La Coalition canadienne de l’archivage Web
The Canadian Web Archiving Coalition

An update

Carole Gagné, Bibliothèque et Archives nationales du Québec
Nicholas Worby, University of Toronto
Corey Davis, COPPUL
Outline

1. Overview
2. Trends and initiatives
3. Collections
4. Update from BAnQ and LAC
La Coalition canadienne de l’archivage Web (CCAW)  
Canadian Web Archiving Coalition (CWAC)

“...established in 2017 to develop an inclusive community of practice within Canadian libraries, archives, and other memory institutions engaged or otherwise interested in web archiving, in an effort to identify gaps and opportunities best addressed by nationally coordinated strategies, actions, and services...”
Trends

- Information sharing
- Capacity building
- Community of practice
- Bringing together LAC and BAnQ w/ academic libraries and archives
Initiatives

- National Archive-It license
- DuraCloud sans Amazon (CANARIE-funded)
- Webinars
Initiatives

- Copyright
- Bi-annual web archiving survey (based on NDSA)
- Bi-annual meeting
- Collections (over to you Nicholas!)
Collections

- Information sharing to foster collaboration
  - Compiled collection & other web archive related policy documentation
  - Information sharing on curatorial methodologies
  - Identify gaps & topics of shared interest
- Tools, workflows, resources to do collaborative collecting
  - Pilot national collaborative collection
  - Internship opportunities for web archiving
Why a Canadian community of practice?

- Most national forums are organized along disciplinary/occupational lines (e.g. archives, libraries, museums, galleries)
- Very little interaction between English & French institutions
Enabling informal collaboration

Leonard Cohen
1934-2016
Web Archiving at
Bibliothèque et Archives nationales du Québec (BAnQ)

IIPC Web Archiving Conference – Zagreb, June 6, 2019

Carole Gagné
Librarian, Legal Deposit and Acquisitions
Bibliothèque et Archives nationales du Québec (BAnQ) – Who Are We?

➢ A national library, national archives and a major public library (the Grande Bibliothèque)
➢ 12 facilities throughout Québec
➢ Established in 2006
➢ Only provincial national library in Canada
➢ A French-language institution by law
BAnQ’s Web Site Harvesting Programme

➢ Falls within the responsibility of the National Library

➢ In collaboration with the National Archives and the Public Library

➢ Creation of a Web Site Selection Committee:
  ▪ Includes librarians and archivists from the different departments of BAnQ
BAnQ’s Web Archiving Programme: Context

- Began in 2009, as a pilot projet, with harvests on Québec’s government Web sites
- Limited resources and no regular funding
- 2 part-time employees
- Minimal IT and technical support
2018 Temporary resource: Québec government subsidy

- A Programmer-analyst dedicated
- Improvement of tools and workflow
- Increase of the number of authorized harvested Web sites made available online (quadrupled)

- Reduce backlog
Québec legal deposit does not apply to Web sites

Canadian Copyright Act

➢ BAnQ’s interpretation of Copyright Act requires that we obtain authorization to provide access to Web sites.
➢ Authorization licence

Exception: Québec government sites:

➢ Sites for which we have no licences can be viewed within BAnQ facilities.
Harvesting Web Sites: selective approach

BAnQ takes a **selective approach**

- Size of the body of materials to be collected, given BAnQ’s limited resources
- Québec does not have an official domain name widely used
- Legal aspects (obtain a license)

**Selection criteria**
For example:

- Government sites (.gouv.qc.ca, .quebec)
- Québec provincial and municipal elections
- Short-lived and media sites
Tools

➢ Web crawler: Heritrix, version 3.3 and Webrecorder
  • Very large majority of CWAC members have an Archive-it subscription
➢ Database of authorizations and of harvested Web sites
➢ Internal harvest manager (several modules)
➢ OpenWayback, version 2.3.2
Tools: Recent developments

Important changes:

- Transition from a process that was very dependent on IT services to curator’s autonomy at every step
- From harvesting to making the content available

- Flexibility
- Google Chrome Extensions
- Archives Unleashed (in-house)
## A decade of harvesting

<table>
<thead>
<tr>
<th>Category</th>
<th>2009-2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of harvests</td>
<td>13,096</td>
</tr>
<tr>
<td>Number of domains (URL)</td>
<td>2,270</td>
</tr>
<tr>
<td>Number of organizations whose website is made available</td>
<td>1,434</td>
</tr>
<tr>
<td>Size of archives</td>
<td>31 Tb</td>
</tr>
</tbody>
</table>
Accessing Web Sites

Since 2014

- Search by name, URL or subject
- Full-text search not provided
- Open access (90%) or in-house access (10%)


Bibliothèque et Archives nationales du Québec
Goals

➢ Expand the online content
➢ Improve collaboration

Our experience shows how important it is to obtain regular financial support for technical development to maintain the normal performance of a Web archiving programme.
Montréal area:  
514-873-1100

Toll-free elsewhere in Québec:  
1-800-363-9028

banq.qc.ca

Montréal
Grande Bibliothèque
475 boulevard De Maisonneuve Est
Montréal, Québec H2L 5C4

BAnQ Rosemont–La Petite-Patrie
2275 rue Holt
Montréal, Québec H2G 3H1

BAnQ Vieux-Montréal
535 avenue Viger Est
Montréal, Québec H2L 2P3

National Capital
BAnQ Québec
Pavillon Louis-Jacques-Casault
Campus de l’Université Laval
1055 avenue du Séminaire
Québec, Québec G1V 4N1

BAnQ Gaspé
80 boulevard de Gaspé
Gaspé, Québec G4X 1A9

BAnQ Gatineau
855 boulevard de la Gappe
Gatineau, Québec J8T 8H9

BAnQ Rimouski
337 rue Moreault
Rimouski, Québec G5L 1P4

BAnQ Rouyn-Noranda
27 rue du Terminus Ouest
Rouyn-Noranda, Québec J9X 2P3

BAnQ Saguenay
930 rue Jacques-Cartier Est, bureau C-103
Saguenay, Québec G7H 7K9

BAnQ Sept-Îles
700 boulevard Laure, bureau 190
Sept-Îles, Québec G4R 1Y1

BAnQ Sherbrooke
225 rue Frontenac, bureau 401
Sherbrooke, Québec J1H 1K1

BAnQ Trois-Rivières
225 rue des Forges, bureau 208
Trois-Rivières, Québec G9A 2G7
Update from LAC

- Tom J. Smyth
  Manager, Digital Integration
  Digital Preservation and Migration Division
  [tom.smyth@canada.ca](mailto:tom.smyth@canada.ca)
Thanks!

Carole Gagné, Bibliothèque et Archives nationales du Québec  
[carole.gagne@banq.qc.ca](mailto:carole.gagne@banq.qc.ca)

Nicholas Worby, University of Toronto  
[nicholas.worby@utoronto.ca](mailto:nicholas.worby@utoronto.ca)

Corey Davis, COPPUL  
[corey@coppul.ca](mailto:corey@coppul.ca)