

SUCHO

Saving Ukrainian Cultural Heritage Online



Tomasz Naskręt



Who?

Group of more than 1,300 cultural heritage professionals working together to identify and archive at-risk sites, digital content, and data in Ukrainian cultural heritage institutions while the country is under attack.

- librarians,
- archivists,
- researchers,
- programmers

Financial supporters

- [Association for Computers and the Humanities \(ACH\)](#),
- [European Association for Digital Humanities](#)
- [Europeana Foundation](#) (emergency project grants)
- [OpenCollective](#) page

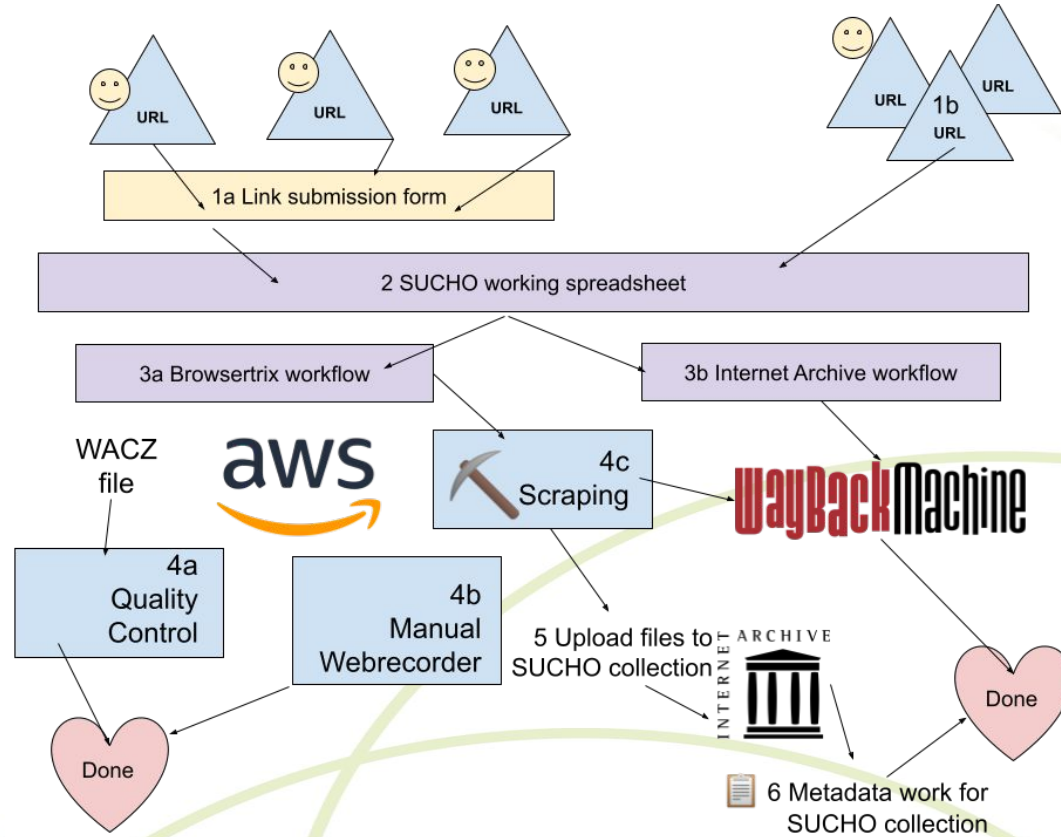
hosting server infrastructure pro bono

- Amazon Web Services
- Digital Ocean
- Wasabi

Coordinators

- Quinn Dombrowski, Stanford University
- Anna Kijas, Tufts University
- Sebastian Majstorovic, Austrian Center for Digital Humanities & Cultural Heritage

Workflows



Browsertrix

Installation:

```
docker pull webrecorder/browsertrix-crawler
```

Crawling:

```
docker run -v $PWD/crawl-config.yaml:/app/crawl-config.yaml -v $PWD/crawls:/crawls/  
webrecorder/browsertrix-crawler crawl --config /app/crawl-config.yaml --generateWACZ
```

crawl-config.yaml file:

```
collection: "sgiaz-uamuseum-com"
```

```
workers: 8
```

```
saveState: always
```

```
seeds:
```

```
- url: http://sgiaz.uamuseum.com/  
  scopeType: "domain"
```

Internet Archive's Wayback Machine

INTERNET ARCHIVE Explore more than 682 billion [web pages](#) saved over time

Wayback Machine

BROWSE HISTORY

Find the Wayback Machine useful? [DONATE](#)



Tools

[Wayback Machine Availability API](#)

Build your own tools.

[WordPress Broken Link Checker](#)

Banish broken links from your blog.

[404 Handler for Webmasters](#)

Help users get where they were going.



Subscription Service

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. [Visit Archive-It to build and browse the collections.](#)



Save Page Now

SAVE PAGE

Capture a web page as it appears now for use as a trusted citation in the future.

Only available for sites that allow crawlers.

A single copy of the Internet Archive library collection occupies 99+ Petabytes of server space (and we store at least 2 copies of everything). We are funded through [donations](#), grants, and by providing web archiving and [book digitization services](#) for our partners. As with most libraries we value the [privacy of our patrons](#), so we avoid keeping the IP (Internet Protocol) addresses of our readers and offer our site in https (secure) protocol.

Current State

- 30TB of scanned documents, artworks and many other digital materials from 3,500+ websites of Ukrainian museums, libraries and archives.
- 2624 task completed
- 128 completed archives

Other Archiving Efforts

- [Archive-It](#) has a team of people capturing Ukrainian government sites, and the .ua domain in general
- [Library of Congress: archiving websites from Ukraine and the Russian-annexed areas.](#)
- Ivy Plus SEEES affinity group: adding endangered sites from the region (including Ukraine) to an IPLC digital collection called [Eastern Europe and Former Soviet Union \(on Internet Archive\)](#).
- UC Berkley: [documenting Ukraine Crisis 2022 on Archive-IT](#); it is a continuation of their Crimean Crisis Web Archive. Contact is Liladhar Pendse.
- Internet Archive Global Events: discussion is under way for a [web archiving project on the war in Ukraine](#).
- American Folklore Society: backups of research and collection data for folklorists, ethnologists, and heritage institutions in Ukraine. We welcome colleagues in need to email: storage@sucho.org.
- University of Alberta's U-ART project: data backup for archival institutions, research institutions, individual scholars in all disciplines. [Contact U-ART](#).
- ["Backup Ukraine"](#) project by the Danish UNESCO National Commission and Blue Shield Denmark. Volunteers can digitally preserve important architecture, landmarks, and monuments using an app that creates a 3D scan.

Conclusions

- lot of manual work, probably some data will be lost anyway
- the last resort solution, but prone to failure in a scenario where the internet / power is cut off
- correct approach for sites, data, etc. defined as cultural heritage ?

- estimated cost of operation :

AWS S3 Standard for Over 100 TB / Month $0.021 \times 100000 = 2100$ \$/m

S3 Glacier Deep Archive*** - For long-term data archiving that is accessed once or twice in a year and can be restored within 12 hours $\$0.00099 \times 100000 = 99$ \$/m