REPORT

Project Acronym: Europeana Newspapers

Grant Agreement number: 297380

Project Title: A Gateway to European Newspapers Online

_____

Report on Europeana Newspapers Information day at the Berlin State Library (SBB)

_____

Revision: **1.0**

Authors: **Ulrike Kölsch (SBB), Hans-Jörg Lieder (SBB)**

Contributions: **Marieke Willems (LIBER)**

Revision History

Revision	Date	Author	Organisation	Description
0.1	March 2014	Ulrike Kölsch	SBB	Frist draft
0.2	April 2014	Hans-Jörg Lieder	SBB	Revision
0.3	16.04.2014	Ulrike Kölsch	SBB	Draft final version
0.4	22.04.2014	Marieke Willems	LIBER	Revision
1.0	22.04.2014	Ulrike Kölsch	SBB	Final version

Table of Contents

1 Executive Summary

Information days are an important part of the dissemination and exploitation activities in the project and highlight the project's objectives and outcomes at the respective national levels.

Specifics of the event are entirely subject to the respective partner's specific intentions. Thus local organizers can discuss specific issues and invite appropriate speakers and an appropriate audience.

Generally speaking, Information days increase the overall awareness of the project's activities and center around any of the following key messages:

1. The value of collaborating at a national and European level to make newspaper content available online (social, economic, cultural, research, technical etc.);
2. How to address the technical issues associated with making digitised newspaper content available;
3. The value of online digitised newspaper collections (i.e. highlight available content).

The third and final year of the project period started for the Berlin State Library, coordinator of the Europeana Newspapers Project, with a national Information day on 27 / 28 February, 2014.

The Information day in Berlin successfully brought together actors, mainly from the larger libraries of the country, from corresponding service providers and from the scientific communities, and the experts of the ENP.

The Information day was scheduled and designed to tie in particularly with the current activities of the German Research Foundation (Deutsche Forschungsgemeinschaft / DFG) with respect to their national scheme for the coordinated digitization of newspapers from German libraries.

2 Event Overview and Context

Since the beginning of this decade major German libraries and the DFG have talked about coordinated newspaper digitisation in Germany on the national level. It was well understood in Germany that the digitization of the very rich newspaper holdings of the country could only be reasonably undertaken in a national effort. Currently five major German libraries participate in a pilot phase, during which various aspects of newspaper digitization are being investigated in detail. The Berlin State Library (SBB), being the owner of one of the largest newspaper holdings in the country, plays a leading role in these national endeavours. SBB maintains the "Zeitschriftendatenbank" (Union Catalogue of Serials / ZDB), which is the world's largest data base regarding serials, including newspapers, and covers the holdings of more than 4.400 German libraries. SBB's major role in the mentioned context is, next to its own digitization projects (cf. http://zefys.staatsbibliothek-berlin.de/), the optimization of the ZDB for the specific purpose of allocating analogue holdings destined for digitization to individual libraries.

The German Information day therefore mainly aimed at making the expertise of the ENP available to the national German context. Due to the extended scope of both the ENP and the national agenda of German libraries, the Information day was organized as a two-day event, so there would be sufficient time for the German projects to present their activities, for the ENP colleagues to present in detail their approaches, methods and best practice recommendations, and for resulting discussions. The participation of three WP leaders (SBB, KB, UIBK) and a technical partner (CCS) of the project ensured that specific tasks in the project could be presented and discussed in great detail.

During the Information day many technical, organizational and legal challenges were discussed. Important topics included, amongst others, the selection and use of digital objects, aspects of long-term digital preservation, retro-digitization within the frameworks of different projects and programmes, metadata granularity, the use of standards, the development of online presentation functionalities and legal aspects with regard to newspaper holdings.

One further context of the German Information day is worthy of mentioning: Since the year 2000 the Federal Ministry of Education and Research (BMBF)[1] and the initiative "Science in Dialogue"[2] have promoted the "Year of Science"[3] in Germany. These Science Years highlight the exchange between the public and research communities. The motto of the Science Year 2014 is "The Digital Society"[4], and touches all aspects of the changing "information society". The Year of Science is accompanied by a variety of events in Germany and the ENP Information day could also take advantage of this and fitted nicely into that context as well.

3 Audience

The invitation to the German Information Day was an open one and directed at members of cultural heritage institutions, digital service providers and, since the event was coordinated with the German DFG-activities, the research community. Representatives of the major German libraries participating in the current DFG pilot phase were invited individually, and representatives from all these projects actually attended the conference.

Overall approximately 50 high-level participants attended the Information day.

.

---

[1] http://www.bmbf.de

[2] http://www.wissenschaft-im-dialog.de/

[3] http://www.bmbf.de/de/1758.php

[4] http://www.bmbf.de/en/23173.php

4 Presentations and Panel

Day 1:

The Information day was held in German. General information about the ENP was provided by Hans-Jörg Lieder (SBB), who gave an overview of the project's partners and workpackages, methods and practices, challenges and solutions, aims and objectives[5]. The latest project results were presented and the newspaper content shown online via the newspaper browser prototype developed by The European Library (TEL) (cf. http://www.theeuropeanlibrary.org/tel4/newspapers).

Hans-Jörg Lieder (SBB) gave an overview of the Europeana Newspapers Project.

Prof. Dr. Thomas Bürger from the Sächsische Landesbibliothek – Staats- und Universitätsbibliothek in Dresden[6] (State- and University Library of Saxonia, Dresden) presented an overview of the newspaper digitization landscape in Germany with particular focus on the DFG context.[7]

The director of the Staats- und Universitätsbibliothek Bremen[8] (State- and University Library, Bremen) could unfortunately not attend the information day and was represented by Dr. Maria Hermes, project coordinator for newspaper digitisation in Bremen. She talked about "Digitising all

---

[5] http://www.slideshare.net/Europeana_Newspapers/europeana-newspapers-german-infoday-32391360

[6] www.slub-dresden.de

[7] http://www.slideshare.net/Europeana_Newspapers/enp-id-berlinslubbuerger

[8] http://www.suub.uni-bremen.de/home-english/

German newspapers of the 17th century: A project with complexity! "[9], a DFG pilot project supported by the DFG. She highlighted the particulars of this project, namely the digitization from secondary forms (microfilm and/ or microfiche), specific quality traits (paper, fonts, print quality, faded ink, etc.) and challenges involved in refining the digital images with OCR.

Kay Heiligenhaus from the company semantics[10] presented another DFG pilot project that is undertaken at the Universitäts- und Landesbibliothek Sachsen-Anhalt[11] in Halle (University- and State Library of Saxony-Anhalt)[12]. The focus of this lecture was on cost-reducing software tools that could be used in digitization workflows. A very lively discussion was ensured and centred around possibilities and limitations of automated software support in the digitization of newspapers. It was universally agreed that without employing supporting software tools, costs for the digitization of newspapers would simply be prohibitive.

Clemens Neudecker (KB), leader of Workpackage 2, giving the last presentation of the day, talked at length about the OCR processing in the ENP[13]. Many of the practical details, e.g. technical parameters of the digital images, conversion and binarisation of image files for OCR etc., met a great interest and it was universally acknowledged that the ENP specifications are of great help also to other institutions.

Clemens Neudecker (KB), talked about the OCR processing in the ENP.

---

[9] http://www.slideshare.net/Europeana_Newspapers/europeana-newspapers-infoday

[10] http://www.semantics.de/index.html

[11] http://bibliothek.uni-halle.de/

[12] http://www.slideshare.net/Europeana_Newspapers/europeana-newspapers-german-infoday-32391406

[13] http://www.slideshare.net/Europeana_Newspapers/enp-id-berlinkbneudecker

Day 2:

Clemens Neudecker continued on day 2 of the Information day with a detailed presentation of Workpackage 3 "Evaluation and Quality Control"[14]. This presentation created a lot of interest, because the librarians present at the conference saw the technical uncertainties of digitization projects as a major obstacle for the successful organisation of such projects. Being able to predict results in a very structured, i.e. mathematical way, was recognized as a great help to decision makers in libraries.

Günter Mühlberger (UIBK) presented WP5 with a focus on structural metadata for historical newspapers[15]. The shown concept includes the introduction of three basic classifications levels – Newspaper Content Units, Newspaper Content Sections, Newspaper Structural Elements – and corresponding lower level classifications of content. It has to be stated that the ENP concept for structural newspaper metadata is the first detailed articulation of such concept and therefore meets a lot of interest. The concept was discussed in detail and a dedicated workshop about structural metadata that is forseen by the ENP consortium for the second half of this year was advertised. Various non-partner libraries expressed the desire to participate in this workshop.

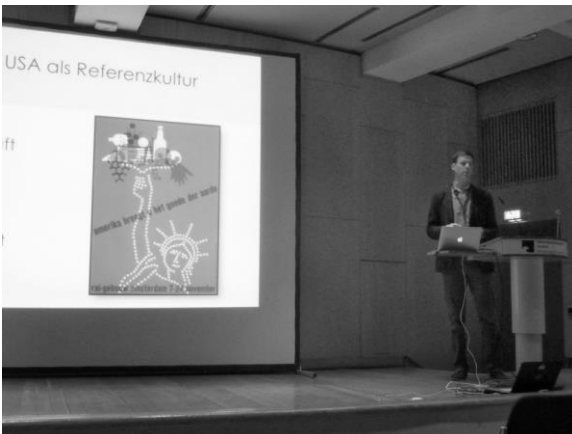*Günter Mühlberger (UIBK) focused on structural metadata for historical newspapers*

---

[14] http://www.slideshare.net/Europeana_Newspapers/europeana-newspapers-german-infoday-quality-assessment

[15] http://www.slideshare.net/Europeana_Newspapers/europeana-newspapers-german-infoday

"Optical Layout Recognition (OLR): Generation and Use of structural data" [16] was the subject of a talk by Claus Gravenhorst (CCS), technical service provider in the ENP.

He presented the OLR workflow in the project, the analysis of layout and structural elements in newspapers, and successfully argued the case for structural matadata, showing the benefits of such metadata both for S&R and presentation functionalities.

The last presentation of the Information day was delivered by Dr. Pim Huijnen, Falculty of Science at the University of Utrecht. Here the perspective of a historical researcher who uses fairly large digital text corpora as a source for his research[17]. Huijnen pointed out that newspapers and journals are of particular interest to scholars of history and other subjects. Though we are necessarily experimenting in text mining, the possibilities seem almost infinite. In different scenarios he showed various possibilities of the application and evaluation of strategies of large scale text analysis. He also talked about difficulties and limitations of the source material, e.g. data is not representative, data is incoherent and faulty due to poor OCR and/ or OLR results.

Dr. Pim Huijnen, Falculty of Science at the University of Utrecht. Presented the perspective of a historical researcher who uses fairly large digital text corpora as a source for his researc.h

The Information day closed with a panel: Hans-Jörg Lieder (SBB), Clemens Neudecker (KB), Kay Heiligenhaus (semantics), Günter Mühlberger (UIBK), Pim Huijnen (University of Utrecht) and Claus Gravenhorst (CCS) answered questions from the audience and discussed various aspects and purposes of digitizing newspapers.

---

[16] http://www.slideshare.net/Europeana_Newspapers/enp-id-berlinccsgravenhorst

[17] http://www.slideshare.net/Europeana_Newspapers/europeana-newspapers-german-infoday-digitale-zeitungsarchive-als-quellen-digitaler-geschichtsforschung

Among the interesting results of this discussion was the unanimous view that libraries should rather engage in mass digitization projects than indulge in deep levels of cataloguing in small scale digitization projects. It is also noteworthy, that the participants agreed that libraries should rather make data directly accessible to researchers (e.g. via APIs) than implement intricate service layers in their respective websites.

Information day was deemed to be a very successful event both by organisers and participants.

Discussion panel first picture from left to right: Günter Mühlberger (UIBK), Kay Heiligenhaus (semantics) and Claus Gravenhorst (CCS).
Second picture from left to right: Hans-Jörg Lieder (SBB), Clemens Neudecker (KB) and Pim Huijnen (University of Utrecht).

5 Conclusions

Since the beginning of this decade major German libraries and the DFG have talked about coordinated newspaper digitisation in Germany on the national level. The German Information day therefore mainly aimed at making the expertise of the ENP available to the national German context.

The Information day in Berlin successfully brought together over 50 participants from the larger German libraries, corresponding service providers, scientific communities and ENP experts. The Information day was scheduled and designed to tie in particularly with the current activities of the DFG with respect to their national scheme for the coordinated digitization of newspapers from German libraries.

During the day, four main learning points were noted:

- Libraries should rather engage in mass digitization projects than indulge in deep levels of cataloguing in small scale digitization projects.
- Without employing supporting software tools, costs for the digitization of newspapers would simply be prohibitive.
- Librarians present at the conference saw the technical uncertainties of digitization projects as a major obstacle for the successful organisation of such projects.
- Newspapers and journals are of particular interest to scholars of history and other subjects, where experimenting in text mining show almost infinite possibilities.

ANNEX I: Agenda

Europeana Newspapers Project - Agenda Information Day - Berlin 27.02. - 28.02. 2014		
PROJECT NUMBER	297380	
PROJECT ACRONYM	Europeana Newspapers Project (ENP)	
Veranstaltungsort	Staatsbibliothek zu Berlin, Simon-Bolivar-Saal, Potsdamer Str.33-35, 10785 Berlin	
Donnerstag 27.02. 2015	**Titel**	**Sprecher**
12.00 - 13.00	Registrierung/Welcome coffee	
13.00 - 13.45	Vorstellung und Project Overview Europeana Newspapers	Hans-Jörg Lieder, Staatsbibliothek zu Berlin, ENP
13.45 - 14.15	Zeitungsdigitalisierung in Deutschland im Kontext des DFG-Programms Digitale Transformation	Thomas Bürger, Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden (SLUB)
14.15 - 14.45	Kaffeepause	
14.45 - 15.30	Digitalisierung der vollständigen deutschsprachigen Zeitungen des 17. Jahrhunderts: Ein Projekt mit Komplexität!	Maria Elisabeth Müller, Staats- und Universitätsbibliothek Bremen
15:30 - 16:15	Der Einsatz automatisierter Verfahren zur vertieften Erschließung historischer Zeitungsbestände in einem DFG-Pilotprojekt der ULB Sachsen- Anhalt	Kay Heiligenhaus, semantics, Halle
16:15 - 16:45	Kaffeepause	
16:45 - 17.15	Digitale Zeitungen - OCR Verarbeitung in Europeana Newspapers	Clemens Neudecker, Nationalbibliothek der Niederlande, ENP
Freitag, 28.02. 2014	**Titel**	**Sprecher**
9.00 - 9.30	Welcome coffee	
9.30 - 10.00	Evaluierung und Qualitätskontrolle	Clemens Neudecker, Nationalbibliothek der Niederlande
10.00 - 10.45	Strukturelle Metadaten für historische Zeitungen. Überlegungen zu einem Standardmodell	Günter Mühlberger, Universität Innsbruck, ENP
10.45 - 11.15	Optical Layout Recognition (OLR): Generierung und Nutzung von Strukturdaten	Claus Gravenhorst, CCS-Content Conversion Specialists Hamburg, ENP
11.15 - 11.45	Kaffeepause	
11.45 - 12.30	Digitale Zeitungsarchive als Quellen (digitaler) Geschichtsforschung	Pim Huijnen, Faculty of Science, University of Utrecht
12.30 - 13:15	Abschlusspanel	

ANNEX II: List of participants

Teilnehmerliste

Project: A Gateway to European Newspapers Online

Project reference no: 297380

Datum: Donnerstag, 27.02.2014

Ort: Staatsbibliothek zu Berlin

No.	Name		Institution	Unterschrift	Newsletter per email
1.	Susann	Solberg	Deutsche Nationalbibliothek		
2.	Jessica	Hubrich	Deutsche Nationalbibliothek		
3.	Chrisoph	Albers	Staatsbibliothek zu Berlin		
4.	Christian	Mathieu	Staatsbibliothek zu Berlin		
5.	Thomas	Kees	Saarländische Universitäts-und Landesbibliothek		
6.	Michael	Lörzer	Thüringer Universitäts-und Landesbibliothek Jena		
7.	Jens	Peters	Walter Nagel GmbH & Co. KG		
8.	Klaus	Lütkehölter	Walter Nagel GmbH & Co. KG		

No.	First name	Last name	Organisation	Signature	
9.	Stefanie	Geißler	ULB Düsseldorf		
10.	Steffen	Hankiewicz	intranda GmbH		
11.	Robert	Sehr	intranda GmbH		
12.	Antoine	Cauty			
13.	Klaus	Zaepke	Universitaetsbibliothek der HU Berlin		
14.	Martha	Ganter	Freie Universität - Bibliothek		
15.	Philipp	Berger	HU Berlin		
16.	Ludger	Syré	Badische Landesbibliothek		
17.	Hartmut	Janczikowski	CCS Content Conversion Specialists GmbH		
18.	Carsten	Strieker	Universitätsbibliothek der HU Berlin		
19.	Doris	Schweißgut	Staatsbibliothek zu Berlin		
20.	Joaquín	Moreno	FU Berlin		
21.	Jorinde	Singh			
23.	Wolfram	Fiedler	Herrmann und Kraemer GmbH		
24.	Yong-Mi	Rauch	Humboldt-Universität zu Berlin, Universitätsbibliothek		

No.	First name	Last name	Organisation	Signature	
25.	Holger	Schie	MIK-CENTER GMBH		
26.	Ines	Kolbe	Stiftung Deutsche Kinemathek		
27.	Birgit	Umathum	Stiftung Deutsche Kinemathek		
28.	Annett	Arnold	Staatsbibliothek zu Berlin		
29.	Bodo	Brennicke	Staatsbibliothek zu Berlin		
30.	Bernd	Augustin	Staatsbibliothek zu Berlin		
31.	Karin	Czwinkalik	Staatsbibliothek zu Berlin		
32.	Timon	Gehrhardt	Staatsbibliothek zu Berlin		
33.	Andrea	Stei	Staatsbibliothek zu Berlin		
34.	Carmen	Thomas	Staatsbibliothek zu Berlin		
35.	Thorsten	Siegmann	Europeana Collections 1914-1918		
36.	Lis	Mielke	Justus-Liebig-Universität Gießen		
37.	Ulrike	Kammer	Justus-Liebig-Universität Gießen		
38.	Astrid	Großgarten	ImageWare Components GmbH		
39.	Hans-Christian	Pust	WLB Stuttgart		
40.	Matthias	Schulz	Berlin-Brandenburgische Akademie der Wissenschaften		

41.	Eva-Maria	Breuer	Staatsbibliothek zu Berlin		
42.	Katrin	Goldschmidt	Universität Potsdam	*signature*	
43.	Ulrike	Reichhoff	Berlin-Brandenburgische Akademie der Wissenschaften		
44.	Hans-Jörg	Lieder	Staatsbibliothek zu Berlin	*signature*	
45.	Ulrike	Kölsch	Europeana Newspapers	*signature*	
46.	Claus	Gravenhorst	CCS Content Conversion Specialists GmbH	*signature*	
47.	Clemens	Neudecker	Nationalbibliothek der Niederlande	*signature*	
48.	Maria	Hermes	Staats- und Universitaetsbibliothek Bremen	*signature*	
49.	Kay	Heiligenhaus	semantics, Halle	*signature*	
50.	Pim	Huijnen	Universität Utrecht	*signature*	
51.	Günter	Mühlberger	Universität Innsbruck	*signature*	
52.	Thomas	Bürger	Sächsische Landesbibliothek Staats- und Universitätsbibliothek Dresden	*signature*	
53.	Joachim	Bauer	CCS Content Conversion Specialists GmbH	*signature*	
54.	Max	Kaiser	Österreichische Nationalbibliothek	*signature*	
55.	Viola	Rosenau	Die Kulturgutscanner	*signature*	

56.	Knut	Lohse	Staatsbibliothek zu Berlin		
	Manfred	Nölte	SuUB Bremen	*signature*	
	Sandra	Kobel		*signature*	