

FROM CAPTURE TO REPLAY: WEB ARCHIVING WITH WEBRECORDER TOOLS

Ilya Kreymer

*Webrecorder
United States
ikreymer@gmail.com
ORCID: 0000-0001-9632-8059*

Emma Dickson

*Webrecorder
United States
emma.jk.dickson@gmail.com
ORCID: 0000-0001-8218-4322*

Lorena Ramírez-López

*Webrecorder
United States
DaleLoreNY@gmail.com
ORCID: 0000-0003-1297-1990*

Abstract – Web archiving can take many forms, but usually involves fully capturing websites, storing the captured content, and faithfully reproducing, or replaying the archived sites. Since the web at its core consists of HTTP requests and responses, this process is often done at the HTTP traffic level to ensure full fidelity.

Traditionally, as pioneered by the Internet Archive and others, capture has been done by a web crawler which would automatically ‘crawl’ sites by making HTTP requests and discovering new links.

Another approach is to capture content through the web browser itself, capturing content right before it gets to the browser (or even right after). This type of web archiving can be done by using the suite of open source tools to capture interactive websites and replay them at a later time as accurately as possible on one’s own computer.

Webrecorder project builds tools that specialize in a ‘user-driven’ form of web archiving, where the user is able to direct the archiving process through their browser. The goal of Webrecorder has been to build quality open source tools to enable ‘web archiving for all’ to allow anyone with a browser to create their own web archives, and to accurately replay them at a later time.

Webrecorder project intends to support the creation of lots of small, decentralized archives by individuals and institutions. Rather than supporting a single centralized silo, a key goal remains to support web archives in a variety of environment and storage scenarios. From users’ personal laptops, to google drive, to IPFS, Webrecorder tools will be designed to support creating and accessing web archives wherever they may be found.

Through this workshop participants will gain a working knowledge on how web archiving works by using a simple web archiving process from capture to replay on a user’s local computer by using two open-source tools from Webrecorder: `archiveweb.page` and `replayweb.page`.

`Archiveweb.page` is an open-source web archiving capture tool that turns your browser into a full-featured interactive web archiving system. It is available as an extension for any Chrome or Chromium based browsers and as a standalone app.

`Replayweb.page` is an open-source browser-based viewer that loads web archive files provided by the user and renders them for replay in the browser. It functions as a serverless (client-side) web archive replay tool.

Both these tools combined create high-fidelity web archives using a basic workflow from capture to replay that participants can control, save, and embed onto other platforms. All the code can be viewed and accessed via the Webrecorder github: <http://github.com/webrecorder/>

Attendees will benefit from this workshop by gaining the ability to create high fidelity captures to make collections that they’ll have full autonomy over and can manage on their own devices and services. This session includes step-by-step installation help for both `archiveweb.page` and `replayweb.page` from the Webrecorder team. There will be a tutorial on how to analyze the website before capturing pages as well as breakdown to better understand components of the practice of web archiving and explore how they can fit into existing and emerging digital preservation workflows.

The target audience members for this type of workshop include: researchers, archivists, librarians, other practitioners working with digital materials, and developers. No prior experience with web archiving will be required of attendees.

Keywords – **web archiving, capture, replay, open-source tools, web collecting; personal digital archiving, high fidelity web archives, managing web archives, skills building, hands-on workshop**

Conference Topics – **Scanning the New Development; Building the Capacity & Capability**

- I. INTRODUCTION TO WEB ARCHIVING
 - A. DEFINITION
 - B. WEB ARCHIVING LIFE CYCLE MODEL
 - C. WEB OBJECT ENCAPSULATION COMPLEXITY

- II. INTRODUCTION TO WEBRECORDER
 - A. ABOUT WEBRECORDER
 - B. WEB ARCHIVING WITH WEBRECORDER

- III. ARCHIVEWEB.PAGE
 - A. ABOUT
 - B. INSTALLATION
 - C. CAPTURING WEBSITES
 - D. CREATING WEB ARCHIVES

- IV. REPLAYWEB.PAGE
 - A. ABOUT
 - B. INSTALLATION
 - C. UPLOADING WEB ARCHIVES
 - D. REPLAYING WEB ARCHIVES