Using Web archives for researchers: experience, challenges, issues
Valérie Schafer

WWW
1- When
2- What
3- Why

Conference Saving the Web
Brussels, 18 October 2019
When did I come to Web Archives?

- I started by studying the Web without Web Archives
- 2011 = discovery of Internet Archive thanks to Niels Brügger
- The French WEB90 project (2014-2018)
- ASAP Project = a turn to DSN like Twitter (2016)
- *Negotiating the Web of the Past* (paper, 2016); *Do Web Archives have Politics?* (presentation, 2017); *Qu’est-ce qu’une archive du Web?* (collective book, 2019)

→ Through WARCnet (funded by the Danish Research Council): comparing Web domains / analysing transnational events (Terrorist attacks, Refugee crisis, EU parliamentary elections, Olympic Games in London….)
What do we do with Web Archives?

- Study of Web archiving processes and heritagization of our digital past
- Web archives as illustrations
- Experience of the Web of the past / « re-enactment »
- Getting some missing information to rebuild the digital landscape
- Crossing Web archives with other sources (printed, digitized, born-digital...)
Qualitative studies:

Focused studies on some webpages;
diachronic study of websites;
Mapping networks and
Web landscape on a precise topic …
The virtual exhibit by Anat Ben-David
20 years of Israeli Websites

Trying to get a whole picture of the Web of the Past

Probing a Nation’s Web Domain — the Historical Development of the Danish Web

**Purpose:** The project aims at a) mapping the development of the Danish web domain .dk, b) developing the research infrastructure needed to conduct the study, c) become a blueprint for a European comparative study.

**Expected outcomes:** The outcome of the project will be: a) an analysis of the Danish web’s historical development, b) a description of the needed research infrastructure, including descriptions of the different steps of the analysis, c) an outline of how the project can be implemented in other European countries.

Quantitative Studies: Big (or fat) data and computational tools ...
Why do National Web Archives matter for researchers?

- 8 years ago: starting with the Wayback Machine and search by URLs
- An evolving field which moved fast during the research (search by key words or time stamps in the Wayback machine)
- A growing collaboration with French national Web Archives from 2016
The value of national collections

→ More « national » content
→ Searchable content at BnF available from 2016
→ Mediations and transparency
Introducing the web-archives: Luxembourg

August 2016 crawl

- 70 million total
- 25 million English
- 2 million Luxembourgish
- 9 million others

URLs per language (top 9)

- English
- French
- Luxembourgish
- Spanish
- Dutch
- Finnish
- German
- Danish
- Greek

DH Benelux, Utrecht, 3-7 July 2017

Presentation “Research opportunities for the archived web in the Benelux” at DH Benelux by Sally Chambers, Peter Mechant, Kees Teszelszky, Yves Maurer
The value of national collections

- Understanding the shaping of Web Archives
- Dedicated tools
- Statistics and metadata
- Sustainability (permalink, preservation of some corpora, etc.)

- In return: common projects
  allowing to disseminate the collections’ value, meetings to better understand researchers’ needs and uses (trading zone)...

- Next steps: Transnational studies / Engagement in developing literacy (MAHEC Winter School with BnL and Publication Offices of the EU) / Cross-media approaches / Digital Public History ...