Snapshot of Web harvesting in Canada

Joanne Paterson, University of Western Ontario
Has this ever happened to you?

Ahhhhhhhhhhhh! This page doesn't exist
Not to worry. You can either head back to our homepage, or sit there and listen to a goat scream like a human.
404 Page not found
A ninja stole this page.

You must return when the moon has friends and the fox is borrowed.
What if the website was of national significance?

www.ararcommission.ca
www.ararcommission.ca today
Hello Folks!

Our lovely government is at it again apparently.

The current victim? The Maher Arar Commission website.

It’s nowhere to be found! It used to be here:

http://www.ararcommission.ca but alas, in attempting to answer a reference question this morning, I find the website is gone. At least we still have: http://www.maherarar.ca, but can we say with certainty that all of the previous website’s contents are available there? I’d like to know whether the Commission’s contents have been captured by the National Library. Isn’t it legally obligated to be there? I certainly hope so...
Commission of Inquiry into the Actions of Canadian Officials in Relation to Maher Arar

Commission d'enquête sur les actions des responsables canadiens relativement à Maher Arar

Arar Commission

About the Inquiry

The Inquiry was established February 5, 2004 under Part I of the Inquiries Act, on the recommendation of the Deputy Prime Minister and Minister of Public Safety and Emergency Preparedness to investigate and report on the actions of Canadian officials in relation to Maher Arar (Factual Inquiry). The Commission is also mandated to recommend an arm's length review mechanism for the activities of the Royal Canadian Mounted Police with respect to national security (Policy Review).

The Honourable Dennis R. O'Connor, Associate Chief Justice of Ontario has been appointed Commissioner of the Inquiry.
Web Harvesting

Involves using software that captures and copies web information, allowing that information to be displayed in context and act as if it were a live website.

Web Archiving is the process of selecting, harvesting, preserving and offering access to captured websites.
LAC Government of Canada domain harvest *.gc.ca began in 2005

- LAC makes this harvested data openly accessible via the *Government of Canada Web Archive* (GCWA, 2006)
  - Decommissioned federal websites emergency-harvested between domain crawl periods
    - ~170 major departmental websites since 2009
  - Thematic collections from the open Canadian web
    - ~15, built with increasingly complex collaborative relationships
History of the RCMP

Born out of a need for a national police force to implement the law in Canada's newly acquired western territories, the Royal Canadian Mounted Police has...
Government of Canada Web Archive

Origins of the RCMP

Sir John A. Macdonald | The N.W.M.P. | The March West | The American
Sioux | Settlement | Policing the Railway | Servants of the Public | The North
West Rebellion | Free Land | Klondike Gold Rush | North to the Arctic Shores
| The Royal Canadian Mounted Police

Visit our new RCMP History pages for interesting facts about

Historical Notes, Uniform and Dress, Games...
HARVEST

ALL THE WEBSITES
Admissions and Student Resources
Campus Information, News and Events
Academic Programs, Faculties and Colleges
Research and Scholarship
Libraries and Information Resources
Administration and Services
Alumni and Friends
Western in the Community
LAC Thematic / Selective Harvest

• Federal and Provincial Elections
• Royal Commissions & Commissions of Inquiry
• Canada’s Participation in Olympic & Paralympic Games
• State Funerals
• Transitions in Federal Organizations
• Decommissioned Federal Organization/Websites
• Visits to Canada from British Royals
• Change of Governors General
• 100th Anniversary of the Calgary Stampede
• Commemoration of the War of 1812
Collaboration is key

• With LAC, two major stakeholders, the University of Alberta and U of T, run their own harvesting programs using Archive-It
• Curation, Selection, Creating seed lists, Quality assurance all take thoughtful human intervention, time-consuming.
Welcome to Archive-It!
Attend a live informational webinar and demo to learn more about the service.

Explore Collections

Wikileaks Document Release Collection
By Internet Archive Global Events
A collection of websites, news coverage, and commentary surrounding the Wikileaks releases.

Jasmine Revolution - Tunisia 2011
By Internet Archive Global Events
A collection of websites, news coverage, and commentary surrounding the 2011 Jasmine Revolution.

North Africa & the Middle East 2011
By Internet Archive Global Events
A collection of websites, news coverage, and commentary. Includes the most recent events in Libya, Egypt.
Archiving the web

The University of Toronto Libraries subscribes to Archive-It, a service created by the Internet Archive. Archive-It allows individual institutions to create small, focused web collections, and to harvest data from websites at regular intervals.

We have created four distinct web archives called "collections". All of these collections are full-text searchable and available to interested researchers.

- **Canadian Government Information**
- **Canadian Labour Unions**
- **Canadian Political Parties & Political Interest Groups**
- **University of Toronto Web Archives**
Gazette Archives

The York University Gazette was published on campus from 1970 to 2002, usually on a weekly or biweekly basis. It was succeeded in September, 2002 by YFile, York's daily electronic newsletter. These electronic archives begin in 1995.

- May 22, 2002
- May 8, 2002
- Apr. 17, 2002
- Apr. 3, 2002
- Mar. 20, 2002
- Mar. 6, 2002
- Feb. 20, 2002
- Feb. 6, 2002
- Jan. 23, 2002
- Jan. 9, 2002

- 2001 Editions
- 2000 Editions
- 1999 Editions
- 1998 Editions
- 1997 Editions
- 1996 Editions
- 1995 Editions

Search Gazette Archives
Collection policies

are extensions of and add value to traditional collection policies

“We archive web content that contributes to learning, teaching and research”
UVic

“We collect and archive web content that is of interest to both current and future researchers, academics and scholars at the University of Alberta, as well as the general public. We endeavor to acquire quality materials that are appropriate in meeting the needs of the wide range of subject areas taught and researched at the University of Alberta.” UofA

“University of Toronto Web Archives collection is an extension of the holdings of the University Archives which is responsible for the appraisal, acquisition, preservation and use of University records of permanent value and the private records of individuals and organizations associated with Canada's largest university. The Archive includes sites created by the University community in the utoronto domain.” UofT
What has been collected so far...
using ARCHIVE-IT

UVic: Academic Calendars

UBC: Asian Library Historical Websites, UBC Press and Scholarly publishing sites

Uof A: Oil Sands, Alberta Flood 2014, Alberta Education Curriculum; Idle No More, Health Science grey literature, U of A websites, Prairie Provinces Politics

Region of Waterloo Archives: Region of Waterloo transit planning, Elmira maple syrup festival, Grand River Accessibility Advisory Committee

United Church of Canada Archives: individual church websites

U Manitoba: Human rights, Ebola virus, University Website, Manitoba Arts and culture sites
Home grown Heritrix

LAC: GCWA

Bibliothèque et Archives nationales du Québec (BAnQ) Political party sites, ministerial sites, Municipal sites

York: Profiles, alumni magazine, Faculty newsletters, Special collections online exhibits
Thank you!

Joanne Paterson jpmkt22@uwo.ca