

Web archive-based research and web archives: a necessary but biased collaboration

Niels Brügger
THE CENTRE FOR INTERNET STUDIES
NETLAB

THE TENSION OF WEB ARCHIVING

- › Web archives need to collaborate with its users, because of the openness and complexity of web archiving
- › The collaboration with web scholars is inherently biasing the web archiving process, thus threatening to make the web archive a mirror of the work of the involved scholars

AGENDA

1. Why web archives and why web archive-based research?
2. Why collaborations?
3. How collaborations — the Danish case
4. The need for digital tools — specific and generic tools

1. WHY WEB ARCHIVES?

- › The importance of the web continues to grow
- › Part of the cultural heritage
- › The web will be an important historical source in the future

WEB ARCHIVE-BASED RESEARCH?

- › Web histories: Studies of the web itself
- › Cultural, social, political... histories: Studies of activities outside the web, but integrated with the web — political parties/movements, youth culture...

The web and historiography

- › Digital history *vs.* Web history
- › web as research tool and/or platform for dissemination/debate *vs.* web as historical source

2. WHY COLLABORATIONS?

Digital collections

- › analog original
- › selection of *what* to digitize
- › digitization is transparent
- › identical copies

Web archives

- › ephemeral original
- › selection of *what* to archive and of *how* to archive
- › web archiving is less transparent — technical deficiencies and dynamics of updating
- › versions without original

2. WHY COLLABORATIONS?

Digital collections

- › one copy of each
- › systematic, register
- › hyperlink is add-on
- › consistent

Web archives

- › too little, too much
- › unsystematic, no register
- › hyperlink is inherent
- › inconsistent in terms of time and space
- › accumulated heterogeneity over time

2. WHY COLLABORATIONS?

- › A wide and complex array of possibilities of archiving
- › The 'malleable' web archive
- › Someone has to make a number of choices
- › Choices have an impact on subsequent use

Therefore collaborations

- › the web archive needs to get information about possible use scenarios
- › the researchers need to get involved in the web archiving with a view to ensuring the usefulness

2. WHY COLLABORATIONS?

The tension of web archiving

- › Web archives need to collaborate with its users, because of the openness and complexity of web archiving
- › The collaboration with web scholars is inherently biasing the web archiving process, thus threatening to make the web archive a mirror of the work of the involved scholars

2. WHY COLLABORATIONS?

- › It is necessary that web scholars influence the web archive, but in doing this they bias the web archive

A collaboration:

- › which is constantly negotiated
- › where scholar's immediate needs should be evaluated in the light of preserving as broad an archiving as possible — today's scholars do not know the needs of tomorrow's scholars
- › with a history — to some extent the history of the web archive and of the web archive-based research

3. HOW — THE DANISH CASE

- › **1998, Jan.:** Legal deposit of static web (pdf etc.)
- › **2000, Sept.:** The Centre for Internet Studies (CFI)
 - › “promote research on the social and cultural functions and meanings of the internet”
 - › work for the establishing of a national web archive
- › **2001, Jan.:** meeting between CFI and the State and University Library (Aarhus), the Royal Library (Copenhagen)
- › **2001, June:** international conference 'Preserving the Present for the Future — Strategies for the Internet', Copenhagen, organised by the two libraries and CFI

3. HOW — THE DANISH CASE

- › **2001-02**: netarkivet.dk — a pilot project, the two libraries and CFI
- › **2002, June**: netarkivet.dk, Final report — the three strategies
- › **2002-04**: Preparation of the revision of the legal deposit law (the Royal Library & CFI involved)
- › **2004, Dec.**: law passed
- › **2005, July**: Netarkivet established, joint venture between the two national libraries

3. HOW — THE DANISH CASE

- › **2006, Jan.:** the advisory group for Netarkivet is established, members appointed by the Minister for Culture — web archive, scholars, content providers
- › **2006, Feb.:** CFI establish a small working group 'Analytical Software' — no succes in persuading our Department to support this

3. HOW — THE DANISH CASE

A critical rear-view mirror perspective:

- › aware of our own role in the creation of Netarkivet
- › tried to get scholars from other disciplines involved

To a certain extent Netarkivet mirrors our approach as media scholars — the consequence of this is to be judged in the future

3. HOW — THE DANISH CASE

The present and the near future — new perspectives
›**2012, Jan.:** Digital Humanities Lab Denmark established — a national digital research infrastructure project

NetLab

- ›A research infrastructure project for the study of internet materials — online and archived
- ›Close cooperation with Netarkivet
- ›NetLab Forum — brings users together (e.g. musicology)

4. THE NEED FOR DIGITAL TOOLS

- › Archived web material can be accessed and used analytically in a variety of ways — the 'malleable' web archive
- › Access and analytical use is a function of the characteristics of the archived web material
- › The archived web is not accessible as 'raw material' — it has to be formed, prepared for use
- › Digital tools for access, analysis and presentation are needed

4. THE NEED FOR DIGITAL TOOLS

- › Again collaborations are key: The tools must be adapted to the characteristics of the web archive as well as developed to suit actual researcher needs
- › A digital research infrastructure: the intermediary layer between the collections and the research activities
- › Issues of financing, organising...
- › The Danish case: NetLab — scholars, web archives and 'tinted' IT developer

4. THE NEED FOR DIGITAL TOOLS

- › Tools can serve many purposes — from search and visualization to a variety of analytical aims
- › The continuum of tools: specific—generic

4. THE NEED FOR DIGITAL TOOLS

Specific tools

- › Closely related to a specific research field, project, or question
- › Limited user group

Generic tools

- › Search — URL, free text, links...
- › Delimitation — timespan, web material
- › Duplicates — identify, examine, remove/select
- › Documentation — from collection to individual web page

QUESTIONS? COMMENTS?
