Archives Unleashed

June 2018 Newsletter

Happy Spring and Early Summer!

Let's take a closer look at what the Archives Unleashed team and community has been up to!

Highlights from the recent Toronto datathon are also included in this issue.

What's been happening?

- Released AUT 0.16.0.
- Co-Hosted Toronto Datathon
- Continued development and testing of Archives Unleashed Toolkit (AUT) and Cloud (AUK)
- Continued development to support PySpark DataFrames
- Sustainability planning
- Conducted research on incorporation processes
- Initiated preliminary analysis and cost modeling of sample WARC collections
- Updated documentation
- Released blog posts detailing AUT activities
- Initiated planning for Vancouver Datathon

Toronto Datathon
26-27 April 2018

The Archives Unleashed team recently co-hosted a datathon with Nich Worby and the [University of Toronto Libraries](#)!

Checkout out the [datathon summary](#).

Featured Articles

On Cloud Number 9: Accessibility, Usability, and Functionality with the Archives Unleashed Cloud (AUK)

Imagine a world where you could access and analyze a web archives file in as few as three clicks, with no coding necessary. Oh wait, there's a portal for that: the Archives Unleashed Cloud. ([Read More...](#))

To DataFrame or Not, that is the Questions: A PySpark DataFrames Discussion

The AUT team engaged with users in a discussion around alternative solutions to RDD that are now available, including PySpark DataFrames, and what this would mean for accessibility, impact, and uptake. ([Read More...](#))

Sustainability Measures with the Archives

Archives Unleashed Celebrates Its Sweet 0.16.0

Unleashed Team

Sustainability can give off an "elephant in the room" vibe— everyone knows it's there, but it can be an intimidating situation to tackle. The question most projects face is, "how do we plan for and ensure a sustainable future?" ([Read more...](#))

In April 2018, [AUT 0.16.0](#) was released! The toolkit has seen quite a bit of development and testing since the beginning of the year, so let's take a closer look into what's been done and the individuals who've helped to make it happen. ([Read more...](#))
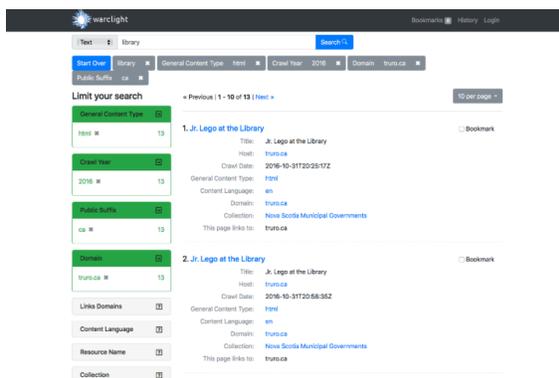
The Cost of Working with WARCs?

Have you ever wondered about the true costs of data analysis? The Archives Unleashed team has been analyzing the processing statistics for sample web archiving collections to understand the true costs of working with WARCs. ([Read more...](#))

Check It Out 👓

[Warclight](#) is a [Project Blacklight](#) based [Rails engine](#) that supports the discovery of web archives. We have some amazing [WALK Partners](#) who have contributed institutional data for indexing.

Read all about Warclight in our [latest blog posts](#).

Upcoming Events 📅

Archives Unleashed Datathon #Vancouver
1-2 November 2018
Harbour Centre, Vancouver

We are pleased to be co-hosting our second Canadian datathon with Rebecca Dowson, Simon Fraser University Library and KEY (SFU's Big Data Initiative)!
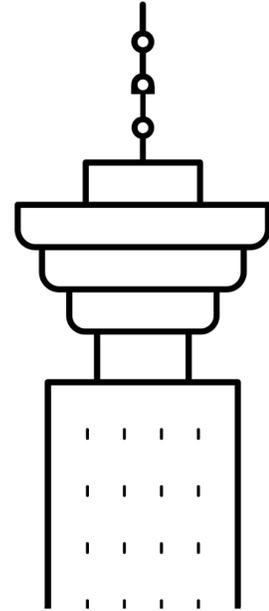
The Call for participation is now open.

We invite all archivists, researchers (from the humanities, social sciences, and beyond), librarians, computer scientists, and web archiving enthusiasts to join us over the course of two days to collaboratively work with web collections and explore cutting-edge research tools through hands-on experience.

Submissions Due: **Midnight (EST) on 16 July 2018.**

Webpage: https://archivesunleashed.org/vancouver/
Share our Call for Participation**!**

Get Involved

Interested in getting involved with the Archives Unleashed Project? Connect with our team and help grow our community

- Follow us on [Twitter](#)
- Join our [Slack](#) group
- Participate on [Github](#)
- Subscribe to our [newsletter](#)
- Submit to our [datathon](#)
- Share our news with colleagues and friends

(🐦) Twitter   (○) GitHub   (🔗) Website   (✉) Email

The Archives Unleashed Project, aim to make petabytes of historical internet content accessible to scholars and others interested in researching the recent past. Supported by a grant from the Andrew W. Mellon Foundation, the project will develop web archive search and data analysis tools to enable scholars and librarians to access, share, and investigate recent history since the early days of the World Wide Web.

The Archives Unleashed Project
200 University Avenue West | Waterloo, ON N2L 3G1