

Immediate Action for Information Rescue: How We Learned More Than We Wanted about COVID-19 through Web Archiving

Tom J. Smyth

Manager, Digital Preservation
Digital Preservation and Migration
Division

Library and Archives Canada



Library and Archives
Canada

Bibliothèque et Archives
Canada

Canada

How COVID changed LAC's Curation Methodology



- LAC has been conducting “events-based” harvesting since 2013:
 - Lac-Megantic Rail Disaster, Shootings at War Memorial and Parliament Hill, Fort McMurray Wild Fires, Ottawa and Humboldt Broncos Bus Accidents
- While we did have rapid reactions to major events, they were ad hoc, and consisted of scrambling to get the best media resources –a means to get ahead of events (that by nature could not be planned or anticipated) was needed.
- To address this, in 2016 we began daily harvest of a broad range of media.
- This broad range began picking up information on COVID-19 when it emerged in the news in December 2019.
- Automating media collection freed curators to expand topic/resource selection.

COVID-19 News Media

- 48 total, 34 done daily:
- 35 English, 10 French, 11 Indigenous, 2 multicultural
- Acadie Nouvelle
- Bloomberg Canada
- CBC
- City News Toronto
- CTV News
- Financial Post
- Global News
- Globe and Mail
- Huffington Post Canada
- iPolitics
- Journal de Montreal
- Journal de Quebec
- La Presse
- Le Devoir
- Le Droit
- National Post
- Rabble.ca
- Radio-Canada
- Toronto Star
- Vancouver Sun
- Whitehorse Daily Star
- Winnipeg Free Press
- Macleans - COVID-19

COVID-19 Collection Scope

- Public health information (federal and provincial, emphasis on public health comms);
- Impact on business and the economy;
- Impact on charity and non-profit sector;
- Health science and medicine (e.g., research and vaccine efforts)
- Indigenous perspective;
- Social and cultural aspects (religion, art, cultural expression including satire);
- Alternative perspectives (Anti-mask and anti-COVID sentiment);
- Impacts on families, children, and education;
- Impact on professional sports (NHL, MLB, soccer, etc.);
- Curated social media related to COVID-19.

Documenting COVID-19 in Canada



Activity:	Total:
Total news/media websites crawled for COVID-19, YTD	48
Total non-media resources selected for COVID-19, YTD	1,458
Total digital assets collected for COVID-19, YTD	163,587,756
Total data collected for COVID-19, YTD	6.5 TB
Tweets captured for COVID-19, YTD	721,852

Curator Burnout...

- In the course of the work, curators came to know most of the intimate details and current stasis of COVID
- Curator trauma and burnout is a real and major concern.
- Strategies for wellbeing have included:
 - Rotating the curation of high impact subjects (health, etc.)
 - Providing the option to change to less stressful subjects
 - Switching to a more technical role (quality control, metadata)
 - Rotating out of COVID project to another project entirely

Program Future State



- Continue collecting information related to COVID-19 in Canada (as long as it takes...)
- Deploy a new Government of Canada Web Archive access portal in FY 2020-21.
- Expand to Canadian domain crawls (all of .ca) in the future



Questions/comments welcome

Library and Archives Canada

www.bac-lac.gc.ca

Telephone: 613-996-5115 or 1-866-578-7777 | TTY: 613-992-6969 or 1-866-299-1699

National
Capital Region

Vancouver
British Columbia

Winnipeg
Manitoba

Halifax
Nova Scotia



Library and Archives
Canada

Bibliothèque et Archives
Canada

Canada