University of British Columbia Library
Web Archiving FAQ

General

Why is The University of British Columbia Library archiving websites?

The UBC Library advances research, learning and teaching excellence by connecting communities, within and beyond the University, to the world's knowledge. The purpose of the University of British Columbia Web Archiving initiative is to ensure that web content that contributes to that mission is preserved and accessible over time. This includes online content that may constitute the institution's corporate memory.

I was contacted via e-mail by The University of British Columbia Library about the archiving of my site. Why?

A reasonable effort is made to notify content owners of the inclusion of their websites in the University of British Columbia Web Archives. Your website was selected because it was considered to have material that contributes to research, learning, and teaching at our institution. For more information about the collection selection criteria, please see our Web Archiving Collections Development Policy.

Who do I contact at the University of British Columbia Library regarding web archiving?

Web archiving is managed out of the Digital Initiatives Unit; specific web archiving questions and project proposals can be directed to digitization.centre@ubc.ca.

Content and Access

What kind of websites are within the scope of The University of British Columbia Library's Web Archives Collection?

We aim to archive content that is considered in the interest of the University; websites that contribute to its research, learning, and teaching, as well as online content that may constitute the institution's corporate memory.

The following types of websites are considered within the scope of our collection:

- Research, public or governmental interest
- Historical or geographically local significance
- Complementary to relevant existing collections
- Content produced by the university or affiliated organizations

Priority will be given to websites at risk of disappearance, those with unique born digital content, and frequently updated websites. For more information about the collection selection criteria, please see our Web Archiving Collections Development Policy.

I don't want my website available at the University of British Columbia Web Archives. Can you exclude my website from your collection?

If you would like to complete a takedown requisition, please complete our Request for takedown of online materials form. Please note that UBC is able to remove websites from the UBC Web Archives Collections only; this will not remove your website from the Wayback Machine's archived collections.

How can I suggest a website for the University of British Columbia Web Archives Collection?

Suggestions from the academic community and website owners are welcomed. To propose a website, please see our guidelines for proposals.

Who can access the University of British Columbia Web Archives?

The University of British Columbia Web Archives Collections are publicly available to all users with a connection to the internet. There is no plan to collect restricted or private information during website crawls. Collections are automatically discoverable through the Archive-it site as they are indexed as part of the Archive-it subscription. They may be browsed, searched, and accessed through:

- The University of British Columbia Library's Archive-it collection page: https://archive-it.org/organizations/734.
- The Wayback Machine website: https://archive.org/web/.

Do you own the copyright of the material collected to The University of British Columbia Web Archives?

The University of British Columbia Library does not assert ownership rights over the intellectual property of the contents included in the web archive collection. 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.

Can I use material from a website in your collection in my work?

Following the "Fair Dealing Exception" under the Copyright Act, if deemed fair, use is allowed for purposes of research, private study, education, satire, parody, criticism, review, or news reporting without the copyright holder's permission. For more information, please refer to the Copyright Guidelines of University of British Columbia, available at: http://copyright.ubc.ca/guidelines-and-resources/copyright-guidelines/.

How frequently will my website be crawled?

The default frequency of capture will be one time only, except for websites that are being updated on a regular basis. In those cases the frequency will be defined on a case-by-case basis.

Will you eventually stop crawling my website at regular time intervals?

Reappraisals will be performed observing the current collection development policy and captures of a website may be discontinued if: the website is no longer valuable for the Library's mission and its user's demands; the value of the website is limited to a specific time period (e.g. a temporary one-time event); the website did not suffer any significant changes for three consecutive years; or archived versions of a website are exhibiting severe technical issues impeding proper access to their content.

Technical aspects

What technology does the University of British Columbia use for harvesting websites?

The University of British Columbia Library subscribes to Archive-it, a subscription service from the Internet Archive, which allows institutions to build, manage and search their own web archive. For more information about Archive-it please visit https://archive-it.org/learn-more/.

Will the archived version of my website look exactly like the live one?

In order to maintain the "look and feel" of the online content harvested, images, audio, and video files will be captured along with text. However, the capture of the following types of content may be restricted due to crawling technology limitations with such formats as javascript files, streaming and downloadable media, database driven content, password protected content, etc. For more information, please visit: https://webarchive.jira.com/wiki/display/AITH/5+Challenges+of+Web+Archiving.

What if I have robots.txt exclusions in my website?

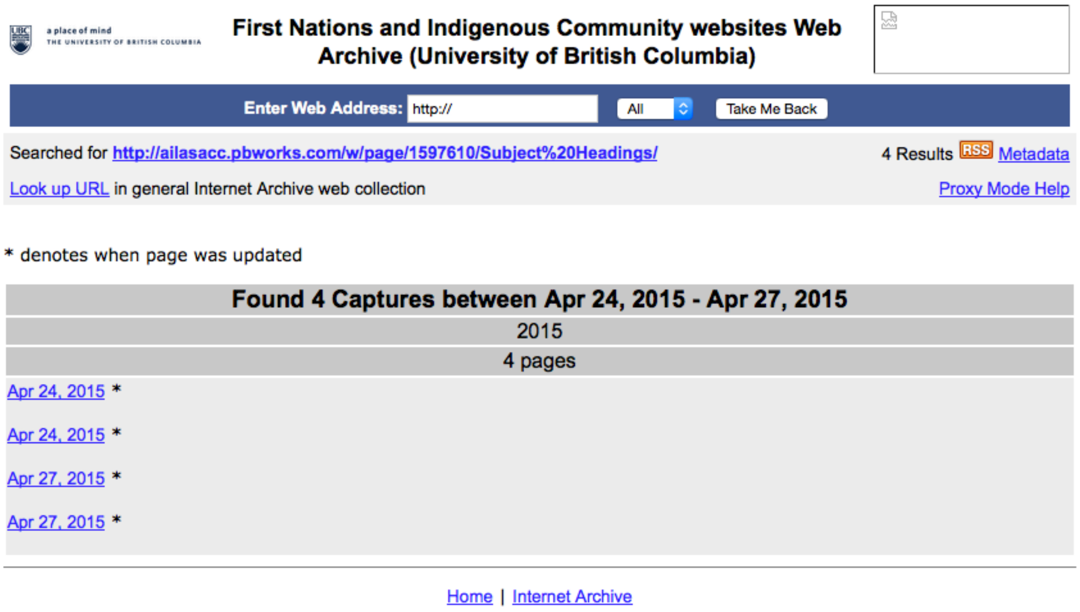
Observing the recommendations of the Copyright Act (http://laws-lois.justice.gc.ca/eng/acts/C-42/FullText.html) regarding digital locks, website-specific directives preventing archiving will be obeyed, whether expressed in machine-readable format using the robots.txt exclusion standard, password protection or in reasonably discoverable human-readable text. In these cases, crawling will only proceed if permission from the content owner is obtained.

How do I know if I am seeing the archived version or the live version of a website?

Archived web pages will feature a banner at the top of the page indicating the date and time the content was captured, as well the collection and institution responsible for its archiving.

What does it mean when there's an asterisk (*) next to a capture date on the Wayback Machine page?

In order to keep up-to-date versions on web pages, we harvest content at regular time intervals where appropriate. The asterisk (see example below) indicates that the content captured on that date has been updated from the previously archived copy. If there is no asterisk, the content on the archived page in that crawl is identical to the previously archived copy.


a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA
Enter Web Address: http://
All
Take Me Back
RSS

First Nations and Indigenous Community websites Web Archive (University of British Columbia)

Searched for http://ailasacc.pbworks.com/w/page/1597610/Subject%20Headings/

4 Results Metadata

Look up URL in general Internet Archive web collection

Proxy Mode Help

* denotes when page was updated

Found 4 Captures between Apr 24, 2015 - Apr 27, 2015

2015

4 pages

Apr 24, 2015 *

Apr 24, 2015 *

Apr 27, 2015 *

Apr 27, 2015 *

Home | Internet Archive

Errors and troubleshooting

Why can’t I see the images on a site?

If you see a small red “x” instead of the image, it means that it was not captured due to technical issues. Grayed out images means that their capture was blocked by a robots.txt exclusion set up by the content owner.

I got an error message, what does it means?

Below is a list of common error messages you may see while navigating an archived website:

- **Not in Archive:** The page you are looking for was not archived in or collection.
- **Robots.txt Query Exclusion:** A robots.txt file exclusion put on a site by its owner is preventing crawling. We respect robots.txt files and will not crawl such pages without permission.
- **Failed Connection:** The server where the information is stored is down. This is usually a temporary error.