Klaus Ceynowa, Matthias Groß, Andreas Kahl, Gabriele Meßmer

Linked Open Data geht in die Fläche: Der B3Kat stellt seine Daten frei

Einleitung

Der Bibliotheksverbund Bayern (BVB) und der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) haben 2011 ihren gemeinsamen Verbundkatalog *B3Kat* als Open Data im Format MARCXML und als Linked Open Data im Format RDF/XML veröffentlicht. Enthalten sind die Beschreibungen zu über 23 Millionen Medien aus 180 wissenschaftlichen Bibliotheken in Bayern, Berlin und Brandenburg. Die Daten werden im Internet zur allgemeinen Nutzung unter der *CC0 Public Domain Dedication* bereitgestellt. Damit handelt es sich um das umfangreichste Datenpaket, das bisher von deutschen Bibliotheken und Bibliotheksverbünden als Linked Open Data zugänglich gemacht wurde. Zusammen mit den Datenfreistellungen des Hochschulbibliothekszentrums des Landes Nordrhein-Westfalen und mehrerer Bibliotheken weiterer Bundesländer steht damit ein erheblicher Teil der bibliographischen Titel- und Bestandsdaten wissenschaftlicher Bibliotheken in Deutschland ohne Einschränkung zur Verfügung.

Im Folgenden werden die Motivation zur Freistellung der Daten bei den beteiligten Partnern sowie der konkrete Weg der Entscheidungsfindung und -umsetzung in den verantwortlichen Verbünden dargestellt. Weiterhin werden die technische Realisierung des Datenangebotes sowie die sich daraus ergebenden Nutzungs- und Selektionsmöglichkeiten beschrieben und die Resonanz auf das seit Dezember 2011 verfügbare Datenangebot skizziert.

Kooperative Datennutzung in Bibliotheken

Elektronische Bibliothekskataloge bieten ihren Benutzern den Nachweis der in der jeweiligen Bibliothek bzw. im jeweiligen Bibliotheksverbund vorhandenen Medien. Für Literaturrecherchen bieten sie in den Titelaufnahmen neben einer bibliographisch zuverlässigen Beschreibung zumeist Normdatenverknüpfungen und Sacherschließungselemente (Schlagwörter und Notationen) sowie zunehmend Links auf Kataloganreicherungen (Inhaltsverzeichnisse, Klappentexte etc.) an. Bei einer ständig wachsenden Anzahl von Aufnahmen ist bereits der direkte Zugang zu den Volldigitalisaten oder Volltexten durch entsprechende Hyperlinks gegeben.

Unauthenticated
Download Date | 12/5/16 3:11 PM

Für die Katalogisierung reduziert sich durch die kooperative Erschließung im Verbund und die Nutzung von Fremddaten der Arbeitsaufwand erheblich. Viele Projekte im bibliothekarischen, wissenschaftlichen und auch kommerziellen Umfeld sind auf qualitativ hochwertige Metadaten angewiesen (z.B. Projekte der Deutschen Forschungsgemeinschaft, Google-Digitalisierung, WorldCat-Einbindung). Das „Give-and-Take" von Metadaten ist in und zwischen den deutschen Bibliotheksverbünden bereits seit langem selbstverständliche Praxis: Zum Beispiel können aus den Katalogisierungsdatenbanken der Verbünde via Z39.50 von Katalogisierern aus allen deutschsprachigen Verbünden Datensätze abgeholt und nachgenutzt werden.

Vom Bibliotheksverbund Bayern wurden in der Vergangenheit bereits größere Datenmengen als Batch-Lieferung an eine Vielzahl von Institutionen abgegeben, unter anderem:

- Datenbankabzug aus dem Katalog der Bayerischen Staatsbibliothek für DFG-geförderte Digitalisierungsprojekte: Reduzierte Datensätze im SISIS- und im MAB-Format. Die Daten werden halbjährlich neu selektiert und bereitgestellt.
- Heritage of the Printed Book (HPB)-Datenbank des Consortium of European Research Libraries (CERL): Komplette Datensätze des Altbestands der Bayerischen Staatsbibliothek aus dem Verbundkatalog.
- OCLC (WorldCat): Komplette Datensätze der Bayerischen Staatsbibliothek, der bayerischen Hochschulbibliotheken und der regionalen staatlichen Bibliotheken des Freistaates im MARC21-Standard-Format aus dem Verbundkatalog.
- OCLC (Virtual International Authority File): Alle Titeldatensätze des Bibliotheksverbundes Bayern, die eine Verknüpfung zu den Personennamen-Datensätzen der nationalen Gemeinsamen Normdatei (GND) haben.
- Europeana: Reduzierte Datensätze aller Digitalisate der Bayerischen Staatsbibliothek im Metadatenformat der Europeana, den Europeana Semantic Elements (ESE)[1] aus dem Verbundkatalog.
- Zentrales Verzeichnis Digitalisierter Drucke (zvdd): Reduzierte Datensätze aller Digitalisate im ESE-Format (Dublin Core) aus dem Verbundkatalog.
- Google und Wikipedia: MARCXML-Daten aus dem Katalog der Bayerischen Staatsbibliothek.

Die Kosten der Erschließungsleistung, die aus öffentlichen Mitteln in die Erstellung und Pflege bibliographischer Katalogdatenbanken investiert wurden und

1 Das Europeana Semantic Elements (ESE) set basiert auf Dublin Core und ist um Europeana-spezifische Felder erweitert. Informationen s. hier: http://www.europeana.eu/schemas/ese/ (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

werden, sind bedeutend. Sie bemessen sich nicht nur am Personal, sondern auch am Transfer herkömmlicher Kartenkataloge in elektronische Kataloge (Retrokonversion). Ohne die praktizierten weitreichenden Kooperationen wäre der insgesamt erforderliche Aufwand noch um ein Vielfaches höher, um annähernd vergleichbare Ergebnisse bei Nachweis und Zugänglichkeit der Bibliotheksbestände zu erreichen. Die bibliographischen Daten der Bibliotheks- und Verbundkataloge stellen also einen hohen Wert dar. Je intensiver die Daten nachgenutzt und verknüpft werden, desto wertvoller werden sie. Es besteht also durchaus ein öffentliches Interesse daran, diese Daten, die den Zugang zu dem riesigen Corpus der in Bibliotheken verfügbaren wissenschaftlichen Literatur und Fachinformation (in gedruckter und digitaler Form) eröffnen, möglichst uneingeschränkt und in heterogenen Nutzungsszenarien wirksam werden zu lassen. Eine entsprechende „Öffnung“ der Bibliotheks- und Verbundkataloge schafft damit die Möglichkeit für vielfältige „Metadatendienste“ als vom klassischen Online-Katalog unabhängige Services für Forschung, Lehre und Studium.

Open Data und Open Catalog

Das kostenfreie Bereitstellen von sogenannten Roh-Daten für die Allgemeinheit zur von Rechten unbelasteten Weiternutzung wird in vielen Bereichen (Schwerpunkt Software und wissenschaftlich verwertbare Inhalte) seit längerem praktiziert. Diese Form der Open-Data-Bewegung ist Teil der Open-Access-Bewegung, die sich als Gegengewicht zur Kommerzialisierung von Wissen und Information positioniert.

Open Catalog ist ein relativ neuer Teilbereich der Open-Data-Bewegung. Dabei werden die bibliographischen Daten eines Kataloges - in der Regel inklusive Norm- und Lokaldaten - zum Komplettdownload zur Verfügung gestellt. Open Catalog ist damit eine Fortsetzung des im Bibliotheksbereich seit langem praktizierten Gedankens der Kooperation und des Datenaustausches zwischen Bibliotheken und Bibliotheksverbünden.

Die direkte Freigabe professionell erstellter Katalogdaten kann unter anderem folgende Ziele unterstützen:

- Sichtbarmachen lokaler bibliographischer Datensammlungen weltweit und direkt im World Wide Web (nicht versteckt im „Deep Web“).
- Nutzbarmachen dieser Daten zu speziellen wissenschaftlichen Zwecken (z.B. Unterstützung bibliographischer Forschung).
- Aufwerten und Ergänzen etablierter Webangebote Dritter (z.B. Wikipedia).

Unauthenticated
Download Date | 12/5/16 3:11 PM

- Bereitstellen der Grundlagen für neue oder neu zu entwickelnde Informationsdienstleistungen.
- Integration der Katalogdaten in das Semantic Web, verbunden mit der Möglichkeit, von Ergänzungen und Aufwertungen Dritter an den eigenen Daten zu profitieren.
- Bestreben, mit öffentlichen Mitteln erstellte Inhalte (Volltexte) kostenfrei zugänglich zu machen: Open Catalog leistet den bibliographischen Beitrag zum Open-Access-Gedanken. Damit kann unter anderem einfacher vermieden werden, dass Überschneidungen mit anderen Digitalisierungsprojekten entstehen.
- Förderung der Kooperation bibliothekarischer Einrichtungen national und international, zum Beispiel mit dem Ziel, eine kostenfreie, weltweit zugängliche Rechercheplattform zu schaffen.

Open Catalog: Modelle der Freistellung

Wenn Katalogdaten als Open Catalog zur Verfügung gestellt werden, sind vor allem die rechtlichen Rahmenbedingungen der Datenfreistellung zu klären. Es gibt dafür grundsätzlich zwei Modelle: die Bereitstellung der Daten unter einer bestimmten Lizenz oder die völlige Freigabe, das heißt der Verzicht auf jegliche Rechte an den Daten.

Eine Übersicht über Standard-Lizenzverträge findet sich auf den Seiten der Open Data Commons[2] und - angepasst an das deutsche Recht - auf den Seiten der Creative Commons Deutschland[3].

Beispiele für Lizenzen sind:

- *Attribution-Lizenz*
  Sie verlangt die Namensnennung des Urhebers der Daten.
- *Non-Commercial-Lizenz*
  Die Daten werden unter der Auflage freigegeben, keine kommerzielle Nutzung zu verfolgen, oder unter den Vorbehalt einer vorherigen Zustimmung gestellt.
- *Share-Alike-Lizenz*
  Die Daten müssen auch nach Veränderungen unter derselben Lizenz, unter der die Ausgangsdaten stehen, weitergegeben werden.

2 http://www.opendatacommons.org/licenses/ (27.06.2012)
3 Was ist CC? http://de.creativecommons.org/was-ist-cc/ (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

– *CC0 Public Domain Dedication*
  Die Daten werden ohne Einschränkungen freigegeben.[4]

Im Unterschied zu den bereits bekannten Creative-Commons-Modellen handelt es sich bei der CC0 1.0 Universal Public Domain Dedication nicht um eine Lizenz im engen Sinn, sie sagt vielmehr aus, dass das mit CC0 versehene „Datum" gemeinfrei ist. Um dem Gedanken von Linked Open Data voll zu entsprechen, ist es grundsätzlich erforderlich, die Daten ohne Einschränkung zur Verfügung zu stellen.[5] Damit erspart man sich zudem jeglichen Verwaltungsaufwand zur lizenzrechtlichen Kontrolle der Anwender. Die Alternative ODC PDDL[6] wurde zum Zeitpunkt der Entscheidung noch nicht diskutiert. Es ist jedoch durchaus vorstellbar, in Zukunft die OpenDataCommons-Lizenzen zu verwenden, sollten kommende Entwicklungen zeigen, dass diese für Metadaten besser geeignet sind.

Open Catalog und Linked Data

Seinen vollen Nutzen entfaltet ein als Open Catalog verfügbar gemachter Bibliotheks- oder Verbundkatalog dann, wenn die zur freien Verwendung bereitgestellte Datenbasis zugleich als Linked Data für Nutzungsszenarien des Semantic Web aufbereitet wird.[7] Die heute gängige Informationsvermittlung, bei der sich der Nutzer Informationen besorgt, indem er z.B. eine Katalog- oder Suchmaschinenrecherche absetzt, wird im Semantic Web zunehmend dadurch abgelöst, dass kontextbezogen, also zu bestimmten Fragestellungen, weiterführende und ergänzende Informationen automatisiert angeboten werden. In dieses Semantic Web sollten unbedingt auch Katalogaufnahmen einbezogen werden, die auf die reichen Bestände von Bibliotheken in ihren umfassenden Wissensspeichern zielen. Traditionell enthalten die Katalogdaten bereits eine Reihe von Voraussetzungen für maschinenlesbare Verlinkungen (Verknüpfungen zwischen hier-

4 http://creativecommons.org/publicdomain/zero/1.0/ (27.06.2012)

5 Vgl. Tim Berners Lee: Linked Data. http://www.w3.org/DesignIssues/LinkedData.html (27.06.2012), besonders den Abschnitt „Is your linked open data 5 star?". Zur näheren Definition des Begriffes „frei" bzw. engl. „open" vgl.: Defining the open in open data, open content and open services. http://opendefinition.org/ (27.06.2012)

6 Open Data Commons: Public Domain Dedication and License: http://opendatacommons.org/licenses/pddl/

7 Zur Definition der Begriffe „linked data" und „open data" vgl. W3C Incubator Group: Library Linked Data Incubator Group final report. 2011. Kapitel 1 http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Scope_of_this_report (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

archischen Werken, Verknüpfungen zu Normdaten, z.B. Personennamen, Orten, Sachbegriffen). Durch Linked-Data-Verfahren können diese bereits vorhandenen Verlinkungen in ganz neue Szenarien einbezogen werden. Dies trifft insbesondere für Anwendungen zu, die nicht bibliotheksspezifisch sind, oder die die Daten mit Daten aus ganz anderen Bereichen zusammenführen.[8] Der Nutzer dieser Angebote kann zum Beispiel spezielle Fragestellungen (etwa nach Konzepten und Zusammenhängen) beantworten, und automatisch zusätzliche Informationen aus DBpedia oder speziellen Fachdatenbanken mit in die Abfrage einbeziehen. Er ist hierbei nicht auf die Abfragemöglichkeiten des OPACs (mit seinen vorkonfigurierten Suchmaschinen-Indices) beschränkt.

Solche Mashups mit anderen Angeboten im Semantic Web (z.B. mit Geolocation-Diensten oder Sozialen Netzwerken) können neue, intelligente Dienste ermöglichen. Zudem können die Nachnutzer der als Linked Data bereitgestellten offenen Daten ihre eigenen Angebote qualitativ verbessern. Wikipedia nutzt beispielsweise die Titeldaten der Bayerischen Staatsbibliothek zur Anreicherung der Literaturhinweise in der Online-Enzyklopädie. Das Beispiel DBpedia zeigt den Weg zu einer neuen Generation von Rechercheanwendungen und Dienstleistungen für Wissenschaftler und Bibliotheksbenutzer (sog. Intelligente Filtersuche). Antworten auf komplexe Fragestellungen werden hier von Linked-Data-Anwendungen generiert, indem strukturierte Informationen aus Wikipedia-Artikeln gezogen und zusammen mit anderen frei zugänglichen Daten zu einer neuen Datenbasis verbunden werden. Eine große Zahl von Anwendungsszenarien findet sich auf der „Use cases"-Seite der Library Linked Data Incubator Group des World Wide Web Consortium (W3C).[9]

Als Kernkomponente des semantischen Netzes gilt das Resource Description Framework (RDF), das als Datenmodell sogenannte Tripel verwendet. RDF-Tripel bestehen jeweils aus einem Subjekt (der beschriebenen Ressource), einem Prädikat (einer Eigenschaft) und einem Objekt (einer Zeichenkette oder einer verlinkten Ressource). Dabei werden Subjekt und Prädikat immer in Uniform Resource Identifiers (URI) ausgedrückt, das Objekt kann entweder als URI oder als Zeichenkette dargestellt werden. Das folgende Beispiel stellt ein solches Tripel dar:

- http://lod.b3kat.de/title/BV035280616 ist der URI des Subjekts, im konkreten Fall der Titel des Werkes *Geschichte Europas : von 1815 bis zur Gegenwart*,
- http://purl.org/dc/elements/1.1/creator ist der URI des Prädikats, hier die Verfasserangabe (creator) und

8 Ebd. Kapitel 3.1.1-3.1.3: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Library_data_is_not_integrated_with_Web_resources (27.06.2012)

9 W3C: http://www.w3.org/2005/Incubator/lld/wiki/UseCases (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

– http://d-nb.info/gnd/124838464 ist der URI eines Objekts, hier der Link zum Normdatensatz des Verfassers *Rainer Liedtke* in der Gemeinsamen Normdatei (GND).

BVB BibliotheksVerbund Bayern — kobv

Geschichte Europas at lod.b3kat.de

http://lod.b3kat.de/title/BV035280616

Property	Value
isbd:P1006	von 1815 bis zur Gegenwart
isbd:P1016	Paderborn [u.a.]
geonames:countryCode	DE
dcterms:creator	<http://d-nb.info/gnd/124838464>
dcterms:description	Rainer Liedtke
frbr:exemplar	<http://lod.b3kat.de/bib/DE-11/item/BV035280616> <http://lod.b3kat.de/bib/DE-12/item/BV035280616> <http://lod.b3kat.de/bib/DE-127/item/BV035280616> <http://lod.b3kat.de/bib/DE-188/item/BV035280616> <http://lod.b3kat.de/bib/DE-19/item/BV035280616> <http://lod.b3kat.de/bib/DE-20/item/BV035280616> <http://lod.b3kat.de/bib/DE-29/item/BV035280616> <http://lod.b3kat.de/bib/DE-355/item/BV035280616> <http://lod.b3kat.de/bib/DE-384/item/BV035280616> <http://lod.b3kat.de/bib/DE-473/item/BV035280616> <http://lod.b3kat.de/bib/DE-521/item/BV035280616> <http://lod.b3kat.de/bib/DE-703/item/BV035280616> <http://lod.b3kat.de/bib/DE-824/item/BV035280616> <http://lod.b3kat.de/bib/DE-Aug4/item/BV035280616> <http://lod.b3kat.de/bib/DE-Di1/item/BV035280616> <http://lod.b3kat.de/bib/DE-M352/item/BV035280616>
dcterms:extent	301 S. Ill., Kt.
foaf:homepage	<http://worldcat.org/oclc/492091667>
bibo:isbn	9783506765796 9783825232054
dcterms:issued	2010 (xsd:int)
dcterms:language	<http://id.loc.gov/vocabulary/iso639-2/ger>
rdagr1:placeOfPublication	<http://id.loc.gov/vocabulary/countries/gw>
dcterms:publisher	Schöningh
is owl:sameAs of	<http://lod.b3kat.de/isbn/9783506765796> <http://lod.b3kat.de/isbn/9783825232054>
owl:sameAs	<http://d-nb.info/991844459> <http://www.culturegraph.org/about/DNB-991844459>
dc:subject	900 (dcterms:DDC) 940 (dcterms:DDC) Geschichte 1815-2008
dcterms:subject	<http://d-nb.info/gnd/4015701-5> <http://dewey.info/class/900/about> <http://dewey.info/class/940/about> <http://lod.b3kat.de/rvk/NK1500> <http://lod.b3kat.de/rvk/NK1600> <http://lod.b3kat.de/rvk/NP1200>
dcterms:title	Geschichte Europas
rdf:type	bibo:Book

Diese Seite zeigt Informationen aus dem SPARQL endpoint http://lod.b3kat.de/sparql.
As Turtle | As RDF/XML | Browse in Disco

Abbildung 1: Beispieltitel zur Illustration des RDF-Datenmodells.

Unauthenticated
Download Date | 12/5/16 3:11 PM

Der B3Kat mit seinen 23 Millionen Katalogdatensätzen wurde auf diese Weise in mehr als 600 Millionen Tripel zerlegt, wobei bei Bedarf durch feinere Datenmodellierung noch wesentlich mehr Tripel erzeugt werden können.

Mit diesem universellen und vom W3C standardisierten[10] Datenmodell lassen sich die im Netz verstreut liegenden Informationen bündeln und zusammen verarbeiten. Nach einem Abgleich bzw. Mapping der den RDF-Tripeln zugrundeliegenden Ontologien können echte Mehrwerte erzielt werden (Semantic Enhanced Retrieval, Linked Data Enrichment), da die Abfragen dann nicht mehr auf eine bestimmte fachliche Domäne oder Community beschränkt sind.

Nicht zuletzt, um eine nahtlose und automatische Zusammenführung aus unterschiedlichsten Quellen zu ermöglichen, ist die oben beschriebene offene Lizenzierung möglichst großer Datenmengen notwendig.

Der B3Kat als Linked-Open-Data-Service

Der B3Kat, der 23 Millionen Datensätze aus 180 wissenschaftlichen Bibliotheken Bayerns, Berlins und Brandenburgs umfasst, ist der gemeinsame Verbundkatalog des Bibliotheksverbundes Bayern (BVB) und des Kooperativen Bibliotheksverbundes Berlin-Brandenburg (KOBV) und wird technisch vom BVB betrieben. Die Entscheidung zur Freistellung des B3Kat als Linked Open Data erfolgte bereits Anfang 2011 auf der Grundlage einer Empfehlung der gemeinsamen *Kommission für Erschließung und Metadaten (KEM)* von BVB und KOBV. Die Kommission hat ihre Empfehlung in acht Punkten zusammengefasst:

1. Bibliographische Metadaten können einen wesentlichen Beitrag für das sogenannte Semantic Web liefern.
2. Open Bibliographic Data erhöhen wesentlich die Sichtbarkeit von Bibliotheken und Bibliotheksbeständen.
3. Linked Open Data haben das Potential für innovative und bessere Informationsangebote.
4. Die bisher entstandenen Kosten, die in den Verbundkatalog eingeflossen sind, sind hoch, aber der Wert der Katalogdaten erhöht sich durch Linked Open Data.
5. Die bilaterale Abgabe von Katalogdaten an Dritte ist bisher schon üblich; mit Open Data würde diese Form der Kooperation mit vergleichsweise geringerem Aufwand konsequent fortgeführt und auf jedermann ausgeweitet.

10 W3C: Resource Description Framework (RDF): “Concepts and abstract syntax“. http://www.w3.org/TR/rdf-concepts/ (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

6. Die Kommission empfiehlt deshalb, die Daten des B3Kat als Open Data bereitzustellen. In einem ersten Schritt können Datensätze und -elemente, für die eine endgültige rechtliche Klärung noch nicht vorliegt, ausgeklammert werden. Die einzelnen B3Kat-Teilnehmerbibliotheken sollten die Möglichkeit haben, ihre Bestände von der Freistellung auszunehmen.
7. Da offene Daten nur dann ihren vollen Wert entfalten, wenn sie uneingeschränkt zur Verfügung stehen, schlägt die Kommission vor, die Daten aus fachlicher Sicht ohne rechtliche Einschränkungen freizustellen.
8. Die Kommission empfiehlt des Weiteren, die Daten in Form von RDF-Tripeln und zusätzlich in einem klassischen Bibliotheksformat (vorzugsweise MARCXML) bereitzustellen.

Auf der Grundlage dieser Empfehlung wurde die Freistellung des B3Kat als Linked Open Data von den Gremien und Kommissionen im BVB und KOBV einhellig befürwortet und von der Verbundzentrale des BVB sowie dem Stabsreferat Informationstechnologie der Bayerischen Staatsbibliothek in zwei Teilprojekten bis zum Dezember 2011 umgesetzt. Im Teilprojekt 1 *Open Data* erfolgte die Bereitstellung des B3Kat im international standardisierten bibliothekarischen Format MARCXML zur freien Nachnutzung. Schon mit Blick auf die Metadatennutzung im Rahmen internationaler Bibliothekskooperationen und kommerzieller wie nicht-kommerzieller Webangebote (Google, Wikipedia etc.) schied das im deutschsprachigen Raum gebräuchliche Format MAB für diesen Zweck aus. Die MAB-Ausgangsdaten mussten deshalb zunächst möglichst komplett ins MARC-Format konvertiert und dann in MARCXML umgesetzt werden. Zusätzlich zu den bibliographischen Feldern wurden auch die Bibliothekskennzeichen der besitzenden B3Kat-Bibliotheken in die MARC-Datensätze übernommen.

Die Basislieferung mit circa 23 Millionen Titeldatensätzen wurde in drei Paketen auf einem Web-Server des BVB verfügbar gemacht, neue und geänderte Datensätze können kontinuierlich aus einem OAI-PMH-Repository abgeholt werden. Der MARCXML-Datenpool stellt zudem die datentechnische Ausgangsbasis für das Teilprojekt 2 *Linked Open Data* zur Umwandlung der Daten in das dort erforderliche Datenformat RDF/XML dar. Die Basislieferung wird halbjährlich neu bereitgestellt. Das Angebot ist unter www.bib-bvb.de/open-data.html abrufbar. Abbildung 2 zeigt die Optionen zur Nutzung und zum Download der B3Kat-Daten.

Von der Freistellung ausgenommen wurden und werden die URLs zu Kataloganreicherungen (Umschlagseiten, Tables of Content, Abstracts etc.), die von den Firmen Aux Amateurs de Livres und Casalini bezogen werden, sowie die Bestände einiger kleinerer Verbundbibliotheken, die aus verschiedenen Gründen eine Veröffentlichung nicht wünschen.

Unauthenticated
Download Date | 12/5/16 3:11 PM

Im Teilprojekt 2 *Linked Open Data* erfolgte die Freistellung der als Open Data vorliegenden B3Kat-Verbunddatenbank als Linked-Open-Data-Service. Zu diesem Zweck wurden frei verfügbare Dateien erzeugt, die die Katalogdaten im RDF/XML-Format enthalten. Daneben wird eine hoch verfügbare und performante SPARQL-Schnittstelle[11] angeboten, die Abfragen auf diese Daten erlaubt.

Open Data: Der BVB und der KOBV öffnen ihren Katalog

Der gemeinsame Verbundkatalog B3Kat des Bibliotheksverbundes Bayern (BVB) und des Kooperativen Bibliotheksverbundes Berlin-Brandenburg (KOBV) umfasst über 23 Millionen bibliografische Datensätze der Hochschulbibliotheken und weiterer Bibliotheken Bayerns, Berlins und Brandenburgs.

Die Eckdaten:

- Der Katalog mit Stand 12.06.2012 wird im Standard-Datenformat MARC-XML bereitgestellt.
- Alle seit dem 13.06.2012 neu aufgenommenen und geänderten Daten stehen zum Harvesting als OAI-PMH-Repository unter http://bvbr.bib-bvb.de:8991/aleph-cgi/oai/oai_opendata.pl?verb=ListRecords&set=OpenData&metadataPrefix=marc21 zur Verfügung.
- Die Veröffentlichung erfolgt unter der Creative Commons License CC0.

Die direkten Links zum herunterladen des B3Kat:

Open Data B3Kat Export 2012 Teil 1
Open Data B3Kat Export 2012 Teil 2
Open Data B3Kat Export 2012 Teil 3
Open Data B3Kat Export-Beispiel-Daten

Abbildung 2: B3Kat – Open Data Service.

Die Realisierung des B3Kat als Linked-Open-Data-Service erfolgte konkret in vier Arbeitsschritten:

- Vergabe persistenter *URIs* für jeden Katalogdatensatz. Dafür wurde die Verbund-ID als vergleichsweise bester verfügbarer Identifier ausgewählt. Um die allgemeine Zugänglichkeit der URIs zu erhöhen, wurden zusätzlich Links von auf ISBNs basierenden URIs zu den eigentlichen Titel-URIs erzeugt.
- Inhaltliche Zuweisung der Kategorien der Katalogdatensätze zu Eigenschaften (RDF-Prädikaten). Hier werden die wichtigsten MARC-Tags in RDF-Tripel gemappt. Für das Mapping wurde nach zwei Prinzipien vorgegangen. Erstens wurden möglichst etablierte *Ontologien* verwendet; zweitens wurden, wo immer möglich, Links auf andere Daten (z.B. Normdaten) der Ausgabe von

11 *SPARQL Protocol and RDF Query Language* ist die vom W3C entwickelte definierte Abfragesprache für RDF-Daten.

Unauthenticated
Download Date | 12/5/16 3:11 PM

Verbalisierungen vorgezogen. So können die Daten sehr gut mit bereits in RDF vorliegenden Daten vernetzt werden. Zur Zeit werden unter anderem die Gemeinsame Normdatei (GND), WorldCat, Library of Congress Subject Headings (LCSH) und die Dewey Decimal Classification (DDC) verlinkt. Außerdem werden Links auf die entsprechenden Titelaufnahmen anderer Verbünde und auf culturegraph.org hinzugefügt.

– Auswertung der Bestandsnachweise der B3Kat-Bibliotheken und deren Abbildung in RDF, so dass automatisch abgefragt werden kann, welche Bibliotheken die beschriebenen Medien besitzen.
– Bereitstellung der Daten als Datei-Dumps für umfangreiche Analysen oder lastintensive Anwendungen, und in einem *Triplestore* über dessen SPARQL-Schnittstelle für Mashups und maschinelle Abfragen.

Das Angebot ist unter http://lod.b3kat.de abrufbar. Abbildung 3 zeigt die Möglichkeiten zur Nutzung des B3Kat als Linked Open Data, Abbildung 4 die verwendeten Ontologien.

Abbildung 3: B3Kat – Linked Open Data Service.

Entsprechend der Empfehlung der Kommission für Erschließung und Metadaten wird der B3Kat sowohl als Open Data wie auch als Linked Open Data unter der CC0 1.0 Universal Public Domain Dedication freigegeben und steht damit zur

Unauthenticated
Download Date | 12/5/16 3:11 PM

völlig uneingeschränkten Nachnutzung zur Verfügung, explizit auch in kommerziellen Anwendungsszenarien. Seit seiner Veröffentlichung im Dezember 2011 wurde der B3Kat als Open Data im MARCXML-Format mehr als 650 mal komplett oder in Teilen heruntergeladen, als Linked Open Data in RDF knapp 1300 mal.

Mapping der MARC-Felder auf RDF-Prädikate

Liste der zur Zeit verwendeten Ontologien

- rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
- owl: http://www.w3.org/2002/07/owl#
- dc: http://purl.org/dc/elements/1.1/
- dcterms: http://purl.org/dc/terms/
- bibo: http://purl.org/ontology/bibo/
- frbr: http://purl.org/vocab/frbr/core#
- foaf: http://xmlns.com/foaf/0.1/
- geonames: http://www.geonames.org/ontology#
- marcrel: http://id.loc.gov/vocabulary/relators/
- rdagr1: http://rdvocab.info/Elements/

Weitere, hilfsweise verwendete Ontologien

- dcmitype: http://purl.org/dc/dcmitype/
  (Für die Medientypen (rdf:type) als RDF-Object verwendet)
- skos: http://w3.org/2004/02/skos/core#
  (Für die hilfsweise angelegten RVK-URIs, die mit dem Prädikat dc:subject verwendet sind. Nicht bei den eigentlichen Titeldaten.)

Abbildung 4: B3Kat-LOD-Projekt – verwendete Ontologien.

Inzwischen wird die SPARQL-Schnittstelle zum Beispiel von recensio.net[12], der Rezensionsplattform für die europäische Geschichtswissenschaft verwendet. Recensio-Redakteure müssen auf ihrer Plattform die Rezension selbst und die ISBN/ISSN des rezensierten Werkes eingeben. Die Metadaten für das rezensierte Werk werden in einem halbautomatischen Prozess aus dem Triplestore in die Recensio-Datenbank geladen.

Open Data Bayern und Open Catalog

Die Bereitstellung des B3Kat als Open Data und als Linked Open Data wurde vom bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst mit Sondermitteln unterstützt und ist Teil der Open-Data-Initiative der Bayerischen Staatsregierung. Das im Rahmen dieser Initiative im Herbst 2011 eröffnete Portal *OpenData Bayern*[13] bietet eine Vielzahl von Datensets und Applikationen staatlicher Behörden und Einrichtungen des Freistaates zur Nachnutzung an, unter anderem Geodaten, Umweltdaten und statistische Daten. Der B3Kat ist seit seiner

12 http://www.recensio.net (27.06.2012)
13 http://opendata.bayern.de (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

anderem Geodaten, Umweltdaten und statistische Daten. Der B3Kat ist seit seiner Freistellung als Linked Open Data ebenfalls über dieses Portal erreichbar. Die Open-Data-Inititiative der Bayerischen Staatsregierung ist explizit dem Gedanken der möglichst uneingeschränkten Nachnutzung von mit öffentlichen Mitteln erstellten Daten verpflichtet. Auf der Homepage von OpenData Bayern heißt es hierzu: „Frei zugängliche Daten bergen ein großes Innovationspotential, wenn sie auch außerhalb der Verwaltung genutzt werden können, z.B. für neue Analysen und Geschäftsmodelle, für journalistische Zwecke und in der Wissenschaft. Oder wenn Bürgerinnen und Bürger sie nutzen, um hieraus selbst neue Anwendungen und Dienste zu entwickeln. Nicht zuletzt tragen frei zugängliche Daten aber auch dazu bei, die öffentliche Verwaltung selbst transparenter zu machen.“[14]

Abbildung 5: Preisverleihung durch Bundesinnenminister Dr. Hans-Peter Friedrich (Apps für Deutschland/Foto: Oliver Feist).

Das Angebot des B3Kat mit 23 Millionen Datensätzen als maschinenlesbarem Linked-Open-Data-Service stellt derzeit das umfassendste Open-Data-Projekt im deutschsprachigen Raum dar und wird so auch in der Öffentlichkeit wahrgenommen. Im Rahmen des ersten nationalen Programmierwettbewerbs „Apps für Deutschland“ wurde das Angebot auf der CeBIT 2012 von Bundesinnenminister

14 http://opendata.bayern.de/ueber.html (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

Dr. Hans-Peter Friedrich in der Kategorie „Daten“ mit dem 2. Preis ausgezeichnet. „Besonders erfreulich“, so Daniel Dietrich von der Open Knowledge Foundation Deutschland, „war die Einreichung der Bibliotheksverbünde von Bayern, Berlin und Brandenburg. Sie zeigt, dass Länder- oder Behördengrenzen kein Hindernis sind, wenn der Nutzen für die Bürgerinnen und Bürger im Vordergrund steht.“[15]

Fazit

Innerhalb der Bibliothekswelt wurden und werden bibliographische Daten in vielfältiger Weise ausgetauscht, nachgenutzt, miteinander abgeglichen und angereichert. Dies geschieht im Regelfall jedoch aufgrund von direkten, zumeist bilateralen Vereinbarungen. Die Bereitstellung von Datenselektionen ist aufgrund der zunehmenden Zahl von Anfragen und Nutzungsszenarien, aber insbesondere auch durch die speziellen Wünsche hinsichtlich Selektionskriterien und Datenformaten, ressourcenintensiv und für den Endanwender oft vergleichsweise unkomfortabel.

Für eine weit stärkere Sichtbarkeit der Bibliotheken und ihrer Bestände im Internet und eine bruchlose Einbindung in das schnell wachsende semantische Netz ist es notwendig, die Daten aus ihren „Datensilos“ im Deep Web zu befreien und direkt, ohne Wechsel in proprietäre Anwendungen, zugänglich zu machen. Linked Data bietet ein sehr großes Potential, die qualitativ hochwertigen bibliographischen Metadaten für innovative und bessere Informationsangebote zu nutzen. Sehr erleichtert wird dies, wenn die Daten den Nachnutzern als Open Data zur Verfügung gestellt werden.

Mit der Bereitstellung des B3Kat als Linked Open Data unter der CC0 Public Domain Dedication haben die Bibliotheksverbünde BVB und KOBV die Möglichkeit genutzt, ihre reichhaltigen bibliographischen Datenbestände außerhalb der begrenzten Möglichkeiten klassischer Bibliothekssysteme in den vielfältigen Anwendungsszenarien des Semantic Web wirksam werden zu lassen. Die Vorteile dieses Angebotes – insbesondere die weltweite Verfügbarkeit der Daten und die verstärkte Homogenisierung der Metadatendienste – übersteigen dabei aus Sicht der am B3Kat teilnehmenden Bibliotheken bei weitem den „Preis“, der für jede Freigabe von Daten als Linked Open Data zu zahlen ist: den „Kontrollverlust“ über die eigenen Daten“assets“ sowie die – zumindest tendenzielle – „Invisibility“ des ursprünglichen Datenherstellers im Zuge iterierter Nachnutzungen.

15 http://apps4deutschland.de/ (27.06.2012)

Unauthenticated
Download Date | 12/5/16 3:11 PM

Zitierte Webseiten

Alle Links wurden am 27. Juni 2012 aufgerufen.

Apps für Deutschland. Open-Data-Wettbewerb. http://apps4deutschland.de/
Creative Commons: CC0 1.0 Universal Public Domain Dedication. http://creativecommons.org/publicdomain/zero/1.0/legalcode
Creative Commons Deutschland: Was ist CC? http://de.creativecommons.org/was-ist-cc/
Defining the open in open data, open content and open services. http://opendefinition.org/
Europeana: Europeana Semantic Elements (ESE). http://www.europeana.eu/schemas/ese/
OpenData Bayern. http://opendata.bayern.de
Open Data Commons. Legal tools for open data. http://www.opendatacommons.org/licenses/
Open Data Commons: Public Domain Dedication and License: http://opendatacommons.org/licenses/pddl/
Recensio. http://www.recensio.net
Tim Berners Lee: Linked data. http://www.w3.org/DesignIssues/LinkedData.html
W3C Incubator Group: Library Linked Data Incubator Group final report. 2011. http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/
W3C: Resource Description Framework (RDF): concepts and abstract syntax. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/
W3C: Use cases / case studies. http://www.w3.org/2005/Incubator/lld/wiki/UseCases
W3C: SPARQL protocol for RDF. http://www.w3.org/TR/rdf-sparql-protocol/
W3C: SPARQL query language for RDF. http://www.w3.org/TR/rdf-sparql-query/

Unauthenticated
Download Date | 12/5/16 3:11 PM