Thematic web archives as an emerging collections strategy

Jean Blackburn and Dana McFarland (VIU)

Susan Paterson and Larissa Ringham (UBC)

CRKN 2022

Web archiving is the process of **collecting** portions of the World Wide Web, preserving the **collections** in an archival format, and then serving the archives for access and use.

https://netpreserve.org/web-archiving/

Web archiving as collection development

- Local
- Relevant
- Open Access
- Collaborative
- Community-engaged
- Compatible with collections policy & practice
- Active & Activist

"The implications for an archival project like Documenting Ferguson are **transformative** with respect to helping information professionals **re-imagine collections and services** through born digital objects and **collaborative outreach and research practices**…a model that many institutions might consider as **collecting activities become more open and community driven**… Washington University Libraries was able to document a politically charged event as it unfolded and **develop new relationships with the surrounding community** placing the university in a new light..the project allowed for progressive librarians to **promote and collect new content that informs those who create the scholarship** as well as participate in pedagogical processes of curriculum creation and instruction." - Makiba J. Foster & Meredith R. Evans

"Since 2013, the Web and Social Media Preservation Program team at Library and Archives Canada has developed web archival curation methodologies for reacting to and documenting what we formerly called **"Events-based collections"**. However, it quickly became apparent that these events-based collections could **absorb considerable unplanned and unforeseen resources** and we needed new ways to create efficiencies. COVID-19 has unfortunately proven us right in the need to create efficiencies for documenting events, and has also definitively demonstrated that web archiving truly is an **immediate action for information rescue and preservation of critical born-digital documentary heritage** -- which will become a primary source for future research."

- Tom J. Smyth, "Managing Evolving Scopes and Elaborating our Events-Based Program Methodologies"

Poll

Are you undertaking web archiving at your institution? (yes / no / thinking about it but haven't started yet)

If yes, what are you web archiving? (institutional content, thematic, other?)

Where does web archiving work happen in your organization?

- Collections
- Digitization or other technical department
- Both
- Other

UBC Library Web Archiving Service

University of British Columbia
Archive-It Partner Since: Jul, 2013
Organization Type: Colleges & Universities
Organization URL: http://www.library.ubc.ca/

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Institutional content

On-demand requests

Thematic collections

2013 - 2022:

4.2 TB of archived web content

Web archiving @ UBC: then and now

2013

1 collection

Web content captured: 32 GB

2022

20 collections

Web content captured: 1.6 TB
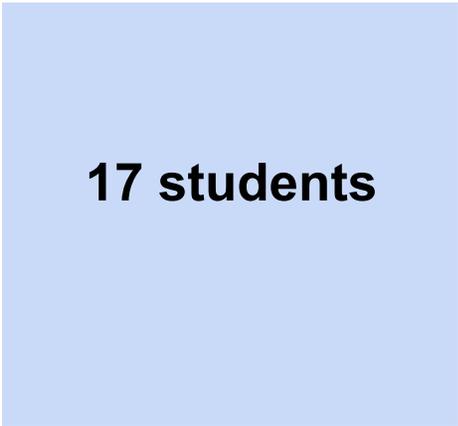
Web archiving as a service

Web archiving team

Student community engagement
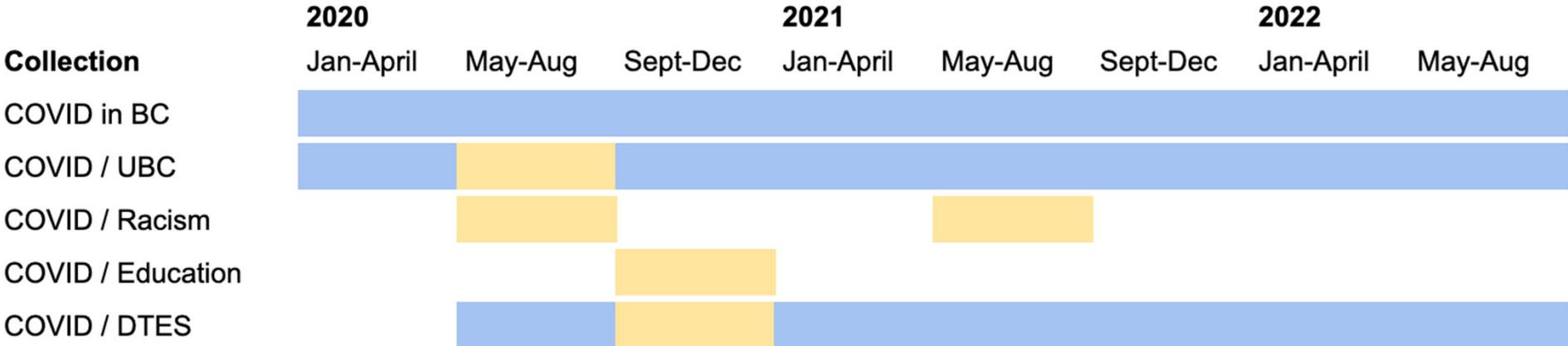
UBC iSchool

University of Alberta iSchool

University of Toronto iSchool

17 students

credit courses | class projects | student co-op terms
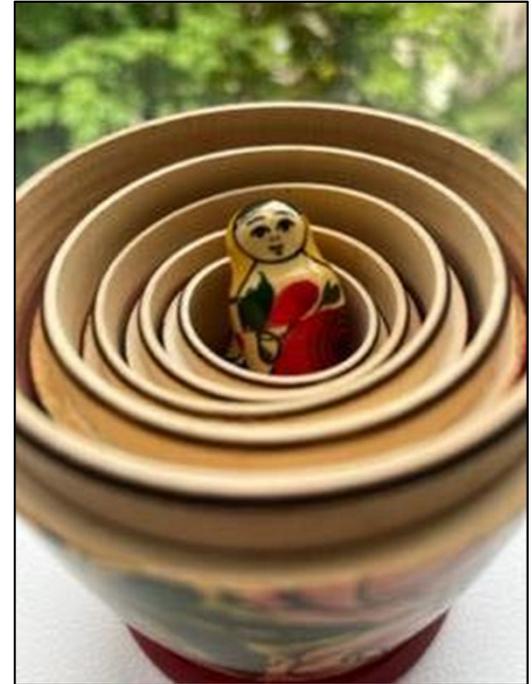
COVID web archiving: collecting timeline

Collection	2020 Jan-April	May-Aug	Sept-Dec	2021 Jan-April	May-Aug	Sept-Dec	2022 Jan-April	May-Aug
COVID in BC	▓	▓	▓	▓	▓	▓	▓	▓
COVID / UBC	▓	▒	▓	▓	▓	▓	▓	▓
COVID / Racism		▒			▒			
COVID / Education			▒					
COVID / DTES		▓	▒	▓	▓	▓	▓	▓

602 URLS ; 1.9 TB

▓ Library team

▒ Student project support

Why create smaller collections?

- More manageable focused collections
- Collections become publicly available sooner
- Students enable us to achieve this work
- Outreach to faculty and librarians for recommendations
- Allows for collaboration with partners on content identification
- Ensures local community content (ephemeral, endangered) captured
- Ensures multiple perspectives are captured

Approaches to Building Web Archive Collections

- Independent approach
  - BC Extreme Weather Events (2021)
  - Trans Mountain Pipeline
  - Marijuana Legalization
  - B.C. Hydro's Site C Project
  - Vancouver's Broadway Corridor Project (upcoming)
- Collaborative approach
  - 2013 Federal Government Websites
  - B.C. Wildfire Collections
  - COVID-19
  - 2022 BC Municipal Government Election

Collaboration is Key

[T]he truth is that web archiving has always been a collaborative effort. [...] What I've learned over the last 25 years is that only by collaborating with hundreds of libraries and organizations, employing thousands of domain experts and archivists, and coordinating with a global consortia of tool builders have we been able to build thousands of web collections and the Wayback Machine, a web archive that has proven useful to researchers, journalists and millions of people."

– Brewster Kahle. "Building Web Collections: Cooperation Past and Future."
Against the Grain 33:6. https://bit.ly/3eGDj46.

Positive Outcomes

- Aligns with the UBC Library Strategic Plan

- Advance research, learning & scholarship

- Engage with communities

- Create and deliver responsive collections

- Expand collaborations and partnerships

- Materials are openly available

- Materials support future scholarship

- Focuses our selection and collection policies

- Develop much-needed documentation
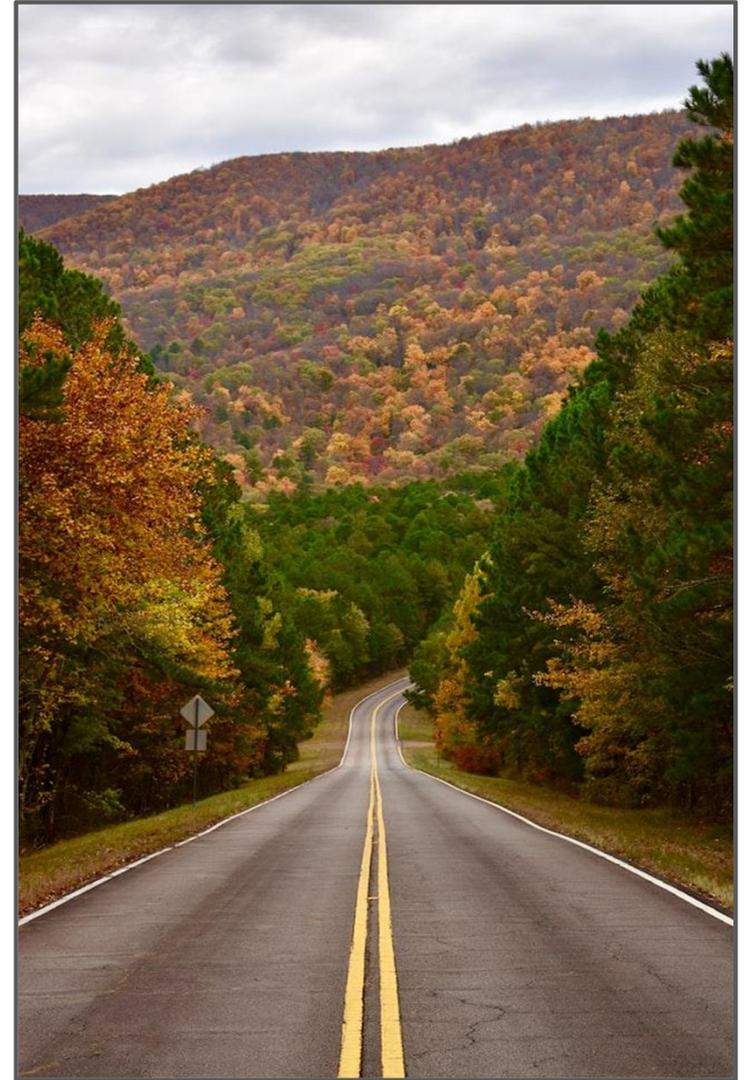
- Institutional support

Challenges

- Time constraints → trying to capture a lot of information quickly

- Limited capacity within the Library for this work

- Capacity with faculty members to partner with us

- Many collections are ongoing → expensive in terms of resource funding

- Scope creep

- Increasing the visibility of the web archive collections

- Ensuring inclusion and equity in the archives

- Technical challenges

- Partnerships

Going forward

- Increase capacity for active collection building with Web Archiving Team
- Identify university / community partners
- Engage with community engagement librarians
- Increase awareness of collections

Photo by Brendan Steeves on Unsplash

VIU planning and examples, cases

"Non-traditional collections" & web archives

Alexandro Malaspina Research Centre

BC Studies 2017: (Un)Settling British Columbia

COVID-19: Vancouver Island, BC (Central & North)

Commemorating Ye'yumnuts

Compass Rose: Explorations in Thought

Duncan-North Cowichan Referendum 2018

International Journal of UNESCO Biosphere Reserves, vol. 1, no. 1

Project REEL Life

The Homeroom. British Columbia's History of Education Website

The Navigator [student newspaper]

COVID-19: Vancouver Island, BC (Central & North)
Collected by: Vancouver Island University

Archived since: Mar, 2020
Description: In March 2020 the Canadian Web Archiving Coalition (CWAC) stated: "The deadly flu outbreak of 1918-19 is often called the 'forgotten pandemic.' Our responsibility now is to ensure the lessons of COVID-19 are not forgotten. Our collective efforts to capture and preserve the essential online elements of this unprecedented event are critical." (CARL-ABRC). The COVID-19: Vancouver Island, BC (Central & North) Web Archive reflects community experiences and responses to the pandemic in the region served by Vancouver Island University, documenting and supporting diverse aspects of scholarly inquiry, creativity, and
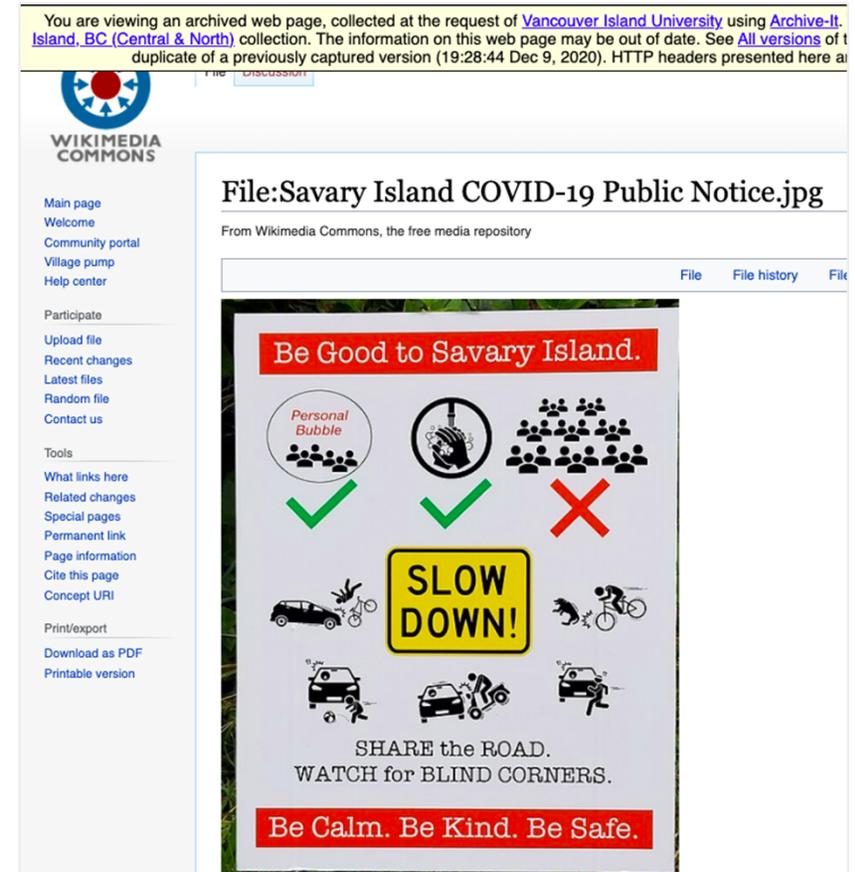
- "…community experiences and responses to the pandemic in the region served by Vancouver Island University, including and supporting diverse aspects of scholarly inquiry, creativity, and community life."

- "…to support current and future scholars of the region"

coordination with the University of Victoria's COVID-19 Collection and the Canadian Web Archiving Coalition. Contact us at research.help@viu.ca with questions or for more information about content included in the Web Archive. [Working description 2021 April 6]
Subject: Spontaneous Events, Society & Culture, Science & Health, COVID-19 (Disease), Vancouver Island
Creator: Dana McFarland (Vancouver Island University), Dalys Barney (Vancouver Island University)
Publisher: Vancouver Island University Library
Coverage: British Columbia–Vancouver Island
Date: 2020-

Maemura et al. advocate for

understanding of both the individual decisions that shape a collection through the web archiving process and their situated nature within specific and evolving organizational and technical contexts (2018)

By thinking about web archives as collections, in our collections contexts, we better understand our intentions, and also perceive gaps and opportunities with respect to representation.

Communities & opportunities for collecting the Web: What's happening?

- [Internet Archive Community Webs](#) or simply [save a page to the Wayback](#)
- [CWAC](#)
- [Canadian Government Information Digital Preservation Network (CGI-DPN)](#)
  - Depository Services Program (DSP) E-Collection
  - Government of Canada websites
  - Federal Elections
  - Provincial Elections
  - Thematic Collections
  - COVID-19, Opioid crisis
  - Rapid response collections
- [COPPUL WWAC](#)
- Local and regional e.g.
  - BC Wildfires 2017
  - [Vancouver Island](#), [Manitoba](#), other pandemic collections
  - BC Municipal Elections 2022
  - Old Growth Logging Protests? TBC

Remembering is a Form of Honouring

"**The lived experiences of most of the population have to be intentionally preserved for the future**. If we are not careful, the very same social inequities that are now hampering our ability to fight COVID-19 will determine whose lives will be remembered—the memories of the wealthy, the White, the powerful, will be privileged over those of the racialized, working people, and **those living ordinary lives in extraordinary times**." - Esyllt Jones et al., "Remembering is a Form of Honouring".

Jones E., Sweeney S., MIlligan I., Bak G. and McCutcheon J. (2021). "Remembering is a Form of Honouring: Preserving the COVID-19 Archival Record." *Facets* 6:545-568. doi:10.1139/facets-2020-0115

Resources & Examples

British Columbia Wildfires 2017 Collection

COVID-19 Collection (UVic) ; COVID-19: Vancouver Island, BC (Central & North)

CARL-ABRC (Canadian Association of Research Libraries). (2020, April). Community Call for Sharing and Coordinating Canadian COVID-19 Web Archiving Efforts. Retrieved April 15, 2020, from http://www.carl-abrc.ca/mini-site-page/community-call-covid-19-web-archiving/

Foster, Makiba J., and Meredith R. Evans. (2016). "Libraries Creating Sustainable Services during Community Crisis: Documenting Ferguson." *Library Management*, 37(6/7), 352-362.

Maemura, E., Worby, N., Milligan, I., & Becker, C. (2018). If these crawls could talk: Studying and documenting web archives provenance. Journal of the Association for Information Science and Technology, 69(10), 1223-1233. doi:10.1002/asi.24048. Open access preprint available from https://tspace.library.utoronto.ca/handle/1807/82840

Rollins, Caron, Matt Huculak, and Dana McFarland. (2018 June 5). Web archiving and public knowledge: Collaborating with Vancouver Island Archive-It partners. *Archive-it Blog.*

Smyth, Tom J. (2022). Managing Evolving Scopes and Elaborating our Events-Based Program Methodologies. https://netpreserve.org/ga2022/wac/abstracts/. IIPC Web Archiving Conference Abstract.

Vancouver Island University

Jean Blackburn
jean.blackburn@viu.ca

Dana McFarland
dana.mcfarland@viu.ca

https://archive-it.org/home/VIU

University of British Columbia

Susan Paterson
susan.paterson@ubc.ca

Larissa Ringham
larissa.ringham@ubc.ca

https://archive-it.org/organizations/734