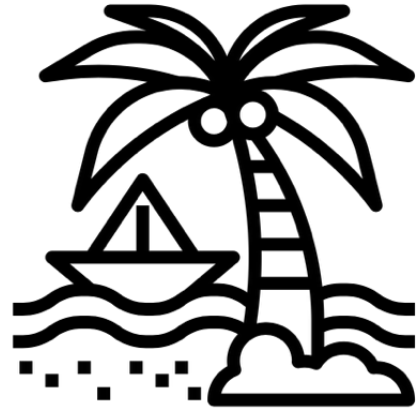# Archives Unleashed

Summer 2021 Newsletter

Welcome to the summer edition of the Archives Unleashed newsletter!

The heatwaves have not slowed down our work! As we move into September, let's catch up with the latest Archives Unleashed news.

Coming Soon!

Here's a sneak peak at the Archive Research Compute Hub (ARCH), our under-development interface to work with Internet Archive Archive-It collections to create datasets. Stay tuned!
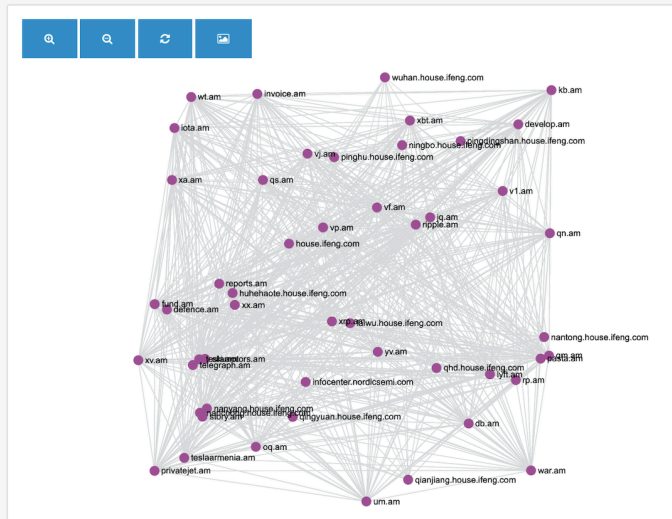
ARCH

Datasets

Learn More: ARCH Documentation

Collection Analysis

Collection Name	Public Collection	Recently Created	Last Created	Size ▼
2013 Supreme Court Hearings on DOMA/Prop 8	Yes	-	-	1.8 TB
Decentralized Web Summit	Yes	-	-	1.8 TB
SOPA Blackout	Yes	-	-	21.8 GB
2016 U.S. Presidential Election Results	Yes	-	-	402.7 GB
Dakota Access Pipeline	Yes	-	-	468.7 GB
United States Government Shutdowns	Yes	-	-	541.3 GB
#blacklivesmatter Web Archive	Yes	-	-	6.6 TB
Burning Man	Yes	-	-	64.8 GB
Total Solar Eclipse 2017	Yes	-	-	890.5 GB
Occupy Movement 2011/2012	Yes	-	-	905.7 GB

Top Hosts Sample

What does this graph show?

Dataset

We've detected macOS, and there is a known issue with Safari and Finder extracting gzipped text.
Please use The Unarchiver to extract domain-graph.csv.gz.

File name: domain-graph.csv.gz
File size: 5.3 MB
Result count: 591,343 lines
Date completed: 2021-07-30
Description: A CSV with the following columns: crawl date, source domain, target domain, and count.

Download dataset with
wget / **curl** ?

What's been happening? ⚙️

- **Sunset of the old Archives Unleashed Cloud** - In June 2021, we said goodbye to the Cloud. We're now a step closer to transitioning the Cloud's conceptual framework and analytical functions to the Archive-It environment.
- **Developed the ARCH interface** - The team has developed a new interface to analyze web archive collections by creating over a dozen datasets. The prototype is in beta, and we've moved it onto new clusters for powerful data processing and optimization. Based on initial user testing, the team has implemented additional user interface improvements.
- **Writing Documentation** - We've written user documentation that outlines the features and functions of the ARCH interface, provides use cases for dataset files and supporting materials for researchers to explore datasets with external tools.
- **Conducted UX Testing** - Our team has connected with internal and external candidates to test the ARCH prototype's efficiency and design aspects - and we've learned a lot!
- **Launched Cohort Program** - In July, we were pleased to host a virtual opening event to welcome our inaugural five Cohorts. Mentorship meetings are underway. In working with Archive-It partners, teams have begun to experiment with web archive collections in ARCH. We will be sharing progress throughout the year, so stay tuned!
- **Participated in spring conferences** - We were pleased to connect with friends and colleagues at virtual conferences and meetings. Team members participated in:
  - 2021 WAC - the 2021 Web Archiving Conference was hosted online by the National Library of Luxembourg.
  - RESAW 2021 - organized by Luxembourg Centre for

Contemporary and Digital History (C²DH) / University of Luxembourg

○ IFLA WLIC 2021 - special invitation to speak at IFLA's World Library and Information Congress in collaboration with the Big Data Special Interest Group.

AU Cohort Program

Cohort projects are in full swing and are excited to see all the planning around methods of analysis and the various web archive collections being used. Over the course of the year we will be providing updates on team progress.

Check out our introduction to teams and a short summary of projects in our recent Medium post: Introducing Archives Unleashed Cohorts.

One of our cohorts, led by Tim Ribaric from Brock University's Digital Scholarship Lab is exploring Crisis Communication in the Niagara Region during the COVID-19 Pandemic. They have recently launched their project website and first blog post, which you can check out: https://brockdsl.github.io/archives_unleashed/

Recent Presentations

WAC 2021 (Virtual Conference)
https://netpreserve.org/ga2021/

It was a pleasure to connect with the IIPC community this year! This year's Web Archiving Conference (WAC) was hosted online by the National Library of Luxembourg. The Archives Unleashed team delivered two presentations focused on community building and engagement, as well as designing accessible and usable research infrastructure.

All presentations from the conference are available through the IIPC YouTube Channel.

Building community through Archives Unleashed datathons: lessons learned
Presentation | Q&A Session | Open Access Article

Accessible web archives: Rethinking and designing usable infrastructure for sustainable research platforms
Presentation

RESAW 2021 (Virtual Conference)

Following IIPC's General Assembly and Web Archiving conference, the team was pleased to meet with friends, collaborators, and peers at the 4th RESAW conference to share a summary of our project's progress during year one of integration with Archive-It.

Integrating Archives Unleashed with the Internet Archive
Presented by Ian Milligan (University of Waterloo) and Helge Holzmann (Internet Archive)

Presentation Slides

IFLA WLIC 2021 (Virtual Conference)

Project manager Samantha Fritz was invited to speak on a panel at IFLA's World Library and Information Conference (WLIC), sponsored by IFLA's Big Data Special Interest Group.

Web Archives as Big Data: Building Tools and Community to Support Access and Usability
Presentation Slides

Reading List

Don't forget to check out the Archives Unleashed Medium site!

- [So Long, Farewell to the Archives Unleashed Cloud](#)

- [Introducing Archives Unleashed Cohorts](#)

- ["Building Community at a Distance" makes the DOLS Short List](#)

Upcoming External Events

There are a number of events coming up in the broader web archiving and digital preservation communities that might be of interest. Be sure to keep an eye out as events requiring registration have opened:

2021 Virtual Archive-It Partner Meeting
September 29, 9:00am – 2:00pm Pacific Time

This annual conference is dedicated to connecting Archive-It partners and sharing information about web archiving. The program will include technology updates, guest speakers, and moderated discussion.

Registration is required for this event and space is filling up fast!

https://archive-it.org/blog/2021-archive-it-virtual-partner-meeting/

iPres2021

19-22 October 2021

Hybrid Conference (Online & Beijing, China)

The 17th International Conference on Digital Preservation will use a hybrid model to promote wider communication among the preservation communities. Registration is open, with early bird rate in effect until September 15.

Find out more and register at https://www.ipres2021.ac.cn/en/web/index/11690

Digital Preservation 2021: Embracing Digitality (#DigiPres21)

- DLF Forum (November 1-3)
- Learn@DLF (November 8-10)

Registration is open until October 25, 2021, or until the event reaches capacity. These events are all held online.

Find out more and register at https://ndsa.org/conference/

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

This email was sent to <<Email Address>>