Europeana Newspapers Project and Beyond

Berlin, January 31, 2014
Hans-Jörg Lieder, Project Coordinator - Berlin State Library
Staatsbibliothek zu Berlin
Overview

1. Why newspapers
2. Core mission ENP
3. Presentation on the European level
4. Services
5. The 20th century
Why newspapers? (and why not yet)

- Not meant to be preserved
- Stubborn, if not wicked material (paper, ink, fonts, layout, library binding, wrongly sorted or missing editions, …)
- Sheer quantity is intimidating

- Reflects immediacy and fast pace of public discourse
- Covering (almost) every aspect of the world (global, regional, local news, culture, sports, science, politics, economy, advertising, obituaries, …)
- Vast cross-domain interest
Europeana Newspapers Project Partners

| 01. State Library Berlin, Germany | 10. CCS Content Conversion Specialists GmbH |
| 02. National Library of the Netherlands | 11. Stichting LIBER, Netherlands |
| 06. State and University Library Hamburg, Germany | 15. University of Innsbruck |
| 07. National Library of France | 16. State Library Dr. Friedrich Tessmann, Italy |
| 08. National Library of Poland | 17. The British Library, UK |
| 09. University of Salford | 18. Europeana Foundation, Netherlands |
Who

Blue – Providing Content

Yellow – Providing Technical Services

Green – Associate Partners
## Content

<table>
<thead>
<tr>
<th>Country</th>
<th>Partner</th>
<th>Titel</th>
<th>Startdate (overall)</th>
<th>Enddate (overall)</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>BnF</td>
<td>80</td>
<td>1814</td>
<td>1944</td>
</tr>
<tr>
<td>Germany</td>
<td>SBB</td>
<td>6</td>
<td>1872</td>
<td>1940</td>
</tr>
<tr>
<td></td>
<td>SUB-HH</td>
<td>16</td>
<td>1721</td>
<td>1945</td>
</tr>
<tr>
<td>Netherlands</td>
<td>KB</td>
<td>203</td>
<td>1618</td>
<td>1900</td>
</tr>
<tr>
<td>Italy</td>
<td>LFT</td>
<td>15</td>
<td>1813</td>
<td>1949</td>
</tr>
<tr>
<td>Estonia</td>
<td>NLE</td>
<td>43</td>
<td>1852</td>
<td>1944</td>
</tr>
<tr>
<td>Finland</td>
<td>NLF</td>
<td>11</td>
<td>1900</td>
<td>1910</td>
</tr>
<tr>
<td>Latvia</td>
<td>NLL</td>
<td>117</td>
<td>1868</td>
<td>1955</td>
</tr>
<tr>
<td>Poland</td>
<td>NLP</td>
<td>118</td>
<td>1914</td>
<td>1939</td>
</tr>
<tr>
<td>Turkey</td>
<td>NLT</td>
<td>22</td>
<td>1818</td>
<td>1928</td>
</tr>
<tr>
<td>Austria</td>
<td>ONB</td>
<td>275</td>
<td>1686</td>
<td>1945</td>
</tr>
<tr>
<td>Serbia</td>
<td>UB</td>
<td>45</td>
<td>1830</td>
<td>1944</td>
</tr>
</tbody>
</table>

- 20 languages
- ca. 950 titles
- ca. 10 m pages
- refined
- 8 m OCR
- 2 m OLR
- 2 m NER
- vast amount of newspapers from 1914-18

This project is partially funded under the ICT Policy Support Programme (ICT PSP) as part of the Competitiveness and Innovation Framework Programme by the European Community [http://ec.europa.eu/ict_psp](http://ec.europa.eu/ict_psp)
Further to content

1. Defining metadata standards (images, texts, structural metadata, enrichments)

2. Developing quality evaluation and prediction tools to estimate OCR and OLR results

3. Defining aggregation and delivery mechanisms to transfer data from local providers to TEL and Europeana

More at German Information Day: February 27/28, Berlin
The European Newspapers Browser

First Iteration:
• Basic text search
• Filtering of results by date, country, newspaper, language, library
• OCR results shown
• Zoom able version of full image
• Clickable links between full text and image (where possible)
• Link to newspaper source library

Second Iteration
• Added information about particular titles
• See what was published on a particular day
• Search over titles (not just text)
• Other browse-able visualisations of publication and library source
• Search / browse via entities
The European Newspapers Browser

Search Newspapers:

[+] Further Search Options...

Browse by Issue Date

Explore Newspapers

Browse by Newspaper Title

- Wiener Zeitung (29,018 issues)
- Bote für Tirol (24,088 issues)
- Bozner Zeitung (18,420 issues)
- Margareten Zeitung (255 issues)

Browse by Country

This project is partially funded under the ICT Policy Support Programme (ICT PSP) as part of the Competitiveness and Innovation Framework Programme by the European Community http://ec.europa.eu/ict_psp
The European Newspapers Browser
• Full texts will be available at www.theeuropeanlibrary.org

• Images and metadata will be available at Europeana

• In future?
A word about services

- Collection of texts = knowledge base
- Quality of texts and level of aggregation crucial for level of services
- Size matters – big and small
- Open services to be customized by users
- Text search, search in structured texts, data mining (frequency analysis, semantic world field analysis, searches for „networks“, geographical searches, …)
- Librarian resources: islands of knowledge in an ocean of information
- Non-librarian resources: the world as the world sees it
- Think visual!
Mind the gap: towards the contemporary

Challenges:
• Legal stuff (IPR, „personal rights“)
• Business models and uncertainties

Options:
• Seek understanding with rights holders
• Cooperate with authors associations
• Cooperate with publishers (e.g. digitization projects)
Thank you!

Contact:

hans-joerg.lieder@sbb.spk-berlin.de
ulrike.koelsch@europeana-newspapers.eu

For more information, please see:

www.europeana-newspapers.eu
Twitter (@eurnews)
Facebook: https://www.facebook.com/EuropeanaNewspapers