

What happens if you publish the National Bibliography under a CC0 license? – Experiences of the German National Library (DNB)

Anke Meyer-Heß

Digital Services, German National Library, Frankfurt, Germany

E-mail address: a.meyer@dnb.de

Jochen Rupp

Digital Services, German National Library, Frankfurt, Germany

E-mail address: j.rupp@dnb.de



Copyright © 2016 by Anke Meyer-Heß and Jochen Rupp This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

Since 1 July 2015 all bibliographic data of the German National Library and the authority data of the Integrated Authority File (GND) are provided free of charge and can be freely re-used under "Creative Commons Zero" (CC0 1.0) terms.

This includes about 11 million authority records and over 14 million bibliographic records that can be downloaded via different online interfaces or APIs.

The data can be obtained via [online interfaces](#) ([Data Shop](#), [SRU](#) or [OAI-PMH interface](#) etc.) after an initial registration and authorization (registration free of charge).

The basic procedure for users to obtain metadata from the German National Library is for them to download the data themselves in standard formats via online interfaces.

Provision charges will be applied by the German National Library for active provision of data via FTP or WWW server which are not included in the basic service.

This change in the business model was made in order to move customers from individual data provision to automated interfaces and allow them to set their own filters and update preferences.

Since this change and the change to a CC0 license, the number of new registrations per month to get access via the SRU or OAI interface has quadrupled. The proportion of foreign users is increasing, however mostly from the German speaking countries Switzerland and Austria.

In addition already existing customers who used to apply the simple formats like DNB Casual (oai_dc) or RDFxml that were always free of charge changed to the format MARC21-XML.

These customers are primarily major suppliers of software that now provide their customers the ability to access the data of the German National Library.

Among the new customers are also university libraries, smaller municipal and public libraries and software vendors. Going forward, we expect that this trend will continue.

The paper will explain the decision to change to a CC0 license, the organizational and structural changes that had to be made and the experiences made in the first year.

Keywords: CC0, National Bibliography, Business Model

Content

What happens if you publish the National Bibliography under a CC0 license? – Experiences of the German National Library (DNB).....	1
1 Introduction.....	3
2 Legal basis	3
3 Content of the DNB Data service	3
3.1 National bibliographic data and further data sets	4
3.1.1 National Bibliography.....	4
3.1.2 New Release Service.....	5
3.1.3 German Union Catalogue of Serials (ZDB).....	5
3.1.4 Integrated Authority File (GND)	5
3.2 Range of services	5
3.2.1 Self- service	5
3.2.2 Selection services.....	6
3.3 Metadata selection	6
4 Change of the Business Model	7
4.1 Politico-economic reasons	7
4.2 Data reasons	7
4.3 Cause of actions	8
4.4 Period of transition.....	8
4.4.1 2010.....	9
4.4.2 2012.....	9
4.4.3 2015 Free of charge and CC0	9
5 After publishing in CC0.....	10
5.1 Customer reactions.....	10
5.2 Later developments.....	10
5.3 Guarantee	11
5.4 Provision	11
6 Conclusions.....	11
7 References.....	12

1 INTRODUCTION

Since July 1 2015 all bibliographic data of the German National Library (DNB) and the authority data of the Integrated Authority File (GND) are provided free of charge and can be freely re-used under "Creative Commons Zero" (CC0 1.0) terms. The road we took to arrive there was not free from issues and controversial discussions. This paper will explain the decision to change to a CC0 license, the organizational and structural changes that had to be made and the experiences made in the first year.

2 LEGAL BASIS

The DNB is a federal institution with legal capacity under public law. The law regarding the DNB¹ from 22 June 2006 (BGB1 I p. 1338) defines the tasks, obligations and organs of the federal institute. The law confers widespread autonomy on the DNB; it comes under the jurisdiction of the Federal Government Commissioner for Culture and Media.

“The Library is entrusted with the task of collecting, permanently archiving, **bibliographically classifying and making available to the general public** all German and German-language publications from 1913, foreign publications about Germany, translations of German works ...“ and “... responsible for ... **production, marketing and distribution of national bibliographic services** ...”

The DNB provides a comprehensive range of **Digital Services** including data transfer (export), catalogue enrichment, persistent identification of online resources and long-term preservation.

3 CONTENT OF THE DNB DATA SERVICE

The DNB provides

- bibliographic information for independent and (especially in the electronic field) non-independent publications
- content information for independent (and in the future also for non-independent) publications, authority data for the fields: person, subject heading, corporate body, uniform title
- and catalogue enrichment

in defined standards (scope, format, transfer method/interface, cataloguing level, provision cycles).

Most of the records are linked with the Integrated Authority File (GND, authority records) which contains defined authorised terms and names including additional information such as forms of reference, broader and narrower terms, source information, dates of birth and death etc. which are based on RDA-rules. All the data records are given persistent identifiers.

Supplementary information to the pure bibliographic data is also offered through catalogue enrichment. In terms of text this includes e.g. tables of contents, abstracts, blurbs (printed on the back or rear dust-jacket of a book), extracts, bibliographies and lists of illustrations, indices and cover information. This additional information is linked in each case with the corresponding title data.

¹Law regarding the German National Library, see: <http://www.dnb.de/law>

The data provided by the DNB is manually or automatically generated; here the DNB is bound by international rules and national standards. It develops these rules and standards in partnership with national and international bodies, and guarantees homogeneous use as far as possible within its own environment. Any changes to these usage standards which have a lasting impact on the structure and quality of the cataloguing data are clearly documented and published on a regular basis.

3.1 National bibliographic data and further data sets

3.1.1 National Bibliography

At <http://dnb.dnb.de> anyone may research into the database of the German National Bibliography (Deutsche Nationalbibliografie) free of charge via public networks. The Deutsche Nationalbibliografie records the collections of the DNB and fulfils its legal mandate to index all German and German-language publications, foreign publications about Germany and translations of German publications. This database serves as the original bibliographic indexation. The continuity of the Deutsche Nationalbibliografie is secured through the data safeguarding conception and the commitment of the DNB in the area of long-term availability.

Apart from the freely accessible catalogue the Data Service provides also various national bibliography services. These are designed to meet the needs of customers who wish e.g. to select new data records at certain intervals, or who wish to continue receiving a particular format. They are offered as value-added services in addition to the use of the database.

The Deutsche Nationalbibliografie includes the following series:

- Series A - Monographs and periodicals from the publishers' book trade. Books, magazines, non-music recordings, further AV media, microfiches and electronic publications → 109,757 records in 2015.
- Series B - Monographs and periodicals from outside the publishers' book trade. Books, magazines, non-music recordings, further AV media, microfiches and electronic publications. Issued weekly → 27,349 records in 2015.
- Series C - Maps. Issued quarterly → 3,628 records in 2015.
- Series H - University Publications. Dissertations and postdoctoral thesis at German universities and dissertations and postdoctoral thesis in German language at universities abroad. Issued monthly → 18,652 records in 2015.
- Series M - Printed Music. Issued monthly → 2,978 records in 2015.
- Series O - Online Publications. Issued monthly. → 246,976 records in 2015.
- Series T - Recorded Music. Issued monthly → 4,676 records in 2015.

The Deutsche Nationalbibliografie is provided as a free online journal in PDF format in the publicly accessible catalogue, via the OAI-PMH interface, the Data Shop and via the Data Service (provision is charged).²

² See also <http://www.dnb.de/EN/nationalbibliografie>

3.1.2 New Release Service

The New Release Service (ND) is based on the publishers' reports to the *Verzeichnis Lieferbarer Bücher* (German Books in Print, VLB) and is offered by the DNB as an information service for advance announcements and newly-released titles. All titles that are reported to the VLB are automatically passed on from the MVB *Marketing- und Verlagsservice des Buchhandels GmbH* to the DNB. In the DNB the titles are categorized according to the Deutsche Nationalbibliografie, which is based on the Dewey Decimal Classification (DDC) and lists them weekly in the New Release Service.

The New Release Service replaces the Series N advance announcements for monographs and periodicals (CIP), which have ceased publication with issue 51, 2002. It is not part of the Deutsche Nationalbibliografie.

The DNB provides the New Release Service as a free online journal in PDF format in the publicly accessible catalogue, via the Data Shop and via the Data Service.

Series ND. Issues monthly → 118,645 records in 2015

3.1.3 German Union Catalogue of Serials (ZDB)

The German Union Catalogue of Serials³ is the world's largest database dedicated for serials of all kinds, particularly journals and newspapers. As a service and competence centre for serials, the ZDB provides data and services centrally and thus is the basis for a variety of information and services provided by the library networks and individual libraries in the German-speaking world. It currently contains more than 1.8 million bibliographic records for publications in all languages from the 16th century onwards and more than 14.7 million holding records of approximately 4,160 German and Austrian libraries.

3.1.4 Integrated Authority File (GND)

The Integrated Authority File (GND)⁴ is an authority file for Persons, Corporate Bodies, Conferences and Events, Geographic Information, Topics and works. It is used above all for the cataloguing of literature by libraries, but it is also increasingly being deployed in archives, museums, projects and web applications. It is operated cooperatively by the DNB, all German-speaking library networks and the associated libraries, the German Union Catalogue of Serials (ZDB) and numerous other institutions. Contributions to the GND are made either via the networks or in direct agreement with the DNB.

3.2 Range of services

3.2.1 Self- service

The DNB makes the bibliographic data available in structured form for download.

This basic service includes the possibility to select and transfer any amount of title and authority data at any time intervals, in defined formats, via the Data Shop, the SRU and OAI-PMH interfaces, both as individual data records and as new or updated records. Catalogue enrichment which supplements the bibliographic data records can be freely downloaded via a link in the bibliographic data.

³ <http://www.dnb.de/EN/zdb>

⁴ <http://www.dnb.de/EN/gnd>

3.2.2 Selection services

If the compilation of specific data or a different data format to the one provided free of charge are required, these constitute supplementary services which are more labour-intensive than the standard services and therefore are charged.

The costs are based on the calculated total cost for the development and operation of the relevant service and can be taken from the latest versions of the relevant cost overviews.⁵

3.3 Metadata selection

The metadata can be transferred into several exchange formats:

- MARC 21/XML
- Dublin-Core-Elements (extract, only bibliographic records)
- RDF serialisations (RDF/XML, JSON-LD, Turtle)
- CSV (comma-separated values, only bibliographic records)
- PDF: New Release Service, series of Deutsche Nationalbibliografie

Full copy (complete dump)	Provision (update)	Formats	Number of records/ size of data (zipped)
DNB bibliographic records	February, June, October	MARC 21	approx. 14 million/ approx. 3.3 GB
		MARC21-xml	
DNB bibliographic records	February, June, October	RDF (RDF/XML) & RDF (Turtle)	approx. 12.5 million/ approx. 1.4 GB
Integrated Authority File (GND) authority records	February, June, October	MARC 21	approx. 11.3 million/ approx. 1.3 GB
		MARC21-xml	
		RDF (RDF/XML) & RDF (Turtle)	approx. 0.9 GB
German Union Catalogue of Serials (ZDB) bibliographic records	March and September	MARC 21	approx. 1.8 million/ approx. 465 MB
		MARC21-xml	
		RDF (RDF/XML), RDF (Turtle) & RDF (JSON-LD)	approx. 215 MB
German Union Catalogue of Serials (ZDB) holdings records	March and September	MARC 21	approx. 15.2 million/ approx. 1.2 GB
		MARC21-xml	

⁵ More under <http://www.dnb.de/dataservice>

Address data of the German ISIL and Library Code Agency	March and September	RDF (RDF/XML), RDF (Turtle) & RDF (JSON-LD)	approx. 16,000/ approx. 25 MB
---	---------------------	---	----------------------------------

4 CHANGE OF THE BUSINESS MODEL

4.1 Politico-economic reasons

In 2010 DNB decided to publish its data free of charge in the future.

We "... did so because we, the librarians of a publicly funded library, existing to serve the public, strongly believe that serving the public today means opening up as many services, holdings and data as (legally, technically, organizationally ...) possible, making them not only accessible via the World Wide Web but making our data an integral part of the World Wide Web!"⁶

Previously the DNB regarded supplying interlibrary associations and individual users with bibliographic data for re-use in their catalogues and databases as one of its key roles. Now, however, it is increasingly important to respond to the challenges arising from a globally networked and more tightly interlinked information infrastructure (linked data, semantic web). The more the system is dominated worldwide by a heterogeneous range of portals and service providers which can only be influenced to a limited degree, the more the national bibliography and other data services of the DNB needed to change, including their quality and user orientation.

"What is even more important is that they [libraries] are able to really and seriously reach out to many more users in typical internet 'places' like social networks, portals, and search engines. This will not only help them fulfil their public mission but also increase traffic back to the library's site, helping them to monitor and prove their relevance in today's internet based information universe."⁷

4.2 Data reasons

Even before the 1 July 2015 the DNB provided its bibliographic data free of charge for research purposes, yet applied a charge if the data was to be re-used. In return it guaranteed a certain degree of cataloguing quality which conformed to the relevant cataloguing codes.

Now however, the DNB was and is making more and more data available, some of which it no longer generates itself, but integrates data from third parties. The DNB transfers all the data it provides into MARC 21 and other export formats, ensuring uniformity for the re-use of the data. In addition, the profile of the integration-compatible services was considerably expanded, providing a useful supplement to the existing services for requesting and transferring data.

⁶ Niggemann 2012: "The importance of open data to national libraries", <http://www.ifla.org/past-wlic/2012/181-niggemann-en.pdf> p. 2

⁷ Niggemann 2012: "The importance of open data to national libraries", <http://www.ifla.org/past-wlic/2012/181-niggemann-en.pdf> p. 1

4.3 Cause of actions

At the start of 2010 the DNB really began work on completely changing its business model with the aim of providing its data free of charge.

Until then, all data created and maintained by the DNB was licensed against a fee. Data production and preparing for download required human interaction and no self-service interface was offered.

In preparation two topics were considered to be crucial: the loss of income and fear that someone else could take the data and earn money with it. For an institution that in 2010 made approximately 750.000 euro per year from the sales of bibliographic and authority data, the loss of this income stream would indeed have a noticeable impact on the balance sheet.⁸

“Giving up a reliable source of income is not easy psychologically speaking and even harder when funding or administrative bodies have to be convinced. In the case of the DNB, the decision to stop selling metadata, but to do it gradually, over a period of five years, was made in the end, because, after a long discussion, the conviction prevailed, that this market option would vanish anyway during that time.”⁹

Thus it was decided to move gradually from a charged service towards providing its data free of charge for downloading and processing in the following five years.

The free service, which was to be introduced step-by-step, includes the possibility to select and transfer any number of individual sets of bibliographic data, or new or updated data sets, in the formats offered by the DNB via self-service interfaces, the Data Shop, the SRU or OAI-PMH interfaces at any time intervals.

Other forms of data transfer than those offered in the self-service interfaces are and will be charged as additional services, as they are more labour-intensive on the part of the library. By taking this step the DNB is supporting the free flow of data, especially amongst providers of bibliographic data, and is helping to realise the vision of an emancipated, freely accessible “web of data”.

The conditions of use for the free re-use of the data were to be based on the Creative Commons and the British Crown licence.¹⁰

4.4 Period of transition

A staged transition to the new business model was implemented starting in 2010.

Accordingly the fees for the bibliographic services were reduced in stages over the coming years, whereas the fees for data provision were raised step by step. Especially this change in the business model was made in order to move customers from individual data provision to automated interfaces and allow them to set their own filters and update preferences.

However as Niggemann says “It is not enough to set our data ‘free’ in a legal or a financial sense, we also have to make them easily discoverable, reusable, sharable in a very concrete,

⁸ See: Svensson 2013: “Licensing Library and Authority Data Under CC0: The DNB Experience“
http://www.w3.org/2013/04/odw/odw13_submission_57.pdf

⁹ Niggemann 2012: “The importance of open data to national libraries“, <http://www.ifla.org/past-wlic/2012/181-niggemann-en.pdf> p. 2

¹⁰ <https://data.gov.uk/terms-and-conditions>

technical, internet-based sense”¹¹, we also started improving our access methods and interfaces over the next five years.

4.4.1 2010

In 2010 the change started with a provision of certain data sets under a CC0 licence:

- bibliographic data in a form which is usable by the Semantic Web (in RDF format)
- bibliographic data on electronic publications in the bibliographic series O (automatic cataloguing, no manual linking with authority data or other data, agreed standardised data format)
- data from the GND
- all additional information on the bibliographic data such as tables of contents

Further we also introduced new automated interfaces for self-service download as SRU and OAI PMH.

If this data was collected in the available formats via the automated interfaces offered as described above, no fees were charged.

The fees for the retrieval of structured cataloguing data on physical items (manual cataloguing by expert staff for subdomains, manual linking with authority data and in some cases with other data, high level conformity with rules, agreed standardised data format) were still charged.

4.4.2 2012

In 2012 another major step towards our goal of CC0 data was made. We announced the use of CC0 for all data except the last and the current bibliographic year (authority data was already fully CC0).

Metadata produced in the last two years were not free of charge and not CC0. The charge for that metadata was reduced every year, so that the library could compare the loss of income.

By that time the biggest part of the national bibliography was also published as Linked Open Data.

4.4.3 2015 Free of charge and CC0

From 1 July 2015 on all bibliographic data of the DNB and the authority data of the GND are provided free of charge and can be freely re-used under “Creative Commons Zero” (CC0 1.0)¹² terms.

This includes about 11 million authority records and over 14 million bibliographic records that can be downloaded via different online interfaces.

Now the basic procedure for users to obtain data from the DNB is for them to download the data themselves in standard formats via online interfaces.

¹¹ Niggemann 2012: “The importance of open data to national libraries“, <http://www.ifla.org/past-wlic/2012/181-niggemann-en.pdf> p. 2

¹² <http://creativecommons.org/publicdomain/zero/1.0/>

In preparation we changed the registration procedure for the interfaces, as no contracts etc. are needed any more. However we still ask for some contact information.

5 AFTER PUBLISHING IN CC0

5.1 Customer reactions

The 1 July 2015 was a quieter day for us than expected. After we notified our customers, that now finally the data was free of charge it took some hours until the first reactions reached us.

These reactions can be categorized into two groups:

The first group was made of mainly already existing customers, of whom many asked whether they understood us correctly. Many of these customers still downloaded the data from FTP server every week (which is still charged) and thus decided to move to the OAI-PMH or SRU interfaces to save money.

The other group consisted of new customers, mainly software suppliers and search engines. These customers were only waiting for the data to be free of charge and CC0 and had already prepared an implementation to start to work in July and only contacted us to get access to the interfaces.

5.2 Later developments

Several studies about consequence of the EU regulations (PSI Directive - Re-use of Public Sector Information)¹³ were carried out. The studies¹⁴ have shown that the use and re-use of digital information have increased dramatically following the reduction or abolition of fees, that new users have been attracted and that innovative areas of re-use have been created, meaning that "... the social and economic advantages far outweigh the short-term financial benefit of cost-effectiveness in cases where no or only low charges [...] are levied".¹⁵

For the DNB this is definitely true. The number of new registrations per month for access via the SRU or OAI-PMH interface has quadrupled after 1 July 2015. Since then we got up to six new customers every week in comparison to six new customers a month before. We have not noticed any slowdown of demand so far.

In addition already existing customers who used to apply the simple formats like DNB Casual (oai_dc) or RDFxml, which were always free or charge, changed to the formerly charged format MARC21-XML.

Our customers now are from all over the world including academic and public libraries, library networks, software producers, research institutes, publishers, booksellers, reference management software providers, service providers for libraries and individuals. The

¹³ Notification from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the re-use of public sector information - review of the directive 2003/98/EC - (SEC(2009)597, KOM(2009) 212, final version 7.5.2009, p 9.

¹⁴ See Dapp et al (2016): "Open Data. The Benefits" http://www.kas.de/wf/doc/kas_44906-544-1-30.pdf?160418125028

¹⁵ Notification from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the re-use of public sector information - review of the directive 2003/98/EC - (SEC(2009)597, KOM(2009) 212, final version 7.5.2009, p 9.

proportion of foreign users is definitely increasing, however mostly from the German speaking countries Switzerland and Austria. So this is a wide range of communities and we are still listening to their needs.

5.3 Guarantee

As mentioned above the reasons for the new business model also included a change in data quality. The records are provided, as available, with no guarantee to its being continuous, punctual, error-free or complete, or to its not infringing the rights of third parties (e.g. personal rights and copyright).

The DNB now re-uses data which it receives from third parties and treats them the same as data which it has generated itself.

5.4 Provision

The basic procedure for users to obtain data from the DNB is for them to download the data themselves in standard formats via online interfaces.

Although now we offer the data free of charge, for some interfaces we still require credentials. This means we require data consumers to register with our Digital Services department – who will create a customer account.

We do this to keep some track of the data use and in order to build customer relations. When we know who our customers are, we have better possibilities to interact with them and to adapt our services to their needs.

Service	Charge	Registration needed
Online Catalogue (OPAC)	No	No
Data Shop	No	Yes
SRU interface	No	Yes
OAI-PMH interface	No	Yes
ISBN, ISSN selection of metadata	Yes	No
Data service for dumps	Yes	No

6 CONCLUSIONS

After one year of publishing the national bibliography under CC0 and free of charge we can make the following conclusions:

The use of CC0 has had a very positive echo and has rendered us some very good publicity. Now on a regular basis, we get requests from small software companies planning to program apps or websites using our data fulfilling the hope of Niggemann to reach out to more users over the internet.¹⁶

¹⁶ See Niggemann (2012): “The importance of open data to national libraries“ <http://www.ifla.org/past-wlic/2012/181-niggemann-en.pdf> p. 14

Since all the data contains links to the DNB catalogue, the visitors of the catalogue and more important the Google ranking has increased a lot in the last years.

“We now share the dilemma most cultural institutions find themselves in: How can we share metadata widely and still not lose track of the reuse? “¹⁷ This is especially important, if you want to prove the impact of your data.

We find it is very hard to track who uses what parts of our data. That’s why until now, we ask institution to register, to give us some indication of the usage.

Further we actually save effort in writing bills and accounting, which is of course connected to a loss of quite some income, which is definitely noted in the overall budget of the library.

With a good deal more customers the work in counselling and support however has risen, so that we cannot complain about being out of work.

In the near future we are planning some improvements for our customers and for us as well:

1. We plan to implement a new customer relationship management system (CRM) in order to serve our customers better and to reduce the time we spent on answering questions.
2. We plan to implement more tracking tools to find out what features of our data are the most used. Further, having the contact data we plan to do user surveys on a regular basis.

So while the significance of the DNB-data has increased in the world, the budget actually has decreased.

7 REFERENCES

Dapp et al (2016): “Open Data. The Benefits” http://www.kas.de/wf/doc/kas_44906-544-1-30.pdf?160418125028

Niggemann (2012): “The importance of open data to national libraries“ <http://www.ifla.org/past-wlic/2012/181-niggemann-en.pdf>

Svensson 2013: „Licensing Library and Authority Data Under CC0: The DNB Experience“ http://www.w3.org/2013/04/odw/odw13_submission_57.pdf

Notification from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the re-use of public sector information - review of the directive 2003/98/EC - (SEC(2009)597, KOM(2009) 212, final version 7.5.2009

¹⁷ Niggemann(2012): “The importance of open data to national libraries“ <http://www.ifla.org/past-wlic/2012/181-niggemann-en.pdf> p. 4