Making Historic Newspapers Available Online: Why, Where and How

IFLA Newspaper Pre-Conference

14 August 2014, Geneva

Hans-Jörg Lieder, Staatsbibliothek zu Berlin – Preußischer Kulturbesitz | Berlin State Library

Why Newspapers?

Cons:

- Originals are cumbersome objects
  - Prone to damage and destruction due to paper quality
  - Missing issues and pages
  - Difficult to deal with from a catalogueing point of view
  - Poor bindings
  - Funny fonts and fading ink
- Microforms may also be cumbersome objects
  - Skewed images, text loss
  - More missing issues and pages, plus duplicate pages

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

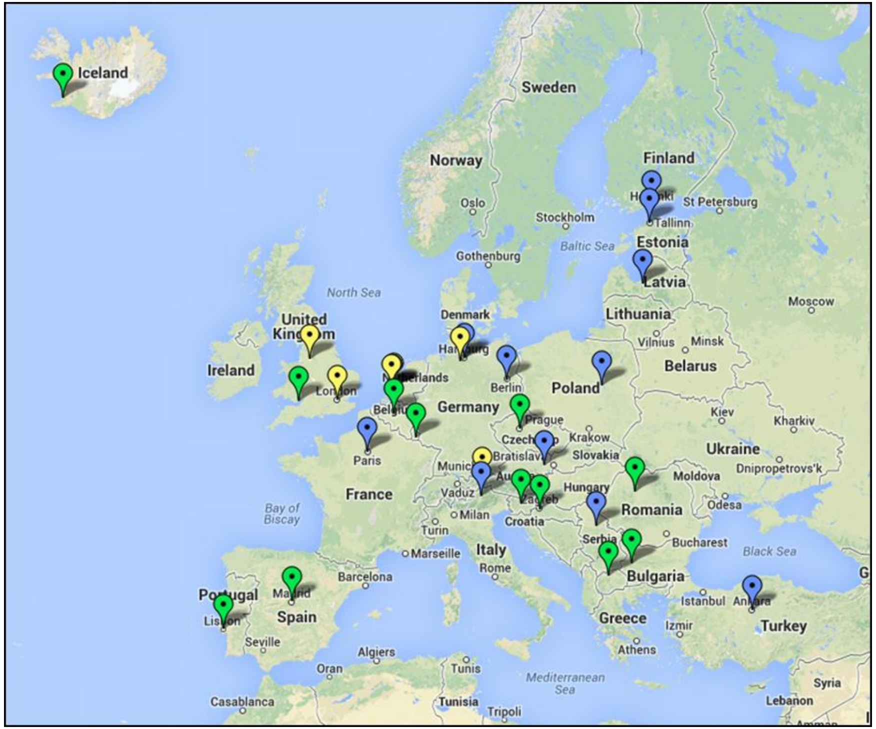
That‘s Why!

Pros:

- “Newspapers are the second hand of history”
  - Provide insights into history’s microstructure
  - Unlimited thematic scope
  - Interesting for all fields of scholarship, but also for the layman
  - Massive digital newspaper text corpora allow for new ways of research
  - A European perspective: significant contribution to the shaping of identities of peoples and individuals

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 Europeana Newspaper Project – Who?



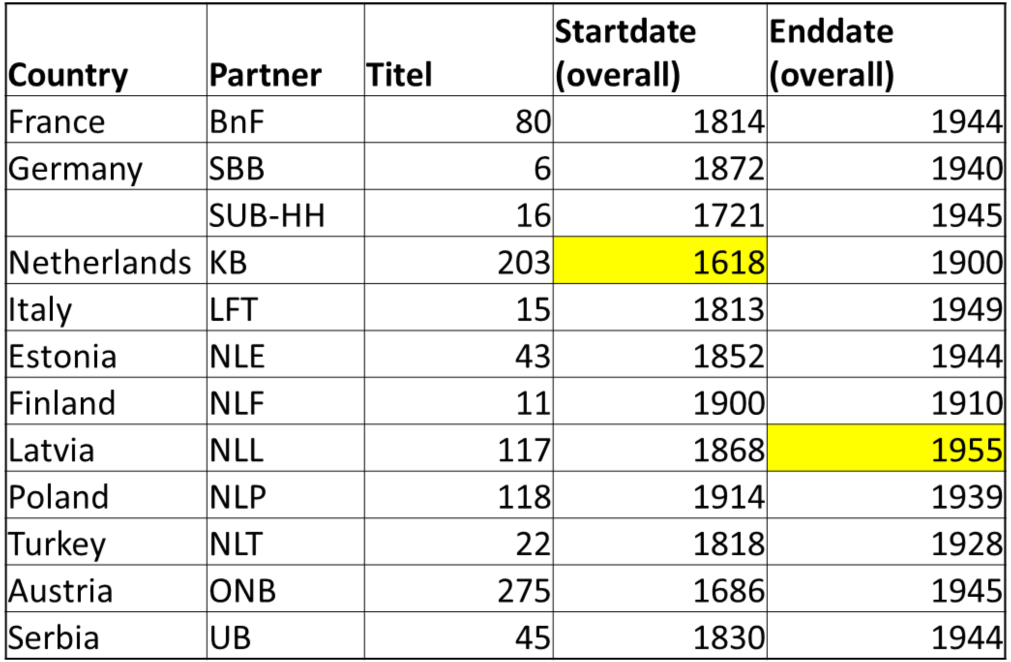
Blue – Content Providers

Yellow – Service Providers

Green – Associated Partners

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 Europeana Newspaper Project – What?



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
Finland	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 languages

ca. 950 titles

ca. 10m pages refined

- 8m OCR
- 2m OLR
- 2m NER

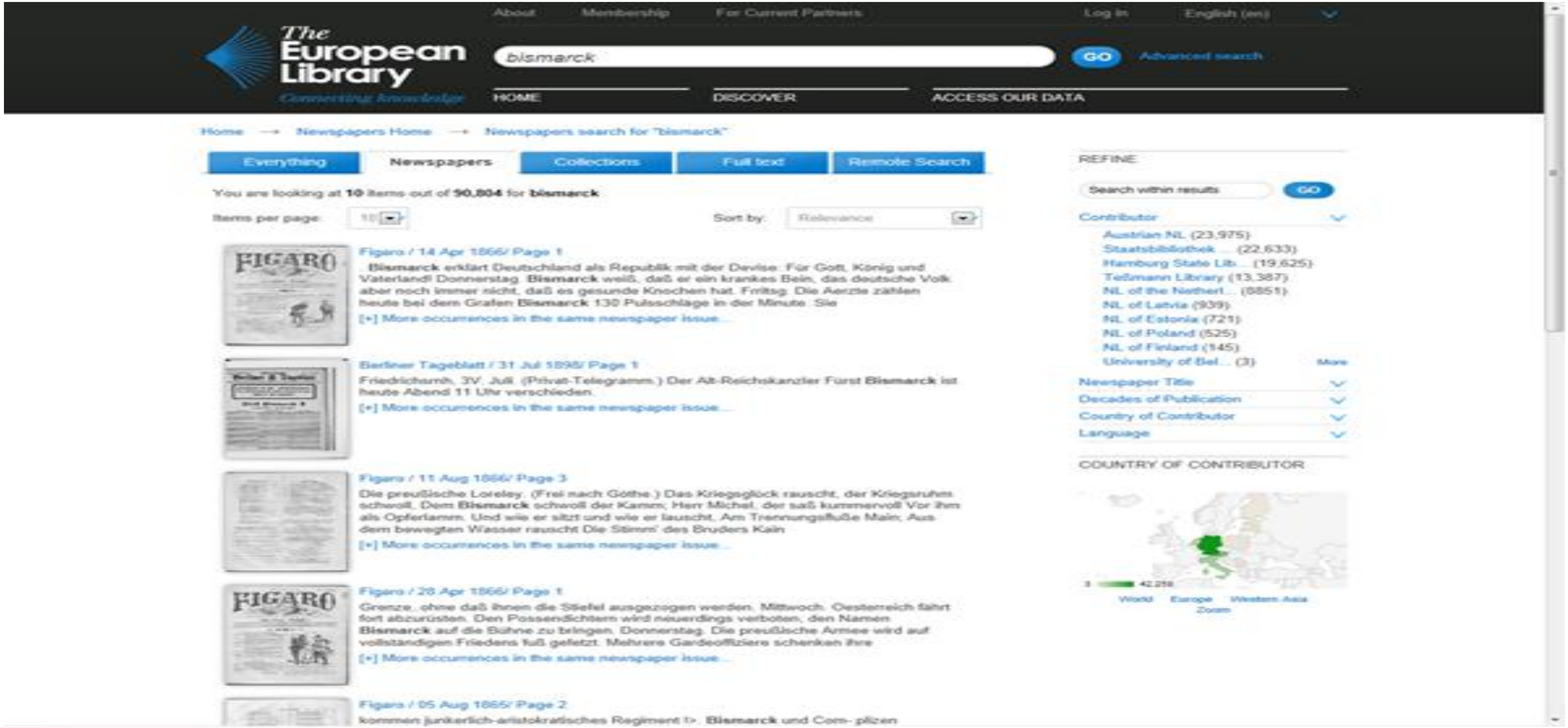
This project is partially funded under the ICT Policy Support Programme (ICT PSP) as part of the Competitiveness and Innovation Framework Programme by the European Community
http://ec.europa.eu/ict_psp

The Europeana Newspaper Project – What else?

- Tools for informed selection of newspapers for digitisation
- Specifications and tools for the creation and validation of OCR-ready images
- Large-scale, highly automated workflows for refinement (OCR, OLR, NER)
- Metadata best practice recommendations
- Transmission of data to European Portals and the Union Catalogue of Serials
- Presentation of results

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

What does it look like … in TEL?



Home → Newspapers Home → Newspapers search for "bismarck"

You are looking at **10** items out of **90,804** for **bismarck**

Items per page: 10

Sort by: Relevance

Figaro / 14 Apr 1866/ Page 1

Bismarck erklärt Deutschland als Republik mit der Devise: Für Gott, König und Vaterland! Donnerstag. **Bismarck** weiß, daß er ein krankes Bein, das deutsche Volk aber noch immer nicht, daß es gesunde Knochen hat. Freitag. Die Aerzte zählen heute bei dem Grafen **Bismarck** 130 Pulsschläge in der Minute. Sie

[+] More occurrences in the same newspaper issue...

Berliner Tageblatt / 31 Jul 1898/ Page 1

Friedrichsruh, 3V. Juli. (Privat-Telegramm.) Der Alt-Reichskanzler Fürst **Bismarck** ist heute Abend 11 Uhr verschieden.

[+] More occurrences in the same newspaper issue...

Figaro / 11 Aug 1866/ Page 3

Die preußische Loreley. (Frei nach Göthe.) Das Kriegsglück rauscht, der Kriegsruhm schwoll, Dem **Bismarck** schwoll der Kamm; Herr Michel, der saß kummervoll Vor ihm als Opferlamm. Und wie er sitzt und wie er lauscht, Am Trennungsfluße Main; Aus dem bewegten Wasser rauscht Die Stimm' des Bruders Kain

[+] More occurrences in the same newspaper issue...

Figaro / 28 Apr 1866/ Page 1

Grenze, ohne daß ihnen die Stiefel ausgezogen werden. Mittwoch. Oesterreich fährt fort abzurüsten. Den Possendichtern wird neuerdings verboten, den Namen **Bismarck** auf die Bühne zu bringen. Donnerstag. Die preußische Armee wird auf vollständigen Friedens fuß gefetzt. Mehrere Gardeoffiziere schenken ihre

[+] More occurrences in the same newspaper issue...

Figaro / 05 Aug 1865/ Page 2

kommen junkerlich-aristokratisches Regiment !>. **Bismarck** und Com- plizen

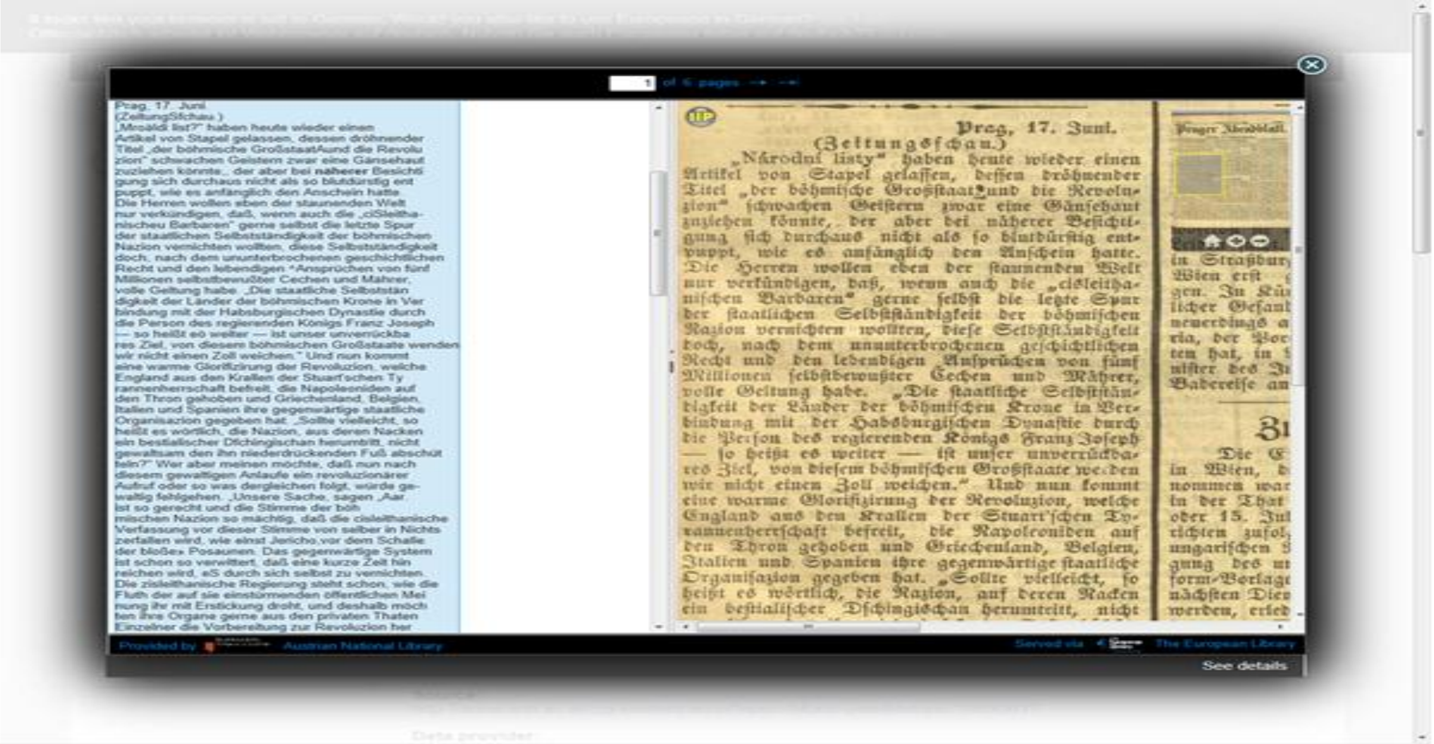
Contributor

Austrian NL	(23,975)
Staatsbibliothek ...	(22,633)
Hamburg State Lib...	(19,625)
Teßmann Library	(13,387)
NL of the Netherl...	(8851)
NL of Latvia	(939)
NL of Estonia	(721)
NL of Poland	(525)
NL of Finland	(145)
University of Bel...	(3)

More

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

What does it look like … in Europeana?



Prag. 17. Juni
(ZeitungSfchau.)
„Mroäldi list?" haben heute wieder einen Artikel von Stapel gelassen, dessen dröhnender Titel „der böhmische GroßstaatAund die Revolu zion" schwachen Geistern zwar eine Gänsehaut zuziehen könnte„ der aber bei näherer Besichti gung sich durchaus nicht als so blutdürstig ent puppt, wie es anfänglich den Anschein hatte Die Herren wollen eben der staunenden Welt nur verkündigen, daß, wenn auch die „ciSleitha-nischeu Barbaren" gerne selbst die letzte Spur der staatlichen Selbstständigkeit der böhmischen Nazion vernichten wollten, diese Selbstständigkeit doch, nach dem ununterbrochenen geschichtlichen Recht und den lebendigen *Ansprüchen von fünf Millionen selbstbewußter Cechen und Mähner, volle Geltung habe. „Die staatliche Selbststän digkeit der Länder der böhmischen Krone in Ver bindung mit der Habsburgischen Dynastie durch die Person des regierenden Königs Franz Joseph — so heißt eö weiter — ist unser unverrückba res Ziel, von diesem böhmischen Großstaate wenden wir nicht einen Zoll weichen." Und nun kommt eine warme Glorifizirung der Revoluzion, welche England aus den Krallen der Stuart'schen Ty rannenherrschaft befreit, die Napoleoniden auf den Thron gehoben und Griechenland, Belgien, Italien und Spanien ihre gegenwärtige staatliche Organisazion gegeben hat „Sollte vielleicht, so heißt es wörtlich, die Nazion, aus deren Nacken ein bestialischer Dfchingischan herumtritt, nicht gewaltsam den ihn niederdrückenden Fuß abschüt teln?" Wer aber meinen möchte, daß nun nach diesem gewaltigen Anlaufe ein revoluzionärer Aufruf oder so was dergleichen folgt, würde ge-waltig fehlgehen. „Unsere Sache, sagen „Aar ist so gerecht und die Stimme der böh mischen Nazion so mächtig, daß die cisleithanische Verfassung vor dieser Stimme von selber in Nichts zerfallen wird, wie einst Jericho vor dem Schalle der bloße» Posaunen. Das gegenwärtige System ist schon so verwittert, daß eine kurze Zeit hin reichen wird, eS durch sich selbst zu vernichten. Die zisleithanische Regierung steht schon, wie die Fluth der auf sie einstürmenden öffentlichen Mei nung ihr mit Erstickung droht, und deshalb möch ten ihre Organe gerne aus den privaten Thaten Einzelner die Vorbereitung zur Revoluzion her

Prag, 17. Juni.
(Zeitungsſchau.)
„Národní listy“ haben heute wieder einen Artikel von Stapel gelaſſen, deſſen dröhnender Titel „der böhmiſche Großſtaat und die Revoluzion“ ſchwachen Geiſtern zwar eine Gänſehaut zuziehen könnte, der aber bei näherer Beſichtigung ſich durchaus nicht als ſo blutdürſtig entpuppt, wie es anfänglich den Anſchein hatte. Die Herren wollen eben der ſtaunenden Welt nur verkündigen, daß, wenn auch die „cisleithaniſchen Barbaren“ gerne ſelbſt die letzte Spur der ſtaatlichen Selbſtſtändigkeit der böhmiſchen Nazion vernichten wollten, dieſe Selbſtſtändigkeit doch, nach dem ununterbrochenen geſchichtlichen Recht und den lebendigen Anſprüchen von fünf Millionen ſelbſtbewußter Čechen und Mährer, volle Geltung habe. „Die ſtaatliche Selbſtſtändigkeit der Länder der böhmiſchen Krone in Verbindung mit der Habsburgiſchen Dynaſtie durch die Perſon des regierenden Königs Franz Joſeph — ſo heißt es weiter — iſt unſer unverrückbares Ziel, von dieſem böhmiſchen Großſtaate werden wir nicht einen Zoll weichen.“ Und nun kommt eine warme Glorifizirung der Revoluzion, welche England aus den Krallen der Stuart'ſchen Tyrannenherrſchaft befreit, die Napoleoniden auf den Thron gehoben und Griechenland, Belgien, Italien und Spanien ihre gegenwärtige ſtaatliche Organiſazion gegeben hat. „Sollte vielleicht, ſo heißt es wörtlich, die Nazion, auf deren Nacken ein beſtialiſcher Dſchingiskhan herumtritt, nicht

Prager Abendblatt.

in Straßburg
Wien erſt
gen. Jn Kü
licher Geſand
neuerdings a
ria, der Bor
ten hat, in
niſter des Jn
Badereiſe an

Zi

Die E
in Wien, d
nommen war
in der That
oder 15. Jul
richten zufol
ungariſchen
gung des un
form-Vorlage
nächſten Dien
werden, erled

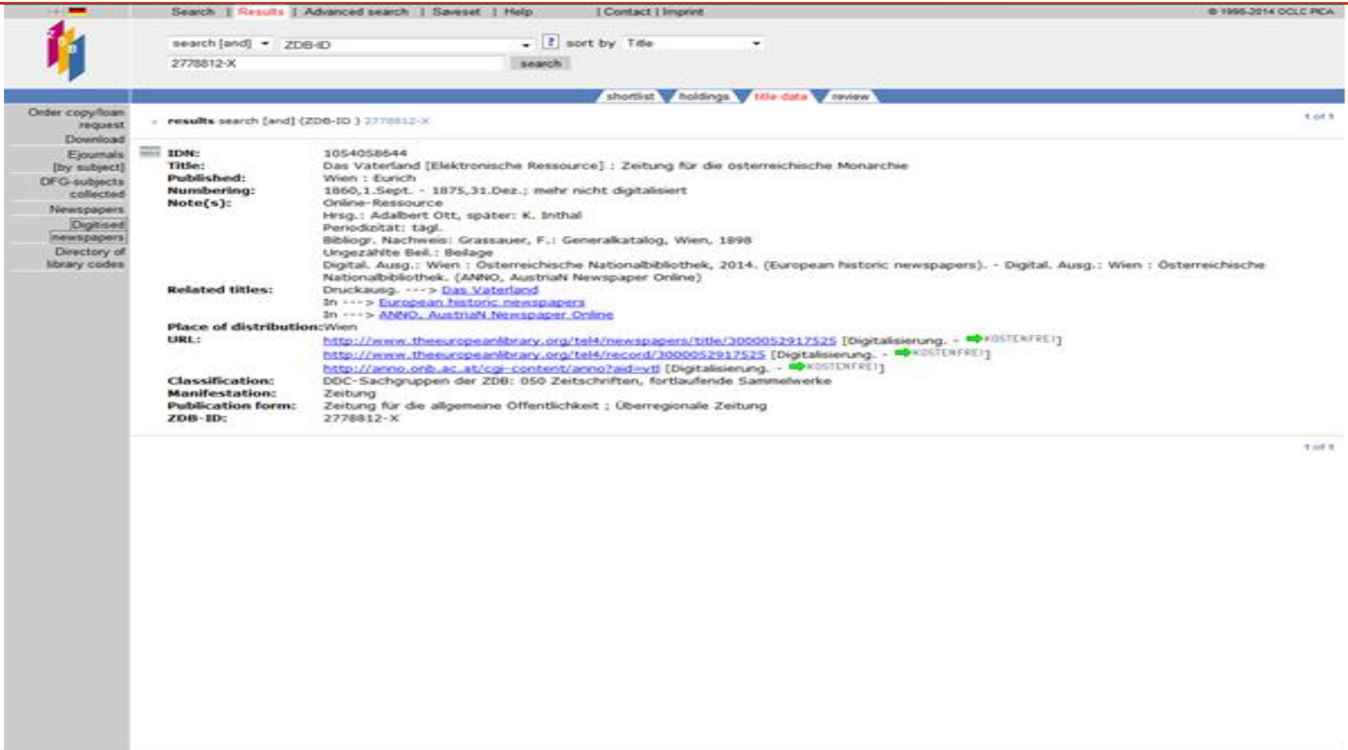
Provided by Austrian National Library
Served via The European Library
See details
Data provider

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

ICT PSP
ICT POLICY SUPPORT PROGRAMME

What does it look like … in the Union Catalogue of Serials?



results search [and] (ZDB-ID) 2778812-X

IDN:	1054058644
Title:	Das Vaterland [Elektronische Ressource] : Zeitung für die österreichische Monarchie
Published:	Wien : Eurich
Numbering:	1860,1.Sept. - 1875,31.Dez.; mehr nicht digitalisiert
Note(s):	Online-Ressource Hrsg.: Adalbert Ott, später: K. Inthal Periodizitat: tägl. Bibliogr. Nachweis: Grassauer, F.: Generalkatalog, Wien, 1898 Ungezählte Beil.: Beilage Digital. Ausg.: Wien : Österreichische Nationalbibliothek, 2014. (European historic newspapers). - Digital. Ausg.: Wien : Österreichische Nationalbibliothek. (ANNO, AustriaN Newspaper Online)
Related titles:	Druckausg. ---> Das Vaterland In ---> European historic newspapers In ---> ANNO, AustriaN Newspaper Online
Place of distribution:	Wien
URL:	http://www.theeuropeanlibrary.org/tel4/newspapers/title/3000052917525 [Digitalisierung. - KOSTENFREI] http://www.theeuropeanlibrary.org/tel4/record/3000052917525 [Digitalisierung. - KOSTENFREI] http://anno.onb.ac.at/cgi-content/anno?aid=vtl [Digitalisierung. - KOSTENFREI]
Classification:	DDC-Sachgruppen der ZDB: 050 Zeitschriften, fortlaufende Sammelwerke
Manifestation:	Zeitung
Publication form:	Zeitung für die allgemeine Öffentlichkeit ; Überregionale Zeitung
ZDB-ID:	2778812-X

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

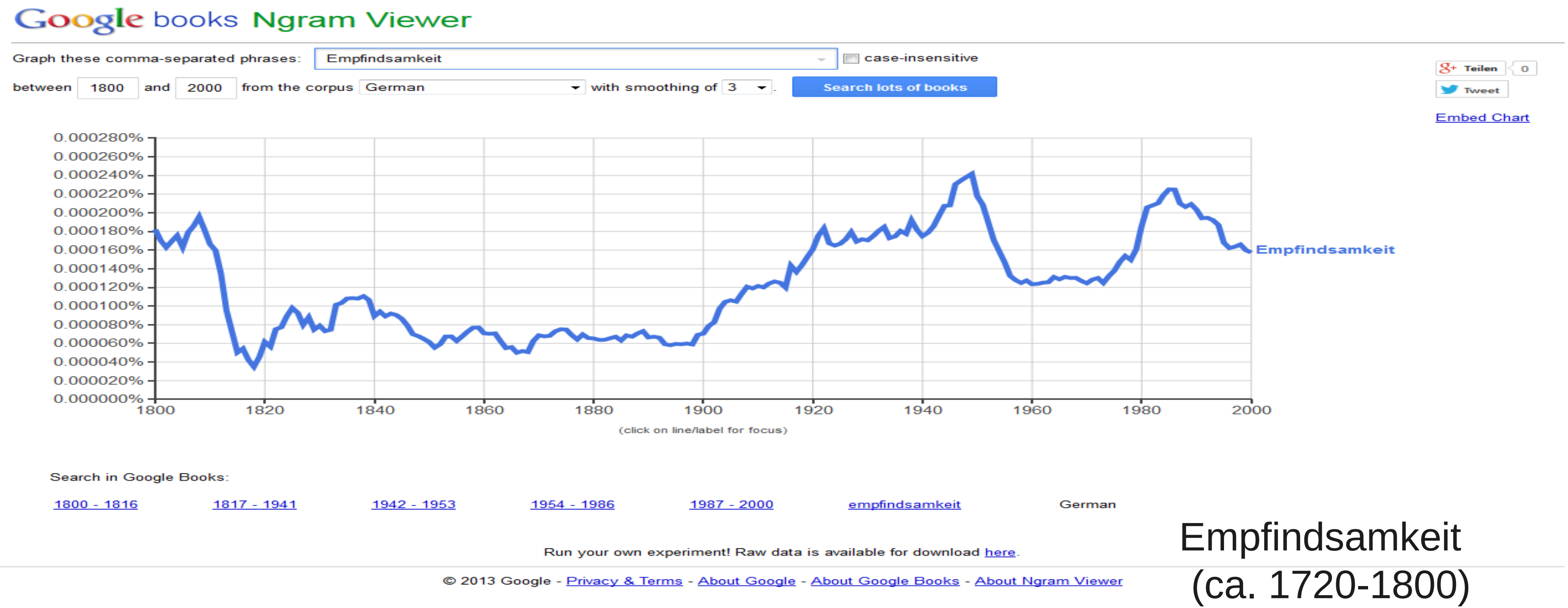
ICT PSP
ICT POLICY SUPPORT PROGRAMME

What about Services?

- Richest service portfolio available at local web pages (if you‘re lucky)
  - Calendar navigation, search in texts
  - filters to narrow down queries or result sets
  - mark-ups, annotations, links to other information resources, etc.
- Services at TEL
  - Calendar navigation, search in texts
  - Filters for searches: title, date, owning library
  - Filters for results: title, date, owning library, country, language

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

What about further Services? – An Example



Graph these comma-separated phrases: Empfindsamkeit □ case-insensitive

between 1800 and 2000 from the corpus German with smoothing of 3.

Search lots of books

Embed Chart

Search in Google Books:

1800 - 1816 1817 - 1941 1942 - 1953 1954 - 1986 1987 - 2000 empfindsamkeit German

Run your own experiment! Raw data is available for download here.

© 2013 Google - Privacy & Terms - About Google - About Google Books - About Ngram Viewer

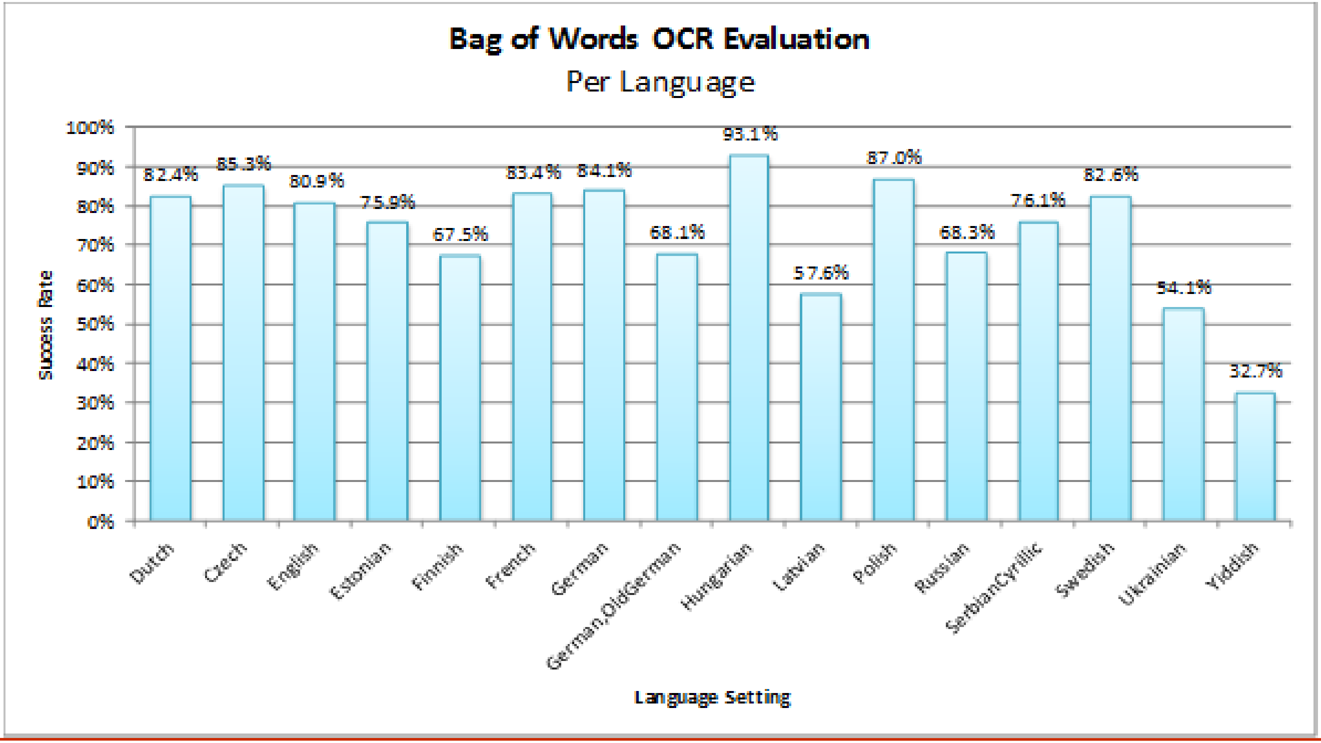
Empfindsamkeit
(ca. 1720-1800)
=
Sentimentalism

What about further Services?

- Natural language processing
- Text mining
- Visualisations
- Cross-media linking
- Semantic field analysis
- Links to other resources, librarian and non-librarian
- …
- LIBERATE YOUR DATA AND LEARN FROM YOUR USERS!

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

Digital Text Corpora: The Inconvenient Truth



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

What About Digital Text Corpora?

- Provide possibilities for corrections where data is presented
- Options for improvement
  - Automated corrections (index and page level)
  - Software aided corrections
  - Crowdsourcing
- Challenges: data synchronisation, update intervals, versioning…

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

Thank you for your attention!

IFLA Newspaper Pre-Conference

14 August 2014, Geneva

Hans-Jörg Lieder, Staatsbibliothek zu Berlin – Preußischer Kulturbesitz | Berlin State Library