NetLab and RESAW — the Danish/European case

———

Niels Brügger
Head of NetLab &
of the Centre for Internet Studies
Agenda

1. Scholars, web archives — a personal account
2. What is NetLab?
3. What is RESAW?
4. What we do right now in NetLab:
   › Scholarly needs? A matrix for a generic approach
   › Probing a nation’s web sphere
1. Scholars, web archives — a personal account

Web historian — historical interest in the web started in the late 90ies
1. Scholars, web archives — a personal account

What I was missing in the late 90ies:

1. A preserved version of my object of study, the web of the past
2. Methodological and theoretical reflections about the scholarly use of archived web
3. A research infrastructure and analytical tools
1. Scholars, web archives — a personal account

From a personal solution to a national web archive

Jyllands-Posten
7 September 2000
1. Scholars, web archives — a personal account

1998: Legal deposit of static web (pdf etc.)

2000: 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: international conference 'Preserving the Present for the Future — Strategies for the Internet'
1. Scholars, web archives — a personal account

2001-02: netarkivet.dk — a pilot project, the two national libraries and CFI — three archiving strategies

2002-04: Preparation of the revision of the 1998-legal deposit law (the Royal Library & CFI involved)

2005: Netarkivet established, joint venture between the two national libraries
1. Scholars, web archives — a personal account

From vague ideas to more unfolded methodological and theoretical reflections
Conference 'Preserving the Present for the Future — Strategies for the Internet', 2001

The last page of the Internet?

Niels Brügger
Center for Internet Research, Denmark

2005

... and later several articles
1. Scholars, web archives — a personal account

The history of dr.dk 1995-2005

Started in 2007 — still ongoing
2. What is NetLab?

› An internet research infrastructure within the Danish research infrastructure for the humanities Digital Humanities Lab
› Established 2012
› Based on the work in the Centre for Internet Studies
› Research(er) driven development of research infrastructure
3. What is RESAW?

'A Research Infrastructure for the Study of Archived Web Materials' — established in late 2012

National web archives delimit the borderless information flow on the web by national barriers.

Promote the establishing of a collaborative transnational European research infrastructure for the study of archived web materials
3. What is RESAW?

IT developers (e.g. L3S)
3. What is RESAW?

› A larger network of relevant institutions and researchers, European as well as international (app. 40 participants)
› The basis for an application to EU’s Horizon 2020 within the topic ‘Integrating and opening existing national and regional research infrastructures of European interest’
3. What is RESAW?

The group coordinates a number of activities in 2014-15, including:
› a PhD seminar, Aarhus, 11-12 June 2014
› a seminar, London, 3-4 December 2014
› an international conference, Aarhus, 8-10 June 2015 entitled 'Web Archives as scholarly Sources: Issues, Practices, and Perspectives' — call out soon
› small pilot projects (e.g. how the internet domain .eu can be archived, Eurovision Song Contest...)
› resaw.eu
4. Scholarly needs? a matrix for a generic approach

Which tools, procedures, and policies do scholars need?
› The phases of research
› Transversal preconditions
The phases of research

Corpus creation

Analysis

Dissemination

Storage

Search

Duplicates

Select

Isolate

Identify

Evaluate

Select/remove/combine
The phases of research

Corpus creation

Analysis

Dissemination

Storage

Analytical tools

Visualization
The phases of research

- Corpus creation
- Analysis
- Dissemination
- Storage

- Legal issues
- Referencing
The phases of research

Corpus creation  Analysis  Dissemination  Storage

Legal issues  Access  Long term preservation

Corpora  Tools
Transversal preconditions

Workspace

Documentation

Data management

Collaboration
4. Scholarly needs? a matrix for a generic approach

<table>
<thead>
<tr>
<th></th>
<th>Workspace</th>
<th>Documentation</th>
<th>Data management</th>
<th>Collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus</td>
<td>Search</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duplicates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Select</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Isolate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Tools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Visualization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissemination</td>
<td>Legal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>referencing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>Legal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Access</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long term</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Probing a nations web sphere

The project: to study the historical development of a nation's web — the Danish web

Similar: Big UK Domain Data for the Arts and Humanities
4. Probing a nation's web sphere

Why study the development of a nation's web sphere?
› a backcloth for all other types of web entities and activities within the national web area
› identify some of the patterns of the developments of the web and thereby qualify discussions of what is happening on the web today
› compare web developments with phenomena outside the web
4. Probing a nations web sphere

An outline of an analytical design — a gross list of possible 'probes':
› Size — Bytes
› Space — Geolocation
› Structure — Networks of hyperlinks
› Vivid — Domain names and updating
› Content — Closedness, file and software types, language, and semantics