Managing e-legal deposit of electronic versions of daily newspapers in Germany

Susann Solberg

Collection, Acquisition, Descriptive Cataloguing and Standardisation, Deutsche Nationalbibliothek, Frankfurt a. M., Germany.
s.solberg@dnb.de

Abstract:

When the law (DNBG) regarding the Deutsche Nationalbibliothek (German National Library, DNB), passed in 2006, came into force, the Deutsche Nationalbibliothek received the additional task of collecting, cataloguing, indexing and archiving also non-physical media works (online publications).

The German National Library has been implementing a fully-automated workflow for the acquisition, cataloguing and archiving of all kinds of online material since then.
We can handle e-books, online theses, digitised content, music files, audio books, e-journals, digital newspapers and we started archiving websites.

Newspaper publishers are increasingly distributing not only printed editions but also digital versions of the same content featuring the same layout – so-called e-papers – for their subscribers to access via the Internet. The German National Library has been collecting these e-paper editions of daily newspapers since 2010.
In conjunction with a service provider, the library developed an automated process in order to handle the large amounts of data. This is now put into routine action and used to collect the editions of 930 daily and 18 Sunday newspapers from the publishers' servers, to convert them into PDF/A format which is suitable for long-term preservation, and to enter them into the German National Library's catalogue and archive systems. In June 2014 users were able to access roughly 520.000 issues containing over 22 million pages online in the reading rooms. Every month approximately 19,000 new issues are added. They are made available to the general public within 8 days in the Leipzig and Frankfurt reading rooms.
Accordingly, from 2011 the practice of microfilming daily newspapers started by the German National Library in the 1960s has now been dispensed with in cases where an e-paper which corresponds in full to the printed edition is available.

Since the project has been successful the German National Library plans to extend the collection of e-papers to about 1,200 e-paper titles, representing the majority of the approximately 1,500 daily newspapers currently published in Germany.

As an example of a fully-automated workflow for the acquisition, cataloguing and archiving of online material the paper will present the German National Library's workflow for e-papers which corresponds in full to the printed edition..

Keywords: e-paper, electronic newspaper, legal deposit, automated workflow.

1. Introduction

1.1 The German National Library (DNB)

The German National Library (Deutsche Nationalbibliothek; abbreviation: DNB) is entrusted with the task of collecting, permanently archiving, fully documenting and bibliographically indexing all publications issued from 1913 onwards which have a connection to Germany. It is also responsible for making them accessible to the general public.
The DNB was preceded by two institutions: the Deutsche Bücherei, founded in Leipzig in 1912, and the Deutsche Bibliothek Frankfurt am Main, established in 1946. The reunification of Germany saw them amalgamated into a single institution. In 2006 the legal brief of the library was expanded and it was given a new name: Deutsche Nationalbibliothek (German National Library).
The DNB of today is based at two locations: in Leipzig and in Frankfurt (Main).

1.2 The legal basis

In order to fulfil the legal mandate, the DNB has been given legal deposit for the Federal Republic of Germany. The collection mandate is based on the Law regarding the German National Library dated June 2006[1]. Since then, tasks have been extended to the collection of media works in "immaterial form", or to put it simply, online publications.
All commercial and non-commercial publishers in Germany are obliged to submit two (mandatory) copies of their works to the DNB. In the case of online publications, only one copy needs to be provided.
There is one exception: In the case of newspapers, independent of the media-version, only <u>one</u> copy has to be provided and that on <u>request</u> of DNB.

---

[1] DNBG 22 June 2006 (BGBI. I P. 1338)

2. Newspaper collections in Germany

Germany is a country with a rich journalistic culture. In the early years of newspapers in the 17th century about 200 newspaper titles were published. Today you will find about 1.500 daily national, regional and local newspapers.
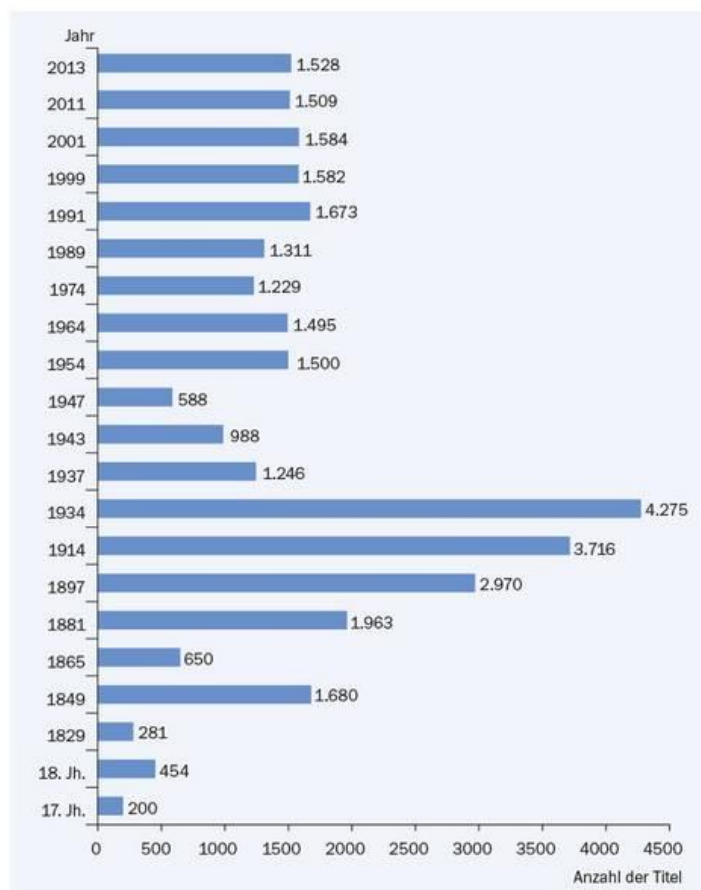

Figure 1: Development of newspapers since the 17th century – amount of titles on national, regional and local levels (Mitteilungen des Mikrofilmarchivs, Schütz / Pankratz, Federation of German Newspaper Publishers)

Germany's history as a federation is one of the reasons why there is not just <u>one</u> national newspaper collection in Germany. Until 1912 (the founding year of the DNB), no library was specifically charged with collecting <u>all</u> printed materials published in Germany, and no library fulfilled the tasks of a national library. In the different German states a couple of important libraries acted as state libraries for their particular state.

This encouraged the development of a variety of well-stocked libraries – but no national library. The national newspaper collection is spread over several libraries and archives throughout the country.
Some German libraries agreed to take up specific responsibilities for the newspaper collection of a particular time segment.

These are:

1450 - 1600	Bayerische Staatsbibliothek München / Bavarian State Library
1601 - 1700	Herzog August Bibliothek Wolfenbüttel (Home of Avisa, one of the first news-periodicals in the world)
1701 - 1800	Niedersächsische Staats- und Universitätsbibliothek Göttingen / Goettingen State and University Library
1801 - 1912	Staatsbibliothek zu Berlin - Preußischer Kulturbesitz / Berlin State Library
1913 ff.	Deutsche Nationalbibliothek (DNB) / German National Library

Besides the national collection mandate of the DNB regional legal deposit has been given to several regional libraries (state libraries or in German "Landesbibliothek") for particular states or regions.

3. The DNB newspaper collection

3.1 Overview

The DNB is a relatively young national library. Its newspaper collection covers a period of one hundred years from 1913 until today. Talking about newspapers means those paid-for newspapers, issued daily or at least three times per week, containing news and other items of current and often local interest to a general readership.

Since its foundation, the DNB's mandate to collect newspapers has changed throughout its history of one hundred years. It has always been a selection of the wide range of newspapers published in Germany.

The newspaper collection consists of national, regional and local newspaper editions. The circulation figures and an equal regional dissemination were examples of helpful selection criteria.

Different media types of newspapers have been collected:

- newspapers in their original hard copy format (1913 – 1990 in Leipzig),
- only microfilmed newspapers (since 1969 in Frankfurt),
- only microfilmed newspapers (since 1991 in Leipzig and Frankfurt),
- e-paper newspapers (since 2010/11).

Only at the Leipzig library branch newspapers in the original hard copy format have been collected. On the contrary, at the Frankfurt location hard copies of daily newspapers have never been collected. Instead of this, a program for microfilming newspapers started in Frankfurt in the late 1960s. All daily newspapers selected for the newspaper collection (in Frankfurt) have been microfilmed. The hard copies were discarded after being microfilmed. The decision in favor of the microfilm format was made for economic reasons. Handling and archiving of newspapers in hard copy format is far more expensive and requires considerably more space than dealing with microfilms.

After the merger of the DNB predecessor institutions in 1990, there was an important change in the collection policy: Hard copies of daily newspapers were not collected anymore in Leipzig.

From 1991 until 2010 the number of newspaper titles collected regularly by DNB was stable with 450 titles. The collection covered 30 percent of 1.500 daily newspaper titles in Germany today (q.v. figure 1).

3.2 The e-paper collection of daily newspapers

Since the extension of the library's legal mandate in 2006, online publications are collected. Newspaper publishers are increasingly distributing not only as printed editions but also as digital versions of the same content featuring the same layout – so-called e-papers – for their subscribers to access via the web. An e-paper is different to newspaper websites such as for example the website of the Frankfurter Allgemeine Zeitung (FAZ)[2].
Especially since the rapid expansion of tablet computers e-papers are popular and the circulation figures in Germany started to rise enormously: between the second quarter of 2012 and 2013 the number of sold e-papers has been duplicated[3] to approximately 400.000 issues per day. In comparison to the number of 22 million sold hard copies per day this is still a small number[4]. Nevertheless it is an attractive part of the newspaper publishers' portfolio with a considerable increase in the future. The DNB has identified about 1.200 of 1.500 currently daily newspapers which are published as e-papers in parallel to the printed version.

As part of a project the DNB has been collecting e-paper editions of daily newspapers since 2010. In conjunction with a service provider (outermedia GmbH, Berlin, former iSquare, Berlin), the library developed an automated process in order to handle the large amounts of data. This is now put into routine action and used to collect the editions of 930 daily newspapers including 18 Sunday newspapers from the publishers' servers, to convert them into PDF/A format which is suitable for long-term preservation, and to enter them into DNB's catalogue and archive systems. At present (June 2014), users are able to access roughly 520.000 issues containing over 22 million pages online in the reading rooms. Every month approximately 19.000 new issues are added. They are made available to the general public within 8 days in the library's reading rooms in Leipzig and Frankfurt.
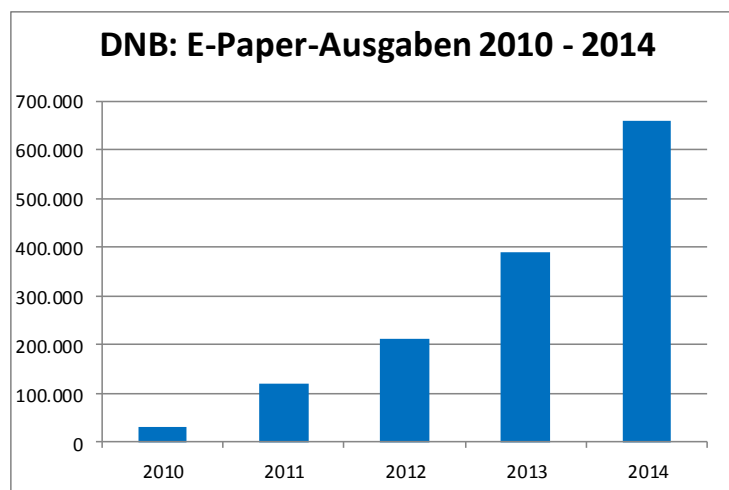

Figure 2: Deutsche Nationalbibliothek: E-paper issues 2010 - 2014 [5]

---

[2] Website of FAZ s. http://www.faz.net/

[3] Keller, Dieter; Eggert, Christian (2013): Ein starkes Medium – Zur wirtschaftlichen Lage der deutschen Zeitungen. In: Zeitungen 2013/14, p. 45 and 72, publ. Bundesverband Deutscher Zeitungsverleger, Berlin 2013, p. 21 - 98

[4] Pasquay, Anja (2013): Zur wirtschaftlichen Lage der Zeitungen in Deutschland 2013. Berlin, 30. Oktober 2013. Bundesverband Deutscher Zeitungsverleger / Federation of German Newspaper Publishers, http://www.bdzv.de/markttrends-und-daten/wirtschaftliche-lage/artikel/detail/zur_wirtschaftlichen_lage_der_zeitungen_in_deutschland_2013/.

[5] Jockel, Kathrin; Schneider, Kurt (2014): E-Paper-Sammlung in der Deutschen Nationalbibliothek: Eine Zwischenbilanz. In: Dialog mit Bibliotheken, 1, p. 21 - 24

In consequence, the newspaper collection policy has changed again: accordingly, from 2011 the practice of microfilming daily newspapers started in the 1960s, has been dispensed within cases where an e-paper which corresponds in full to the printed edition is available.
On the one hand economic reasons again were important for decision making: an automated workflow to collect e-papers is less expensive than dealing with microfilms. On the other hand, the users' needs played an important role: reading e-papers is much more user-friendly than reading microfilms. Last but not least the economic aspect is one of the reasons why the DNB has been able to expand the collection of daily newspapers from 450 titles continuously to nowadays about 930 titles. Since the project has been successful, the DNB plans to extend the collection of e-papers to the whole number of 1.200 currently published e-paper titles in Germany. Only a few titles are still microfilmed because no e-paper edition exists. With nearly 1.300 newspaper titles the DNB's collection is representing about 87 % of the approximately 1.500 daily newspapers currently published in Germany .
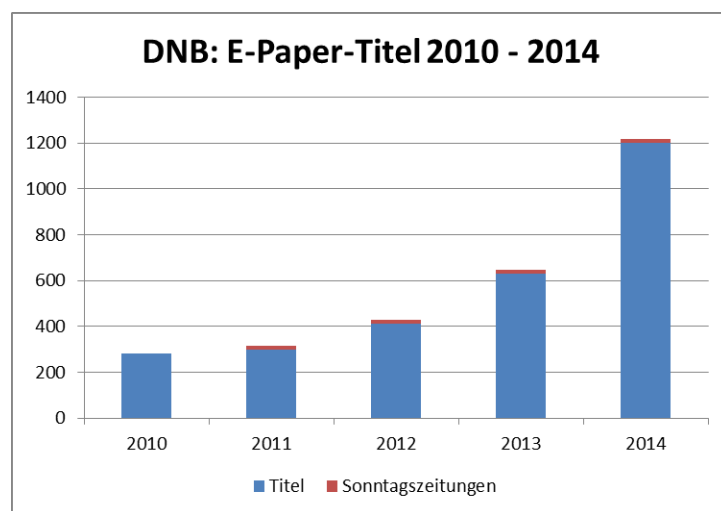

Figure 3: Deutsche Nationalbibliothek: E-paper titles 2010 - 2014[6]

4. The workflow

What are the preconditions to collect and archive e-papers by means of an automated workflow?
First of all an appropriate data format is needed. At present the Portable Document Format, well known as PDF, is the most suitable format. The DNB only collects e-papers provided as a download PDF.

Advantages of the PDF:
- The format is suitable for preserving text, image <u>and</u> layout and, converted in PDF/A-1b format, for long-term preservation.
- It combines information <u>and</u> layout (unlike ePub or other text-focused formats).
- A full text search in one or over all e-paper issues and further automatic processing is possible (article separation, entity recognition or it could be read out for the visually impaired).
- It ensures up-to-date availability.

[6] Jockel, Kathrin; Schneider, Kurt (2014): E-Paper-Sammlung in der Deutschen Nationalbibliothek: Eine Zwischenbilanz. In: Dialog mit Bibliotheken, 1, p. 21 - 24

- The document quality is far better than the quality of microfilms (e.g. colours, pictures).

There are a few disadvantages:
- Preservation of e-papers is more complex than e.g. of microfilm: PDF consists of thousands of pictures and fonts and different text-parts and the format itself is quite complex.
- The establishment of a particular workflow for acquisition, collecting, cataloging, archiving and preservation of e-papers is necessary.

Publishers are legally obliged to deliver online publications to the DNB. Therefore the DNB offers three interfaces, one for the submission of smaller quantities via a web form and two machine-based interfaces for the automated submission of a larger number of objects (OAI-PMH, Open Archive Initiative Protocol for Metadata Harvesting; sFTP, secure File Transfer Protocol).
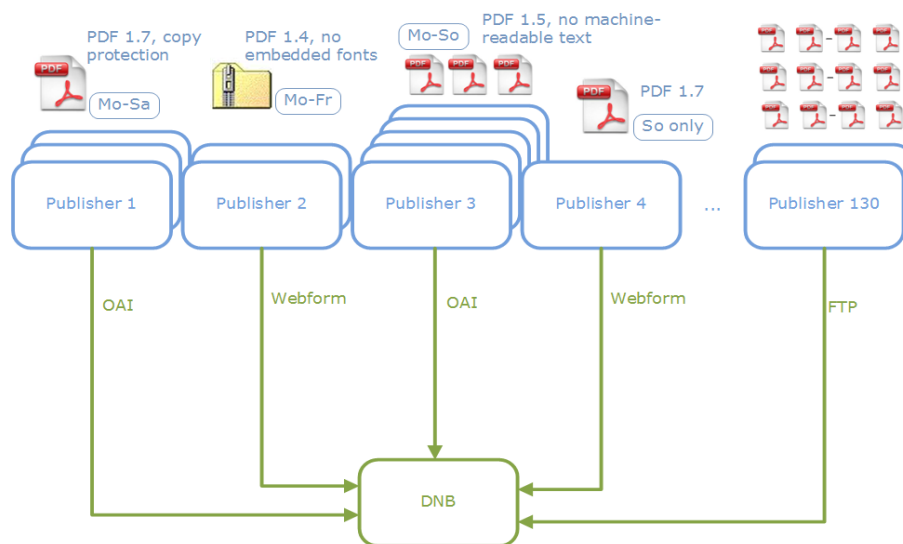

Figure 4: Deutsche Nationalbibliothek

Configuration and distribution of e-papers differ a lot among publishers and in consequence the task of collecting them is complex.
There are differences in the PDF-version used, like PDF 1.3 to PDF 1.7 but also in details such as usage restrictions or not embedded fonts. E-papers also differ in their composition. At best the complete e-paper is just one PDF file, in some cases it is built up from several PDFs, e.g. one PDF for the front section, one for the local section and a varying number of PDFs for supplements or one PDF for every page.
E-papers also differ in the frequency of publication. Some appear from Monday to Saturday, some only from Monday to Friday or everyday. Holidays might cause problems in the workflow because the normal frequency of publication is in disorder.
This heterogeneous situation would have caused a lot of work to the newspaper publishers as well as to the DNB. Publishers would have to deliver their e-paper issues regularly (daily) on an automated basis and in a consistent quality to the DNB. The DNB would have to control this process (including all technical problems) and to ensure the completeness of the collection. For example: at present 150 publishers would have to deliver about 930 issues of about 930 e-paper titles on a daily basis via one of the three interfaces offered by the DNB. Considering the large number of e-paper issues per year it seemed to be nearly impossible

that publishers would manage the daily delivery of e-papers in a satisfying way. The German National Library aims to make the submission of online publications as convenient as possible. In order to cope with this situation, to ensure the completeness of the collection and to handle this large amount of data, the DNB decided to choose a different approach and to reduce complexity.

A service provider was found to collect the e-paper editions of daily newspapers. The service provider, at present outermedia GmbH from Berlin, acts as an aggregator for the DNB. In consequence the delivery process is completely standardized and much easier both for the publishers and the DNB.
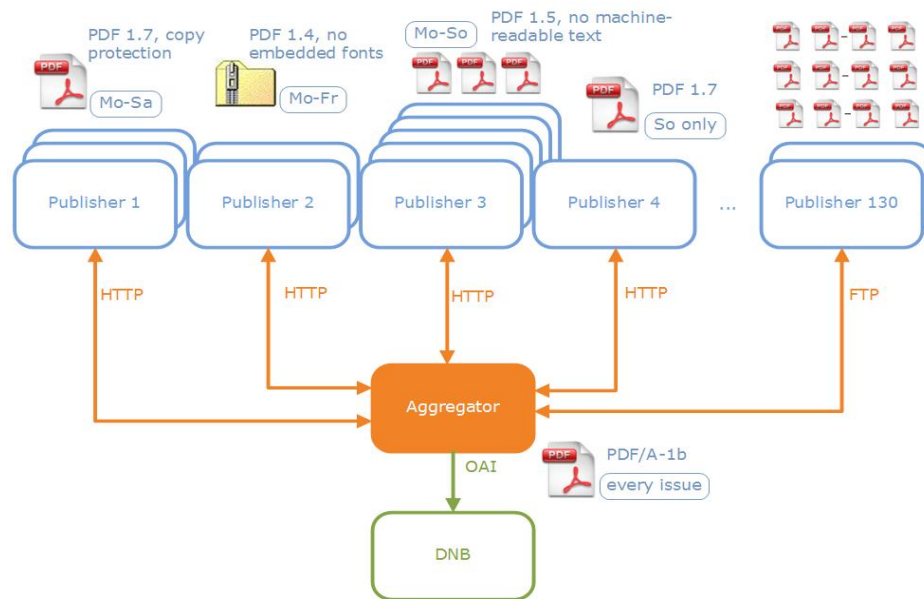

Figure 5: Deutsche Nationalbibliothek

What do the publishers do?
The publishers offer the e-papers already to registered, paying users in a restricted section of their websites. The DNB only needs free access to the e-paper sections of their websites and permission to download the e-papers by the service provider.

What does the service provider (in the following: "aggregator") do?
The aggregator is using software for harvesting the e-papers on the web. The software is configured for every single website and checks them for a new issue daily. In detail it loads the website of the publisher, logs into the e-paper-service and selects the PDF-download or several downloads of the correct issue, combines the downloaded PDFs (1 or more per issue) to one PDF and removes restrictions, processes and converts the PDF-files to PDF/A-1b, validates the PDF/A-file for e.g. correct title, date, number of pages, generates the metadata needed for ingest and provides the files including consistent metadata via an OAI-PMH interface.
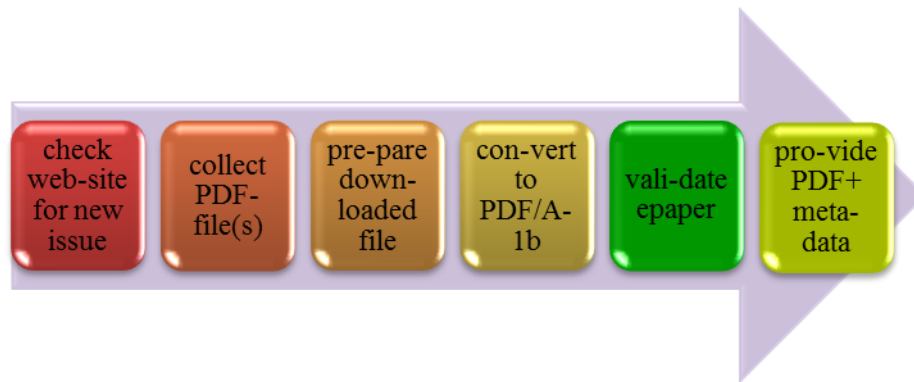
Figure 6: Deutsche Nationalbibliothek

What does the DNB do?

Although the main part of the daily collection of e-papers is done by the aggregator, the DNB is active, too.

At the beginning, before the aggregator starts the configuration, some preparations by the DNB are necessary. The DNB contacts publishers and requests login details, checks the e-paper for collection principles, describes which parts of the e-paper are relevant, defines criteria for quality checks and creates bibliographic descriptions.

On this basis the aggregator configures the software and downloads first issues.

After the aggregator has finished the configuration, DNB checks the results for a period of 14 days. Finally, DNB approves the e-paper title for OAI-harvesting, so the new issues become available to the general public in the reading rooms of DNB.

In the everyday workflow the quality of current titles is assured by making spot tests.

The completeness of the collection is monitored to make sure that every published issue has been collected.

The library imports all objects and metadata via OAI-PMH and handles errors in the import process. Dealing with mistakes and workflow problems is supported by quality management tools used by the aggregator and the DNB.

5. Indexing and use at the DNB

First and foremost, the range which can be accessed by users in Leipzig and Frankfurt am Main has improved significantly: they can now make searches in individual editions of e-papers or access them directly at the click of a mouse from the catalogue. The waiting period before an issue can be accessed has been significantly reduced in comparison to the microfilm version – from 3 to 4 months for filmed newspapers to one week for the digital versions. The automatic collection of electronic editions, including all associated metadata, has considerably improved the bibliographic information about daily newspapers.

KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK

Gesamter Bestand	Musikarchiv	Exilsammlungen	Buchmuseum

→ Suchformular zurücksetzen

tit all "welt berlin" and vlg all "springer"   Finden → ☑ Expertensuche ?

eingeschränkt auf
- Materialarten: Zeitschriften/Serien, Online Ressourcen

Ergebnis der Suche nach: *tit all "welt berlin" and vlg all "springer"*
← Zurück zur Trefferliste

Treffer 2 von 3   ⇐ ← → ⇒

Link zu diesem Datensatz	http://d-nb.info/99714484X
Titel/Bezeichnung	[Die Welt <Hamburg> / Ausgabe Berlin] Die Welt : unabhängige Tageszeitung für Deutschland Ausgabe Berlin [Elektronische Ressource] [Zeitschrift/Serie]
Verleger	Essen : Springer; Berlin : Springer
Erscheinungsverlauf	Nachgewiesen 2001,Jan. -
Umfang/Format	Online-Ressource
Parallele Ausgabe(n)	Sonntagszeitung: [Welt am Sonntag / Ausgabe B] Druckausg.: [Die Welt <Hamburg> / Ausgabe Berlin]
Anmerkungen	Gesehen am 07.06.2012 Periodizität: tägl. Verbreitungsorte: Hamburg ; Berlin ; Berlin Langzeitarchivierung gewährleistet
Weitere Titelhinweise	Auch in: [Die Welt <Hamburg>] Beilage zu: Die Literarische Welt
URL	http://d-nb.info/99714484X http://epaper.welt.de/
Sprache(n)	Deutsch (ger)
gehört zu	E-Paper-Sammlung der Deutschen Nationalbibliothek
Sachgruppe(n)	070 Nachrichtenmedien, Journalismus, Verlagswesen
Zugehörige Bände	1230 Publikationen 1. *17.06.2014. 28 S.* *In: [Die Welt <Hamburg> / Ausgabe Berlin]*

Figure 7: Deutsche Nationalbibliothek

Previously, the existence of the microfilm in the catalogue was documented by just one data record per title. Now, every e-paper title is indexed in the German Catalogue of Serials (Zeitschriftendatenbank, ZDB) and by contrast, there is in addition one data record for each issue of the e-papers held by the DNB.

KATALOG DER DEUTSCHEN NATIONALBIBLIOTHEK

Gesamter Bestand	Musikarchiv	Exilsammlungen	Buchmuseum

→ Suchformular zurücksetzen

partOf=99714484X   Finden → ☑ Expertensuche ?

Ergebnis der Suche nach: *partOf=99714484X*
← Zurück zur Trefferliste

Treffer 3 von 1228   ⇐ ← → ⇒

Link zu diesem Datensatz	http://d-nb.info/1052338585
Persistent Identifier	urn:nbn:de:101:1-201406202873
URL	Archivserver der Deutschen Nationalbibliothek
In:	[Die Welt <Hamburg> / Ausgabe Berlin], 12.06.2014. 32 S.
Erscheinungsjahr	2014

Figure 8: Deutsche Nationalbibliothek

This data record indexes the main metadata such as title, date of issue, persistent identifier for quoting the electronic resource (URN), and number of pages. But above all, it allows the particular document to be specifically searched for and accessed directly in the reading rooms in Leipzig and Frankfurt.

The service is available to all users of the library in the reading rooms; it cannot be offered to external users for copyright reasons. Nor is it possible for users to save an e-paper, or parts of it, in digital form and take it home – also for copyright reasons.

6. Conclusion

At present the results of the e-paper project are convincing. The automated workflow has been put in routine action. Nevertheless it also creates some work but much less then previous workflows would demand. It has enabled the DNB to expand the newspaper collection within an acceptable budget. This has been made possible, despite the economic restrictions, by the reduction of filming activities and the resulting savings in labor and costs. At the end of 2014 the current newspaper collection of the DNB will cover about 87 % of all newspapers published in Germany. This is an excellent result and we are proud of it.

7. Outlook

As mentioned above, regional legal deposit has also been given to several regional libraries (state libraries or in German "Landesbibliothek") for particular states or regions. By now the tasks of some of these libraries[7] have been extended to the collection of online publications, too. The DNBs' e-paper collection contains numerous titles which should be collected by the state libraries, too. It is obvious to discuss a co-operative workflow, access and use of the e-paper collection, of course by keeping copyright and legal deposit matters in mind.

For publishers and the legal deposit libraries it would be attractive to simplify the delivery procedures as much as possible. The aim is that newspaper publishers are obliged to submit only one mandatory e-paper copy to the DNB and the DNB ensures secure access to the e-papers for the particular state library which has been given legal deposit for these e-paper titles. The DNB and a group of regional libraries have started to discuss a closer co-operation. The future development remains challenging.

References

Bundesverband Deutscher Zeitungsverleger (BDZV; Federation of German Newspaper Publishers)
http://www.bdzv.de/
http://www.bdzv.de/home-engl/bdzv/the-federation-of-german-newspaper-publishers/

Deutsche Nationalbibliothek (DNB; German National Library)
www.dnb.de

---

[7] Regional legal deposit mandate extended to the collection of online publications in German states: Baden-Württemberg, Thüringen, Hamburg, Sachsen-Anhalt, Brandenburg, Hessen, Nordrhein-Westfalen.

Jendral, Lars (2013)
Die elektronische Pflicht in den Bundesländern. In: Bibliotheksdienst , vol. 447, no. 8-9, p. 592-596

Jockel, Kathrin; Schneider, Kurt (2014)
E-Paper-Sammlung in der Deutschen Nationalbibliothek: Eine Zwischenbilanz. In: Dialog mit Bibliotheken, no. 1, p. 21 – 24

Keller, Dieter; Eggert, Christian (2013)
Ein starkes Medium – Zur wirtschaftlichen Lage der deutschen Zeitungen. In: Zeitungen 2013/14, p. 45 and 72, publ. Bundesverband Deutscher Zeitungsverleger, Berlin 2013, p. 21 – 98

Pasquay, Anja (2013)
Zur wirtschaftlichen Lage der Zeitungen in Deutschland 2013. Berlin, 30. Oktober 2013. Bundesverband Deutscher Zeitungsverleger / Federation of German Newspaper Publishers, http://www.bdzv.de/markttrends-und-daten/wirtschaftliche-lage/artikel/detail/zur_wirtschaftlichen_lage_der_zeitungen_in_deutschland_2013/

Schneider, Kurt (2010)
E-Paper und ihre Sammlung durch die Deutsche Nationalbibliothek. In: Zeitschrift für Bibliothekswesen und Bibliographie, vol. 57, no. 3-4, p. 175-182

(about 3.400 words)