europeana
newspapers

SEVENTH FRAMEWORK PROGRAMME

ICTPSP
ICT POLICY SUPPORT PROGRAMME

Europeana Newspapers Projekt

Deutsches Kulturerbe auf dem Weg in die Europeana II,
Berlin, 20.3.2014

Ulrike Kölsch, Projektkoordinatorin –
Staatsbibliothek zu Berlin – Preußischer Kulturbesitz

Überblick

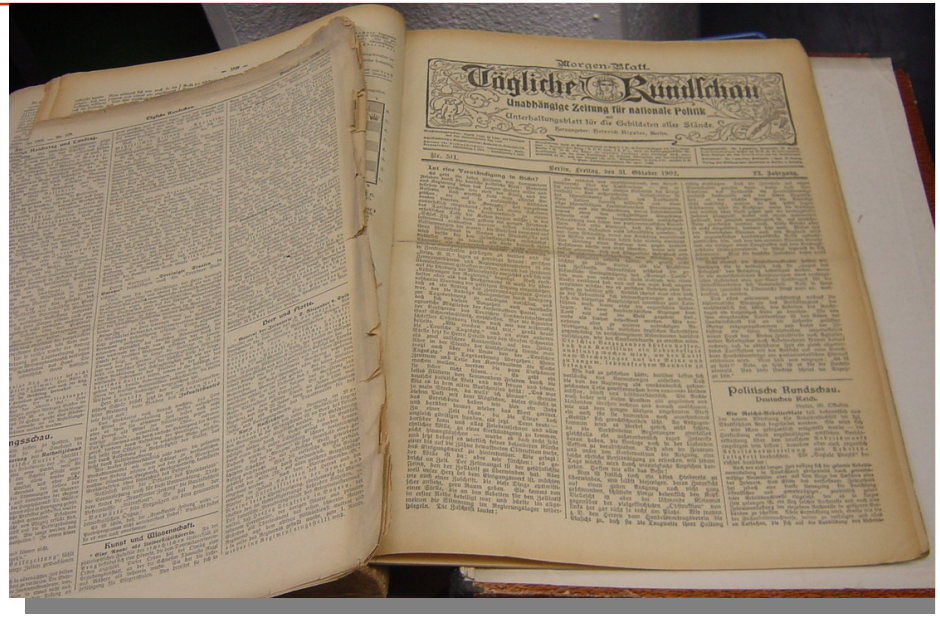
1. Warum Zeitungen?
2. Projektpartner
3. Arbeitspakete
4. Ergebnisse

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

Herausforderungen……

Brüchiges Papier

- *Fehlende Ausgaben*
- *Unvollständige Beilagen*
- *Bindung mit Textverlust*
- *Druckqualität*
- *Drucktypen*
- *Mengen an Material*
- *Rechtliche Unsicherheiten bei quasi-zeitgenössischem Material*



europeana newspapers

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

SEVENTH FRAMEWORK PROGRAMME

Nutzen

„Newspapers represent normal life in its own native locality"

"the newspaper and the historian" Lucy Maynard Salmon (1923)

„Die Zeitungen sind der Sekundenzeiger der Geschichte. Derselbe aber ist meistens nicht nur von unedlerem Metalle als die beiden anderen, sondern geht auch selten richtig."

(Arthur Schopenhauer)

Image: Nationaal Archief The Netherlands

europeana newspapers

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

SEVENTH FRAMEWORK PROGRAMME

ICTPSP ICT POLICY SUPPORT PROGRAMME

Nutzen

- Unmittelbarkeit und Geschwindigkeit der öffentlichen Debatte
- Thematisiert wird (fast) alles (Politik, Kultur, Sport, Wirtschaft, Wissenschaft, Werbung, etc.)
- Deshalb: großes, interdisziplinäres, spartenübergreifendes Interesse

Image: Nationaal Archief The Netherlands

europeana newspapers

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

SEVENTH FRAMEWORK PROGRAMME

ICTPSP ICT POLICY SUPPORT PROGRAMME

Projektpartner

01. State Library Berlin, Germany
02. National Library of the Netherlands
03. National Library of Estonia
04. National Library of Austria
05. National Library of Finland
06. State and University Library Hamburg, Germany
07. National Library of France
08. National Library of Poland
09. University of Salford
10. CCS Content Conversion Specialists GmbH
11. Stichting LIBER, Netherlands
12. National Library of Latvia
13. National Library of Turkey
14. University Library of Belgrade
15. University of Innsbruck
16. State Library Dr. Friedrich Tessmann, Italy
17. The British Library, UK
18. Europeana Foundation, Netherlands

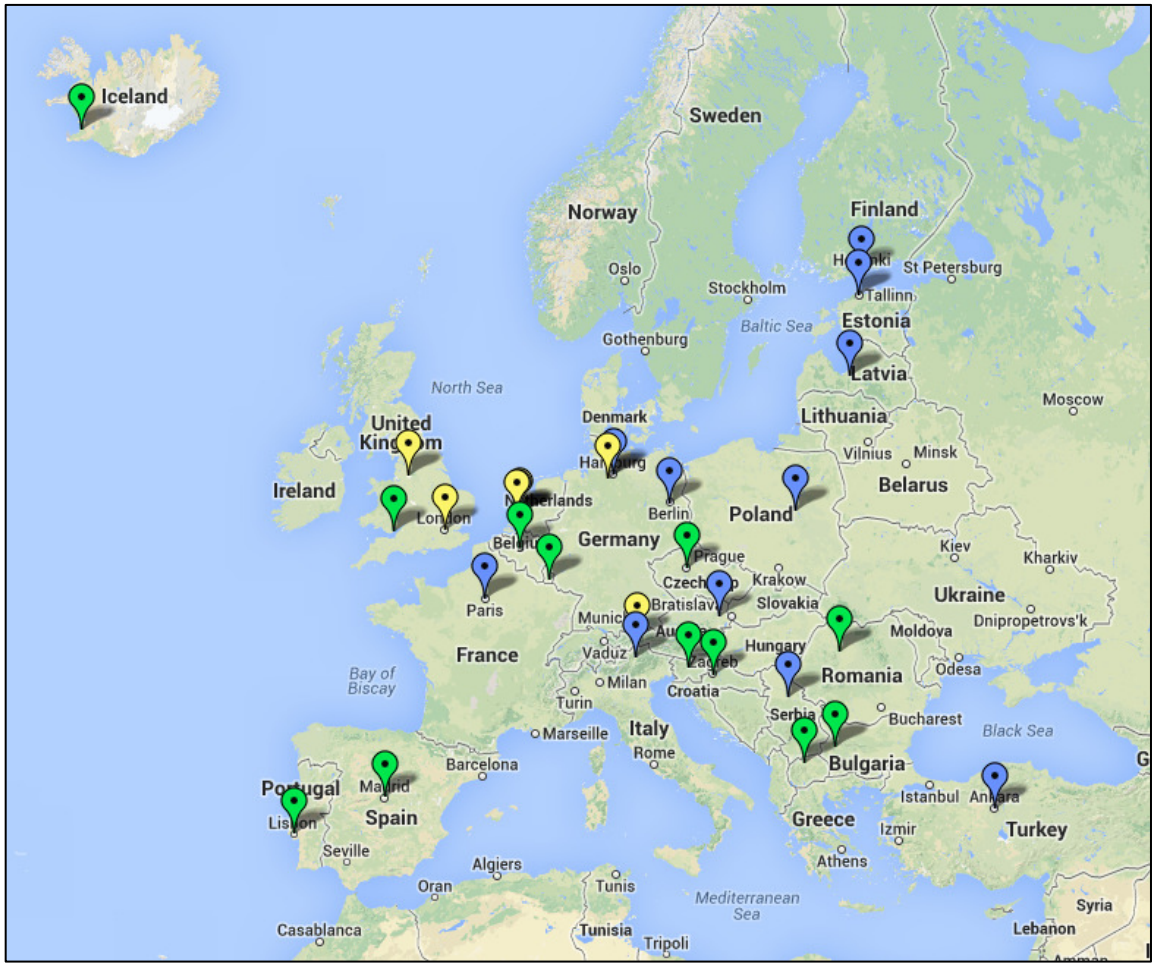
europeana newspapers

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

SEVENTH FRAMEWORK PROGRAMME

ICT PSP
ICT POLICY SUPPORT PROGRAMME

Projektpartner und Assoziierte Partner



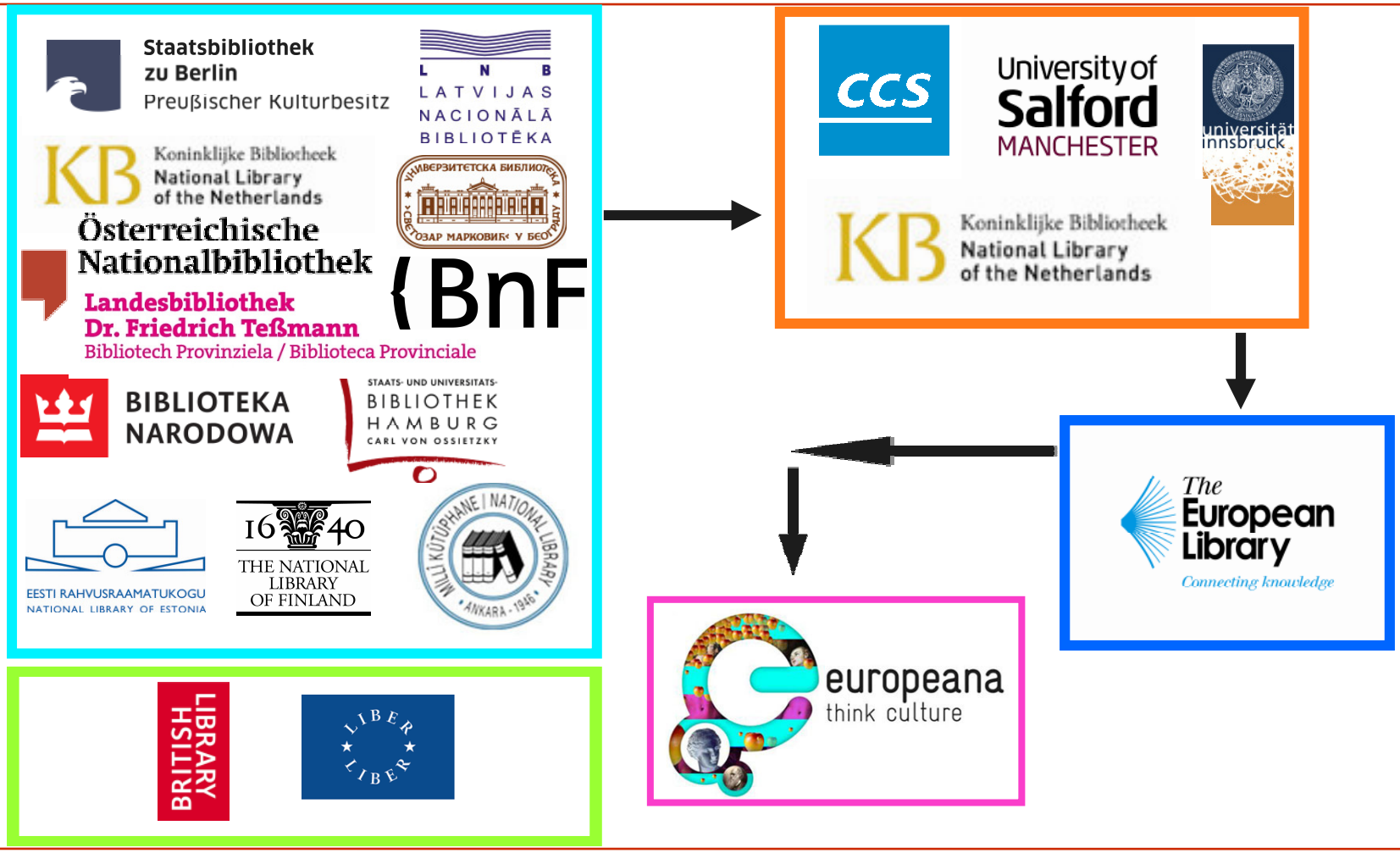
Blau – Daten

Gelb – Technik

Grün – Assoziierte Partner

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

Who



europeana newspapers

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

SEVENTH FRAMEWORK PROGRAMME
ICTPSP

Europeana Newspapers: Kernaufgaben

Übersicht der Arbeitspakete

WP 1	Project Coordination and Management
WP 2	Refinement of Digitized Newspapers
WP 3	Evaluation and Quality Assessment
WP 4	Aggregation and presentation of digitized newspapers for Europeana
WP 5	Metadata best practice recommendations
WP 6	Dissemination and Exploitation

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

Europeana Newspapers: Arbeitspaket 2 - Refinement of Digitized Newspapers

1. Digitale Abbildungen liegen lokal vor
2. Erzeugung von Volltexten (OCR) für 8 Mio Seiten
3. Erzeugung von Volltexten und Layoutanalyse (OLR) für 2 Mio Seiten
4. Named Entity Recognition (NER) für 2 Mio Seiten (Niederländisch, Französisch, Deutsch)
5. Entwicklung einer zeitungsspezifischen S&R-Umgebung bei The European Library für alle entsprechenden Daten
6. Metadatennachweis in Europeana

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

Europeana Newspapers: Arbeitspakete 3 und 5

WP 3 Evaluation and Quality Assessment

- Software-Ergebnisse ermöglichen eine a priori-Evaluierung zu erwartender Ergebnisse durch OCR- und OLR-Prozesse

WP 5 Metadata best practice recommendations

- Definition von Metadaten-Standards (Images, Volltexte, Anreicherungen, Strukturdaten)

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

Europeana Newspapers: Arbeitspaket 4 Aggregation and presentation of digitized newspapers for Europeanab

- Titel wurden von Bibliotheken ausgewählt (Kriterien: Verfügbarkeit der Images, Qualität der Images, Bereitschaft Metadaten unter einer CC0-Lizenz an Europeana zu liefern (vgl. Europeana Data Exchange Model), Beachtung rechtlicher Einschränkungen, Relevanz für Nutzer)
- Bearbeitete Daten werden für TEL und Europeana aggregiert
- Bibliotheken ermöglichen verschiedene Zugänge zu den Materialien:
  - (1) Text und Image befinden sich bei TEL und werden angezeigt
  - (2) Wie (1), aber Text und Image befinden sich auf einem bibliothekseitigen Imageserver
  - (3) Snippet view: Nutzer sieht nur einen Ausschnitt des Images
  - (4) Wie (3), aber Text und Image befinden sich auf einem bibliothekseitigen Imageserver
  - (5) Plain text view: Der Nutzer sieht den Volltext, aber nicht das Image
  - (6) Nur Metadaten: Der Nutzer sieht lediglich deskriptive Metadaten und wird für die Ansicht der Inhalte an eine lokale Website übergeben (ggf. fallen Gebühren an)

Weiterer Content: Metadaten aus europäischen Nationalbibliotheken und anderen größeren Einrichtungen (European Registry of Digitized Newspapers)

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

Zeitungen im Überblick

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

20 Sprachen

ca. 950 Titel

ca. 10 Mio Seiten bearbeitet

8 Mio OCR
2 Mio OLR
2 Mio NER

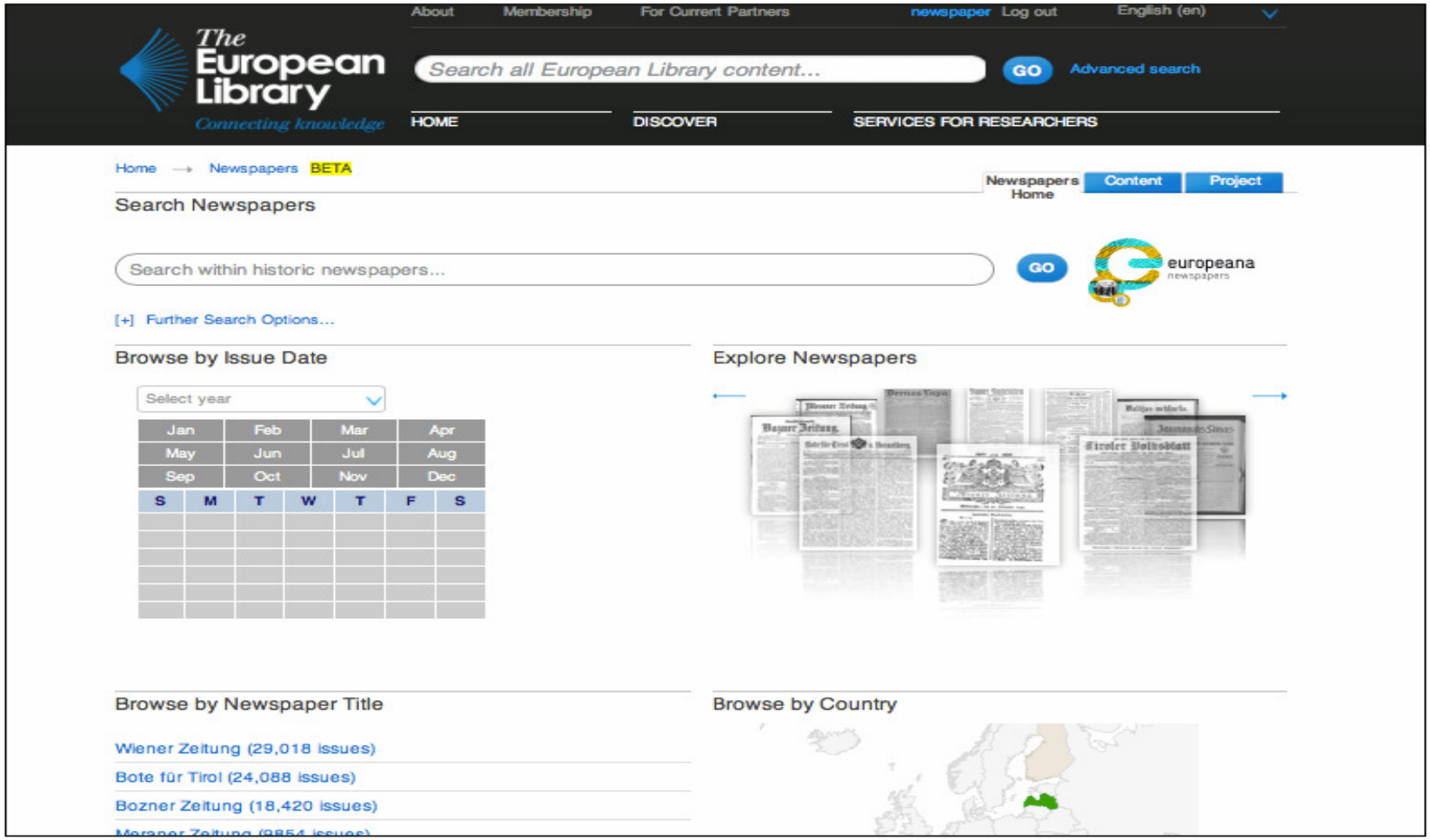
This project is partially funded under the ICT Policy Support Programme (ICT PSP) as part of the Competitiveness and Innovation Framework Programme by the European Community http://ec.europa.eu/ict_psp

Europeana Newspapers: Arbeitspaket 4 (Aggregation and presentation of digitized newspapers for Europeana)

- Entwicklung eines “European Newspapers Browsers”:
  - Erste Entwicklungsstufe:
    - Einfache Textsuche
    - Ergebniseinschränkung (Zeit, Land, Zeitungstitel, Sprache, besitzende Institution)
    - OCR-Ergebnisse werden dargestellt
    - Images sind zoomfähig
    - Aktive Links zwischen Images und Volltexten (wenn möglich)
    - Link zur jeweiligen besitzenden Institution
    - Zweite Entwicklungsstufe
    - Zusätzliche Informationen über bestimmte Zeitungstitel
    - Suche in Titeln (nicht ausschließlich in Texten)
    - Weitere navigierbare Visualisierungen
    - Suche / Stöbern nach automatisch erkannten Entitäten

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



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 Library / Europeana

- Volltextergebnisse unter www.theeuropeanlibrary.org
- Images und Metadaten unter www.europeana.eu
- Wie sieht die Zukunft aus?

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

Mind the gap

Herausforderungen:

- Rechtliches (IPR, "Persönlichkeitsrechte")
- Geschäftsmodelle und Unsicherheiten

Optionen:

- Übereinkünfte mit den Rechteinhabern
- Zusammenarbeit mit Autorenverbänden
- Zusammenarbeit mit den Verlagen/Verlegern (z.B. für weitere Digitalisierungsprojekte)

europeana newspapers

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

SEVENTH FRAMEWORK PROGRAMME

ICTPSP ICT POLICY SUPPORT PROGRAMME

THANK YOU!

Contact:

ulrike.koelsch@europeana-newspapers.eu

hans-joerg.lieder@sbb.spk-berlin.de

For more information, please see:

www.europeana-newspapers.eu

Twitter (@eurnews)

Facebook:

https://www.facebook.com/EuropeanaNewspapers

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