



europæana
newspapers



The Europeana Newspapers Project

A Gateway to European Newspapers Online

Berlin, 18.06.2012

Lisabet Mielke, Staatsbibliothek zu Berlin, Germany



Content

- Aims
- Consortium
- Structure
- Areas of activity

Why newspapers?

„Zeitungen sind die Sekundenzeiger der Geschichte“
(Newspapers are the sweep hands of history)

Arthur Schopenhauer

- Relevant to all citizens
- Highly relevant to European policies incl. Europeana
- Newspapers in libraries – between
 - Heaven = solid and complete originals, excellent microfilm copies
 - and Hell = frail and crumbly originals, missing editions, incomplete supplements, poor microfilm copies, legal uncertainties with contemporary material

Aims & Objectives

1) Selection, Refinement & Aggregation of content

- Make Europeana the largest provider of pan-European newspaper collections
- Provision of more than **18 million newspaper pages** to Europeana, many of those with full-texts

2) Analysis of existing newspaper collections

- Survey of newspaper holdings in Europe

3) Quality Assurance & Best Practice recommendations

- Contribute to optimised workflows and data aggregation infrastructures
- Provide best practice recommendations for digitisation, refinement, workflows, metadata etc. and evaluation tools

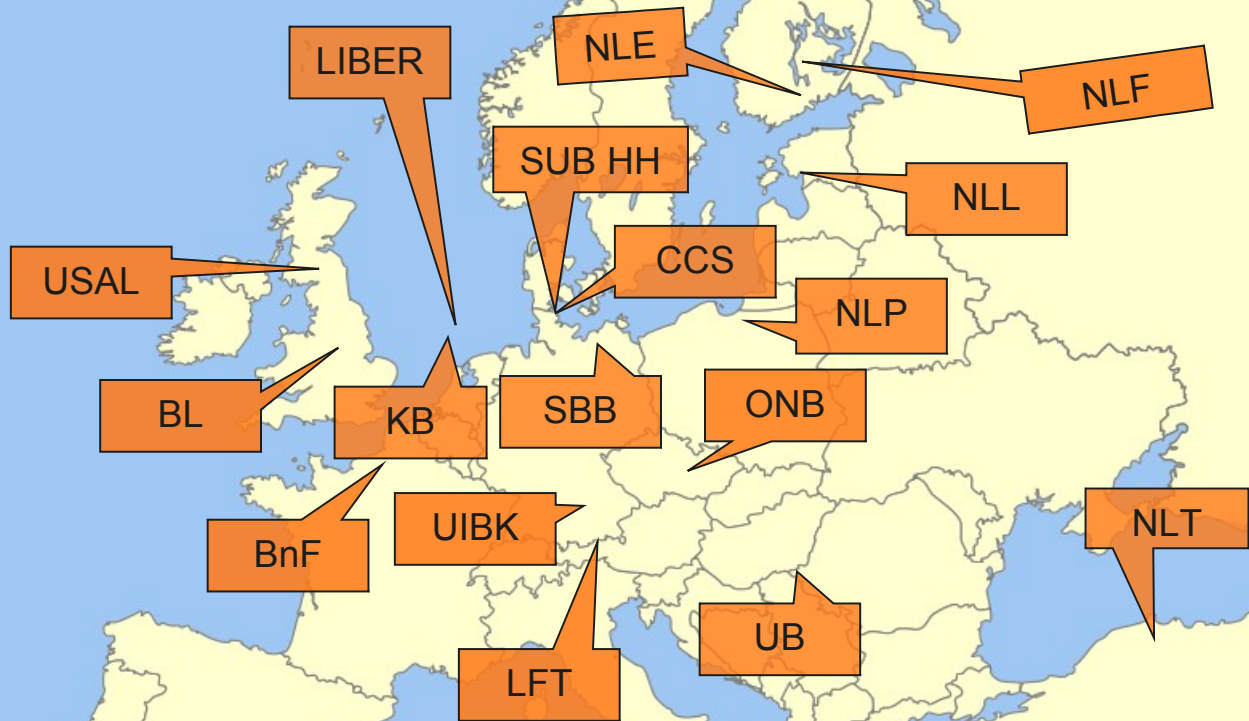
4) Presentation and full-text search

- Improve access to newspaper collections within Europeana

Consortium & Stakeholders

- 17 partners from 12 countries within the consortium
 - National and University libraries
 - Research institutes
 - Software Company
- External partners and stakeholders:
 - Involvement of libraries outside the project consortium
- Framework:
 - funded as a Best Practice Network in the ICT-PSP of the European Commission
 - Project Duration: February 2012 – January 2015

Europeana Newspapers Consortium



Consortium Partners

1. Staatsbibliothek zu Berlin
(project co-ordinator)
2. National Library of the Netherlands
3. National Library of Estonia
4. Österreichische Nationalbibliothek
5. National Library of Finland
6. Staats- und Universitätsbibliothek Hamburg
7. Bibliothèque nationale de France
8. National Library of Poland
9. University of Salford
10. CCS Content Conversion Specialists GmbH
11. Stichting LIBER
12. National Library of Latvia
13. National Library of Turkey
14. University Library of Belgrade
15. University of Innsbruck
16. Landesbibliothek Dr. Friedrich Tessmann
17. The British Library

Project Structure

- **Work Package 1: Coordination and Management**
 - Berlin State Library (SBB)
- **Work Package 2: Refinement of digitised newspapers**
 - National Library of the Netherlands (KB)
- **Work Package 3: Evaluation and Quality Assessment**
 - University of Salford (USAL)
- **Work Package 4: Aggregation and presentation of digitised newspapers for Europeana**
 - The European Library (TEL)
- **Work Package 5: Metadata best practice recommendations**
 - University of Innsbruck (UIBK)
- **Work Package 6: Dissemination and Exploitation**
 - Association of European Research Libraries (LIBER)

WP 1: Coordination and Management

Berlin State Library (SBB)

- Project administration
- Financial control
- Project communication
- Project quality assurance
- Risk management

WP 2: Refinement of digitised newspapers

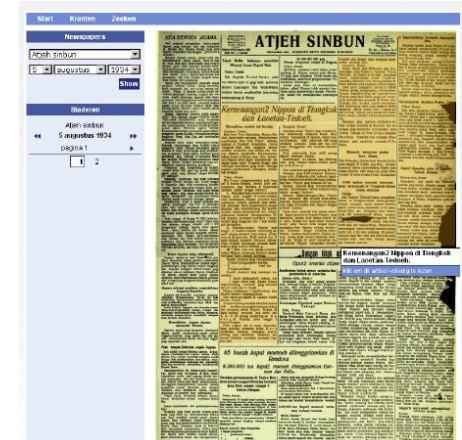
National Library of the Netherlands (KB)

- Analyse and select available digital newspaper collections
- Define digitisation requirements and minimum quality of newspapers
- Coordinate refinement of selected content provided by libraries

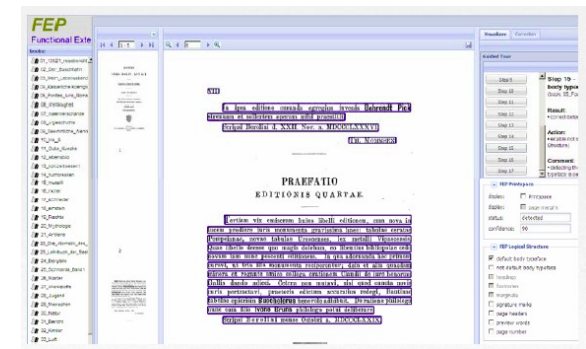
WP2: Refinement of digitised newspapers

National Library of the Netherlands (KB)

- 8 million pages “as is”
- 10 million refined pages: OCR (UIBK, Austria)
- 2 million refined pages: OCR/OLR (article segmentation) (CCS, Germany)
- CCS produces OCR and verification of column recognition, zoning, article segmentation, and page class recognition
- CCS provides libraries with a client technology for manual correction of recognition and segmentation results



CCS: Column recognition, article segmentation

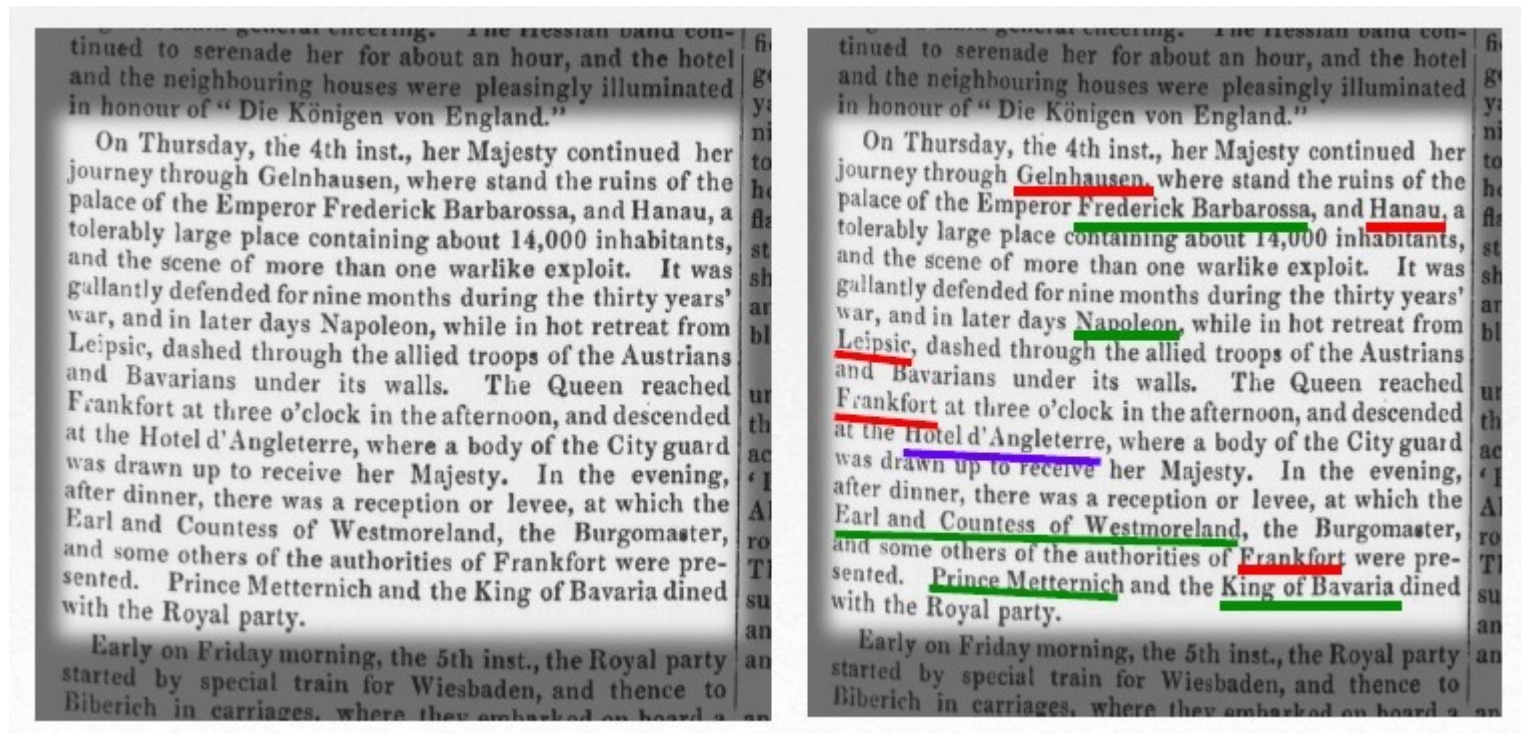


UIBK: Detection of headings, footnotes, etc.
Table of contents extraction

WP 2: Refinement of digitised newspapers

National Library of the Netherlands (KB)

- KB provides named entities recognition (NER) for material from up to three languages (Dutch, English and German)



WP 3: Evaluation and Quality Assessment

University of Salford (USAL)

- Evaluation of use cases by comparing ground truth and the digitised version
- Overview of usability, limitations and potential of existing material
- Identification of bottlenecks and recommendations for improvements
- Recommendations for best practice in digitisation projects

WP 4: Aggregation and presentation

The European Library (TEL)

- Aggregate newspaper metadata from content providers
- Recommendations how to align newspaper metadata to European Data Model (EDM)
- Analysis of public and private digital newspaper collections across Europe
- Creation of a European registry for digitized newspapers
- Creation of a full-text index of newspaper content
- Development of a newspaper content browser

WP 4: Aggregation and presentation

The European Library (TEL)

- Questionnaire to analyse the extent of digitised newspapers collections at their institutions
 - Embedding of results in “Zeitschriftendatenbank” of Staatsbibliothek zu Berlin (Union Catalogue of Serials)
 - Identification of potential new partners for the extension of the network
- If you hold digital newspaper collections and like to participate in the survey → please contact:
survey@europeana-newspapers.eu/

WP 4: Aggregation and presentation

The European Library (TEL)

- Within the lifetime of the project, a content browser will be built within TEL portal so that users can ...
- Search full text, e.g.
 - by search term,
 - by named entities
 - by collections of newspapers
 - by date
- See newspaper images
- Be linked to relevant library sources

Explore Europe's cultural collections

Lewis and Clarke

CAPTAIN LEWIS'S EXPEDITION.

THE following particulars of the expedition of captain Lewis, from the mouth of the Missouri, which empties into the Mississippi at St. Louis, to the Pacific ocean, transmitted to general Clarke, of Kentucky, by his brother, who accompanied the expedition, will be found interesting.

St. Louis, 23d Sept. 1806.

DEAR BROTHER,

We arrived at this place at 12 o'clock to-day, from the Pacific ocean, where we remained during the last winter, near the entrance of the Columbia river. This station we left on the 27th of March last, and should have reached St. Louis early in August, had we not been

WP 5: Metadata best practice recommendations

University of Innsbruck (UIBK)

- Analysis of metadata formats in use by libraries
- Align metadata models with the METS/ALTO standard and release best practise recommendations
- Usability of the recommendation will be tested through an evaluation cycle
- Provide recommendations on best practices for refinement of digitised newspaper collections for Europeana

WP 6: Dissemination

Association of European Research Libraries (LIBER)

- Objectives
 - Establishment of publicity
 - Increasing usage of Europeana
 - Awareness raising among target groups
- Tasks
 - Media Communication
 - Three main dissemination workshops
 - National information days
 - Network extension





Thank you for your attention

Lisabet Mielke, Staatsbibliothek zu Berlin

lisabet.mielke@europeana-newspapers.eu

www.europeana-newspapers.eu