

## Western Libraries Web Archiving Policy

### I. Purpose

Western Libraries collects and preserves born-digital records and publications pertaining to our collection areas. Our web archiving efforts began in 2018 by capturing websites and born digital records and publications created by scholars at Western University and by organizations in the local community. Archives and Special Collections is responsible for managing Western Libraries' web archiving initiatives.

### II. Collection Scope

We collect and archive web content that supports the work of both current and future teaching, research and scholarship needs at Western University and beyond. We endeavor to acquire quality web content that is appropriate in meeting the needs of the wide range of subject areas taught and researched at Western University.

#### *i. Priorities*

The Web Harvesting and Archiving Team will use the following priorities for evaluating potential web archiving collections and projects:

- Complement existing collecting strengths
  - Web archiving is part of a broader collection strategy and the development of new collections should complement and be informed by our stated collection development priorities, the research we support, our institutional history, and our geography.
- Capture at-risk content
  - Although all web content is to some extent at-risk, some web content is more at risk than others, such as time-limited events (for example: elections, disasters, and protests, among others).
- Respect limited resources
  - We must work within the resources available, particularly storage space.
- Consider what others are collecting to prevent duplication
  - We must consult other web archive portals and the Canadian Web Archiving Coalition to prevent the duplication of work and to consider how we can collaborate with other institutions in our harvesting efforts.
- Consider access
  - We must have a plan in place to enable access and discovery to web archived content. Currently, items are hosted on the Archive-it website (<https://archive-it.org/organizations/1273>). We can consult with CMDA and Western Archives to consider cataloguing and description, access, and possibilities for preservation, such as downloading the Warc files to OCUL's Permafrost.
- Consider on-demand requests
  - Requests for on-demand web harvesting will be handled on a case-by-case basis.

#### *ii. Acquisition Method*

Western Libraries' web collections are harvested, archived, and accessed through Archive-it, a web-archiving service provided by the Internet Archive. A web crawl captures a snapshot of a web page and stores a copy in the Internet Archive. This material can be accessed through Archive-It and the WayBack Machine.

#### *iii. Frequency of Crawls*

The frequency of web crawls depends on the frequency and relevancy of updates to the web pages. Web pages that change frequently are crawled more often. Web crawls can be scheduled yearly, quarterly, semiannually, weekly, or one-time depending on the web page.

#### *iv. Scope and Limitation of Crawls*

File formats and types captured include but are not limited to HTML, CSS, JavaScript, PDFs, images: jpegs, png, and videos mpgs, mp3s etc. 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. If warranted WHAT may ask permission of the rights holder to harvest particular sites, or portions of sites.

### **III. Access**

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. Archiving the content without restrictions will ensure long-term access by faculty, staff, students and researchers at Western University, as well as researchers and scholars outside the institution.

### **IV. Ownership**

We do not assert ownership rights over the intellectual property of the contents included in the web archive collection. All rights of ownership remain with the owner(s) identified on the website for the full term of copyright. We are not involved in the creation of the websites and have no oversight for the contents the site owners have communicated to the public by telecommunication. We assume no responsibility for the accuracy or lawfulness of the websites or the contents within.

### **V. Authorization**

All requests to reproduce and use the archived content must be sent to the website owner directly. We cannot authorize use of the material nor will we act as intermediaries to the transaction. Users are responsible for identifying the copyright status of the website's contents, as well as identifying and contacting the appropriate authority for permission.

When reproducing and using content from the web archive, we encourage users to review the website's terms of use, as well as inform themselves of the relevant laws applicable to their country. In doing so, we wish to establish a safeguard to ensure users are made aware of their obligation to deal fairly with the content.

## **VI. Notification, Take-Down, and Opting-Out**

When a website owner authorizes communication of their work to the public by telecommunication without technological restrictions, we view this as their implicit consent to the indexing and caching of their website. Where a website uses technological protection measures to restrict crawling technology, we will not harvest the content without providing notification and/or securing permission.

At the request of a website owner, we 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. Any third parties wishing removal of content must make their request via the website owner. Take-down requests received from third parties will not be honored as only the website owner can authorize its modification.

## **VII. Liability**

We endeavor 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 website 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 websites for copyright compliance. A website 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.

Adapted from University of Alberta Libraries Web Archiving Policy, University of Wisconsin-Madison Archives Web Archiving Collection Policy, University of Toronto's Web Archives Collection Development Policy, Stanford Libraries' Web Archiving Policy, and Library of Congress Collections Policy Statements.