

Actively Archiving: Creating a web archive for future research into vaccine hesitancy, misinformation, and public health

Samantha Abrams¹, MLIS; Barbara Bernoff Cavanaugh², MSLS; Richard James², MSLS; Samantha J. Kaplan³, PhD, MLIS; Hannah Rutledge², PhD, MLIS, AHIP; Megan von Isenburg³, MSLS, AHIP, Lucy Brooks Waldrop³, MA, MLIS

What is a Web Archive?

Web archiving is the process of creating and storing copies (or snapshots) of live websites to ensure that they are preserved and remain publicly accessible for future use. Sites are typically captured using automated web crawlers, which systematically identify and follow hyperlinks on selected websites (referred to as seeds), copying and storing information as they work. This work is done because the web — the world's dominant communication platform — is highly dynamic and its content remains ephemeral and vulnerable. Web archiving use cases include:

- Personal digital archiving;
- Preservation of the scholarly record;
- Institutional archiving and records management;
- Legal deposit;
- Library collection development; and / or
- Research data mining.

The Ivy Plus Libraries Confederation uses Archive-It to crawl websites. Launched in 2006, Archive-It is an end-to-end hosted solution to create, store, and provide access to collections of web content. The service also allows users to add, import, and export descriptive metadata, and allows for public browsing and full-text search via archive-it.org.

Why Vaccination in Modern America?

Background:

The **Vaccination in Modern America Web Archive** was collaboratively developed by librarians at two health sciences libraries whose institutions prioritize vaccine research.

Purpose:

- Capture potentially ephemeral information about vaccination
- Facilitate future research into health information and misinformation by health services researchers, information scientists, sociologists, public health professionals, and others

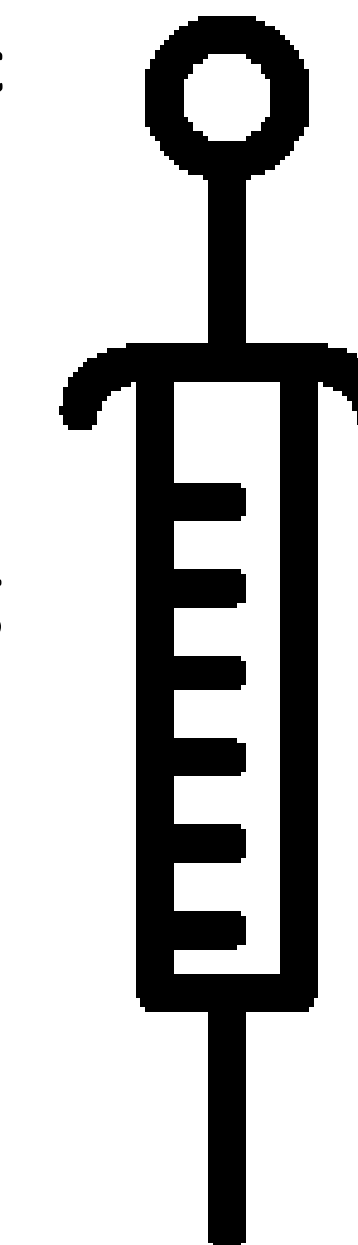
Perspective and Scope:

- Includes both pro- and anti-vaccination sites
- Curators oppose the rhetoric of anti-vaccination websites
- In order to address misinformation, we believe it is essential to catalog the tactics used by the sites seeding misinformation
- While conceived and launched before the COVID-19 pandemic, it has been expanded to include sites specifically addressing COVID vaccination

Selection Criteria

Pro-vaccination sites:

- Federal agency and public health organization websites on vaccination
- Sites that focus on evidence-based parenting
- Sites that address vaccine misinformation



Anti-vaccination sites:

- Sites that indicate opposition or foster uncertainty
- Sites that focus on chemical vaccine adjuvants or natural living as a replacement for vaccination
- Sites with expressed sympathies to vaccine avoidance

Behind the Scenes / Seeds

A **Seed** is an item in Archive-It with a unique ID number. The Seed URL tells the crawler where to go on the live web and acts as an access point to archived content.

Title: Fearless Parent
URL: <https://fearlessparent.org/>
Captured 6 times between Mar 16, 2020 and Mar 1, 2021
Videos: 1 Videos Captured
Subject: Anti-vaccination movement
Group: Anti-vaccination
Creator: Habakus, Louise Kuo
Language: English

Title: COVID-19 Vaccine Media Hub
URL: <https://health-desk.org/vaccines/>
Captured once on Mar 9, 2021
Subject: COVID-19 (Disease), COVID-19 Pandemic, 2020 —, Communication in science, Disinformation
Group: Pro-vaccination
Creator: Australian Science Media Centre, Meedan
Language: English

Metadata Created by Group

- Title
- URL (aka Seed)
- LC Subject Headings
- Group
- Creator
- Language

Metadata Generated by Archive-It

- Captured
- Videos



Accessing the Archive

The Archive is available at <https://archive-it.org/collections/13389>

- Selecting a member site or seed will open the record of captures by year
- Regular crawling and digital archiving of sites enables users to see changes over time
- Seeds are crawled every 6 months
- Once a user selects a specific capture, they can view the archived version of the site from that day

Found 5 Captures between Feb 3, 2020 - Mar 1, 2021	
2020	2021
4 pages	1 page
Feb 3, 2020 *	Mar 1, 2021 *
Feb 3, 2020 *	
Sep 3, 2020 *	
Sep 3, 2020	

Next Steps

The web archive launched in February 2020. Both pro- and anti-vaccination sites were identified. The archive is being expanded to include pro- and anti- COVID-19 vaccination websites.

Viewers can recommend sites here: <https://bit.ly/3aV6601>

About the Ivy Plus Libraries Confederation

The Ivy Plus Libraries Confederation is a voluntary union of 13 sovereign academic libraries: Brown, U. Chicago, Columbia, Cornell, Dartmouth, Duke, Harvard, Johns Hopkins, MIT, Penn, Princeton, Stanford, and Yale.

Author affiliations: 1) Columbia University; 2) University of Pennsylvania; 3) Duke University.