Web Archives Guidelines

1.0 Web Archives Collection Guideline

Introduction

This document outlines the types of websites identified as priorities for collection and preservation, as well as the conditions under which Queen's University Library and Archives will provide access to and manage collected websites.

Scope

Queen's University Library and Archives will capture websites that falls within the following areas:

- University administration and services
- Faculties, Departments, Schools, Research Centres and Institutes
- Student Organizations and Other University-affiliated groups
- University publications
- Private records (as part of fonds acquired in accordance with the Archives' Collection Mandate)

On a case-by-case basis, Queen's University Library and Archives will also capture faculty or researcher websites, comprising an online record of research output or data that cannot otherwise be captured through another Library service, such as Research Data Management.

Queen's University Library and Archives may also negotiate the capture of content in support of the research needs of Queen's University faculty or the history of Kingston and the region, where resources allow.

Access

Where appropriate and feasible, and where not restricted at the request of a website owner, the contents of the websites that are collected will be publicly available. Maintaining the content without restrictions will ensure access by faculty, staff, students and researchers at the Queen's University, as well as researchers and scholars outside the institution.

Ownership

Queen's University Library and Archives respects the intellectual property rights and the proprietary rights of others. Copyright ownership remains with the owner(s) identified on a website and governed by local, national, and/or international laws and regulations. The library assumes no responsibility for the accuracy or lawfulness of the websites or the contents within.

Authorization and Copyright

Queen's University Library and Archives collects this material for research and private study. When reproducing and using content from the website collection, we encourage users to review the individual websites' terms of use. Queen's University users should also refer to the Queen's University Library Copyright & Fair Dealing webpage.

Notification, Take-down, and Opting Out

When a site 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 site. Where a site uses technological protection measures to restrict crawling technology, such as a robots.txt exclusion, we will not harvest the content without providing notification and/or securing permission.

If you are a copyright owner or otherwise have exclusive control over materials presently available through this collection, and do not wish your work(s) or website(s) to be available through this website collection, please let us know. Any third parties wishing removal of content must make their request via the collected website's owner.

Prior to collecting a website, we will inform the site owner of our intent to do so and will provide the site owner with a copy of this guideline.

Liability

Queen's University Library and Archives endeavors to maintain the integrity of the website collection 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 website collection. We do not assume responsibility for verifying the accuracy of the collected content nor are we responsible for monitoring the collected 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 website collection do so at their own risk.

2.0 Web Archives Preservation Action Plan

Introduction

This document describes the preservation plan for web archives content in the Queen's University Digital Library. Most of the web archive content is Queen's University web sites. The preservation plan for web archives content follows from policies and practices described in the Digital Preservation Strategic Plan and the Digital Preservation Implementation Plan. This document explains practical steps that Queen's University Library and Archives takes to preserve the intellectual content of web archives in digital format. It outlines the basic tools, methods, and standards used for the long-term preservation of web archive content.

Content Formats

For the preservation of web archive content, Queen's University Library and Archives require WARC versions of the content, descriptive metadata, and occasionally, screen captures of the archived site. During the ingest process, derivatives are created for full-text searching of web archives. Queen's University Library and Archives continuously monitors developments in file formats to determine if and when formats require migration (see Environmental Monitoring of Preservation Formats).

SIP Format

Queen's University Library and Archives will create their own web archive SIPs (see Definition of SIP) in accordance with Queen's Archives' Digital Preservation procedures.

Analysis on Ingest

Upon ingest, every file in the repository is subject to identification of its file format and validation using FITS. Queen's University Library and Archives records the output of the FITS identification and validation processes to a technical metadata datastream (TECHMD_FITS) that is associated with the object in the repository.

Content Excluded

Queen's University Library and Archives do not ingest files that are not referenced (either as part of a representation or as associated datastreams) in the associated metadata. As the SIP is retained, these files can later be ingested if necessary.

Format Normalization

There is no format normalization if the submitted object is a WARC.

Metadata Normalization

When necessary, Queen's University Library and Archives crosswalk descriptive metadata from MODS to Dublin Core. The repository creates preservation metadata for each file. The preservation level, explained in the Digital Preservation Implementation Plan, is applied to each file upon ingest and recorded in the preservation metadata for each file.

Acceptable Formats

For the Full Preservation level for web archives, currently the only acceptable format is WARC.

Acknowledgements

Adapted from and inspired by: OCUL's Scholars Portal "Collection Policy".

3.0 Web Archives Access Plan

Introduction

This document outlines how Queen's University Library and Archives will provide access to collected web archives.

Storage

Queen's University Library and Archives uses the [Internet Archive's](#) [Archive-It](#) service to collect and store web archives.

Queen's University Library and Archives will also transfer a subset of web archive WARC files to the Queen's University Library Digital Asset Management system or Ontario Library Research Cloud for further preservation management and action in accordance with the Web Archives Preservation Action Plan.

Access Points

Researchers can browse or search the web archives using the Archive-It search interface at the [Queen's University Archive-It page](#).

Queen's University Library and Archives will endeavour to provide further access by linking captured websites to associated archival descriptions stored in [AtoM](#), and by linking the entirety of the web archives collection to the federated Library search interface.

Metadata

To facilitate searching, appropriate Web Archives team members will add Dublin Core metadata to each seed and collection within the web archive.