DELIVERABLE

Project Acronym: Europeana Newspapers

Grant Agreement number: 297380

Project Title: A Gateway to European Newspapers Online

D4.3 Report on European Registry for Digital Newspapers Online

---

Revision: 1.0

Author: **Hans-Jörg Lieder (SBB)**

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	**Public**	**x**
C	**Confidential, only for members of the consortium and the Commission Services**	

Revision History

Revision	Date	Author	Organisation	Description
0.1	15-07-2014	Hans-Jörg Lieder	SBB	Version for internal review
0.2	21-07-2014	Hans-Jörg Lieder	SBB	Update after internal discussions
0.3	23-07-2014	Clemens Neudecker	KB	Internal review
1.0	28-07-2014	Hans-Jörg Lieder	SBB	Final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1. Executive Summary 4
2. Newspaper Data in Europeana Newspapers 5
3. Union Catalogue of Serials / Zeitschriftendatenbank 7
4. Europeana Newspapers Data in the ZDB 11
5. Outlook 13

1. Executive Summary

Following a survey that was undertaken within Task 4.1 Survey on digital newspaper collections in Europe, the Description of Work has foreseen the creation of a European Registry for Digital Newspapers. The European Registry for Digital Newspapers makes available the newspaper metadata collected during the survey as part of the well-established German Union Catalogue of Serials (Zeitschriftendatenbank – ZDB), which is the world's largest dedicated database for serial materials.

During the course of the Europeana Newspapers project, a network of eleven associated partners was established. These associated partners were chosen from institutions that responded to the survey and are supported by small financial contributions that enable them to participate in project events. In turn, associated partners have agreed to provide newspaper (meta)data to the project. In order to make the European Registry for Digital Newspapers a more comprehensive and meaningful service, it was decided to also integrate all metadata from project partners.[1]

As a result the following metadata for ca. 18 million newspaper pages will be made available in the ZDB:

- Title
- Date range
- Language
- Place/region of distribution
- A link to the actual resource (if applicable)

Thus the ZDB will, thanks to Europeana Newspapers, enrich its service portfolio and serve as a central access point to digital and digitised newspapers for international audiences. In order to sustain this service, ZDB will, in cooperation with The European Library (TEL) and Europeana, establish workflows to ensure the continued transfer of data from the European portals after the end of the Europeana Newspapers project.

[1] The extra staff effort to carry out the additional task of integrating metadata from project partners is provided by SBB.

2. Newspaper Data in Europeana Newspapers

Two types of newspaper data are created and/or collected within the Europeana Newspapers project:

1. Digital images with full text created via Optical Character Recognition (OCR) within the project
2. Metadata about digital newspapers collected from the network of associated partners

Ad 1: Partner libraries provide digital images of newspaper pages to technical partners University of Innsbruck (UIBK) or Content Conversion Specialists (CCS) for refinement, i.e. the creation of full text with OCR for around 8 million pages and, in addition, OCR plus Optical Layout Recognition (OLR) for around 2 million pages.

After the refinement process, content and metadata are sent to partner The European Library (TEL) for indexing and ingestion into the TEL newspaper portal.[2] TEL transforms the metadata to meet their internal format requirements and assures that a rights statement, as required by Europeana, is available. Following the ingestion process, TEL forwards the metadata to both Europeana (including thumbnail images) and ZDB (excluding thumbnail images).

Ad 2: The ENP was successful in establishing a network of 11 associated partners. These associated partners provide their newspaper metadata to Europeana Newspapers which in turn forwards this metadata to Europeana and ZDB for ingestion. Thus the project makes available a vast amount of digital newspaper metadata on the scale of around 18 million pages. Europeana Newspapers associated partners are:

- Narodna in univerzitetna knjižnica / National and University Library, Ljubljana, Slovenia
- Landsbókasafn Íslands - Háskólabókasafn / National and University Library of Iceland
- Koninklijke Bibliotheek van Belgie - Bibliothèque royale de Belgique / Royal Library of Belgium
- Nacionalna i sveučilišna knjižnica u Zagrebu / National Library of Croatia
- St. Cyril and Methodius National Library / National Library of Bulgaria
- Biblioteca Nationala a Romaniei / The National Library of Romania
- Bibliothèque nationale de Luxembourg - Nationalbibliothéik / National Library of Luxembourg
- Národní knihovna České republic / National Library of the Czech Republic
- Biblioteca Nacional de España / National Library of Spain
- Biblioteca Nacional de Portugal / National Library of Portugal
- Llfrygell Genedlaethol Cymru / National Library of Wales

It is worth noting that the response to Europeana Newspapers call for associated partnerships was overwhelming, with a huge number of libraries, networks and aggregators expressing a vivid interest in joining the associated partners network. However, due to limited funds available to Europeana Newspapers, only 11 institutions could be accepted, since Europeana Newspapers reimburses some travel costs spent by the associated partners attending project events. In response to many more institutions wishing to benefit from Europeana Newspapers, the consortium decided to establish the option to register as networking partner. The networking

[2] http://www.theeuropeanlibrary.org/tel4/newspapers.

partners do not receive any financial support from Europeana Newspapers, but are directly informed about relevant developments within the project.[3]

Associated partners provide newspaper metadata and thumbnail images to Europeana Newspapers – describing around 8 million pages – which are collected by TEL and then forwarded to Europeana and ZDB. Again, it is noteworthy that the associated partners in many cases offered to provide not only metadata but also images with full text to Europeana Newspapers. Due to financial restraints at TEL and due to the limited project resources, this rich data can presently not be ingested via TEL. However, TEL is working on (business) models to allow for a later ingest. This work will be an addition to the work described in the Description of Work.

[3] For a list of all networking partners see: http://www.europeana-newspapers.eu/consortium/project-partners/#networking.

3. Union Catalogue of Serials / Zeitschriftendatenbank

The ZDB, jointly run by Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (SBB) and the German National Library (Deutsche Nationalbibliothek - DNB), is the world's largest dedicated database for serial materials, mainly journals and newspapers, but also other forms of publications like monographic series and even websites and is a major part of the infrastructure for libraries in Germany and Austria.

The ZDB is a metadata system containing title level records and, attached to these, holding records, that inform about the availability of a specific item in a specific library. Links to available online resources that may either point to public domain materials or to resources that are subject to specific license restrictions are included. However, most data of the ZDB are provided under a CC0 license.[4] ZDB-identifiers (IDs) are widely used persistent IDs for serial materials and play an important role in a number of use scenarios. They are often used by other services, such as the German Library of Digital Serials (Elektronische Zeitschriftenbibliothek - EZB) or, on the international level, OCLC`s WorldCat.[5]

More than 4.400 libraries and other types of institutions (e.g. archives, museums, research institutions) in Germany and Austria participate in the ZDB by cataloguing their bibliographic units and holdings information, making the ZDB a truly national union catalogue for serial materials. Editorial teams of more than 30 staff members at SBB take care of the high quality of the data independently of its origin.

The ZDB is distinguished by a number of features:

- Commonly used standards of cataloguing on the basis of agreed international rules
- Availability of editorial teams at SBB that are responsible for all title level records ensuring the homogeneity of data
- Homogeneous data structured for easy machine-readability
- De-duplicated title level records
- Unique, persistent identifiers (ZDB IDs) for all title level records that therefore have the character of authority data
- Authority data and controlled vocabularies are used in cataloguing
- Basic subject classification capacity
- Extensive information on title changes and journal/newspaper affiliation

Generally speaking, the ZDB acts as a data hub and data transformation platform connecting and merging data from publishers, aggregators, networks and libraries. Therefore the ZDB provides data services and synchronisation facilities for different systems, acting as an integration layer. Data export routines for different formats and dedicated interfaces enable external services of

[4] Around 97% of the participating institutions release their data under a CC-0 licence.

[5] The use of ZDB identifiers was contractually agreed upon by SBB, DNB and OCLC (http://oclc.org/home.en.html) and will be effective as of 2015. For details of the cooperation see: http://oclc.org/news/announcements/2014/ZDB-agreement-optimizes-coverage-of-Germanic-libraries-in-WorldCat.en.html.

different kinds (local, regional, international). A variety of workflows to support services on a national and international scale is implemented around the ZDB, e.g. interlibrary loan, document delivery, acquisition, licensing workflows, quality management and data validation.

A variety of library projects require data from the ZDB as a necessary basis for their work, e.g.

- German newspaper digitisation projects funded by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG)
- VD17 and VD18 are cooperative German library projects aiming at digitising prints of the 17th and 18th centuries; ZDB data is used for the allocation of digitisation quotas to individual libraries regarding serials;
- A variety of initiatives use the ZDB for the cataloguing of websites, partly undertaken within the context of the German "Virtuelle Fachbibliotheken" (virtual subject-specific libraries);
- A number of initiatives, e.g. the "Sondersammelgebiete" (special areas of collection) financed by the DFG, require statistical analysis data from the ZDB.

Currently the ZDB contains around 1.7 million unique bibliographic title records for both conventional and digital serials, and around 13.5 million records describing local holdings. The title records are linked to around 10.1 million authority records of the German National Authority file.[6] This data is made available centrally to everyone via the ZDB-OPAC.[7] The data is also distributed to the German and Austrian library networks who in turn deliver it to local libraries for integration into local OPACs of the participating institutions. Thus, metadata about serials is both available for one-stop searches and for searches in specific library environments where it is tied to local services.

In the following screenshots, some of the core search functionality offered in the ZDB is captured.

[6] Gemeinsame Normdatei (GND), see: http://www.dnb.de/gnd.

[7] http://dispatch.opac.d-nb.de/DB=1.1/.

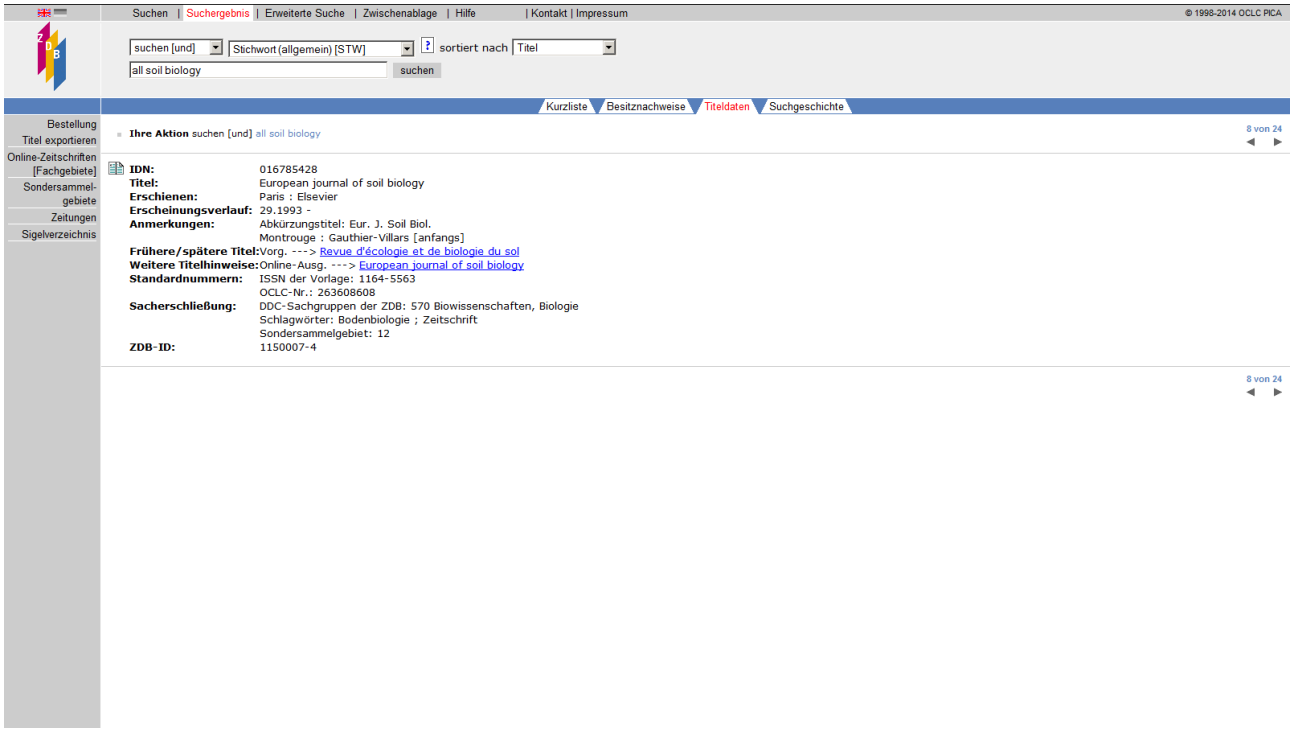

Fig. 1: Example of a bibliographic title record in ZDB

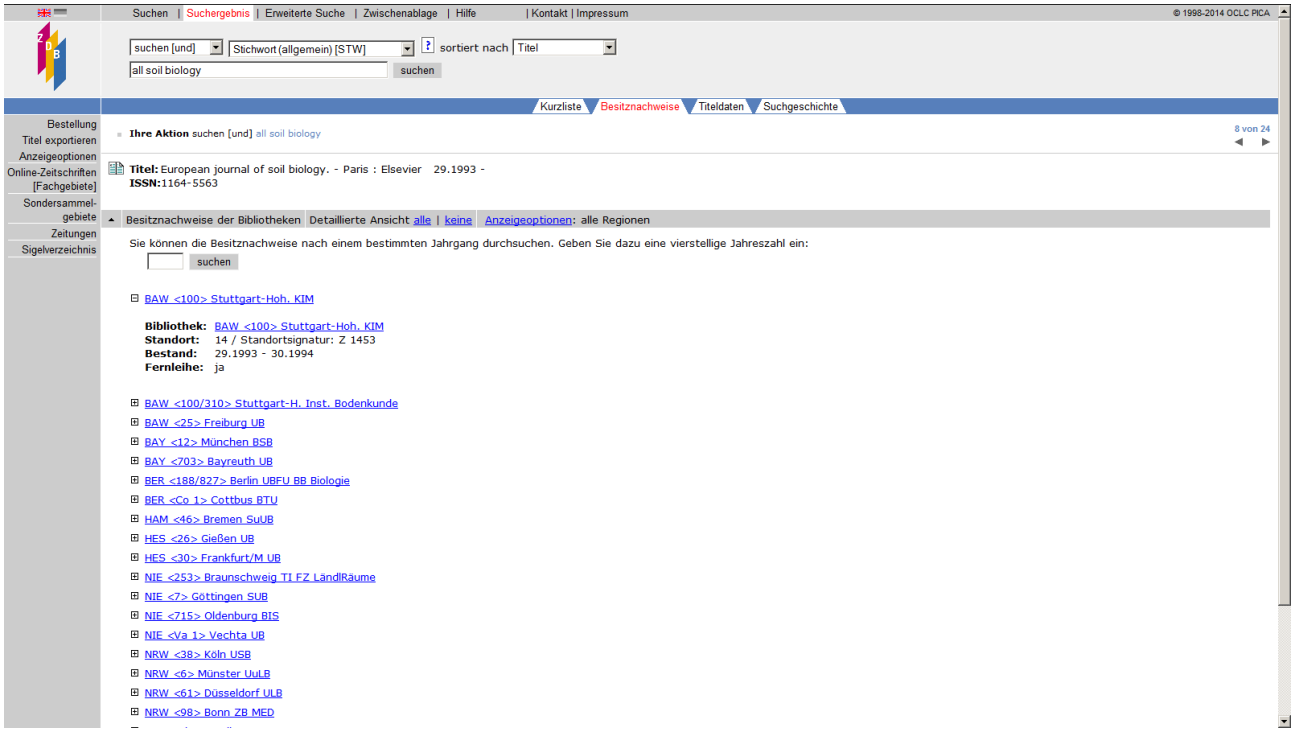

Fig. 2: Example of accumulated holdings information in ZDB

Newspapers in both analogue and digital forms play an important role in the ZDB. German digitisation projects require instant access to relevant data and ZDB therefore offers a specific page dedicated to searches for borne digital and digitised newspapers. Via this page, digitisation initiatives can easily find out which newspapers have already been digitised and/or are foreseen for future digitisation.

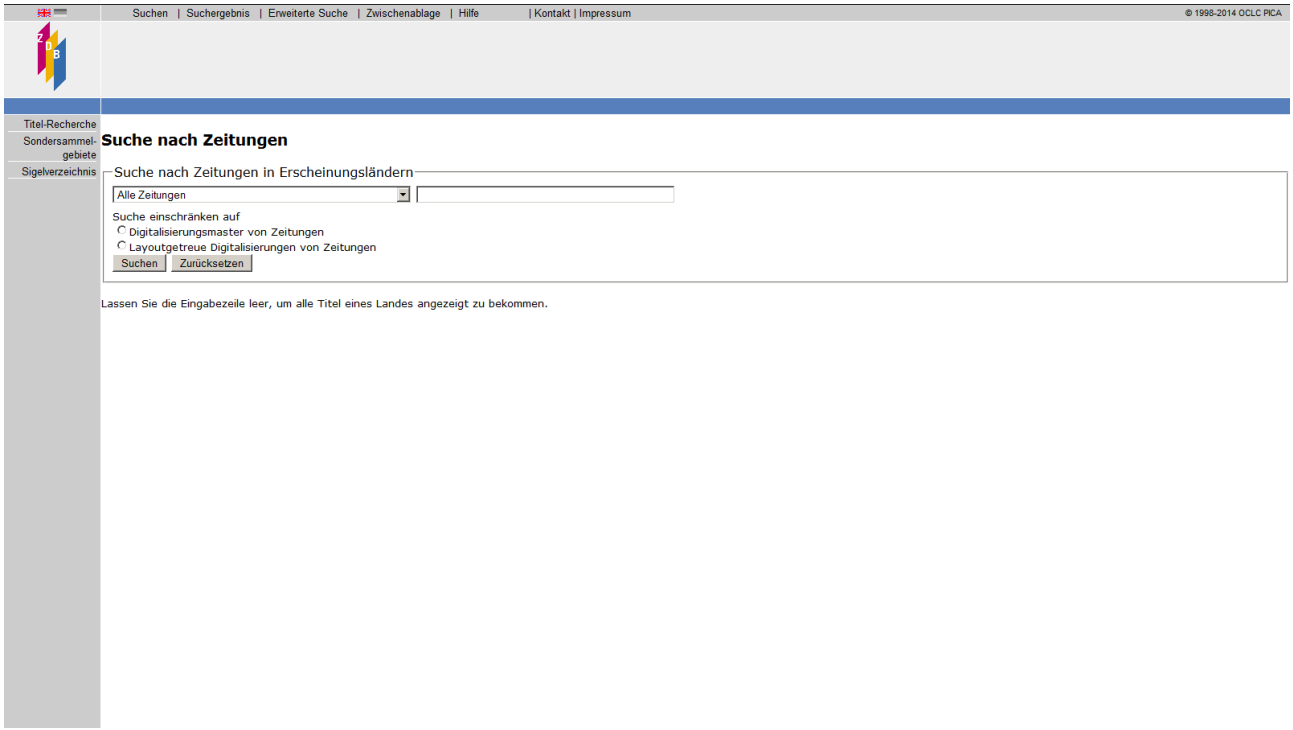

Fig. 3: ZDB search page for newspapers

4. ENP Data in the ZDB

Since the ZDB so far has provided data and services to institutions in Germany and Austria only, its interfaces were available in German exclusively. With the integration of data from Europeana Newspapers, the ZDB has created an English newspaper page designed for international audiences that will go online in August 2014. On that page the ZDB provides title-specific metadata that originates from Europeana Newspapers. To enable searches, the following metadata facets are integrated in the ZDB from Associated Partners:

- title
- country of publication
- language of publication
- holdings details:
  - running time of a newspaper holding
  - name of institution
- access URL:
  - local library environment
  - TEL
  - Europeana

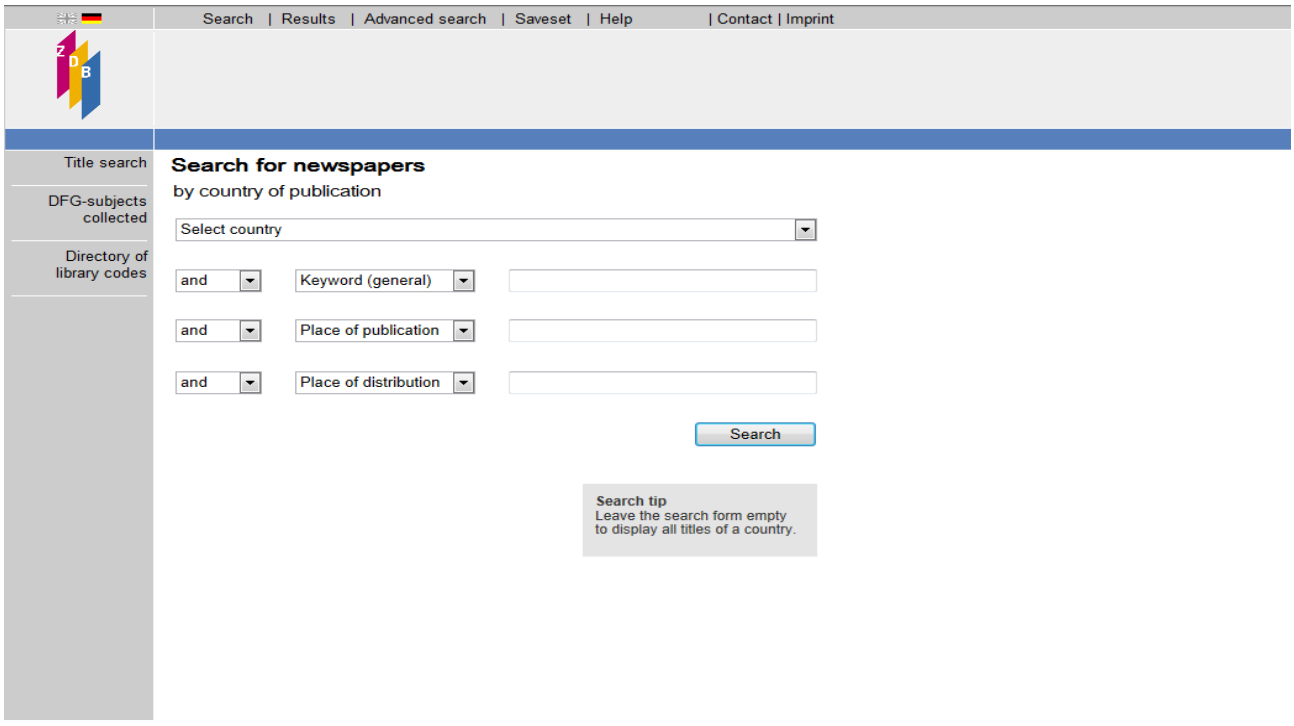

Fig. 4: European Registry for Digital Newspapers (mock-up)

The following search options regarding data about digital and digitised newspapers are provided:

- search by country of publication
- search by newspaper title
- search by place of publication
- search by place of distribution
- search by keyword
- search by publisher
- search by ISSN
- search by person name
- search by DDC classification
- search by ZDB-ID
- search by owning institution
- search by shelf mark

The data to be integrated in ZDB is being forwarded from TEL, ingest schedules are therefore dependant on TEL's ingest schedule that is detailed in D4.2 *Newspaper Aggregation and Indexing Plan*. Note that not only metadata from project partners will be ingested but also, in addition to the work specified by the Description of Work, metadata from associated partners. The metadata is delivered to ZDB in an easy to handle CSV (comma separated values) format. The actual integration of the metadata is, by and large, a manual process: For most newspapers bibliographic title records already exist in the ZDB. The identification of the proper title level records to which specific metadata facets need to be added is, in most cases, an intellectual task that cannot rely exclusively on automated matching routines. Given that the establishment of newspaper titles as such is a complicated endeavour, different cataloguing rules across Europe add to the difficulties. It is to be hoped that the internationalisation of cataloguing rules – RDA in particular[8] – will ease some of these problems in future.

The editorial resources required for ingestion of data from associated partners of Europeana Newspapers are kindly provided by SBB.

[8] Resource Description and Access, see http://www.rda-jsc.org/rda.html.

5. Outlook

Design and functionalities of the search page of the European Registry for Digital Newspapers as outlined in chapter 4 of this document represent standard options of a typical PICA/OCLC OPAC.

However, the ZDB OPAC is currently undergoing a fundamental revision that will result in the abolition of the standard OPAC as used until now. In the framework of a DFG-funded project[9] a new OPAC is currently being developed and will be available in March 2015. The new OPAC aims at moving away from the library-specific peculiarities of the old OPAC and will be much more intuitive and easy to use especially by non-experts. Particular focus is being laid on the visualisation of specific data facets by utilising geographic maps and tools for visualising relations, namely the often complicated ”history” of a specific newspaper with, e.g., changing titles.[10]

The functionalities of the European Registry for Digital Newspapers will be part of the new OPAC and search and browse options will therefore be greatly improved as of next year. Again, the resources required for these improved functions will be borne by SBB.

The internationalisation of the ZDB by integrating international newspaper metadata is of great significance not just within Europeana Newspapers but also as part of the strategic agenda of the ZDB as such. SBB therefore seeks ways to establish a permanent workflow with TEL and Europeana so that in future newspaper metadata that is delivered or made available to either TEL or Europeana can be seamlessly forwarded to SBB for integration in the ZDB. By doing so, the ZDB will serve as a major central access mechanism to digital and digitised newspapers across Europe.

[9] “Further development of the ZDB as steering and research instrument for newspaper digitisation projects and for newspapers”. For more information about this project see: http://www.dnb.de/EN/Wir/Projekte/Laufend/zdbWeiterentwicklung.html (in English).

[10] To give an example of such relations: the “Leipziger Allgemeine Zeitung” (1911-21) and the “Leipziger Handelsblatt und Zeitung für Sachsen“ (1919-21) merged into the ”Neue Leipziger Zeitung” (1921-40) which became the ”Neue Leipzger Tageszeitung” (1941-45). Such ”histories” often are very complicated and are most clearly presented as a graph that is created by automatically interpreting links between data records.