

Julia Hauser

# Der Linked Data Service der Deutschen Nationalbibliothek

## Vom Projekt zum Produkt

Die Deutsche Nationalbibliothek (DNB) stellt seit 2010 ihre Daten im Semantic-Web-konformen Format RDF über den Linked Data Service bereit. Sie schafft damit erstmals für Anwender und Nutzergruppen einen Zugang für die Nachnutzung ihrer Daten, der keine Kenntnisse bibliothekarischer Formate voraussetzt. Anfangs handelte es sich um einen Prototypen, der im Rahmen von Projekten weiterentwickelt wurde. In mehreren Iterationsschritten wurde er auf technischer, inhaltlicher und organisatorischer Ebene optimiert. Mittlerweile ist der Dienst soweit gefestigt, dass RDF als gleichberechtigtes Exportformat angeboten werden kann: Der Linked Data Service hat sich vom Projekt hin zum Produkt entwickelt.

Dieser Beitrag gibt einen Überblick über Motivation, Lizenzfragen, Evolutionsschritte und Inhalte sowie einen Einblick in geplante nächste Schritte.<sup>1)</sup>

## Motivation und Hintergrund

Das Umfeld, in dem Informationsspezialisten wie Bibliothekare, Archivare und Dokumentare arbeiten, hat sich mit der Verbreitung des World Wide Web und den zugrunde liegenden Informationstechnologien grundlegend verändert. Damit ergaben sich neue Möglichkeiten der Datenbereitstellung, der Datennutzung und interdisziplinärer institutionenübergreifender Kooperationsmöglichkeiten. Die vollständige Überführung existierender Arbeitsprozesse und lang etablierter Strukturen hin zu einer modernen webbasierten Informationsinfrastruktur ist ein langer Weg – erste Schritte dahin sind bereits getan; weiteres Potential ist vorhanden und entsteht durch den fortschreitenden technologischen Entwicklungsprozess immer wieder neu. Während Bibliotheksdaten viele Jahre lang in geschlossenen Katalogsystemen isoliert waren, sind

sie heute dabei, zu einem stabilisierenden Element des World Wide Web zu werden. Kontrollierte Vokabulare, standardisierte Metadatenformate und ein persistenter Nachweis der Ressourcen sind Komponenten, die die Arbeit von Informationsspezialisten seit jeher begleiten. Bibliotheksdaten können nun einen Beitrag zur Verlässlichkeit, Beständigkeit und Datenqualitätssicherung des Semantic Web leisten. Zur Bedeutung etwa der Nationalbibliografie im Zeitalter des World Wide Web, aber auch des WWW für die Nationalbibliografie, schreiben Kett et. al.:

»Es sind Datennetze wie die Linked Open Data Cloud, in die Nationalbibliografien integriert werden müssen, um den Erwartungen des World Wide Web gerecht und auch in Zukunft noch als relevant wahrgenommen zu werden.«<sup>2)</sup>

Zukünftig sind weiterführende Schritte notwendig, um das volle Potenzial des Linked Data Modells auszunutzen. Damit es für externe Dienstleister und Portale möglich wird, die bibliografischen Daten in ihre Angebote zu integrieren, sollen beispielsweise vorliegende textuelle Informationen (Literals) in Referenzen auf Entitäten umgesetzt werden, also mit Personen, Körperschaften, Ereignissen und Orten sowie anderen Objekten verknüpft werden. Dies ist ein weiterer Schritt in die Richtung, die Nationalbibliografie als einen Teilgraphen in der Linked Data Cloud zu verankern.<sup>3)</sup>

Viele Bibliotheken haben die darin liegende Chance erkannt, sich bibliotheksfremden Sparten und Kooperationspartnern aus der Forschung zu öffnen und die bibliografischen Daten für weitere Anwendungsfälle attraktiver zu vermarkten. Dies schafft eine bessere Nachnutzbarkeit in anderen Kultur- und Gedächtniseinrichtungen wie Museen und Archiven und ermöglicht eine bessere (technische) Zitierbarkeit im Bereich der Forschung und Wissenschaft – einem Bereich, der zunehmend webbasiert arbeitet und auf verlässliche und persistente Datenquellen im Web angewiesen ist.

Neue Möglichkeiten für Informationsspezialisten durch WWW und Informationstechnologien

Nutzung von Daten über Spartengrenzen hinweg

Je stärker Bibliotheksdaten mit anderen Datenanbietern und Institutionen vernetzt sind, je mehr wird ihre Bedeutung in der heutigen Informationsgesellschaft gestärkt: Der Wert eines Knotens im Geflecht des Semantic Web bestimmt sich danach, wie viele Fäden zu ihm hinführen.

Description Framework)<sup>4)</sup> fortan gleichberechtigt neben weiteren Exportformaten, die von der Deutschen Nationalbibliothek angeboten werden<sup>5)</sup>. Im Folgenden werden die einzelnen Evolutionschritte des Linked Data Service erläutert.

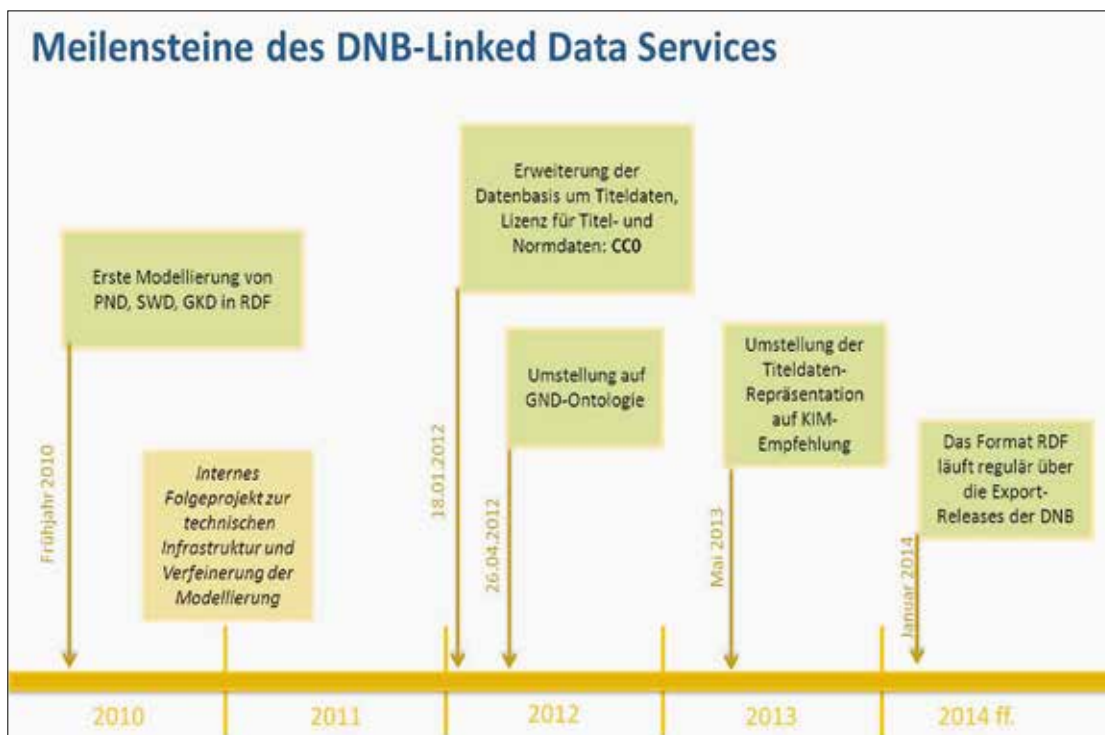
### Erreichte Meilensteine im Linked Data Service der Deutschen Nationalbibliothek

Der Linked Data Service begann als kleines internes Forschungs- und Entwicklungsprojekt im Jahr 2009 mit ersten Überlegungen dazu, wie die Daten der Deutschen Nationalbibliothek im Kontext des Semantic Web bereitgestellt und nutzbar gemacht werden könnten. Funktionalitäten und inhaltliche Modellierungsaspekte wurden in mehreren Iterationsschleifen in Folgeprojekten bearbeitet. Hierbei halfen sowohl DNB-interne Evaluationsrunden sowie Feedback aus der Linked Data Community. Auf diesem Wege wuchs der Linked Data Service Schritt für Schritt aus dem Projektstatus hinaus und wird nun zum Produkt mit verbindlicher Releaseplanung. Hiermit steht RDF (Resource

### Normdaten als Linked (Open) Data

Im Jahr 2010 hat die DNB sich der Semantic-Web-Community geöffnet und erstmals ihre Normdaten als Linked Data veröffentlicht. In der Gemeinsamen Normdatei (GND)<sup>6)</sup> sind Personen, Familien, Körperschaften, Kongresse und Veranstaltungen, Geografika, Schlagwörter und Werke (z. B. die Einheitssachtitel des Deutschen Musikarchivs) enthalten. Die heutige GND bestand damals aus drei getrennten Normdateien: der Personennamendatei, der Schlagwortnormdatei und der Gemeinsamen Körperschaftsdatei. Die Modellierung erfolgte mittels bestehender Ontologien wie dem Friend-of-a-friend-Vocabulary (FOAF)<sup>7)</sup> und dem Elementset der Resource Description and Access (RDA)<sup>8)</sup>. Diese Umsetzung konnte den Datenstrukturen nur zum Teil gerecht werden. Daher wurde im Zuge der Zusammenführung der drei früheren

RDF gleichberechtigt neben weiteren Exportformaten der DNB



Umstellung auf GND-Ontologie

Normdateien zur GND auch die RDF-Repräsentation angepasst: Die DNB spezifizierte eine eigene GND-Ontologie<sup>9)</sup>, die im April 2012 produktiv ging. Die GND-Ontologie wird sukzessive mit Elementen aus anderen, bereits existierenden Vokabularen verknüpft, im ersten Schritt mit dem RDA Element Set und FOAF.<sup>10)</sup> Sie wird seit ihrer Erstveröffentlichung laufend aktualisiert und auch zukünftig an kommende Formatänderungen angepasst.

## Umstellung auf die offene Lizenz CC0

Nutzungslizenz für Titel und Normdaten: CC0

Ein wichtiger Aspekt bei der Datenveröffentlichung im Kontext von Linked Data ist die Frage nach der Nutzungslizenz: Kommerzielle externe Anwendungen und Services können nur dann auf einem Dataset aufbauen, wenn durch dessen Nachnutzung keine rechtlichen Konsequenzen zu befürchten sind, sie also unter einer offenen Lizenz stehen. Daher entschloss sich auch die DNB dazu, ab Januar 2012 sämtliche im Format RDF verfügbaren DNB-Daten unter die Nutzungslizenz Creative Commons Zero (CC0)<sup>11)</sup> zu stellen<sup>12)</sup>. Seit dem 1. Juli 2012 stehen darüber hinaus alle Normdaten und ein Großteil der Titeldaten der DNB inklusive Kataloganreicherungsdaten auch in anderen Formaten kostenfrei zur Verfügung und werden unter CC0 zur freien Nachnutzung angeboten.<sup>13)</sup> Die Entscheidung, die DNB-Daten schrittweise freizugeben und unter die Lizenz CC0 zu stellen, ist eine logische Konsequenz aus den Entwicklungen in der zunehmend globalisierten und vernetzten Informationswelt. Eine freie Lizenz ist die erste Voraussetzung für die Nachnutzung der Daten im Semantic Web.

## Bereitstellung von Titeldaten

Anfang 2012 wurde der Linked Data Service um Titeldaten erweitert - die Voraussetzung, um auch im Kontext von Linked Data auf die Titel der Deutschen Nationalbibliografie zu referenzieren. Der Umfang der konvertierten Titeldaten umfasst derzeit die Deutsche Nationalbibliografie inklusive

der darin enthaltenen fortlaufenden Sammelwerke der Zeitschriftendatenbank (ZDB<sup>14)</sup>). Musikalien, Tonträger und die Bestände des Deutschen Exilarchivs 1933 - 1945 und des Deutschen Buch- und Schriftenmuseums sind aktuell noch nicht umgesetzt. Während für bibliografische Textdaten bereits ein Application Profile durch die DINI-AG-KIM Gruppe Titeldaten<sup>15)</sup> spezifiziert wurde, gibt es für Archivalien und Sonderbestände derzeit noch keinen Standard, der implementiert werden kann. Eine enge Zusammenarbeit mit Fachleuten aus dem Bereich Archiv und Dokumentation ist erforderlich, um eine RDF-Repräsentation für Sonderbestände zu erarbeiten. Aktuell ist eine Kooperation der DINI-AG-KIM Gruppe Titeldaten mit Musikbibliothekaren in Planung, um ein Anwendungsprofil für Musikalien und Tonträger zu erarbeiten. Langfristiges Ziel ist es, sämtliche Daten der Deutschen Nationalbibliothek inklusive kontrollierter Vokabulare als Linked Open Data bereitzustellen. Weitere Schritte sind geplant<sup>16)</sup>.

## Geplante Schritte und Ausblick

Neben bereits erwähnten inhaltlichen Erweiterungen ist vor allem der Ausbau externer Verlinkungen für die Zukunft vorgesehen: Die Titeldaten sollen in kommenden Releases mit Titeln der British Library (BL)<sup>17)</sup> verknüpft werden. Auch die Normdaten sollen weiter mit externen Verknüpfungen angereichert werden, so etwa zu Geonames<sup>18)</sup>, den LCSH<sup>19)</sup> und RAMEAU<sup>20)</sup>.

Die Deutsche Nationalbibliothek arbeitet außerdem in unterschiedlichen Projekten und Initiativen mit, die sich mit der Bereitstellung und Nutzung von Daten im Semantic Web beschäftigen, wie der Deutschen Digitalen Bibliothek<sup>21)</sup> oder der europäna<sup>22)</sup>. Darüber hinaus ist die DNB Mitglied der Early Experimenters Group der von der Library of Congress initiierten Bibliographic Framework Transition Initiative (BIBFRAME)<sup>23)</sup>. Ziel dieser Initiative ist es, das MARC-21-Format durch ein nachhaltiges, RDF-basiertes Rahmenkonzept abzulösen. Die vorhandene DNB-Expertise hinsichtlich Konkordanzen, Formatumstiegen und der Konvertierung sowie Bereitstellung von RDF-Daten sind hierbei wichtige Grundlagen.

Empfehlung: KIM Application Profile für Titeldaten

Ziel, sämtliche Daten der DNB als Linked Open Data bereitzustellen

Internationale Standardisierung von Ontologien für Kultureinrichtungen

## Verstetigung des Services: RDF als gleichberechtigtes Exportformat

Seit 2014: RDF als reguläre Export-Releases in den Datenbanken

Seit Januar 2014 werden die Norm- und Titeldaten in RDF, die über den Linked Data Service bereitgestellt werden, in regulären Export-Releases mit bearbeitet<sup>24</sup>). Mit diesem Schritt hat der Linked Data Service endgültig seinen Projektstatus verloren und wurde in den Regelbetrieb überführt: Von nun an gelten verbindliche Zeitrahmen für die Anforderungserhebung, Tests, Realisierung und die Bereitstellung. Diese Abläufe werden für die Kunden transparent gemacht, sodass ggf. ausreichend Vorlauf für Anpassungen in ihren Systemen zur Verfügung steht. Die DNB schafft damit

eine zuverlässige Struktur für die Nachnutzung der Daten und die Einbindung in Anwendungen durch externe Nutzer. Zukünftig werden verbindlich in den Monaten Januar, Mai und September neue Releases veröffentlicht und anschließend zeitnah Gesamtabzüge (Dumps) in RDF bereitgestellt. Für Rückmeldungen, Fragen und Änderungsankündigungen sowie als allgemeiner Informationskanal wurden Mailinglisten eingeführt für den Linked Data Service im Allgemeinen<sup>25</sup>) und die GND-Ontologie im Speziellen<sup>26</sup>). Anregungen und Kritik zur Verbesserung des Linked Data Services sind willkommen über die Mailinglisten – weitere Kontaktmöglichkeiten sind auf der Webseite genannt<sup>27</sup>).

### Anmerkungen

1 Dieser Text erschien im Dezember 2013 im arbidio-Themenheft zu Linked Open Data. Siehe <<http://www.arbidio.ch>>

2 Kett, Jürgen ; Manecke, Mathias ; Beyer, Sarah: Die Nationalbibliografie im Zeitalter des Internets. In: ZfBB 59 (2012), 2, S. 70

3 Vgl. ebd., S. 71 ff.

4 <<http://www.w3.org/TR/rdf-primer/>>

The advertisement for FAUST 7 features a header with four navigation tabs: FAUST 7, FAUST EntryMuseum, FAUST EntryArchiv, and FAUST iServer. Below the tabs, the title 'Datenbank- und Retrievalsystem' is displayed in red. The central graphic is a word cloud with 'FAUST 7' as the largest text. Other prominent words include 'Archiv', 'Bibliothek', 'Museum', 'Bilder', 'Bücher', 'Datenbank', 'Bildarchiv', 'Musik', 'Videos', 'Zugangsrechte', 'Chroniksuche', 'Reportfunktionen', 'Online Recherche', 'Internet', 'Vorlagendatenbanken', 'Import/Export', 'Erfassungshilfen', 'EAD', 'Flexibel', 'Integrierte', 'OCK', 'Expertenrecherche', and 'Filme'. The background of the word cloud is a colorful, wavy pattern of dots in shades of green, yellow, and red. At the bottom left, there is a QR code and the text 'Alle Infos: [www.land-software.de](http://www.land-software.de)'. At the bottom right, the logo for 'LAND Software Entwicklung' is shown, consisting of a red square with a white stylized 'L' and the text 'LAND Software Entwicklung' next to it. The contact information at the bottom reads: 'Postfach 1126 • 90519 Oberasbach • Tel. 09 11-69 69 11 • [info@land-software.de](mailto:info@land-software.de)'.

- 5 <<http://www.dnb.de/datendienst>>
- 6 <<http://www.dnb.de/gnd>>
- 7 <<http://xmlns.com/foaf/spec>>
- 8 <<http://rdvocab.info>>
- 9 <<http://d-nb.info/standards/elementset/gnd#>>
- 10 Vgl. Haffner, Alexander: Internationalisierung der GND durch das Semantic Web, 2012.  
<[http://www.kim-forum.org/Subsites/kim/DE/Materialien/Dokumente/dokumente\\_node.html](http://www.kim-forum.org/Subsites/kim/DE/Materialien/Dokumente/dokumente_node.html)>
- 11 <<http://creativecommons.org/publicdomain/zero/1.0>>
- 12 Svensson, Lars G.: Licensing Library and Authority Data Under CC0: The DNB Experience, 2013.  
<[http://www.w3.org/2013/04/odw/odw13\\_submission\\_57.pdf](http://www.w3.org/2013/04/odw/odw13_submission_57.pdf)>
- 13 <<http://www.dnb.de/geschaeftsmodell>>
- 14 <<http://www.zeitschriftendatenbank.de>>
- 15 <<https://wiki.dnb.de/display/DINIAGKIM/Titeldaten+Gruppe>>
- 16 Vgl. S. Hartmann, J. Hauser: Die Deutsche Nationalbibliografie in der Linked Data Cloud. In: Bibliotheken: Tore zur Welt des Wissens / 101. Bibliothekartag in Hamburg 2012. Hrsg. von Klaus-Rainer Brintzinger u. a. – Hildesheim ; Zürich ; New York, NY : Olms, 2013 (S. 57-71)
- 17 <<http://www.bl.uk/>>
- 18 <<http://www.geonames.org/>>
- 19 <<http://id.loc.gov/authorities/subjects.html>>
- 20 <<http://rameau.bnf.fr/>>
- 21 <<https://www.deutsche-digitale-bibliothek.de/>>
- 22 <<http://www.europeana.eu>>
- 23 <<http://bibframe.org/>>
- 24 <<https://wiki.dnb.de/display/ILTIS/ILTIS-Change-Management>>
- 25 <<http://lists.dnb.de/mailman/listinfo/lds>>
- 26 <<http://lists.dnb.de/mailman/listinfo/gnd-ontology>>
- 27 <<http://www.dnb.de/lds>>