

# COPPUL Digital Stewardship Network Community of Practice Mixer: Winter Wondrous Web Archiving Workflows

January 29, 2026 @ 12pm Pacific / 1pm Mountain / 2pm Central

Hosted by the [CDSN Community of Practice](#)

Quick links	
<a href="#">Join the Planning Committee</a> and help us organize events!	<a href="#">Sign up for the mailing list</a>
<a href="#">COPPUL Code of Conduct Policy</a>	Event feedback

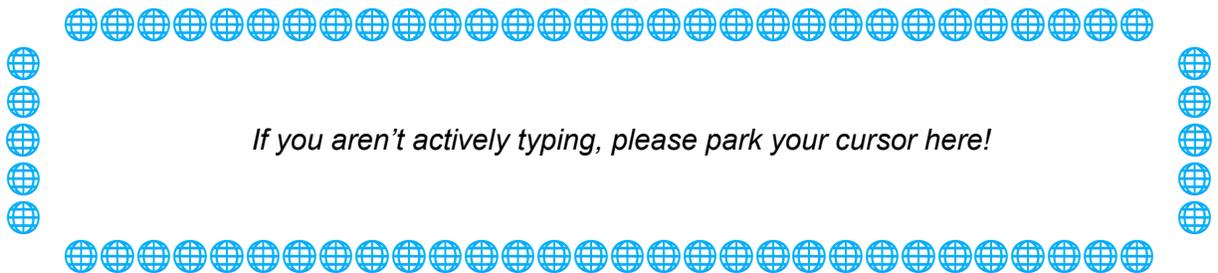
## AGENDA

- Welcome to the event
  - Feel free to introduce yourself in the Zoom chat!
- Community Presentations (recorded)
  - Short Presentation: Kenton Good (Head, Digital Production and Preservation Services, University of Alberta)
  - Lightning Talk: Delaney Sweep (Digital Preservation Technician, University of Calgary)
- Community Roundtable (not recorded)
  - What has worked well in your web archiving workflows? Any recommendations you'd like to highlight?
  - What hasn't worked ideally in your web archiving workflows? Any lessons learned?

## UPCOMING EVENTS

- Knowledge Exchange - April 29, 2026 (tentative date, details TBD)
  - If you would like to participate in planning this event or giving a presentation, lightning talk, demo or workshop on a digital stewardship topic please email us at [cdsn@coppul.ca](mailto:cdsn@coppul.ca)

# Community notes



## **Short Presentation: Kenton Good (Head, Digital Production and Preservation Services, University of Alberta)**

- Web Archiving at the University of Alberta: <https://library.ualberta.ca/web-archiving>
- Active web archiving program since 2009
  - Archive-It subscriber starting in 2009 (<https://archive-it.org/organizations/401/>)
  - Currently 42.6TB of archived data; 12 active collections, 18 inactive collections
- Over the years, have tried many ways to organize the work
  - Single individual, small/medium/large teams, collections focus, IT involvement, CLIR fellow focused on web archiving...
- Current solution is a cross-functional team!
  - Not a passive committee, but a group of colleagues that come together to enable an activity within the library
  - A commitment to do some work within that team
  - Currently 11 members, each with a time commitment of 1-2 hours/week, and 1-hour meeting/month
- UAL web archiving team roles (that are not mutually exclusive, some colleagues perform numerous roles)
  - External collaboration: alignment with external web archiving initiatives
  - Collections: new collection decisions & collection curation
  - Production: create & test seeds, metadata creation, compile seeds lists, WARC preservation
  - Communications & outreach: outreach & promotion of services & uses
- Each collection has a collection curator

- Individuals can be responsible for multiple collections, as applicable
- For new collections:
  - Work with requestor and/or collections group to set scope
  - Work with content production to generate seed list
- Ongoing collection work: responsible for checking in on the collection at least 1-2 times/year
  - Does the proposer/subject specialist think the collection should continue?
  - Responsible for scoping decisions: changes in scope? Any seeds to add/remove?
  - Goal: avoid “zombie” collections!
- Requests are received via shared mailbox so that all team members receive it; it is then triaged by a team member to either:
  - Tag for review: crawl may not be possible, new collection request
    - Team member reviews request, and then it is either:
      - Rejected: crawl not possible or out of scope, advise requestor accordingly
      - Team review: tabled at next team meeting for review & decision
  - Pre-approved requests: production team tests crawl, etc.
  - So far, received requests are at a manageable volume (10-15 emails a month est.)
- Creating new collections is decided collectively by the Team
  - Collection level metadata created using the UAL web archiving - Collection Metadata Guidelines (led by Metadata Librarian)
  - Once collection created, QAd, collection metadata created, and it's ready to be made public, the collection curator should create a short announcement blurb, usually based on the collection description
    - Should be shared with communications teams members for dissemination
- Use Github issues to track work
  - Built-in issues management, integrated with email, transparent
  - Trail of decisions & actions

- Successful so far!

### **Lightning Talk: Delaney Sweep (Digital Preservation Technician, University of Calgary)**

- Uses Archive-it from Internet Archive to collect web-based content connected to the university
- University of Calgary Web Archiving Policies:  
<https://library.ucalgary.ca/frequently-asked-questions/question/201649>
- Web archiving process started with individual completing online request form:  
<https://libguides.ucalgary.ca/webarchive> (Suggest/Add a site)
  - [Explanation of very helpful workflow diagram!]
  - Involves some iterative processes (e.g. QA, patching, re-crawling)
- QA workflow for ongoing crawls
  - Checking new crawls for website seeds that were previously set up & captured to ensure any change to root URL or new issues are caught & corrected
  - Future crawls run using new parameters or scoping rules etc.
- Many requests come from internal sources (e.g. university archives)

#### **FYI from the Zoom chat:**

The Archive-It team at the Internet Archive has an upcoming free webinar called “So you’ve inherited an Archive-It account...” coming up on February 25. Register here:  
[https://us02web.zoom.us/webinar/register/WN\\_rt5aaF5XQRSwvhjAwqOtZQ#/registration](https://us02web.zoom.us/webinar/register/WN_rt5aaF5XQRSwvhjAwqOtZQ#/registration)

### **Group Discussion and Questions**

*What kind of advertisements of the web archiving service have you used?*

- Alberta: Mostly passive on the website until relatively recently
  - New communications & outreach role within the Team is more proactively internally & externally promoting new collections
  - Using some social media channels, e.g. Instagram
  - Recognition of the challenges in communication across academic institutions
- Calgary: Teach a workshop on digital archiving each summer, so mention that this service is offered there
- How do we show value in these collections to researchers & students?

- [ARCH \(Archives Research Compute Hub\)](#): A platform for building research collections, analyzing them computationally, and generating datasets
- When it's a website that's still online, it's awkward to promote the archived version (that could be useful in the future, but maybe not as much now..?)

*Experience exporting WARC files from the Wayback Machine (stemming from a donor request to crawl not only their current live site, but also to preserve the crawls that already exist in the Wayback Machine, with the goal being to preserve previous versions of their website outside of the Wayback Machine (e.g., ingesting WARC files into Archivematica).*

- Calgary: We don't use Archivematica, but do use LibNova; understand theoretically this could be done but haven't looked far yet
- Alberta: Can download & it works as-advertised
  - Also web recording suite of tools
  - VIU: +1 webrecorder / conifer can make some great saves. But not always
- If you can think of these as data, the WARC files / text extracted can be deposited in a repository
  - Probably old news for many examples: [Web Archives for Historical Research Group Dataverse](#)

*How do you capture the student experience?*

- Haven't gone far with social media capture, more student clubs/groups
  - VIU: Social media is challenging to capture (e.g. RIP Twitter)
- It would be helpful to engage them more directly to get insight on where they're recording/capturing
- This collection while not web archiving, was presented & was fascinating: [School of Complex Adaptive Systems Research Datasets](#)

*What has worked well in your web archiving workflows? Any recommendations you'd like to highlight?*

- Don't give up!
  - Things are constantly changing, both with the websites you're trying to crawl, and the tools used
  - It may have been unsuccessful 6 months ago, but maybe it'll work now?
  - Sometimes it's a long game

- Collaborations
  - SFU and VIU (with UBC and UVic) Collaborative Collecting blog post: <https://archive-it.org/blog/collaborative-collecting-for-the-2022-british-columbia-local-general-elections/>
  - Shared portal for BC Elections: <https://archive-it.org/collections/20653>
  - Recognition that collaborative efforts have strengths & challenges/limitations
    - Communication is important

*What hasn't worked ideally in your web archiving workflows? Any lessons learned?*

- Ambiguity around roles
  - Having clear definitions for who does what has made a big difference
- Hard to do “off the sides of your desk”
  - It's so easy to think you'll come back to collections, but it can be challenging to actually do so (& not lose track of 90 day period for test crawls)
- Crawl quality problems for some sites with dynamic contents (e.g. sites created/hosted on sites.google)
  - Turned to external tools such as WebCrawler, but the Wayback Machine won't be able to replay these captured by the external tools..
  - You can create these files, but then what do you do with them? Challenges with access
- Metadata!
- VIU: [Looking back at 2018 post](#), struck by how much the issues remain the same, but we have made some gains!

*How can we make these collections more discoverable?*

- Making catalogue records for some (but not all) sites
- UBC has perhaps done some work to have collection level metadata for web archive collections available through ex libris - e.g. <https://go.exlibris.link/dxrfr4mD>
- VIU has done work to edit them into Wikipedia articles as resources: [https://en.wikipedia.org/wiki/COVID-19\\_pandemic\\_in\\_British\\_Columbia](https://en.wikipedia.org/wiki/COVID-19_pandemic_in_British_Columbia)