

PISA Digital Humanities Summer School, June 2023

Afternoon Exercise, 16:00 to 17:30, 16 June 2023

Potential of: Exploring Web Archives as an example of challenges of working with digital materials.

In this exercise in teams of 3-4 participants you will examine a Web Archive and explore the possible ways that you might use it as researchers.

Exercise 2, Part 1:

Using the Library of Congress Collection Web Archive at <https://www.loc.gov/web-archives/> and UK Web Archive at <https://www.webarchive.org.uk/> compare the way the web archives of September 11 2001 and the “Manchester Arena bombing” are archived and presented.

Example questions might be:

- Can you tell what the motivation was for collecting the web archive?
- How was the web archive collected (broad crawl, target crawl)?
- Is the web archive permissions-based, founded in legal deposit legislation, or “an aggressive crawl”?
- What is the frequency of the crawls?
- What is the prevalence of missing pages, images, broken links (404) errors?
- Can you download material from the web archive?
- Can you re-construct the history of sites from the archive? Can you export material as WARC files?
- Can you tell how the coverage of these events changes over the days surrounding it from the Web Archive.
- If you wanted to download the archive for research, what steps would you need to take to do this.
- How would you validate the authenticity and integrity of the resource?
- As a researcher how would you identify gaps in the web archives?
- What other digital materials might you turn to understand the web archives?

After you have done the analysis we will discuss your findings.

Exercise 2, PART 2: (If we have time we will turn to this exercise).

Choose one of the archives listed below and search for an event or theme that your group is interested in and consider how it is represented in the archive and what can you find out about it.

Consider the coverage of the event in the context of other material about it that is available on the web and draw some conclusions about how valuable or not the web archive is in allowing you to develop an understanding of how the event and the coverage of it developed.

What kinds of tools do you require as researchers to use web archives effectively?

Report Back:

Your groups report back should provide an introduction to the web archive you explored, what kind of archive it is, how it is constructed and how the web resources were collected, how the content it holds is accessed, what information is provided to the users.

Library of Congress

- Introduction: <https://www.loc.gov/programs/web-archiving/about-this-program/> •
Web Archive Service: <https://www.loc.gov/programs/web-archiving/web-archives/> or <https://www.loc.gov/web-archives/>

National Library of Australia

- Introduction: <https://www.nla.gov.au/collections/what-we-collect/archived-websites>
 - Web Archive Service (TROVE): <https://webarchive.nla.gov.au/collection>

National Library of New Zealand

- Introduction: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive> • Web Archive Service: <https://natlib-primo.hosted.exlibrisgroup.com/primoexplore/search?vid=NLNZ> (Integrated into the NLNZ Catalogue)

National Archives UK

- Introduction: <https://www.nationalarchives.gov.uk/webarchive/>
- Web Archive Service: <https://webarchive.nationalarchives.gov.uk/search/>

The British Library

- Introduction: <https://www.bl.uk/collection-guides/uk-web-archive>
- Web Archive Service: <https://www.webarchive.org.uk/>