

Tags: Bibliothek, International, Big Data

Vom Projekt zum Produkt: Der Linked Data Service der Deutschen Nationalbibliothek

Das Umfeld, in dem Informationsspezialisten wie Bibliothekare, Archivare und Dokumentare arbeiten, hat sich mit der Verbreitung des World Wide Webs und den zugrunde liegenden Informationstechnologien grundlegend verändert. Damit ergaben sich neue Möglichkeiten der Datenbereitstellung, der Datennutzung und interdisziplinärer institutionenübergreifender Kooperationsmöglichkeiten.

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

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 (Literale) 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 verankern¹.

Viele Bibliotheken haben die darinliegende 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.

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.

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 Folge projekten 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 Description Framework) [2](#)fortan gleichberechtigt neben weiteren Exportformaten, die von der Deutschen Nationalbibliothek angeboten werden [3](#).

Im Folgenden werden die einzelnen Evolutionsschritte des Linked Data Services erläutert.

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) [4](#) 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) [5](#) und dem Elementset der Resource Description and Access (RDA) [6](#). Diese Umsetzung konnte den Datenstrukturen nur zum Teil gerecht werden. Daher wurde im Zuge der Zusammenführung der drei früheren Normdateien zur GND auch die RDF-Repräsentation angepasst: Die DNB spezifizierte eine eigene GND Ontologie [7](#), 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 vgl. Haffner, Alexander: Internationalisierung der GND durch das Semantic Web, 2012, http://www.kim-forum.org/Subsites/kim/DE/Materialien/Dokumente/dokumente_no-de.html. Sie wird seit ihrer Erstveröffentlichung laufend aktualisiert und auch zukünftig an kommende Formatänderungen angepasst.

Umstellung auf die offene Lizenz CCO

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) [8](#) zu stellen ^{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. 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 [9](#).

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) [10](#). Musikalien, Tonträger und die Bestände des Deutschen Exilarchivs und des Deutschen Buch und Schriftenmuseums sind aktuell noch nicht umgesetzt. Während für bibliografische Textdaten bereits ein Application Profile durch die Titeldaten der DINI-AG-KIM-Gruppe [11](#) 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 ^{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).}

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) [12](#) verknüpft werden. Auch die Normdaten sollen weiter mit externen Verknüpfungen angereichert werden, so etwa zu Geonames [13](#), den LCSH [14](#) und RAMEAU [15](#).

Die Deutsche Nationalbibliothek arbeitet ausserdem in unterschiedlichen Projekten und Initiativen mit, die sich mit der Bereitstellung und Nutzung von Daten im Semantic Weg beschäftigen, wie der Deutschen Digitalen Bibliothek [16](#) oder der europeana [17](#). Darüber hinaus ist die DNB Mitglied der Early Experimenters Group der von der Library of Congress initiierten Bibliographic Framework Transition Initiative (BIBFRAME) [18](#). Ziel dieser Initiative ist es, das MARC21 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.

Verstetigung des Services: RDF als gleichberechtigtes Exportformat

Ab Januar 2014 werden die Norm und Titeldaten in RDF, die über den Linked Data Service bereitgestellt werden, in regulären Export Releases mit bearbeitet [19](#). Mit diesem Schritt hat der Linked Data Service endgültig seinen Projektstatus verloren und wird 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 all gemeiner Informationskanal wurden Mailinglisten eingeführt für den Linked Data Service im Allgemeinen [20](#) und die GND-Ontologie im Speziellen [21](#). Anregungen und Kritik zur Verbesserung des Linked Data Services sind willkommen über die Mailinglisten – weitere Kontaktmöglichkeiten sind auf der Website genannt [22](#).

1 vgl. ebd., S. 71 ff.

2 www.w3.org/TR/rdf-primer/.

3 www.dnb.de/datendienst.

4 www.dnb.de/gnd.

5 <http://xmlns.com/foaf/spec>.

6 <http://rdvocab.info>.

- 7 <http://d-nb.info/standards/elementset/gnd#>.
- 8 <http://creativecommons.org/publicdomain/zero/1.0>.
- 9 www.dnb.de/geschaeftsmodell.
- 10 www.zeitschriftendatenbank.de.
- 11 <https://wiki.dnb.de/display/DINIAGKIM/Titeldaten+Gruppe>.
- 12 www.bl.uk/.
- 13 www.geonames.org/.
- 14 <http://id.loc.gov/authorities/subjects/html>.
- 15 <http://rameau.bnf.fr/>.
- 16 www.deutsche-digitale-bibliothek.de/.
- 17 www.europeana.eu.
- 18 <http://bibframe.org/>.
- 19 <https://wiki.dnb.de/display/ILTIS/ILTIS-Change-Management>
- 20 <http://lists.dnb.de/mailman/listinfo/lids>.
- 21 <http://lists.dnb.de/mailman/listinfo/gnd-ontology>.
- 22 www.dnb.de/lids.



Julia Hauser

Julia Hauser ist Diplom-Bibliothekarin in der IT-Abteilung der Deutschen Nationalbibliothek (DNB). Hier arbeitet sie seit 2009 an Projekten zur Repräsentation der DNB-Daten im Kontext des Semantic Web mit. Julia Hauser verantwortet den Linked Data Service der DNB und koordiniert den DNB-Beitrag zur BIBFRAME-Initiative der Library of Congress.

Abstract

Français

Depuis 2010, la Bibliothèque nationale allemande (DNB) fournit ses données au format RDF, conforme aux besoins du Web sémantique, via un service de Web des données (linked data).

Par cette initiative, les utilisateurs et groupes d'utilisateurs ont eu, pour la première fois, la possibilité de réutiliser les données, sans avoir besoin d'une quelconque connaissance des formats propres aux bibliothèques. Initialement, il s'agissait d'un prototype, lequel a été développé ensuite dans le cadre de projets. Après plusieurs versions, il a été optimisé sur les plans technique, fonctionnel et organisationnel. Parallèlement, RDF s'est imposé comme format d'exportation: le service de Web de données est ainsi passé du projet au produit.

La contribution donne un aperçu touchant les motifs de la démarche, évoque les problèmes de licence, les étapes de développement, rappelle l'évolution des contenus et fournit un aperçu des prochaines étapes.