IFLA International Newspaper Conference

“Newspaper Digitization and Preservation. New prospects. Stakeholders, Practices, Users and Business Models”

11-13 April 2012
BnF, Paris

With the support of:

The European(a) Newspapers Project

A Gateway to European Newspapers Online

Paris, 12.04.2012

Thorsten Siegmann, Staatsbibliothek zu Berlin, Germany

Content

Project Profile

- Aims
- Consortium
- Framework

Areas of activity

Europeana Newspapers – why newspapers?

"Die Zeitungen sind die Sekundenzeiger der Geschichte."
(Newspapers are the sweep hands of history)

Arthur Schopenhauer

Why newspapers?

- 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

Europeana Newspapers: Aims and Objectives

Europeana Newspapers

- aims at the aggregation and refinement of newspapers for *The European Library* and *Europeana*.
- will use refinement methods for OCR, OLR (article segmentation), and named entity (NER) and class recognition
- the libraries participating in the project will provide around 18 million digitised newspaper pages to Europeana
- Further libraries will be encouraged to contribute newspapers to Europeana and TEL by the project

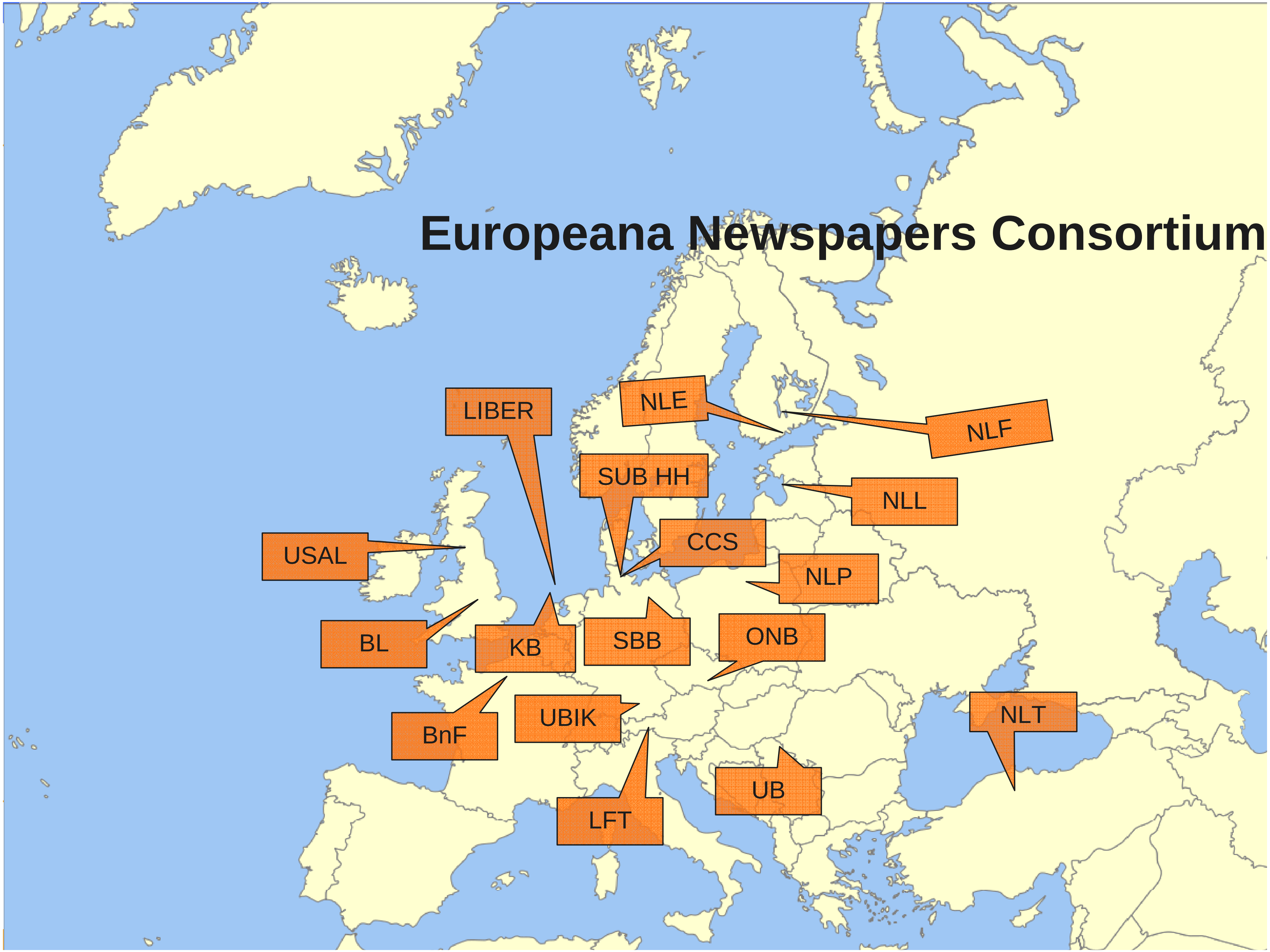
Project Profile: Consortium & stakeholders

- 17 partners from 12 countries within the consortium
  - National libraries
  - University libraries
  - SME

- External partners and stakeholders:
  - Involvement of libraries outside the project consortium

- Framework:
  - Funded as a Best Practise Network in the ICTPSP programme 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 Profile: 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
- Support move from images to texts in Europeana

2) Analysis of existing newspaper collections

- Survey of newspaper holdings in Europe

3) Quality Assurance & Best practise recommendations

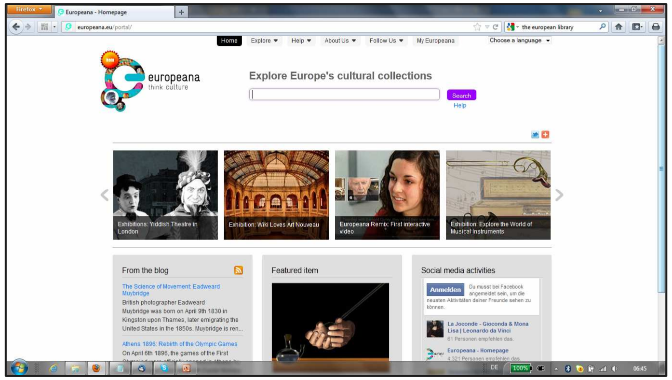
- Contribute to optimised workflows and data aggregation infrastructures
- Provide best practice recommendations for digitization, refinement, workflows, metadata etc. and evaluation tools

4) Presentation and full-text search

- Improve access to newspaper collections within Europeana

1) Selection, Refinement & Aggregation of content

- Aggregation of 18 million pages of digitised newspapers to *Europeana* and to *The European Library*
  - 8 million pages “as is” (content providers)
  - 10 million refined pages: OCR (UIBK, Austria)
  - 2 million refined pages: OCR/OLR (article segmentation) (CCS, Germany)
- Analysis of available digital newspaper collections and selection of subsets suitable for refinement



Explore Europe's cultural collections

From the blog

The Science of Movement: Eadweard Muybridge

British photographer Eadweard Muybridge was born on April 9th 1830 in Kingston upon Thames, later emigrating the United States in the 1850s. Muybridge is ren...

Athens 1896: Rebirth of the Olympic Games

On April 6th 1896, the games of the First

Featured item

Social media activities

Du musst bei Facebook angemeldet sein, um die neusten Aktivitäten deiner Freunde sehen zu können.

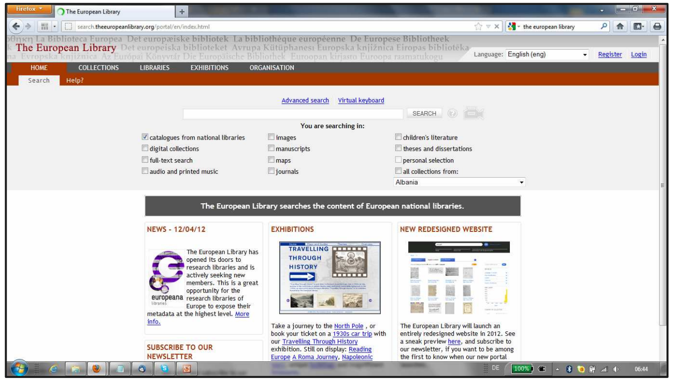
La Joconde - Gioconda & Mona Lisa | Leonardo da Vinci

61 Personen empfehlen das.

Europeana - Homepage

4.321 Personen empfehlen das.

www.europeana.eu/


Firefox
The European Library
search.theeuropeanlibrary.org/portal/en/index.html
the european library
La Biblioteca Europea Det europæiske bibliotek La bibliothèque européenne De Europese Bibliotheek
The European Library Det europeiska biblioteket Avrupa Kütüphanesı Europska knjižnica Eiropas bibliotēka
Evropska knjižnica Az Európai Könyvtár Die Europäische Bibliothek Euroopan kirjasto Euroopa raamatukogu
Register
Login
HOME
COLLECTIONS
LIBRARIES
EXHIBITIONS
ORGANISATION
Search
Help?
Advanced search
Virtual keyboard
SEARCH
TRAVELLING THROUGH HISTORY
europeana
libraries
DE
100%
06:44

Language: English (eng)

You are searching in:

☑ catalogues from national libraries
☐ digital collections
☐ full-text search
☐ audio and printed music
☐ images
☐ manuscripts
☐ maps
☐ journals
☐ children's literature
☐ theses and dissertations
☐ personal selection
☐ all collections from: Albania

The European Library searches the content of European national libraries.

NEWS - 12/04/12

The European Library has opened its doors to research libraries and is actively seeking new members. This is a great opportunity for the research libraries of Europe to expose their metadata at the highest level. More info.

SUBSCRIBE TO OUR NEWSLETTER

EXHIBITIONS

Take a journey to the North Pole , or book your ticket on a 1930s car trip with our Travelling Through History exhibition. Still on display: Reading Europe A Roma Journey, Napoleonic

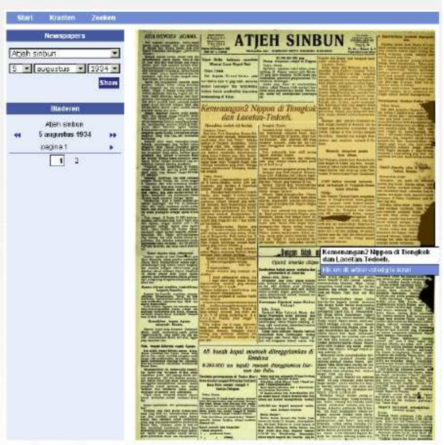
NEW REDESIGNED WEBSITE

The European Library will launch an entirely redesigned website in 2012. See a sneak preview here, and subscribe to our newsletter, if you want to be among the first to know when our new portal

www.theeuropeanlibrary.org/

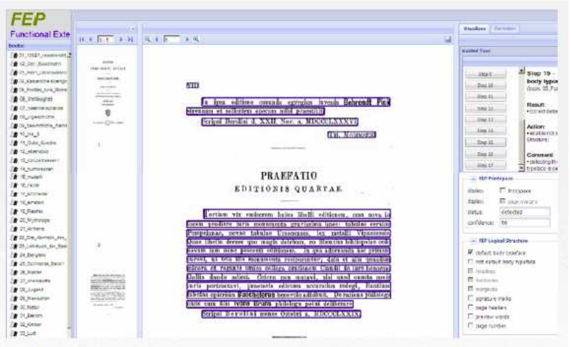
1) Refinement – OCR and OLR

- 10 million refined pages:
  OCR (UIBK, Austria)
- 2 million refined pages:
  OCR/OLR (article segmentation)
  (CCS, Germany)



CCS: Column recognition, article segmentation

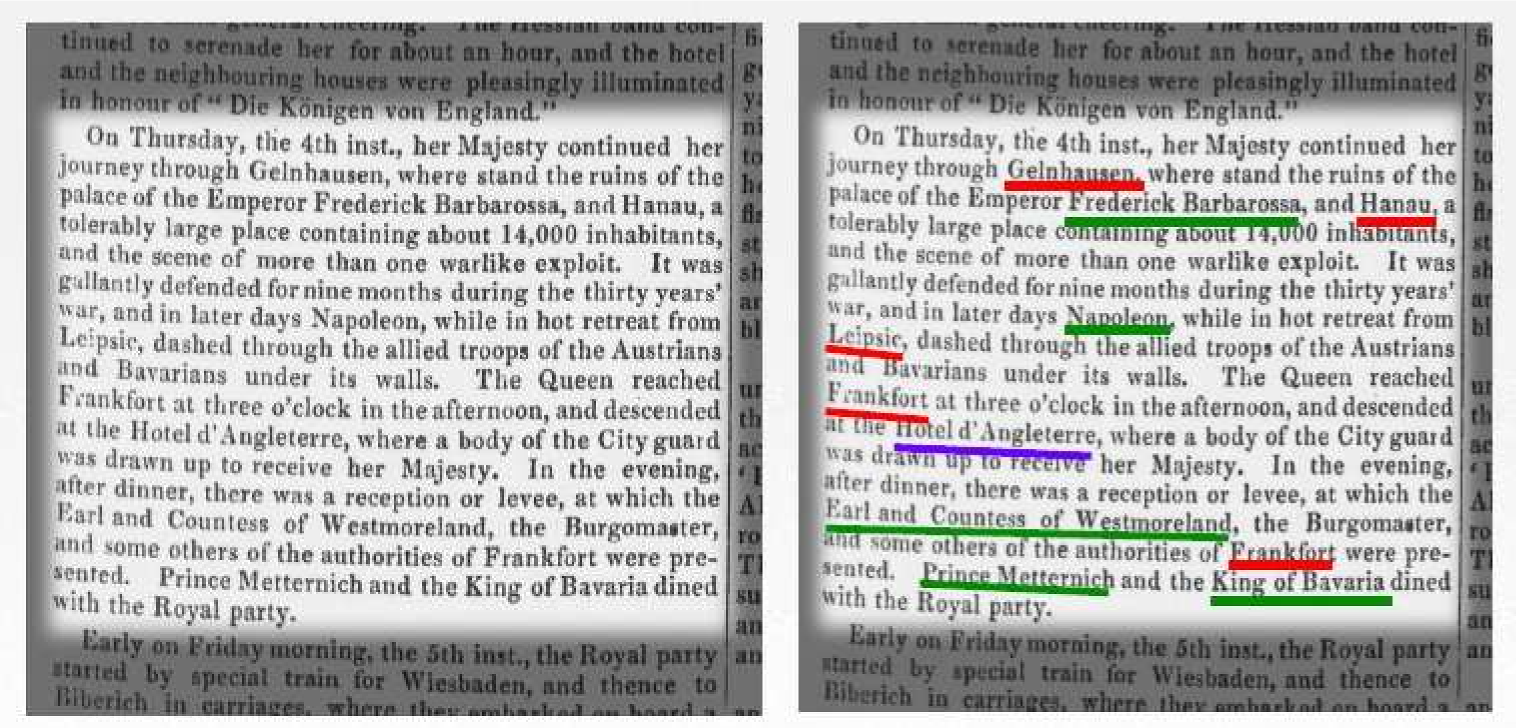
- UIBK enriches the OCR with structural information from their Document Understanding Platform
- CCS produces OCR and verification of column recognition, zoning, article segmentation, and page class recognition
- CCS provides libraries with a client technology for manual correction of recognition and segmentation results



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

1) Refinement - Named Entity Recognition

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



general cheering. The Hessian band continued to serenade her for about an hour, and the hotel and the neighbouring houses were pleasingly illuminated in honour of "Die Königen von England."

On Thursday, the 4th inst., her Majesty continued her journey through Gelnhausen, where stand the ruins of the palace of the Emperor Frederick Barbarossa, and Hanau, a tolerably large place containing about 14,000 inhabitants, and the scene of more than one warlike exploit. It was gallantly defended for nine months during the thirty years' war, and in later days Napoleon, while in hot retreat from Leipsic, dashed through the allied troops of the Austrians and Bavarians under its walls. The Queen reached Frankfort at three o'clock in the afternoon, and descended at the Hotel d'Angleterre, where a body of the City guard was drawn up to receive her Majesty. In the evening, after dinner, there was a reception or levee, at which the Earl and Countess of Westmoreland, the Burgomaster, and some others of the authorities of Frankfort were presented. Prince Metternich and the King of Bavaria dined with the Royal party.

Early on Friday morning, the 5th inst., the Royal party started by special train for Wiesbaden, and thence to Biberich in carriages, where they embarked on board a

general cheering. The Hessian band continued to serenade her for about an hour, and the hotel and the neighbouring houses were pleasingly illuminated in honour of "Die Königen von England."

On Thursday, the 4th inst., her Majesty continued her journey through Gelnhausen, where stand the ruins of the palace of the Emperor Frederick Barbarossa, and Hanau, a tolerably large place containing about 14,000 inhabitants, and the scene of more than one warlike exploit. It was gallantly defended for nine months during the thirty years' war, and in later days Napoleon, while in hot retreat from Leipsic, dashed through the allied troops of the Austrians and Bavarians under its walls. The Queen reached Frankfort at three o'clock in the afternoon, and descended at the Hotel d'Angleterre, where a body of the City guard was drawn up to receive her Majesty. In the evening, after dinner, there was a reception or levee, at which the Earl and Countess of Westmoreland, the Burgomaster, and some others of the authorities of Frankfort were presented. Prince Metternich and the King of Bavaria dined with the Royal party.

Early on Friday morning, the 5th inst., the Royal party started by special train for Wiesbaden, and thence to Biberich in carriages, where they embarked on board a

2) Analysis of existing digitised newspaper collections

- Project partners and others will be contacted until summer 2012 to analyse the extent of digitised newspapers collections at their institutions
    - Results will be embedded in “Zeitschriftendatenbank” of Staatsbibliothek zu Berlin (Union Catalogue of Serials)
    - Potential new partners for the extension of the network will be suggested by survey
- May also be useful to judge technical status of digitised data and as part of gathering descriptive metadata

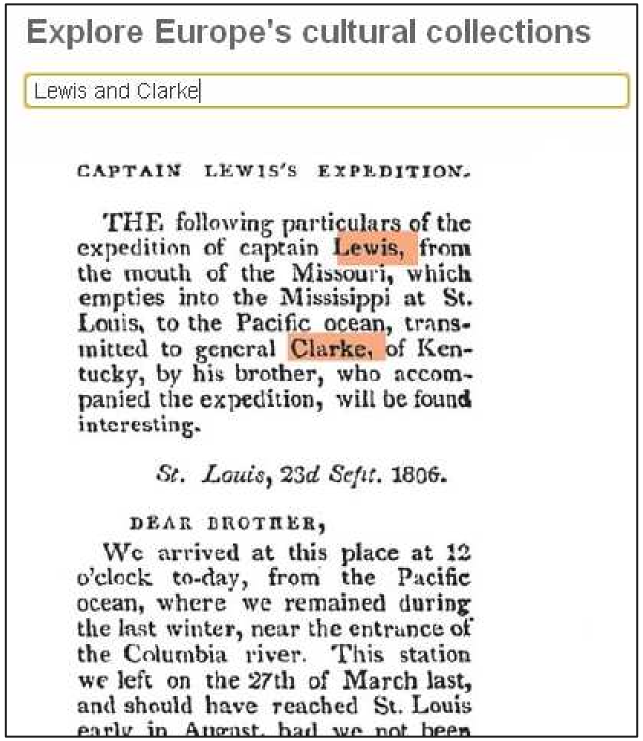
- If you hold digital newspaper collection and like to participate in the survey → please contact: survey@europeana-newspapers.eu/

3) Analysis of work & Best Practise Recommendations

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

4) Presentation & Access to full-texts

- 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
- This browser will be built in TEL during project; and exported to Europeana after the project



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 Missisippi 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, 23*d 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

5) Dissemination

- Objectives:
  - Establishment of publicity
  - Increasing usage of Europeana
  - Awareness raising among target groups
- Tasks:
  1. Media Communication
  2. Workshops and conferences
     - Three main dissemination workshops
     - National information days
     - Network extension
  3. Exploitation

… and more will come soon

- Detailed information will be available soon:

  → http://www.europeana-newspapers.eu/
  (Launch: End of April)

- Participating in the survey on European Newspapers:

  → survey@europeana-newspapers.eu

europeana
newspapers

SEVENTH FRAMEWORK PROGRAMME

ICT PSP
ICT POLICY SUPPORT PROGRAMME

Thank you for your attention!

Thorsten Siegmann, Staatsbibliothek zu Berlin

siegmann@eu1914-1918.eu

www.europeana-newspapers.eu/
(Launch: End of April)