving Policy at the University of Winnipeg Archives, henceforth known as the Archives, within the context of its collection and archives policies, as well as in accordance with principles and techniques of non-intrusive harvesting.

Legal Authority

- 2.1 The Freedom of Information and Protection of Privacy Act C.C.S.M. c. F175; Canadian Copyright Act R.S.C., 1985, c. C-42

Responsibility

- 3.1 The Archives is responsible for the development, administration, and review of this Policy.
- 3.2 This Policy ensures that the Archives respects the intellectual property rights of others. The Archives reserves the right to remove any material that, in our reasonable opinion, may violate copyright or other intellectual property rights. Copyright holders, including third-party copyright holders, who believe their rights have been infringed upon may contact us at archives@uwinnipeg.ca.

Key Definitions

- 4.1 The following definitions apply to terms as they are used in this Policy:

“Access”: The ability to access web content directly through online portals.

“Archive”: The process of copying digital information into a repository for storage, preservation, and access purposes. Often synonymous with “capture”.

“Collection”: A group of archived web documents curated around a common theme, topic, or domain.

“Crawler”: A computer program that explores the web and collects data about its contents. A crawler can also be configured to capture web-based resources. It

starts a capture process from a seed list of entry-point URLs (EPUs)

“Curator”: Anyone responsible for building a collection or collections of web-based resources.

“Dynamic”: A description of web-based content created automatically by software at the web server end. May be (a) personalized for the user based on identification via login or based on cookies stored on the user's computer, (b) tailored to fulfill a specific request made by the user, or (c) code-generated (e.g., using php, jsp, asp, or xml). Information used for personalization or tailoring of pages may be retrieved in real-time from a database or other data store.

“Preserve”: The professional discipline of protecting materials against harm, injury, decay, or destruction to minimize information loss and extend the life of the materials in analogue and digital forms.

“robots.txt”: Files that a site owner can add to their site to keep crawlers from accessing all or parts of it.

“Web archive”: A collection of web-published materials that an institution has either made arrangements for or has accepted long-term responsibility for preservation and access in keeping with an archive's user access policies. Some of these materials may also exist in other forms but the web archive captures the web versions for posterity.

“Web content”: The textual, visual, or aural content that is encountered as part of the user experience on websites. It may include text, images, sounds, videos, and animations.

Collection Scope

- 5.1 The Archives collects and preserves web content of interest to students, researchers, academics, and scholars at the University of Winnipeg and for the general public. The Archives endeavours to acquire quality materials that are appropriate in meeting the needs of the wide range of subject areas taught and researched at the University.
- 5.2 We do not collect web content that requires a password to access except with the express permission of the website owner.

Access

- 6.1 Where appropriate and feasible, and where not restricted at the request of a website owner, the contents of the web archive will be publicly available in the University of Winnipeg's Archive-It collections and permanently added to Archive-It's Wayback Machine.

Rights

- 7.1 The electronic resources are made available by the University of Winnipeg Library for the purposes of research, education, teaching, and private study. All intellectual property rights are retained by the legal copyright holders. The University of Winnipeg does not hold the copyright for all of the content. Formal permission to reuse or republish this content must be obtained from the copyright holder.

Liability

- 8.1 We endeavour to maintain the integrity of the web archive by capturing and presenting the information as it was originally made available to the public. We will not be held liable for the conduct or actions of site owners or those who access and use the web archive. We do not assume responsibility for verifying the accuracy of the archived content nor are we responsible for monitoring the archived sites for copyright compliance. A site owner is responsible for the content they communicate to the public by telecommunication and is solely responsible for securing the necessary permissions and releases when required by law. Those who access and use the web archive do so at their own risk.
- 8.2 Please note that, although we will do our best, the Archives may not be able to preserve the exact form, functionality, and content of sites as they appear on the live web. The following types of content present significant issues for capture and/or display:
 - i. Dynamic scripts or applications such as JavaScript or Adobe Flash;
 - ii. Streaming media players with video or audio content;
 - iii. Password protected material;
 - iv. Forms or database-driven content that requires interaction with the site;
 - v. Exclusions specified in robots.txt files.

Implementation of robots.txt

- 9.1 Website creators can block web crawlers by using robots.txt file script at the top level of the site. Robots.txt is the most widely used method for controlling the behaviour of automated robots on a site (all major robots, including those of Google, Alta Vista, etc. respect these exclusions). It can be used to block access to the whole domain or any file or directory within.

If a website creator would like to remove a site from Wayback Machine: institute robots.txt and submit the site to:

http://pages.alexa.com/help/webmasters/index.html#crawl_site

Opting-Out and Take-Down

- 10.1 When a site owner publishes their work on the Internet without technological restrictions, such as implementing robots.txt, the Archives view this as their implicit consent to the indexing and caching of their site.
- 10.2 At the request of the website owner, the Archives will evaluate requests for the removal of content from the web archive for the full term of copyright, unless a shorter term is agreed upon. The website owner has the authority to make take-down requests. Website owners can contact archives@uwinnipeg.ca for the evaluation process to begin.
- 10.3 Withdrawn sites are hidden from the University of Winnipeg Library's Web Archives Collections. The University of Winnipeg cannot remove withdrawn sites from Archive-It's Wayback Machine. See 9.1 for details.
- 10.4 If you are a copyright holder and believe your intellectual property has been included on a webpage without your consent, please contact the website owners. The University of Winnipeg cannot respond to third-party requests to remove website content. Only the website owner can modify their web content.

Procedures

- 11.1 The Archives will:
 - i. Identify, appraise, select, and manage websites that reflect the collecting interests of the Archives as outlined in 5.1, Collection Scope;
 - ii. Provide descriptions and contextual information for the material and mediate access (via metadata, catalogue records, and an access interface) to facilitate research and retrieval of content;
 - iii. Respect the intellectual property rights of owners.

Relevant Dates

- 12.1 This Policy shall be reviewed by the University Archivist at least once every five years or more frequently as required to reflect changes in legislation or to related University Policies and Procedures.

Approved: 12 December 2019

Revised: _____

Scheduled Review: 12 December 2024