

OPEN CATALOG: EINE NEUE PRÄSENTATIONSMÖGLICHKEIT VON BIBLIOTHEKSDATEN IM SEMANTIC WEB?

Louise Rumpf

Universitätsbibliothek Bamberg / Bibliotheksakademie Bayern

louise.rumpf@uni-bamberg.de

1. Einleitung

1.1 Problemstellung

In den letzten Jahren lässt sich eine zunehmende Integration von Internetangeboten unterschiedlicher Anbieter beobachten. Bibliotheken haben diesen Trend bislang einseitig aufgegriffen, indem sie Angebote Dritter in die OPAC-Anzeige integrieren¹; ihre Katalogdaten sind aber bislang im Internet schlecht auffindbar und für Andere schlecht nutzbar. Durch die immer bessere Verfügbarkeit nicht-bibliothekarischer Informationen im Internet haben Bibliotheken ihre „gatekeeper-Funktion“ beim Zugang zu Informationen verloren², und das Verhältnis zwischen Bibliotheken und ihren Kundeln wandelt sich: Die Kunden sind weniger auf die Bibliothek angewiesen. Zudem haben sich die Suchgewohnheiten immer weiter von den in OPACs vorgesehenen Recherchemöglichkeiten entfernt.³

„Open Catalog“, die Freigabe und Verknüpfung bibliothekarischer Daten im Internet, wird in der bibliothekarischen Diskussion oft als Ausweg aus diesem Dilemma genannt. Verschiedene Bibliotheken und Bibliotheksverbände haben in den letzten Jahren ihre Katalogdaten als **linked open data** (LOD) freigegeben. Auch der Wissenschaftsrat hat jüngst in seinen „Empfehlungen zur Zukunft des bibliothekarischen Verbundsystems in Deutschland“ die Freigabe von Katalogdaten als *linked open*

data als Möglichkeit genannt, der veränderten Informationsinfrastruktur und der gewandelten Nutzeranforderungen gerecht zu werden.⁴ Auch außerhalb des Bibliothekswesen ist die Menge der als LOD bereitgestellten Daten in den letzten Jahren deutlich angewachsen.⁵

In diesem Beitrag soll erläutert werden, was mit „offenen Katalogen“ gemeint ist, welche Unterschiede dabei zu „traditionellen“ OPACs bestehen und welche Chancen und Risiken die Freigabe bibliografischer Daten im Internet für Bibliotheken bietet. Dabei wird ausführlich auf die Datenstruktur eingegangen, die den jeweiligen Modellen zugrunde liegt.

Die Freigabe und Verlinkung bibliothekarischer Daten führt zu Nutzungsmöglichkeiten, die weit über eine veränderte Kataloganzeige für Bibliothekskunden hinausgehen. Vor dem Hintergrund, dass die Rolle der Bibliotheken in der Wissensgesellschaft einem ständigen Wandel unterliegt und bibliothekarische Katalogisierungsarbeit wegen der zunehmenden Metadatenbereitstellung durch Dritte⁶ einem ständigen Rechtfertigungsdruck unterliegt, werden die „neuen Präsentationsmöglichkeiten im Semantic Web“ hier weit definiert: Neben den Anzeigemöglichkeiten für bibliografische Informationen wird auch auf andere Nutzungsmöglichkeiten eingegangen, die sich für Bibliotheken durch die Freigabe und Verlinkung von Daten im Internet ergeben und die dazu beitragen können, in der Wissensgesellschaft sinnvolle Aufgaben von Bibliotheken zu definieren und sie auch als Institutionen zu rechtfertigen. Dabei stehen Katalogdaten im Zentrum der Analyse; es wird aber auch auf die Chancen und Risiken der Bereitstellung anderer bibliothekarischer Daten als LOD eingegangen.

1.2 Aufbau des Beitrags

Nach der Definition zentraler Begriffe (Kapitel 2) wird zunächst aufgeführt, welche Vorteile Bibliotheken durch Bereitstellung bibliothekarischer Daten als *linked open data* haben könnten (Kapitel 3). Im Folgenden werden kurz die technischen Grundlagen von *linked open data* erläutert (Kapitel 4), und es wird ausgeführt, welche technischen, rechtlichen und inhaltlichen Anforderungen Bibliotheksdaten erfüllen müssen, um sinnvoll in die LOD-Cloud eingebunden werden zu können (Kapitel 5). Dabei wird sowohl auf die Unterschiede zur Datenstruktur in MAB/MARC als auch kurz

auf Kompatibilitätsfragen von RDA und RDF eingegangen. Im Folgenden wird dargestellt, welche Präsentations- und Nutzungsmöglichkeiten sich für *library linked data* innerhalb und außerhalb des Bibliothekswesens ergeben, welche LOD-Projekte es im Bibliotheksumfeld bereits gibt und welche Probleme dabei bestehen (Kapitel 6). In Kapitel 7 werden die Ergebnisse zusammengefasst.

2. Zentrale Begriffe

Ausgangspunkt für die Bestrebungen von Bibliotheken, Katalog- und andere Daten freizugeben, war die Idee des *semantic web*⁷: Informationen sollten im Internet auf eine Art und Weise zur Verfügung gestellt werden, dass sie nicht nur von Menschen, sondern auch von Computern interpretiert werden können.⁸ Durch die Verbindung von Informationen aus verschiedenen Quellen werde es möglich, neues Wissen zu erschließen.⁹ Dazu seien zum einen offene und flexible Standards zur Beschreibung der Informationen notwendig, und zum anderen Methoden, um mit Hilfe formaler Logik neue Informationen aus vorhandenen zu generieren.¹⁰

Obwohl die Idee des *semantic web* bereits vor zehn Jahren aufkam und mit RDF (Resource Description Framework) ein allgemeiner Standard zur Informationsbeschreibung etabliert wurde, ist ihre Umsetzung bislang noch nicht besonders weit vorangeschritten. So konstatierten etwa David J. Browne und Richard Boulderstone von der British Library noch 2008:

This [the semantic web] is still a challenge for the future. It may take some ten years before the full effects of the semantic web are felt. It remains a distant vision of some of the leading communication scientists, one which is worth considering in developing strategies for publishers and libraries for mid to long term, but its impact currently remains marginal. [...] [P]arts of the semantic web are being applied selectively in scholarly communication, and over the years the progress being achieved in the development of reliable standards, protocols and procedures will inevitably impact on the scholarly communication system in some way. But when and how remain open questions. The Web 3.0 which will harness the promises of the semantic web are still a glimmer in a few enthusiast's eyes.¹¹

Auch wenn zweifelhaft ist, ob sich die Idee des *semantic web* jemals komplett wird durchsetzen können, steht fest, dass es nur dann überhaupt funktionieren kann, wenn nicht nur die technischen Standards dafür bereitstehen, sondern auch hinrei-

chend große Datenmengen in einer dafür passenden Form – im oben bereits angesprochenen RDF-Format – zur Verfügung stehen¹² und nutzbar sind.

Als **linked data** werden Daten bezeichnet, die nach gewissen Standards aufbereitet im Internet publiziert werden und über URIs (*Uniform Resource Identifiers*, siehe Kapitel 4.1) miteinander verknüpft sind¹³: „Linked Data ist der erste Schritt und das Herzstück des Semantic Webs.“¹⁴ Durch die Nutzung von Standards wie RDF wird gewährleistet, dass die Daten miteinander technisch kompatibel sind.¹⁵

Daten sind dann **open data**, wenn sie unter rechtlichen Bedingungen zur Verfügung gestellt werden, die jedermann ihre kostenfreie Nutzung und Weiterverwendung zu beliebigen Zwecken erlauben („freie Lizenz“). Eine solche freie Lizenz ist die „rechtliche Basis der Nutzung und Kombination verteilter Daten im Netz“¹⁶: Sie ermöglicht eine weitreichende Nutzung der Daten, ist aber bei nicht selbst erstellten Daten nur unter Wahrung der Urheberrechte Dritter möglich.¹⁷

Als **linked open data** (LOD) werden Daten bezeichnet, auf die sowohl die oben beschriebenen *linked data*- als auch die *open data*-Kriterien zutreffen:

‘Linked Open Data‘ sind Daten, die im WWW so zur Verfügung gestellt werden, dass sie öffentlich ohne rechtliche Schranken verfügbar sind (open) und mittels geeigneter Beschreibungssprachen automatisch mit anderen Daten verknüpft werden können (linked). Wenn eine ausreichende Anzahl geeigneter Linked-Open-Data-Sammlungen zur Verfügung steht, kann aus dem WWW ein Semantic Web werden [...].¹⁸

Da die Verlinkung zwischen Daten sich nur dann positiv auswirken kann, wenn genügend verlinkbare Daten vorhanden sind, wird oft gefordert, dass Bibliotheksdaten auch dann als open data zur Verfügung gestellt werden sollten, wenn die jeweilige Bibliothek oder der jeweilige Verbund selbst noch keine konkreten Projekte mit den derart aufbereiteten Daten vorhat.¹⁹

Als **library linked data** werden in diesem Beitrag lediglich von Bibliotheken bereitgestellte Daten verstanden, während die W3C Library Linked Data Incubator Group den Begriff weiter fasst und Daten aus Museen und Archiven mit einbezieht.²⁰ An dieser Stelle ist eine enge Begriffsverwendung jedoch sinnvoll, um Möglichkeiten und Grenzen von *linked data* in Bibliotheken aufzuzeigen.

3. Bibliotheken als Anbieter von linked open data

Für Bibliotheken kann es aus mehreren Gründen sinnvoll sein, ihre Daten als *linked open data* bereitzustellen. Zum einen ergeben sich aus LOD neue Nutzungsmöglichkeiten für Bibliotheksdaten, die in Kapitel 6 ausführlich beschrieben werden. Vor dem Hintergrund, dass einige traditionelle bibliothekarische Arbeitsfelder wie die Katalogisierung im Internetzeitalter an Bedeutung verlieren²¹ und Bibliotheken dadurch unter einen gewissen Rechtfertigungsdruck geraten, liegt eine bessere Nutzung der eigenen Daten und die damit verbundene stärkere Wahrnehmung ihrer Leistungen in ihrem Interesse. Mit *linked open data* kann auch die Vernetzung zwischen den Angeboten verschiedener Kulturinstitutionen verbessert werden, was sich etwa das groß angelegte Europeana-Projekt zum Ziel gemacht hat.

Zum anderen sind Bibliotheksdaten besonders gut für die Integration in die LOD-Cloud geeignet, weil es sich bei ihnen, anders als bei vielen im Internet verfügbaren Daten, um qualitätsgeprüfte Informationen handelt. Schließlich können nicht nur die Nutzer, sondern auch die Bibliotheksmitarbeiterinnen und –mitarbeiter in ihrer täglichen Arbeit von den Möglichkeiten, die sich durch die Datenverlinkung ergeben, profitieren, indem sie auf die Arbeit Anderer zurückgreifen.²²

4. Technische Grundlagen

Im Folgenden sollen einige technische Grundlagen von linked open data beschrieben werden, die notwendig sind, um einordnen zu können, welche Anwendungen mit diesen Daten möglich sind. Dabei wird besonders auf die Unterschiede der Datenstruktur in RDF zu MAB/MARC eingegangen. Beim für die Bereitstellung als linked data notwendigen „Mapping“ bibliothekarischer Daten ins RDF-Format (Resource Description Framework) wird nämlich die Struktur der Katalogdatensätze aufgebrochen.

4.1 Datenstruktur und Standardisierung

Linked open data ist eine Technik, die grundsätzlich inhaltsunabhängig ist, bei der also explizit nicht einzelne Anwendungsmöglichkeiten im Vordergrund stehen.²³ Klassische bibliothekarische Katalogisate haben dagegen den Anspruch, die Funktionen eines Katalogs²⁴ besonders gut zu erfüllen. Dies spiegelt sich auch in der Struktur der

Datensätze wider: Im MARC- oder MAB-Format bilden die einzelnen Katalogisate die „Grundeinheiten“. Innerhalb dieser Datensätze sind dem jeweils beschriebenen Medium klar zuzuordnende Informationen in einer festen Reihenfolge/ in dafür definierten Feldern gespeichert. Die Bedeutung der jeweiligen Informationen (MAB-Feld 100 a: Autor) erschließt sich nur, wenn man die Bedeutung der Feldnummern kennt. Was die Feldnummern inhaltlich bedeuten, ist allerdings nicht Bestandteil der *Daten*. In klassischen bibliothekarischen Datenformaten werden also Informationen über ein bestimmtes Medium gemeinsam mit einem „Schlüssel“ zur Interpretation dieser Informationen – nämlich den Feldnummern – in einem Datensatz gespeichert; der Schlüssel ist allerdings nicht selbsterklärend.

Im RDF-Format wird diese Struktur in zweierlei Hinsicht aufgebrochen: Zum einen werden die Daten nicht mehr in nach Katalogaufnahmen voneinander klar getrennten Datensätzen gespeichert²⁵, sondern in Form von Tripeln aus Subjekt (z.B. *Hinnerke Meier*), Prädikat (z.B. *ist Autor von*) und Objekt (z.B. *Das Buch der 1000 Teesorten*). Aus einem Katalogdatensatz werden also viele RDF-Tripel. Zum anderen bekommen dabei die *Verknüpfungen*, die in MARC/MAB mit Feldnummern ausgedrückt werden, im RDF-Format eine von Maschinen interpretierbare *Bedeutung* zugewiesen – im oben genannten Beispiel „*ist Autor von*“.²⁶ Dieses Tripel-Datenmodell erlaubt gegenüber den Katalogdatensätzen im MARC-/MAB-Format einen deutlich flexibleren Umgang mit bibliografischen Daten, ist offen für nachträgliche Modellerweiterungen und für Verknüpfungen mit anderen Daten.²⁷ RDF ist letztendlich nur „ein abstraktes Modell, in dem Aussagen in Form von Tripeln gemacht werden können“.²⁸ Innerhalb des RDF-Formats können verschiedene Standards verwendet werden, um Daten zu beschreiben²⁹, was einem Bereich mit sehr vielen „Kleinstandards“ wie dem Bibliothekswesen entgegenkommt.³⁰

Dadurch, dass offene Katalogdaten und andere freigegebene bibliothekarische Daten mit RDF in einem allgemeinen Standardformat vorliegen, sind sie deutlich leichter durch Dritte nutzbar als in „traditionellen“ bibliotheksspezifischen Formaten gespeicherte Daten.³¹ Letztere können in der Regel nur mit spezieller Bibliothekssoftware genutzt und über die Z39.50-Schnittstelle, die ebenfalls bibliotheksspezifisch ist und z.B. in Webbrowsern nicht implementiert ist³², in andere Programme exportiert werden.

Aufgrund der oligopolistischen Struktur des Marktes für Bibliothekssoftware und der vielen bibliotheksspezifischen „Kleinstandards“ und Besonderheiten sind die Anreize für Softwareunternehmen, in den „traditionellen Bibliothekssoftware-Markt“ einzusteigen, gering. Die LOD zugrundeliegenden Standards sind dagegen, auch wenn Bibliotheken bei ihrer Entwicklung eine große Rolle gespielt haben, nicht bibliotheksspezifisch. Dadurch ist zum einen die Zahl der Personen, die diese technischen Standards kennen und damit umzugehen wissen – was den Eintritt neuer Anbieter in den Markt erleichtert –, deutlich größer als bei „traditioneller“ Bibliothekssoftware; zum anderen ist es auch deutlich wahrscheinlicher, dass ursprünglich für andere Zwecke gedachte Programme für Bibliotheken nachnutzbar sind.³³ Eine konsequente Umsetzung bibliothekarischer Daten als *linked open data* erhöht also die Wahrscheinlichkeit, dass die Konkurrenz in diesem Markt zunimmt und dadurch der Wettbewerb intensiviert wird, was sowohl dem Preis als auch der Qualität der Bibliothekssoftware zugute kommen könnte.³⁴

4.2 Identifier

Um die einzelnen Bestandteile der Tripel – Subjekt, Prädikat und Objekt – eindeutig zuzuordnen, werden in RDF *http-Uniform Resource Identifiers* (URIs) verwendet. Der IFLA zufolge ist ein Identifier

[e]ine Nummer, ein Code, ein Wort, eine Phrase, ein Logo, eine Vorrichtung etc. die mit einer Entität in Verbindung steht und dazu dient, diese Entität innerhalb des Bereichs, für den der Identifikator definiert ist, von anderen Entitäten zu unterscheiden.³⁵

Bekannte Identifier außerhalb der LOD-Welt sind etwa ISBNs.³⁶ Auch http-URIs werden verwendet, um sicherzustellen, dass verschiedene Dinge auch unterschiedlich bezeichnet werden. Man kann damit allerdings nicht sicherstellen, dass dasselbe Ding auch immer gleich bezeichnet wird, da es keine zentrale Vergabestelle für URIs gibt.³⁷ Damit freigegebene Daten gut miteinander verknüpft werden können, sollten Subjekt, Prädikat und Objekt jeweils aus einem URI bestehen. In RDF ist es allerdings auch möglich, dass das Objekt (und nur dieses!) lediglich aus einer Zeichenkette, einem sogenannten **Literal** besteht, das dann allerdings nicht weiter verlink- und interpretierbar ist.

4.3 Ontologien und Datenabfragesprache

RDFS, SKOS, OWL, BIBO und das Dublin Core Metadata Set (DC) sind Schemata für Metadaten in RDF, sogenannte Ontologien. In Ontologien werden Klassen für Objekte (z.B. „Autor“, „Titel“ oder „Sprache“), die zwischen diesen Klassen möglichen Beziehungen (jemand kann „Autor“ von „Titel“ sein, aber nicht „Autor“ von „Sprache“) und zulässige Schlussfolgerungen aus diesen Beziehungen definiert.³⁸

Der bereits Mitte der 1990er Jahre für die Beschreibung von Internetressourcen entwickelten **Dublin Core**-Standard³⁹ beinhaltet viele für bibliografische Beschreibungen notwendige Elemente wie dc:title, dc:publisher, dc:language, dc:subject. SKOS (**Simple Knowledge Organization System**) ist eine vom W3C verabschiedete Ontologie, die auf bibliothekarischen Anforderungen aufbaut⁴⁰ und mit der man kontrollierte Vokabulare wie Thesauri und bibliothekarische Klassifikationen beschreiben kann⁴¹, in denen es z.B. Begriffshierarchien und verwandte Begriffe gibt. Weitere häufig verwendete Ontologien sind OWL (**Web Ontology Language**), mit der man beispielsweise ausdrücken kann, dass es sich bei verschiedenen im RDF-Format vorliegenden Daten um denselben Inhalt handelt⁴², RDFS (**RDF Schema**), BIBO (**Bibliographic Ontology**) und FOAF (**Friend of a Friend**).

Es gibt verschiedene Formate, in denen diese Ontologien verwendet werden können und die alle RDF-kompatibel sind. RDF/XML, Turtle (Terse RDF Triple Language), N3 und N-Triples sind bekannte Datenaustauschformate.⁴³ Abfragen sind innerhalb von RDF mit der Abfragesprache SPARQL (*SPARQL protocol and RDF query language*) möglich. Aufgrund der Datenstruktur sind damit deutlich komplexere Abfragen möglich als etwa mit SQL.⁴⁴

5. Anforderungen an Bibliotheksdaten

5.1 Open data

Damit Katalogdaten und andere bibliothekarische Daten als *linked open data* bereitgestellt werden können, müssen sie bestimmte Anforderungen erfüllen: Zum einen muss gewährleistet sein, dass die Daten offen im Sinne von *open data* sind, zum anderen müssen sie bestehende *linked data*-Standards verwenden.

Die Daten müssen als *open data* bereitgestellt werden, damit sie für Andere nachnutzbar sind. Dies ist nur dann gut möglich, wenn die verwendete Lizenz erstens die kommerzielle Nutzung der Daten nicht verhindert. Das Verbot einer kommerziellen Nutzung kann für Bibliotheken sinnvolle Software-Weiterentwicklungen verhindern, außerdem ist die Abgrenzung von „kommerziell“ und „nicht kommerziell“ oft schwierig. Zweitens sollte die Lizenz aus Praktikabilitätsgründen keine Nennung des Urhebers erfordern.⁴⁵ Daten können durch den jeweiligen Urheber problemlos freigegeben werden. Wurden sie allerdings von Dritten übernommen, darf eine Freigabe nur mit dessen Einverständnis erfolgen.⁴⁶ Aus diesem Grund stehen einige linked data-Projekte von Bibliotheken bislang nur unter einer Lizenz für nicht kommerzielle Nutzung.⁴⁷

5.2 Linked data

Um Katalogdaten als *linked data* bereitzustellen, müssen sie ins RDF-Format umgewandelt werden („RDF-Mapping“). Dazu müssen http-URIs vergeben werden, mit denen die Daten eindeutig identifiziert werden können. Für die Prädikate sollten URIs aus einer bereits bestehenden Ontologie verwendet werden.⁴⁸ Zwar ermöglicht RDF es auch, neue Ontologien einzubinden. Diese zu erstellen ist allerdings mit deutlich höherem Aufwand verbunden als die Suche nach geeigneten bestehenden Schemata, die für Bibliotheksdaten in der Regel ausreichen sollten.⁴⁹

5.3 RDA und LOD

Wie oben beschrieben können Daten aus „klassischen“ Katalogaufnahmen nach RAK-WB oder auch nach AACR zwar in RDF gemapped werden, aber die Datenstruktur, die sich aus diesen Regelwerken ergibt, ist nicht auf das *semantic web* ausgelegt. Bei der Entwicklung der neuen Katalogisierungsregeln *Resource Description and Access* (RDA) wurde daher eine gemeinsame Arbeitsgruppe mit der Dublin Core Metadata Initiative, IEEE Learning Object Metadata und dem Simple Knowledge Organization System des W3C eingerichtet, die sicherstellen sollte, dass RDA konform zu gängigen *linked open data*-Standards entwickelt wird.⁵⁰ Auch das Rahmenkonzept für bibliografische Daten, das die Library of Congress derzeit mit mehreren internationalen Partnern entwickelt, soll konsequent an den Anforderungen des

semantic web ausgerichtet sein, da diese Konformität mit MARC nicht zufriedenstellend möglich sei:

The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model. The protocols and ideas behind Linked Data are natural exchange mechanisms for the Web that have found substantial resonance even beyond the cultural heritage sector. Likewise, it is expected that the use of RDF and other W3C (World Wide Web Consortium) developments will enable the integration of library data and other cultural heritage data on the Web for more expansive user access to information.⁵¹

Auch wenn eine baldige Einführung von RDA zumindest in den USA nach dem durchwachsenen Ergebnis der Testphase und verschiedenster Kritik⁵² nicht allzu wahrscheinlich ist, ist unzweifelhaft, dass die Erfordernisse für *linked open data* bei den aktuellen Entwicklungen neuer bibliothekarischer Regelwerke und Datenformate eine große Rolle spielen.⁵³ In Anbetracht der üblichen Lebensdauer solcher Regelwerke und Formate scheint es deswegen nicht sinnvoll, *linked open data* als einen weiteren vorübergehenden Hype im Bibliothekswesen abzutun.

6. Anwendungsbereiche

Wenn bibliothekarische Daten als LOD zur Verfügung gestellt werden, sind jederzeit und durch jeden – also nicht nur durch die Bibliotheken, die die Daten im RDF-Format bereitgestellt haben – Verknüpfungen mit anderen RDF-Daten möglich. Um welche *Inhalte* es sich bei diesen handeln kann, ist durch das Datenformat nicht festgelegt. Wird *library linked data* mit anderen Daten verknüpft, muss es sich bei den Verknüpfungen also nicht unbedingt um solche handeln, die Bibliotheken zur besseren Beschreibung des konkreten Werks des jeweiligen „Ursprungsdatensatzes“ für sinnvoll halten⁵⁴ (wie etwa die Anreicherung mit Inhaltsverzeichnissen in „normalen“ OPACs). Auch müssen nicht notwendigerweise Zusammenhänge hergestellt werden, die die Entwickler der miteinander verknüpften Daten ursprünglich zum Ziel hatten.⁵⁵

Es lässt sich also feststellen, dass es keinen fest vorgegebenen Anwendungsbereich für *library linked data* gibt. Dennoch gibt es aufgrund der Inhalte bibliothekari-

scher Datensätze und ihrer Form einige besonders naheliegende Anwendungsmöglichkeiten, die im Folgenden vorgestellt werden sollen. Dabei wird jeweils zunächst allgemein erläutert, welche Anwendungsszenarien es grundsätzlich gibt und wo gegebenenfalls mögliche Probleme liegen. In einem zweiten Schritt werden jeweils bestehende Projekte vorgestellt, wobei der Schwerpunkt auf dem deutschen Bibliothekswesen liegt.

6.1 Normdaten und andere kontrollierte Vokabulare

Kontrollierte Vokabulare, also eindeutig definierte und inhaltlich und formal abgegrenzte Daten, sind für ein RDF-Mapping aus mehreren Gründen besonders gut geeignet: Die Anzahl der Datensätze ist verhältnismäßig klein, es dürfte innerhalb des jeweiligen Vokabulars keine Dubletten geben und die Beziehungen zwischen Subjekt und Objekt sind klar definiert. Bei hierarchisch aufgebauten kontrollierten Vokabularen sind auch die Beziehungen zwischen den verschiedenen Datensätzen eindeutig definiert. Zu dieser Art von Daten gehören etwa die Normdateien SWD, GKD und PND, Fachthesauri, ISO-Sprachencodes, Landesbibliografien und andere bibliothekarische Klassifikationen. Mehrere Gründe sprechen für die Umsetzung kontrollierter Vokabulare aus Bibliotheken als linked open data:

Üblicherweise werden diese nur durch Fachkräfte verwendet, so dass eine gewisse inhaltliche Qualität der Informationen sichergestellt ist. Zudem werden sie in den deutschen Bibliotheksverbänden einheitlich verwendet, während es bei Katalogisaten neben den Katalogisierungsregeln auch noch Konventionen der einzelnen Verbände gibt. Außerdem verbringen Bibliothekarinnen und Bibliothekare seit Jahrhunderten sehr viel Zeit damit, mit Hilfe kontrollierter Vokabulare Medien inhaltlich und formal zu erschließen. Diese Arbeit für die Allgemeinheit besser sichtbar und für verschiedene Zwecke (nach)nutzbar zu machen, kann in einer Zeit, in der Schlagworte für die Recherche nicht mehr zwingend notwendig sind, erheblich zur Rechtfertigung von Bibliotheken beitragen.

Der „Verknüpfungsgrad“ von Normdaten ist schon außerhalb der LOD-Welt verhältnismäßig hoch, und zwar sowohl innerhalb der Normdaten selbst (etwa durch Verweise zwischen Schlagwörtern⁵⁶) als auch zwischen Normdaten und bibliografischen Angaben (durch die Verknüpfungen zu den Normdateien in

Katalogisaten). Ähnliches gilt für Klassifikationen und Fachthesauri. All diese Daten lassen sich daher auch problemlos mit Daten Dritter verknüpfen, was zu einer deutlich erhöhten Nutzung der beschriebenen Ressourcen führen kann. So verlinken etwa viele LOD-Angebote von Bibliotheken zur DBpedia, in der Daten aus Wikipedia im RDF-Format vorliegen.

Die teilweise mit hohem Arbeitsaufwand verbundenen Verlinkungen müssen nicht unbedingt von Bibliothekaren ausgeführt werden. Durch den hohen Mehrwert, der sich durch die Verlinkung der Daten ergibt, finden sich meist Freiwillige dafür, etwa unter den Mitarbeiterinnen und Mitarbeitern der Wikipedia. Beispiele für die gelungene Einbindung von PND-Nummern sind etwa deren Verknüpfung mit Personenartikeln in der Wikipedia⁵⁷, mit Fotos aus dem Bundesarchiv⁵⁸ oder mit Namen in dpa-Meldungen.⁵⁹ Zwar ist es selbstverständlich möglich, dass bei der Zuordnung Fehler entstehen, aber andererseits können auch Fehler in Bibliotheksdaten aufgedeckt und Informationslücken geschlossen werden. Ein Beispiel hierfür ist die Zusammenarbeit der Wikipedia-Autorinnen und –autoren mit der PND-Redaktion der DNB.⁶⁰

Weitere Anwendungsbereiche für LOD im Bereich der kontrollierten Vokabulare sind Crosskonkordanzen zwischen unterschiedlichen Klassifikationen und Thesauri⁶¹ und zwischen verschiedenen nationalen Normdatensystemen, wie etwa durch VIAF⁶² oder zwischen der SWD und dem französischen RAMEAU. Solche als LOD umgesetzten Konkordanzen sowie Fachthesauri erweitern die Suchmöglichkeiten in bibliografischen Daten erheblich.

Innerhalb des Bibliothekswesens liegen die größten Potenziale kontrollierter Vokabulare als LOD vermutlich im Bereich der Erwerbung und der Sacherschließung. „Eigene“ Daten und die anderer Bibliotheken können etwa verwendet werden, um Bestandsschwerpunkte und Aktualitätslücken zu identifizieren oder Systemstellen zu vergeben (entweder auf Basis einer Schlagwortauswertung oder anhand einer Konkordanz, falls der Titel schon in einer anderen Klassifikation mit einer Systemstelle versehen wurde). Durch sprachübergreifende Crosskonkordanzen⁶³ kann die auswertbare Datenbasis dafür deutlich vergrößert werden.

Bisherige Projekte, in denen versucht wurde, anhand semantischer Auswertungen Medien automatisch zu verschlagworten, sind je nach Fachgebiet unterschied-

lich erfolgreich: Während diese Methode bei natur- und ingenieurwissenschaftlichen Titeln zu verhältnismäßig guten Ergebnissen führt, scheint sie bei geistes- und sozialwissenschaftlichen weniger erfolgversprechend zu sein. Eine Anwendung in umgekehrter Richtung, nämlich das Anzeigen ähnlicher Werke, ist einfacher umzusetzen⁶⁴ und könnte in Bibliotheken sowohl in der Erwerbung als auch als Service für die Benutzer verwendet werden.

Neben den oben bereits angesprochenen Normdaten der DNB haben verschiedene andere Nationalbibliotheken, darunter die Library of Congress, ihre Normdaten unter einer freien Lizenz zur Verfügung gestellt.⁶⁵ Verschiedene Klassifikationen wie die Dewey-Dezimalklassifikation, die RVK und die Basisklassifikation stehen ebenfalls als *linked data* zur Verfügung, allerdings aus urheberrechtlichen Gründen teilweise unter Lizenzen, die nur ihre nichtkommerzielle Nutzung erlauben.⁶⁶ Die Zentralbibliothek für Wirtschaftswissenschaften hat den Standardthesaurus Wirtschaft, die GESIS den Thesaurus Sozialwissenschaften in RDF umgesetzt. Beispiele für weitere als *linked data* veröffentlichte Thesauri sind AGROVOC des Welternährungsprogramms⁶⁷ und der Thesaurus for Graphic Materials der Library of Congress.⁶⁸

6.2 Komplett Katalogdatensätze

Während kontrollierte Vokabularien einen guten Ansatzpunkt bieten, um bibliothekarische Daten mit anderen Bereichen des Internet zu verbinden, ergeben sich durch das Mapping kompletter Katalogdatensätze in RDF weitere Anwendungsmöglichkeiten, die vor allem für Bibliotheken selbst interessant sind. Allerdings ist die Umsetzung kompletter Katalogdatensätze wegen der großen Anzahl der umzusetzenden Felder, der großen Datenmenge und der Vielzahl von Dubletten komplizierter als bei Normdaten. Letzteres Problem kann allerdings auch gerade durch die Bereitstellung von Katalogdaten als LOD verringert werden.⁶⁹ Zudem könnte die Zahl der Informationen, die jeweils lokal bereitgehalten werden müssen, verringert werden.⁷⁰

Statistische Auswertungen zu verschiedenen Zwecken sind mit *library linked data* ebenfalls einfacher möglich als mit „klassischen“ bibliothekarischen Datensätzen. Je mehr Bibliotheken ihre Katalogdaten als LOD bereitstellen, desto größer ist dabei die analysierbare Datenmenge. So könnte man etwa untersuchen, welche Nota-

tionen einer Klassifikation besonders häufig oder selten genutzt werden oder anhand von Titelwörtern oder von Autoren vergebenen freien Schlagwörtern neue Forschungsbereiche identifizieren und zeitnah in die Klassifikation einarbeiten.⁷¹ Titeldatenanalysen könnten auch in der Benutzung verwendet werden, z.B. um Suchanfragen automatisch zu ergänzen, ähnliche Titel vorzuschlagen oder automatische Benachrichtigungen für neue Titel mit einem bestimmten inhaltlichen Fokus einzurichten. Wegen der komplexen Abfragemöglichkeiten mit SPARQL wären dabei auch sehr spezielle Suchen möglich.⁷² Auch sind bibliometrische Auswertungen des Publikationsverhaltens mit LOD denkbar.⁷³

In RDF umgesetzte Katalogdaten können außerdem zu einer besseren Vernetzung und Präsentation der Erschließungsleistungen von Bibliotheken und anderen Kulturinstitutionen beitragen.⁷⁴ Die damit verbundenen Anstrengungen lassen sich am Europeana-Projekt eindrucksvoll beobachten, in dem sogar ein eigenes Datenmodell entwickelt wurde. In Europeana als RDF bereitgestellte Daten, nämlich solche zu nationalen Urheberrechtsregelungen nutzt auch der „Public Domain Calculator“ der Open Knowledge Foundation, in dem diese Daten mit Informationen aus Katalogen verknüpft werden, um festzustellen, ob ein Werk nach dem Recht des jeweiligen Landes bereits gemeinfrei ist.⁷⁵

In den meisten Fällen haben sich Bibliotheken bislang jedoch darauf beschränkt, Katalogdaten unter einer freien Lizenz zur Verfügung zu stellen, ohne unmittelbar konkrete Projekte damit durchführen zu wollen. Dafür wurden vielfach eigene „Namensräume“ für http-URIs eingeführt und die Normdaten verknüpft. Das hbz hat in lobid-organisations zudem Adressdaten von Bibliotheken in RDF abgebildet.⁷⁶ Neben dem schwedischen Verbundkatalog⁷⁷ stehen auch die Katalogdaten der DNB und der BNB, des HBZ, der UB der RWTH Aachen sowie der UB Tübingen, der UB Konstanz, der SUB Köln, der ZBMED und des CERN als open data bereit. Im BVB wird gerade an der Umsetzung des B3KAT in *open data* gearbeitet.

Eine weitere auf *linked* (aber nicht open) *data* basierende Anwendung in deutschen Bibliotheken ist die multilinguale Suche „SLUBsemantics“ der Sächsischen Landes- und Universitätsbibliothek Dresden, die Informationen aus Wikipedia-Artikeln nutzt.⁷⁸ Auf die vielfältigen Nutzungsmöglichkeiten von LOD im Publika-

onswesen kann auf dieser Stelle aus Platzgründen nur allgemein hingewiesen werden.⁷⁹

6.3 Nutzungsdaten

Für viele auf Statistiken beruhende Anwendungsmöglichkeiten wäre es sinnvoll, auch Ausleihdaten und andere bestandsspezifische Daten in RDF aufzubereiten. Aus Datenschutzgründen müssten benutzerbezogene Daten selbstverständlich anonymisiert bereitgestellt werden.⁸⁰ Denkbare Anwendungsbereiche finden sich etwa in der Erwerbung (z.B. nach Schlagworten oder Notationen differenzierte Ausleihstatistiken), aber auch in der Benutzung (z.B. Vorschläge für ähnliche Medien⁸¹).

Dadurch, dass RDF ein allgemeiner und kein bibliotheksspezifischer Standard ist, ist die Anzeige für *library linked data* auch nicht an Bibliothekssoftware und deren verbundspezifische Anpassungen gebunden. Werden Katalogdaten im RDF-Format bereitgestellt, können dafür unterschiedliche, zielgruppenspezifische Oberflächen programmiert werden, die jeweils unterschiedliche *linked data* aus anderen Quellen mit einbeziehen könnten. Diese könnte unabhängig vom verwendeten Bibliothekssystem und verbundübergreifend genutzt werden, d.h. Bibliotheken könnten Katalogoberfläche(n) verwenden, die in erster Linie an ihren Bedürfnissen und nicht an den Spezifikationen des jeweiligen Bibliothekssystems ausgerichtet sind. Neben der einfachen Umsetzung verschiedener Anzeigesprachen wären z.B. auch optimierte Kataloganzeigen für Personen mit verschiedenen Sehbehinderungen denkbar.

Bei der Anzeige könnte nicht nur auf unterschiedliche Anforderungen der Bibliothekskunden, sondern auch auf solche der Bibliothekarinnen und Bibliothekare eingegangen werden.⁸² So sieht eine besonders gute Kataloganzeige für Personen, die viel Zeit mit Sacherschließung verbringen, wahrscheinlich anders aus als eine für Bibliothekarinnen und Bibliothekare, die in erster Linie mit der Katalogisierung von Altbestand befasst sind. Die Anforderungen innerhalb dieser Gruppen ähneln sich vermutlich deutlich stärker als die aller MitarbeiterInnen einer Bibliothek oder sogar innerhalb eines Verbundes.⁸³ Je nach Arbeitsschwerpunkt könnten auch unterschiedliche Zusatzinformationen aus der LOD-Cloud in die Anzeige integriert werden, und

durch die verbundübergreifende Nutzbarkeit wäre auch die Zahl der potenziellen Anwender höher.

Weiterhin können als LOD modellierte Ausleihdaten für Vergleiche zwischen Teilbibliotheken verwendet werden und könnten den Erhebungsaufwand für die deutsche Bibliotheksstatistik verringern.⁸⁴ Schließlich können offene Bibliotheksdaten in andere Internetseiten eingebunden werden. Bei einer konsequenten Umsetzung bestandsspezifischer Daten in LOD müssten (potenzielle) Nutzer nicht mehr auf der Seite der jeweiligen Bibliothek nachschauen, ob ein Buch dort gerade ausleihbar ist, sondern könnten zu dieser Information auch zum Beispiel aus einem Wikipedia-Artikel oder einem Rezensionstext gelangen.

6.4 Probleme

Mit der Bereitstellung von Katalog- und anderen Bibliotheksdaten als *linked open data* sind auch einige Probleme verbunden. Neben den oben angesprochenen urheberrechtlichen Fragen, die die Bereitstellung von Katalogdaten als LOD oder die Verknüpfung mit Zusatzinformationen wie Inhaltsverzeichnissen verhindern können, sind vor allem die Qualität der verwendeten Daten und die URI-Vergabe kritisch.

Das Problem der Datenqualität ist in erster Linie ein Dublettenproblem: Werden Katalogdaten einer Bibliothek mit denen anderer Bibliotheken verknüpft, werden Dubletten wegen ihrer unterschiedlichen Katalogaufnahmen unterschiedlich in RDF umgesetzt und somit nicht optimal verknüpft. Dies ist zwar kein grundsätzliches Problem, da diese Gleichheit über `owl:sameAs` ausgedrückt werden kann, aber besser, als die Dubletten nachträglich miteinander zu verknüpfen, wäre eine Katalogbereinigung vor dem RDF-Mapping. Bislang ist die Anzahl der Dubletten zwischen den derzeitigen Bibliotheks- bzw. Verbundkatalogen durch unterschiedliche Katalogisierungsregeln bzw. die unterschiedliche Anwendung derselben Regeln sehr hoch. Auch wenn eine Dublettenkontrolle mit LOD einfacher umzusetzen ist als ohne⁸⁵, ist *library linked data* kein „Allheilmittel“ für die Datenqualität⁸⁶, und einige Konsistenzchecks bei nachträglich miteinander verknüpften Daten sind nicht ohne menschliches Eingreifen machbar.⁸⁷ Zum anderen ist das Problem der Datenqualität allerdings auch ein Regelwerksproblem: Dadurch, dass bestimmte Informationen (wie etwa zweite Herausgeber) nach RAK-WB nicht in die Katalogisate aufgenommen werden,

können sie auch nicht in RDF umgesetzt werden, was z.B. die Ergebnisse bibliometrischer Analysen verfälschen kann. Frühere Befürchtungen, mit LOD sei es nicht möglich, hierarchische Beziehungen abzubilden⁸⁸, haben sich als unbegründet herausgestellt.

Die Probleme bei der URI-Vergabe hängen zum Teil mit dem Problem der Datenqualität zusammen: Werden mehrere URIs für dieselbe Information vergeben, bläht dies den Datensatz unnötig auf, da nicht nur, wie gewünscht, Verbindungen zwischen unterschiedlichen Daten hergestellt werden, sondern auch Informationen „mit sich selbst“, nämlich mit ihren Dubletten, verknüpft werden müssen. Andererseits ist die Wahl der URIs beim RDF-Mapping von entscheidender Bedeutung. URIs sollten dauerhaft gültig sein – was irgendjemand gewährleisten muss –, und sie sollten sinnvolle, eindeutige Bezeichnungen haben – was sichergestellt werden muss, damit es nicht zu Verwechslungen kommt. Nicht nur bei den Beschreibungen der Entitäten mit bestehenden Metadatenschemata, sondern auch bei der URI-Vergabe für die Entitäten selbst ist es wichtig, zunächst bestehende Angebote, bei denen auch die Dauerhaftigkeit der URIs sichergestellt ist, zu nutzen, bevor eigene Standards entwickelt werden.⁸⁹ Das Projekt culturegraph.org der DNB und des hzb setzt an diesem Punkt sowie bei der URI-Vergabe und beim Abgleich von Metadaten an.⁹⁰

7. Fazit

Die Bereitstellung bibliografischer Daten als *linked open data* kann die Sichtbarkeit bibliothekarischer Angebote im Internet deutlich erhöhen. Qualitätsgeprüfte Metadaten, die ein bibliothekarisches „Kernprodukt“ sind, sind sehr gut für die Verlinkung im *semantic web* geeignet; zumindest kontrollierte Vokabulare können auch verhältnismäßig einfach in RDF umgesetzt werden.

Durch die Kombination mit Daten anderer Anbieter und durch formatbedingt bessere Abfragemöglichkeiten ergeben sich für offene Katalogdaten sehr viele Nutzungsmöglichkeiten sowohl innerhalb als auch außerhalb des Bibliothekswesens, die einen großen Mehrwert darstellen. Zudem scheint die Beteiligung an allgemeinen Standards zukunftsweisender zu sein als die Weiterentwicklung bibliotheksspezifischer Sonderformate. Trotz einiger rechtlicher und technischer Schwierigkeiten bei der Umsetzung von Katalogdaten in RDF, und obwohl das Internet sich bislang

noch nicht zum *semantic web* gewandelt hat, ist die Bereitstellung von *library linked data* für Bibliotheken ein sinnvoller Schritt, der weit über „neue Präsentationsmöglichkeiten“ für bestehende Daten hinausgeht.

Endnoten

¹ Dilger & Thompson (2008, S. 40).

² Vgl. Hitzler et al. (2008, S. 9).

³ So spielt die Schlagwortsuche, für die durch die Sacherschließung erhebliche Ressourcen verwendet werden, für „normale“ Benutzer kaum noch eine Rolle. Auch die früher übliche strikte Trennung zwischen bibliografischen Metadaten und Inhalten löst sich durch die zunehmende Verfügbarkeit von Volltexten, Inhaltsverzeichnissen und Kommentarmöglichkeiten im Internet auf: Dilger & Thompson (2008, S. 44–45).

⁴ Wissenschaftsrat (2011, S. 10-11 und 33).

⁵ Eine (mittlerweile nicht mehr sehr übersichtliche) Visualisierung der „Linked-Open-Data-Welt“ ist unter <http://richard.cyganiak.de/2007/10/lod/imagemap.html> zu finden.

⁶ Dilger & Thompson sprechen vom „Verlust des Katalogisierungsmonopols“ und von „ubiquitous cataloging“, das auch eine bessere Integration von Katalogdaten in soziale, kommunikative und Bildungsnetzwerke erforderlich mache. Die zukünftige Rolle der Bibliotheken liege in der Bereitstellung innovativer Dienste und in der Verknüpfung, nicht in der Bereitstellung von Daten: Dilger & Thompson (2008, S. 42 und 45).

⁷ Diese ist wiederum eine Reaktion auf die Defizite des „Web 1.0“: Heterogenität der Informationen in Bezug auf Datenformat und Datstellungsweise, starke Formatabhängigkeit der Daten, fehlende Unterscheidungsmöglichkeiten zwischen gleichlautenden Begriffen und die Notwendigkeit, verstreute Informationen zusammenzusuchen: Hitzler et al. (2008, S. 10–11).

⁸ Während etwa das Ergebnis einer normalen Internetsuche nach „Bank“ sowohl Sitzmöbel als auch Geldinstitute beinhaltet, soll im *semantic web* durch die *Interpretation* der vorliegenden Informationen zwischen diesen beiden Begriffen unterschieden werden können.

⁹ Berners-Lee, Hendler & Lassila 2001, S. 29-37; Hitzler et al. (2008, S. 11); Pohl (2011, S. 2-3); Brown & Boulderstone (2008, S. 271). Das W3C (2011) definiert *semantic web* folgendermaßen: „The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF).“

¹⁰ Hitzler et al. (2008, S. 11–13). Dieses neu generierte Wissen wird oft als „implizites Wissen“ bezeichnet: Hitzler et al. (2008, S. 13). Kittenberger (2010, S. 89-90) merkt dagegen an, dass trotz der neuen Form der Wissensorganisation letztendlich auch im *semantic web* nur explizites Wissen abgebildet werden könne. Zudem bildeten die Kategorien, die den Ontologien zugrundeliegen, stets auch bestehende soziale Strukturen und Wertvorstellungen ab. Dies ist aber bei „traditionellen“ Katalogen noch stärker der Fall, denn dort sind nicht nur die Inhalte, sondern es ist auch das Ausgabeformat festgelegt: Dilger & Thompson (2008, S. 44).

¹¹ Brown & Boulderstone (2008, S. 271 und 274).

¹² Brown & Boulderstone (2008, S. 271-272).

¹³ W3C Incubator Group (2011, Kapitel 1), Pohl (2011, S. 3).

¹⁴ Pohl (2011, S. 3).

¹⁵ Pohl & Ostrowski (2010a, S. 1).

¹⁶ Dies wird z.B. in der Open Knowledge Definition ausgeführt:

<http://opendefinition.org/okd/deutsch/>.

¹⁷ Dazu ausführlich Kreutzer (2011).

¹⁸ Plassmann et al. (2011, S. 206-207).

¹⁹ Bergmann et al. (2011, S. 468).

²⁰ W3C Incubator Group (2011, Kapitel 1). Eine aktuelle Übersicht über die Beiträge der „library linked data group“ in der LOD-Cloud gibt es unter <http://thedatahub.org/group/lld>.

- ²¹ Dunsire & Willer (2011, S. 1-12). Vgl. auch W3C Incubator Group (2011, Kapitel 2). Der Wissenschaftsrat weist in seinen Empfehlungen zur Entwicklung des bibliothekarischen Verbundsystems eindrücklich darauf hin, dass die traditionelle, parallel in mehreren Verbänden stattfindende Katalogisierung ein Auslaufmodell sei und die die „Aufgabe der Zukunft nicht die Erschließung selbst, sondern die Sammlung und Zusammenführung von Katalogdaten“ sei: Wissenschaftsrat (2011, S. 32-33).
- ²² Vgl. auch Wolfe (2008, S. 70-71).
- ²³ Dilger und Thompson betrachten diese Offenheit für verschiedene Nutzungsmöglichkeiten als allgemeine Tendenz im Web, die Innovationen ermögliche: Dilger & Thompson (2008, S. 49).
- ²⁴ Kataloge sollen laut RAK-WB folgende Zwecke erfüllen: Bestandsnachweis einer bestimmten Ausgabe, Nachweis der Werke eines bestimmten Verfassers oder Urhebers, Nachweis der vorhandenen Ausgaben eines bestimmten Werks: Deutsche Nationalbibliothek (2007). Die IFLA nennt in ihren „Richtlinien für die OPAC-Anzeige“ folgende Ziele von Katalogen: Mit den jeweiligen Suchkriterien übereinstimmende Entitäten finden, sie eindeutig identifizieren, nach formalen und inhaltlichen Anforderungen auswählen und Zugang zu der Identität erhalten: Task Force on Guidelines for OPAC Displays (2010, S. 10-11). Dilger und Thompson argumentieren dagegen: „Arguably, connecting users and the works otherwise unknown to them remains the most critical function of library catalogs.“: Dilger & Thompson (2008, S. 42).
- ²⁵ Damit verschiebt sich auch bei der Datenproduktion der Schwerpunkt weg vom der Beschreibung des einzelnen Werks hin zum „big picture“: Wolfe (2008, S. 70-72). Diese Trennung von Format und Inhalt ermöglicht zwar neue Nutzungsmöglichkeiten für bibliothekarische Daten, wird aber nicht einheitlich als positiv angesehen. So ist etwa für Martha M. Yee die Darstellung von Katalogisaten eine der Kernaufgaben von Bibliotheken (“from the catalog user's point of view, cataloging *is* display design“): Yee (2011, S. 129), Hervorhebung im Original.
- ²⁶ Pohl & Ostrowski (2010, S. 263).
- ²⁷ Wolfe (2008, S. 70).
- ²⁸ Pohl & Ostrowski (2010, S. 264).
- ²⁹ Plassmann et al. (2011, S. 207).
- ³⁰ Die W3C Library Linked Data Incubator Group, die die Standardisierung vorantreiben sollte, wurde im Juni 2010 gegründet (Dunsire & Willer (2011, S. 3)) und hat kürzlich ihren Abschlussbericht vorgelegt: W3C Incubator Group (2011).
- ³¹ Bergmann et al. (2011, S. 468).
- ³² Plassmann et al. (2011, S. 206).
- ³³ Bergmann et al. (2011, S. 468). Fehlende Einheitlichkeit der Datenformate ist indes kein Problem, das nur Bibliotheken trifft. So bemängelt etwa Oreste Signore (2008) genau diesen Punkt für Museen und betont, dass gut umgesetzte Thesauri bei der Problembeseitigung eine große Rolle spielen können.
- ³⁴ Vgl. auch Library of Congress (2011).
- ³⁵ International Federation of Library Associations and Institutions (2009, S. 14). Zu den Anforderungen an identifiers siehe Jones (2011, S. 95).
- ³⁶ Jones (2011, S. 97-98) weist darauf hin, dass Identifier von Bibliotheksbenutzern gerne zur Suche verwendet werden, seit dies durch die OPACs möglich wurde.
- ³⁷ Hitzler et al. (2008, S. 37-39 und 47-48).
- ³⁸ Pohl & Ostrowski (2010, S. 264).
- ³⁹ Vgl. dazu Codina, Marcos & Pedraza (2009, S. 16-19).
- ⁴⁰ Dunsire & Willer (2011, S. 8).
- ⁴¹ Plassmann et al. (2011, S. 207).
- ⁴² Codina, Marcos & Pedraza (2009, S. 21).
- ⁴³ Brown & Boulderstone (2008, S. 271); Pohl & Ostrowski (2010, S. 264).
- ⁴⁴ Malmsten (2008, S. 147-148; Hitzler et al. (2008, S. 14); Plassmann et al. (2011, S. 207-208).
- ⁴⁵ Danowski (2010).
- ⁴⁶ Eine ausführliche Übersicht über rechtliche Fragen offener Lizenzen mit Darstellung der verschiedenen Creative Commons-Lizenzen ist: Kreutzer (2011).
- ⁴⁷ Vgl. dazu Eckert (2011).
- ⁴⁸ Zur technischen Umsetzung von Metadaten in RDF siehe die Beiträge in Sicilia & Lytras (2009).
- ⁴⁹ Ausnahme sind bestandsbezogene Daten wie etwa Ausleihdaten und Signaturen, für die es nach Kenntnis der Autorin bislang keine Ontologien gibt.

⁵⁰ RDA beruht auf den *Functional Requirements for Bibliographic Records* und den *Functional Requirements for Authority Data*. Bei der Entwicklung von SKOS wurde explizit auf die Abbildbarkeit von RDA geachtet: Dunsire & Willer (2011, S. 1 und 8).

⁵¹ Library of Congress (2011).

⁵² So befürchtet etwa Martha Yee, dass einige komplizierte Informationsstrukturen mit RDF nicht adäquat abbildbar seien und stellt das RDA-Konzept grundsätzlich in Frage: „Can all bibliographic data be reduced to either a class or a property with a finite set of values? Can everything that catalogers do be reduced to a set of pull-down menus? Cataloging is a discursive art, a kind of descriptive writing. It is not simply the coding of data.“: Yee (2011, S. 129-130).

⁵³ Pohl und Ostrowski erwähnen in ihrem Aufsatz, dass es bereits erste direkte Umsetzungsversuche von RDA in RDF gibt, die allerdings noch nicht besonders anwenderfreundlich seien: Pohl & Ostrowski (2010, S. 266).

⁵⁴ Gerade solche unerwarteten Verknüpfungen von Informationen können aber zur Erschließung neuen, aber trotzdem in den jeweiligen Kontext passenden Wissens beitragen: Dilger & Thompson (2008, S. 42-43).

⁵⁵ Ein Beispiel dafür ist die Verknüpfung von Fotos aus dem Bundesarchiv mit Wikipedia über die PND-Nummern der fotografierten Personen – eine Erweiterung, die den Fotografinnen und Fotografen des Bundesarchivs wahrscheinlich ebensowenig vorschwebte wie den Bibliothekarinnen und Bibliothekaren, die die Personennormdatei entwickelten und nutzen.

⁵⁶ Es gibt verschiedene Verweismöglichkeiten zwischen Schlagwörtern, die auch in Fachthesauri umgesetzt sind.

⁵⁷ <http://de.wikipedia.org/wiki/Hilfe:PND>

⁵⁸ Eine Übersicht über den Ablauf des Projekts findet sich bei Sander (2009), der auch auf die hohen Nutzungszuwächse hinweist.

⁵⁹ Diese Anwendung wird derzeit mit dpa-Pressemeldungen getestet, die noch nicht unter einer freien Lizenz stehen. Das dazu entwickelte Tool findet sich unter <http://toolsserver.org/~apper/dpa/persontagging.php>.

⁶⁰ Die Wikipedia-BenutzerInnen verknüpfen Wikipedia-Artikel über Personen mit PND-Nummern. Dazu gibt es verschiedene Hilfsskripte und Wartungslisten. Über die Seite <http://de.wikipedia.org/wiki/Wikipedia:PND/Fehlermeldung> können seit November 2005 Fehler und Zuordnungen von Werken zu individualisierten PND-Nummern an die DNB gemeldet werden – dies sind zur Zeit monatlich zwischen 150 und 330 Verbesserungsvorschläge. Ein weiteres Beispiel für verbesserte Rückmeldungen durch LOD findet man in Bergmann et al. (2011, S. 468).

⁶¹ Vgl. Lauser et al. (2008).

⁶² Zu den Herausforderungen, die sich bei der Verlinkung von Metadaten aus verschiedenen Sprachen ergeben: Levergood, Farrenkopf & Frasnelli (2008).

⁶³ Levergood, Farrenkopf & Frasnelli (2008).

⁶⁴ Zu „beiden Richtungen“ der automatischen Sacherschließung: Heller (2011).

⁶⁵ Normdaten der Library of Congress inklusive der Library of Congress Subject Headings sowie einige in RDF umgesetzte Sprachcodes finden sich unter <http://id.loc.gov/>, Informationen der DNB zu ihrem linked data-Angebot unter http://files.d-nb.de/pdf/linked_data.pdf. Zur Umsetzung der Library of Congress Subject Headings: Summers et al. (2008).

⁶⁶ Die deutschsprachige Dewey-Dezimalklassifikation wurde von der DNB umgesetzt und steht ebenso wie die von der UB Mannheim in RDF gemappte RVK unter einer Lizenz zur nichtkommerziellen Nutzung. Die vom GBV umgesetzte Basisklassifikation ist frei im Sinne von *open data*. Wegen der hierarchischen Datenstruktur wurde SKOS verwendet.

⁶⁷ Die Projektwebsite ist <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>. Einen Überblick über das Mapping und über Anwendungsmöglichkeiten geben z.B. Lauser et al. (2008).

⁶⁸ <http://id.loc.gov/vocabulary/graphicMaterials.html>

⁶⁹ Dies wird zum Beispiel im Projekt „culturegraph.org“ versucht, das Standardnummern wie die ISBN und die Erstkatalogisierungs-ID zur Dedublizierung nutzt.

⁷⁰ „[I]n this future, the local catalog record will consist simply of one or more widely recognized product identifiers (ISBN, ISMN, etc.) or an OCLC control number (OCN) linked to local management data relating to holdings, location, availability, etc.“: Jones (2011, S. 101).

⁷¹ Aleman-Meza et al. (2009, S. 196).

⁷² Vgl. Malmsten (2008, S. 147-148).

⁷³ Aleman-Meza et al. (2009) zeigen, wie dies grundsätzlich funktionieren kann, nutzen dafür aber keine Bibliotheksdaten, sondern die Datenbank „ACM Digital Library“. Mit Katalogdaten wäre dies – zumindest für Monografien – aber ebenfalls denkbar. Zudem könnten Zusammenhänge zwischen dem Publikationsverhalten und externen Ereignissen (z.B. politischen Umbrüchen) auf einer derartigen großen Datengrundlage besser erforscht werden.

⁷⁴ Beim RDF-Mapping selbst haben Archive und Museen oft ähnliche Schwierigkeiten mit ihren Ausgangsdaten wie Bibliotheken. So spielen etwa in Museen unpräzise Zeitangaben, die schwer in LOD umzusetzen sind, eine große Rolle: Signore (2008, S. 4), und auch dort besteht oft das Problem, dass „collection-level metadata“ in den einzelnen Datensätzen nicht vorhanden sind. Mit diesem Problem befassen sich Renear et al. (2008). Beispiele für (geplante) Anwendungen semantischer Technologien in Museen finden sich in Robering (2008).

⁷⁵ Dieses Tool wurde kürzlich in einem Blogbeitrag vorgestellt:

<http://blog.okfn.org/2011/11/15/work-in-progress-public-domain-calculators/>.

⁷⁶ Pohl 2011, S. 21-22.

⁷⁷ Zur Projektbeschreibung: Malmsten (2008, S. 146).

⁷⁸ Bonte (2011).

⁷⁹ Mögliche Entwicklungsrichtungen werden z.B. von Brown & Boulderstone (2008, S. 273-274) aufgezeigt.

⁸⁰ Die W3C Library Linked Data Incubator Group W3C Incubator Group (2011, Kapitel 1) bezieht sich mit ihren Vorschlägen explizit nicht auf „[d]ata covered by library privacy policies“, sondern nur auf „digital information produced or curated by libraries that describes resources or aids their discovery“;

⁸¹ Auf Ebene der einzelnen Bibliotheken wäre es auch möglich, hier nach Nutzergruppe zu differenzieren, was z.B. für Universitätsbibliotheken sinnvoll sein könnte. Da die Nutzergruppen im Bibliothekssystem sich von Bibliothek zu Bibliothek stark unterscheiden, sind sie aber grundsätzlich für LOD nicht besonders gut geeignet.

⁸² Vgl. Young (2008, S. 83-84).

⁸³ Dies soll nicht in Abrede stellen, dass lokale OPAC-Anpassungen möglich sind. Nähme man LOD als Grundlage für die Kataloganzeige, wären allerdings deutlich größere Anpassungen an die Bedürfnisse der Nutzer möglich, als dies mit solchen Anpassungen der Fall ist.

⁸⁴ Ein Projekt mit diesen Zielen wurde auf der Konferenz „Semantic Web in Bibliotheken 2011“ vorgestellt: Pfeffer (2011).

⁸⁵ Ein Prototyp eines Deduplizierungs-Tools, das Metadaten wie die Erstkatalogisierungsnummer und die ISBN abgleicht, steht unter <http://www.culturegraph.org/demonstrator> zur Verfügung. Mit technischen Aspekten der Dedublizierung beschäftigt sich z.B. Kovačević (2010).

⁸⁶ Pohl & Ostrowski (2010, S. 266).

⁸⁷ Signore (2008, S. 24).

⁸⁸ Yee (2011, S. 129).

⁸⁹ Für „buchbezogene“ Teile der bibliografischen Angabe kann eine eigene URI-Vergabe gegebenenfalls sinnvoll sein. Für Normdaten, Klassifikationen und Thesauri, die schon als LOD umgesetzt wurden, wäre eine neue URI-Vergabe kontraproduktiv, weil die Daten durch die Nutzung bestehender URIs gleich besser verlinkt sind. Vgl. dazu auch Pohl & Ostrowski (2010, S. 266).

⁹⁰ Eine allgemeine Projektbeschreibung gibt es unter <http://www.culturegraph.org/>.

Literatur

- Aleman-Meza, B., Decker, S. L., Cameron, D. & Arpinar, I. B. (2009). Association analytics for network connectivity in a bibliographic and expertise dataset. In J. Cardoso & M. Lytras (Hrsg.), *Semantic web engineering in the knowledge society* (S. 188–205). Hershey / New York: Information Science Reference.
- Bergmann, J., Danowski, P., Pohl, A. & Eckert, K. (2011). Open Data als Innovationsmodell: Experteninterview zur Freigabe bibliothekarischer Daten - Vorteile überwiegen. *Forum Bibliothek und Information*, (6), 466–469.
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 284(5), 29–37.
- Bizer, C., Jentzsch, A. & Cyganiak, R. (2011). State of the LOD cloud: Version 0.3. <http://www4.wiwi.fu-berlin.de/locloud/state/> (Zugriff am 20.11.2011).
- Bonte, A. (2011). Multilinguale semantische Suche in der Closed-Beta: Einladung an Power-User. <http://blog.slub-dresden.de/beitrag/2011/11/24/multilinguale-semantische-suche-in-der-closed-beta-einladung-an-power-user/> (Zugriff am 24.11.2011).
- Brown, D. J. & Boulderstone, R. (2008). *The impact of electronic publishing: the future for publishers and librarians*. München: K.G. Saur.
- Codina, L., Marcos, M.-C. & Pedraza, R. (Hrsg.) (2009). Web semántica y sistemas de información documental. *Bibliotecnología y administración cultural: Vol. 210*. Gijón: Ed. Trea.
- Danowski, P. (2010). Step one: blow up the silo!: Open bibliographic data, the first step towards linked open data. World library and information congress: 76th IFLA general conference and assembly, Göteborg. <http://www.ifla.org/files/hq/papers/ifla76/149-danowski-en.pdf> (Zugriff am 18.11.2011).
- Deutsche Nationalbibliothek (2007). *Regeln für die alphabetische Katalogisierung an wissenschaftlichen Bibliotheken: RAK-WB. 2.*, überarbeitete und erweiterte Auflage, Stand: April 2006 (einschließlich der Aktualisierungen nach der 4. Ergänzungslieferung). [urn:nbn:de:101-2007072711](http://nbn-resolving.org/urn:nbn:de:101-2007072711) (Zugriff am 24.11.2011).
- Dilger, B. & Thompson, W. (2008). Ubiquitous cataloging. In K. R. Roberto (Hrsg.), *Radical cataloging. Essays at the front* (S. 40–52). Jefferson, NC: McFarland & Co.
- Dunsire, G. & Willer, M. (2011). Standard library metadata models and structures for the semantic web. *Library Hi Tech News*, 28(3), 1–12.
- Eckert, K. (2011). Linked Data Service - Häufige Fragen. http://wiki.bib.uni-mannheim.de/linked-data/doku.php?id=faq_de (Zugriff am 20.11.2011).
- Heller, L. (2011). Sacherschließung in und mit der Wikipedia: Idee, Prototyp, Diskussion. Video. Magdeburg. <http://video.gbv.de:8010/asxgen/verko/2011/v8.wmv> (Zugriff am 10.11.2011).
- Hitzler, P., Krötzsch, M., Rudolph, S. & Sure, Y. (2008). *Semantic Web: Grundlagen*. Berlin / Heidelberg: Springer.
- International Federation of Library Associations and Institutions (2009). Erklärung zu den internationalen Katalogisierungsprinzipien. http://www.ifla.org/files/cataloguing/icp/icp_2009-de.pdf (Zugriff am 17.11.2011).

- International Federation of Library Associations and Institutions (2009a). Funktionale Anforderungen an bibliografische Datensätze: Abschlussbericht der IFLA Study Group on the Functional Requirements for Bibliographic Records. Stand: Februar 2009 (2009th ed.). Frankfurt am Main: Deutsche Nationalbibliothek. [urn:nbn:de:101-2009022600](http://nbn-resolving.org/urn:nbn:de:101-2009022600) (Zugriff am 22.11.2011).
- Jones, E. (2011). A question of identity: The role of identifiers in library catalogs. In E. R. Sanchez (Hrsg.), *Conversations with catalogers in the 21st century* (S. 95–103). Santa Barbara: Libraries Unlimited.
- Kittenberger, A. (2010). Faszination: vernetzte Wissensmaschinen. In M. Fugléwicz-Bren, M. Kaltenböck & T. Thurner (Hrsg.), *Zukunft'sWebBuch 2010. Chancen und Risiken des Web 3.0* (S. 85–92). Wien: ed. mono/monochrom.
- Kovačević, A. (2010). Ontology-based data mining in digital libraries. In V. Devedžić & D. Gašević (Hrsg.), *Annals of information systems: Vol. 6. Web 2.0 & semantic web* (S. 163–175). New York [u.a.]: Springer.
- Kreutzer, T. (2011). Open Data - Freigabe von Daten aus Bibliothekskatalogen. <http://www.irights.info/userfiles/open-data-leitfaden.pdf> (Zugriff am 20.11.2011).
- Lauser, B., Johannsen, G. C. C., Keizer, J., van Hage, W. R. & Mayr, P. (2008). Comparing human and automatic thesaurus mapping approaches in the agricultural domain. In J. Greenberg & W. Klas (Hrsg.), *Metadata for semantic and social applications. DC-2008 Berlin. Proceedings of the international conference on Dublin Core and metadata applications, 22-26 September 2008* (S. 43–53). Singapur / Göttingen: Dublin Core Metadata Initiative / Universitätsverlag Göttingen.
- Levergood, B., Farrenkopf, S. & Frasnelli, E. (2008). The specification of the language of the field and interoperability: Cross-language access to catalogues and online libraries (CACAO). In J. Greenberg & W. Klas (Hrsg.), *Metadata for semantic and social applications. DC-2008 Berlin. Proceedings of the international conference on Dublin Core and metadata applications, 22-26 September 2008* (S. 191–196). Singapur / Göttingen: Dublin Core Metadata Initiative / Universitätsverlag Göttingen.
- Library of Congress (2011). A bibliographical framework for the digital age. <http://www.loc.gov/marc/transition/news/framework-103111.html> (Zugriff am 11.2011).
- Malmsten, M. (2008). Making a library catalogue part of the semantic web. In J. Greenberg & W. Klas (Hrsg.), *Metadata for semantic and social applications. DC-2008 Berlin. Proceedings of the international conference on Dublin Core and metadata applications, 22-26 September 2008* (S. 146–152). Singapur / Göttingen: Dublin Core Metadata Initiative / Universitätsverlag Göttingen.
- Open Bibliographic Data Working Group of the Open Knowledge Foundation (2011). Open bibliography principles. <http://openbiblio.net/principles/> (Zugriff am 24.11.2011).
- Pfeffer, M. (2011). Ausleihdaten aus Bibliotheken als Linked Open Data publizieren und nutzen. Vortrag auf der Tagung "Semantic Web in Bibliotheken" in Hamburg am 29.11.2011. <http://swib.org/swib11/vortraege/swib11-magnus-pfeffer.pdf> (Zugriff am 15.06.2011).
- Plassmann, E., Rösch, H., Seefeldt, J. & Umlauf, K. (2011). *Bibliotheken und Informationsgesellschaft in Deutschland: Eine Einführung* (2., gründlich überarbeitete und erweiterte Auflage). Wiesbaden: Harrassowitz.
- Pohl, A. (2010). Open Data im hbz-Verbund. http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2010_open-data.pdf (Zugriff am 24.11.2011).

- Pohl, A. (2011). Linked Data und die Bibliothekswelt. http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_2011_linked-data_ODOK.pdf (Zugriff am 24.11.2011).
- Pohl, A. & Ostrowski, F. (2010). 'Linked Data' - und warum wir uns im hbz-Verbund damit beschäftigen. *B.I.T. online*, 13(3), 259–268.
- Pohl, A. & Ostrowski, F. (2010a). Zur Konzeption und Implementierung eine Infrastruktur für freie bibliographische Daten. http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/pohl_ostrowski_2010_open-data-infrastruktur.pdf (Zugriff am 24.10.2011).
- Renear, A. H., Urban, R. J., Wickett, K. M., Dublin, D. & Shreeves, S. L. (2008). Collection/item metadata relationships. In J. Greenberg & W. Klas (Hrsg.), *Metadata for semantic and social applications. DC-2008 Berlin. Proceedings of the international conference on Dublin Core and metadata applications, 22-26 September 2008* (S. 80–89). Singapur / Göttingen: Dublin Core Metadata Initiative / Universitätsverlag Göttingen.
- Robering, K. (Hrsg.) (2008). *Semiotik der Kultur / Semiotics of culture: Vol. 6. Information technology for the virtual museum: Museology and the semantic web*. Münster: LIT.
- Sander, O. (2009). Bundesarchiv 2.0? Die Kooperation von Bundesarchiv und Wikimedia. Vortragsfolien. Web 2.0 für Gedächtnisinstitutionen, Marburg. http://www.fotomarburg.de/pdf/sander_wikimedia.pdf (Zugriff am 24.11.2011).
- Sicilia, M.-A. & Lytras, M. D. (Hrsg.) (2009). *Metadata and semantics*. New York [u.a.]: Springer.
- Signore, O. (2008). The semantic web and cultural heritage: Ontologies and technologies help in accessing museum information. In K. Robering (Hrsg.), *Semiotik der Kultur / Semiotics of culture: Vol. 6. Information technology for the virtual museum. Museology and the semantic web* (S. 1–31). Münster: LIT.
- Summers, E., Isaac, A., Redding, C. & Krech, D. (2008). LCSH, SKOS and Linked Data. In J. Greenberg & W. Klas (Hrsg.), *Metadata for semantic and social applications. DC-2008 Berlin. Proceedings of the international conference on Dublin Core and metadata applications, 22-26 September 2008* (S. 25–33). Singapur / Göttingen: Dublin Core Metadata Initiative / Universitätsverlag Göttingen.
- Task Force on Guidelines for OPAC Displays (Hrsg.) (2010). *Richtlinien für die OPAC-Anzeige*. IFLA Series on Bibliographic Control: Vol. 40. Berlin / New York: de Gruyter Saur.
- W3C (2011). W3C Semantic Web activity. <http://www.w3.org/2001/sw/> (Zugriff am 24.11.2011).
- W3C Incubator Group (2011). Library Linked Data Incubator Group final report: W3C incubator group report 25 October 2011. <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/> (Zugriff am 10.11.2011).
- Wissenschaftsrat (2011). Empfehlungen zur Zukunft des bibliothekarischen Verbundsystems in Deutschland. <http://www.wissenschaftsrat.de/download/archiv/10463-11.pdf> (Zugriff am 20.11.2011).
- Wolfe, J. (2008). Playing fast and loose with the rules: Metadata cataloging for digital library projects. In K. R. Roberto (Hrsg.), *Radical cataloging. Essays at the front* (S. 69–74). Jefferson, NC: McFarland & Co.
- Yee, M. M. (2011). The single shared catalog revisited. In E. R. Sanchez (Hrsg.), *Conversations with catalogers in the 21st century* (S. 121–136). Santa Barbara: Libraries Unlimited.

Young, J. (2008). Ranganathan's forgotten law: Save the time of the cataloger. In K. R. Roberto (Hrsg.), *Radical cataloging. Essays at the front* (S. 83–84). Jefferson, NC: McFarland & Co.