A NEW INTERFACE FOR WEB ARCHIVES? PROOF OF CONCEPT DEVELOPED BY THE ROYAL LIBRARY

Niels Brügger
PROFESSOR WITH SPECIAL RESPONSIBILITIES, HEAD OF THE CENTRE FOR INTERNET STUDIES, AND OF DIGHUMLAB/NETLAB
A NEW INTERFACE

› Maybe to be implemented as interface to the national Danish web archive Netarkivet
› Developed by Thomas Egense, Netarkivet
› I showed demo videos at web archiving conference in Israel last Sunday
› A lot of interest! The Internet Archive wants to learn more...
THE TEST DATA

› My personal web archive
› 2000-2007
› Archived by the use of Internet Explorer, HTTrack, and similar
› Transformed to WARC-files
› WARC files indexed (solr)
› Interface running on top of this index and the WARC files
› 23GB
### VISION OF A WORKFLOW

<table>
<thead>
<tr>
<th>Anchor</th>
<th>Legal experts</th>
<th>Search</th>
<th>Explorativ e analysis</th>
<th>Corpus creation</th>
<th>Extraction</th>
<th>Analysis</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan and organise the following steps together with researcher</td>
<td>Involved in planning and in the next phases when needed</td>
<td>Full text, URL, crawlog, seed-list, CDX...</td>
<td>On-the-fly analysis of the material to help with search and corpus creation</td>
<td>Final delimitation of what is found</td>
<td>Index file, new WARC, WAT, WANE, LGA, other</td>
<td>Tools supplied by the researcher (AUT, other) or by the web archive</td>
<td>Corpus description and other relevant documentation (public/not-public)</td>
</tr>
</tbody>
</table>

### Face-to-face

**Search and pre-analysis interface**

**Virtual drive**

**LOAR**

---

**Web archive:** provides infrastructural elements and supports these.

**Researcher:** provides research tools and has to learn to use these him/herself (probably some sort of procedure of software approval is needed).
A NEW INTERFACE FOR WEB ARCHIVES? PROOF OF CONCEPT DEVELOPED BY THE ROYAL LIBRARY

Niels Brügger
PROFESSOR WITH SPECIAL RESPONSIBILITIES, HEAD OF THE CENTRE FOR INTERNET STUDIES, AND OF DIGHUMLAB/NETLAB