



Die Deutsche Nationalbibliografie als Linked open Data: Anwendungen und Möglichkeiten

Jürgen Kett

Sarah Beyer

Mathias Manecke

Yvonne Jahns and

Lars G. Svensson
Deutsche Nationalbibliothek
Frankfurt am Main, Germany

*Bearbeitung und teilweise Übersetzung:
Anke Meyer*

Session: 215 — *What is a national bibliography today and what are its potential uses?— Bibliography*

Abstract:

In diesem Papier werden die Anforderungen an eine Nationalbibliografie im 21. Jahrhundert diskutiert. Aufbauend auf den traditionellen Kriterien wie Vollständigkeit und Verlässlichkeit, Aktualität, Zitierbarkeit und Persistenz der Daten ergänzen die Autoren das Kriterium der Integrierbarkeit der Nationalbibliografie ins World Wide Web, da dort der Datenaustausch heutzutage stattfindet. Der Weg ins WWW sollte über Linked Data Technologien führen und die Daten sollten unter einer offenen Lizenz veröffentlicht werden. Als Fallstudie wird die Umsetzung an der Deutschen Nationalbibliothek präsentiert, bei der ca. 70% der Datenbank als Linked Open Data publiziert wurde.

Einleitung

Die Aufgabe einer Nationalbibliografie lässt sich definieren als die vollständige Verzeichnung aller Publikationen aus einem national definierten Raum.¹ Doch die Übertragung dieser Definition auf das Internetzeitalter ist nicht trivial, denn durch das Internet bekam der Begriff Publikation eine neue Bedeutung und der Akt des Publizierens und Recherchierens hat sich grundlegend verändert.

Im Gegensatz zum traditionellen Buch, welches abgeschlossen und somit statisch ist, kommen Publikationen im Netz ohne klare Grenzen aus, und sie sind dynamisch: in ständigem Wandel und interaktiv. Auch haben sich die technischen Möglichkeiten Texte zu verzeichnen geändert: Suchmaschinendienste sorgen für eine grundsätzliche

¹ Cf. Anderson (1974) p. 12

Auffindbarkeit der Informationen. Die Suche im Volltext hat die Nutzung von Katalogisaten und Klassifikationen an den Rand gedrängt. Grund dafür ist aber nicht unbedingt ein funktionaler Vorteile dieser Art der Suche, sondern die Tatsache, dass sie das am leichtesten zu automatisierende Verfahren darstellt: Ressourcen sind im Internet als Volltext verfügbar. Strukturierte Metadaten müssen in den meisten Fällen erst generiert werden. Dazu sind entweder aufwändige technische Verfahren oder manuelle Arbeit erforderlich.

Aber selbst diese Lücke wird an vielen Stellen außerhalb des Bibliothekswesens inzwischen geschlossen: Verlage und Vertriebe generieren bereits von sich aus Metadaten, da sie damit ihre Auffindbarkeit im Netz und damit potentiell auch die Verkaufszahlen steigern. Onlineplattformen wie Wikipedia oder die OpenLibrary² laden die Allgemeinheit ein, Artikel und Metadaten zu Publikationen und deren Kontext zu erzeugen und bilden damit eine perfekte Ergänzung zur Volltextsuche.

Der Anstieg der Massendigitalisierung und OCR veränderte die Situation noch mehr, damit konnte auch (ursprünglich) nichtdigitales Material mit Volltextsuchmaschinentechologie verarbeitet werden. Google hielt dies offenbar für effizienter und effektiver als sich hier auf die traditionelle Art der Erschließung zu stützen – die Suche über Metadaten kommt lediglich als Ergänzung zur Volltextsuche zum Einsatz.³

Die traditionelle Erschließung muss sich fragen lassen, welchen Mehrwert sie in diesem Kontext eigentlich noch bietet bzw. in Zukunft bieten könnte.

Anforderungen an die Nationalbibliografie

Die Nationalbibliografie war traditionell für im Wesentlichen drei Nutzergruppen bedeutsam:

Buchhandel / Verlagswesen, Bibliotheken und Endnutzer (insbesondere Fachwissenschaftler und Literaturwissenschaftler)

Allen drei Nutzergruppen war gemein, dass sie für ihre jeweiligen Verwendungszwecke vor allem auf folgende Grundeigenschaften der Nationalbibliografie bauen mussten:

1. Vollständigkeit und Verlässlichkeit der Daten (Datenqualität)

Für Buchhandel, Bibliothekswesen und Fachwissenschaften war es gleichermaßen von zentraler Bedeutung, dass diese Vollständigkeit weder durch politische noch durch inhaltlich wertende Einschränkungen geschmälert wurde. Dabei spielte die Regelwerkskonformität eine große Rolle: Brüche hatten unmittelbare Konsequenzen auf die Nutzbarkeit und Interoperabilität der Daten.

2. Aktualität

Ohne aktuelle Verzeichnung wäre eine Nationalbibliografie insbesondere für den Buchhandel wertlos gewesen. Dieser Anforderung war der hohe technische Aufwand geschuldet, der insbesondere in der zweiten Hälfte des 20. Jh. betrieben wurde, um trotz ständig steigender Veröffentlichungszahlen die kurzen Veröffentlichungszyklen

² Homepage: <http://openlibrary.org/>

³ Cf. <http://books.google.com/intl/en/googlebooks/about.html>

(wöchentliche Verzeichnisse) durchzuhalten und eine möglichst rasche Erschließung zu ermöglichen.

3. Zitierbarkeit

Hinreichende Vollständigkeit, Verlässlichkeit und Aktualität gaben einer Nationalbibliografie die Fähigkeit, als zentrales Referenzobjekt für das wissenschaftliche Zitieren bibliografischer Nachweise zu fungieren. War ein Werk in einer Nationalbibliografie verzeichnet, dann war es so, wie es dort verzeichnet war, auch existent; wurde eine Titel nicht aufgeführt war es auch sehr wahrscheinlich, dass diese Buch niemals veröffentlicht worden war.

4. Beständigkeit

Die Funktion des Referenzobjektes für wissenschaftliches Zitieren setzte über die genannten Qualitätskriterien hinaus auch ein hohes Maß an Beständigkeit voraus.

Diese Eigenschaft einer Nationalbibliografie war bis zum Beginn des 21. Jh. kein Gegenstand von Diskussionen, weil sie a priori gegeben war. Eine Nationalbibliografie war selbst eine gedruckte Veröffentlichung. Einmal erschienene bibliografische Nachweise konnten zwar in später erscheinenden Kumulationen korrigiert werden, sie blieben aber in der ursprünglichen Version auf Dauer in der Welt und konnten und sollten nicht mehr gelöscht werden.

Die Anforderungen an eine zukünftige Nationalbibliografie müssen weiter gehen und sich an der Datennutzung und Weitergabe im WWW messen.

Das WWW wandelt sich immer mehr in einen Ort für den offenen Datenaustausch: die sogenannte Linked Data Cloud.⁴ Dieses Netz von untereinander verlinkten Datensätzen ist seit 2008 enorm gewachsen, aber andererseits ist wenig bekannt über die Datenqualität und Beständigkeit.

Damit ein semantisches Netz funktioniert, braucht es einen gewissen Grad an Verlässlichkeit, sowohl in Hinsicht auf die Qualität der Informationen, als auch – mindestens ebenso wichtig – in Hinsicht auf deren Haltbarkeit. Dieses Netz kann nur wachsen, wenn die Informationen, über die Aussagen gemacht werden, auch eine Woche später noch existieren. Sie müssen zitierbar sein.⁵

Dies gilt in besonderem Maße für im Internet veröffentlichte Medienwerke und die diese erschließenden Metadaten. Für diesen Teil des Internets die notwendige Verlässlichkeit einzubringen, um dem gesamten Konstrukt eine Stabilität zu geben, könnte die Rolle der Bibliotheken und anderer nicht-kommerzieller Organisationen sein.

Daraus lassen sich folgende Anforderungen ableiten:

⁴ Eine Einführung zu Linked Data und Linked Data Cloud findet man bei Heath and Bizer (2011)

⁵ Cf. Schuster and Rappold (2006)

1. Vollständigkeit/Verlässlichkeit/Interoperabilität

Das Anwachsen des digitalen Publikationsmarktes macht es so gut wie unmöglich alle relevanten Kulturgüter, die für eine Nationalbibliografie relevant sind, zu sammeln. Während der Herausgeber einer traditionellen Nationalbibliographie mit geeigneten Selektionskriterien (z.B. nur selbstständige Publikationen, nur Publikationen mit einem bestimmten Mindestumfang ...) die Schere zwischen dem Anspruch auf Vollständigkeit und der Realität noch klein halten konnte, kann ein solcher Herausgeber im Zeitalter der Netzpublikationen nicht einmal mehr annähernd verlässlich sagen, wie weit die Realität von diesem Anspruch entfernt ist. Stattdessen muss klar sein, dass es völlig neuer Definitionen bedarf, für welchen Teil der Veröffentlichungen Vollständigkeit erreicht werden soll (z.B. Printpublikationen, hochwertige Blogs, bestimmte Hochschulschriften) und bei welchem Teil regelmäßige Snapshots von Webseiten ausreichend sind.

Auch anteilig fehlerbehaftete Daten (wie sie beispielsweise in Prozessen der automatischen Erschließung entstehen können) haben ihre Berechtigung und können für einige Anwendungsfälle großen Nutzen bringen.

Eine neue Anforderung an die Verlässlichkeit ist daher die Herstellung von Transparenz darüber, woher Daten stammen, wie sie erzeugt wurden und für welche Anwendungsfälle sie empfohlen werden.

Die Forderung nach Regelwerkskonformität bezieht sich nicht mehr nur und im immer geringeren Maße auf bibliografische Regelwerke, sondern zunehmend (auch) auf technische Anforderungen an Datenstrukturen, Strukturierungs- und Verknüpfungsregeln, die für die Integration in externe Dienste benötigt werden.

2. Aktualität

Natürlich besteht eine gewisse Erwartungshaltung, dass im Zeitalter des vernetzten Wissens eine relevante Publikation (insbesondere eine Netzpublikation) – wenn sie einmal in der Welt ist – quasi instantan auch in einer Nationalbibliografie gefunden werden kann. In gewisser Weise stehen sich aber der Wunsch nach möglichst aktueller Verzeichnung und die Forderung von Beständigkeit und langfristiger Zitierbarkeit gegenüber. Je früher die Metadaten von noch im Geschäftsgang befindlichen Medien für die Nationalbibliografie "freigeschaltet" werden, desto größer ist die Wahrscheinlichkeit späterer Änderungen und die Notwendigkeit, Versionen der Metadaten zu archivieren.

3. Sicherung der Zitierbarkeit

Eine zentrale Rolle wird die Nationalbibliografie auch in Zukunft als Referenzobjekt für das wissenschaftliche Zitieren bibliografischer Nachweise spielen. Diese Funktion wird insbesondere im Hinblick auf Netzpublikationen vermutlich weiter an Bedeutung gewinnen. Um diesen Service anbieten zu können, muss man eindeutige Identifier für bibliografische Daten anbieten können, die im elektronischen Umfeld funktionieren.

4. Beständigkeit

Die langfristige Sicherung der Integrität bibliografischer Daten ist eine zwingende Voraussetzung für die Sicherung der Zitierbarkeit. Bei dieser Frage geht es auch, aber nicht nur um die permanente Identifizierung der bibliografischen Referenzen. Darüber hinaus spielen Probleme wie Versionskontrolle, Korrektur- und Löschrregularien, Timestamps usw. eine wesentliche Rolle.

Die Struktur der Deutschen Nationalbibliografie im Digitalen Zeitalter

Stellt man sich den im letzten Abschnitt beschriebenen Anforderungen, verbleibt nur eine konsequente Sicht: Die Nationalbibliografie darf nicht neben dem Web stehen.

Die Deutsche Nationalbibliografie wandelte ihre Form von einem Zettelregister zu einer lokalen Datenbank mit immer komplexeren Strukturen, Verknüpfungen zu Normdaten und Außenverknüpfungen. Nun geht sie im WWW auf. Die Nationalbibliografie der Zukunft ist ein integraler Teil des WWW - und das in jeder Hinsicht: strukturell, fachlich, technisch und organisatorisch.

Betrachtet man die Charakteristika, kann man die Nationalbibliografie als Graphen verstehen - als Graphen, der in sich vernetzt, aber auch mit anderen Teilen des WWW untrennbar verbunden ist. Gleiches gilt für alle anderen Verzeichnisse von Kulturgütern und auch für Normdaten, Thesauri und Klassifikationen.

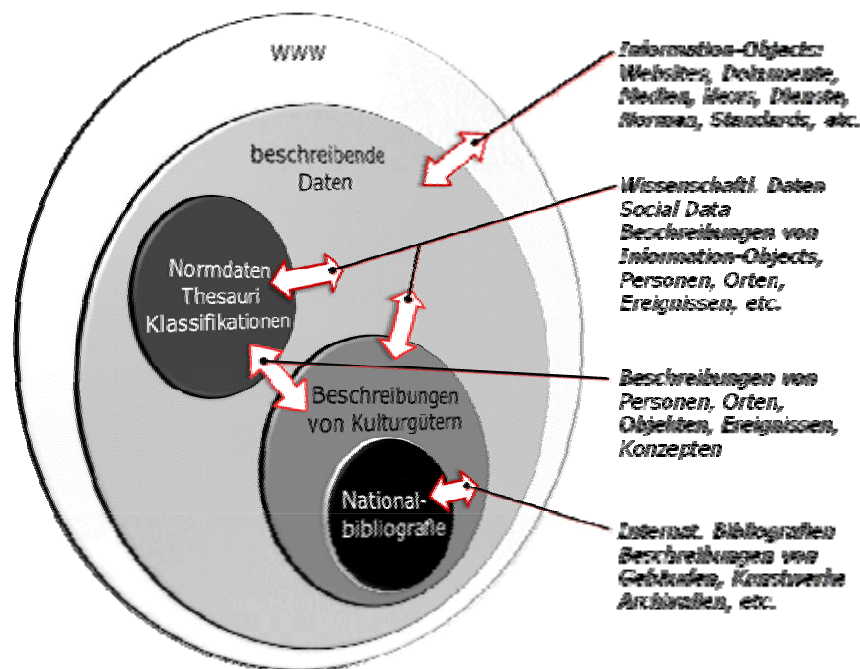


Abb. 1: Die Nationalbibliografie der Zukunft ist ein Graph im WWW. Die Ellipsen stellen Teilgraphen dar. Die Pfeile zwischen den Teilgraphen zeigen beispielhaft mögliche Querverbindungen zwischen den Knoten der Teilgraphen.

Die Nationalbibliografie der Zukunft ist daher kein in sich geschlossenes System und kann aus dem Kontext des WWW nicht mehr verlustfrei herausgeschält werden. Jede druckbare Form der Nationalbibliografie und selbst eine Sicherung aller Daten der Nationalbibliothek wäre ein verlustbehaftetes Derivat, da die Vollständigkeit der Bibliografie auf Daten außerhalb der Datenbank angewiesen ist.

Jeder Knoten dieses Graphen repräsentiert ein Produkt intellektueller bzw. künstlerischer Anstrengungen⁶, im Folgenden vereinfachend „Kulturgut“ genannt. Ein solcher Knoten bündelt signifikante Eigenschaften des Gegenstands, den er repräsentiert, und beschreibt ihn auf diese Weise. Dieses Bündel von Eigenschaften hat mit den Beschreibungen auf Katalogkarten noch Gemeinsamkeiten: Es sind nach vorgegebenen Regeln erfasste Daten. Der Hauptsachtitel einer Publikation ist weiterhin dieselbe Zeichenkette, sowie die ISBN weiterhin dieselbe Zahlenfolge ist. Die meisten anderen Angaben aber sind „Referenzen“, d.h. gerichtete Kanten, die den Knoten mit anderen Knoten im WWW in Beziehung setzen

Diese anderen Knoten können wiederum Metadatenknoten sein: Beschreibungen von Kulturgütern, Personen, Körperschaften, Ereignissen, Orten, Konzepten etc. Viele dieser Knoten liegen redaktionell in der Verantwortung des Bibliothekswesens (beispielsweise die Knoten der GND⁷, RAMEAU⁸ und LCSH⁹). Daneben sind aber auch Referenzen auf ganz andere Knoten möglich: Verweise auf das Kulturgut selbst (sofern es sich hierbei um eine WWW-Ressource handelt) und Verweise auf die Knoten anderer Datenbestände, die von Organisationen außerhalb des Bibliothekswesen gepflegt werden, wie zum Beispiel Links von Personennormdaten zu den entsprechenden Wikipediartikeln.

So gesehen ist die deutsche Nationalbibliografie bereits heute hochgradig vernetzt in einem Graphen, den wir die *Datenbasis der Deutschen Nationalbibliografie* nennen.

Unter dem Begriff Datenbasis ist hierbei die gesamte Graphstruktur zu verstehen, die uns heute und in Zukunft zur Verfügung steht. D.h. die Datenbasis umfasst nicht nur jene Datenbestände, für die die DNB die redaktionelle Verantwortung trägt, sondern auch alle Datenbestände, die mit diesen assoziiert sind. Die Erweiterung der Wissensbasis um Fremddaten wird mit der Öffnung unserer Datasets für das WWW (z.B. durch das Angebot von Linked Data Services) weiter zunehmen.

Die Relationen der aktuellen Wissensbasis sind implizit durch die Anwendung von internationalen Standards wie ISBN, ISSN, ISMN gegeben, über die Erarbeitung von Konkordanzen entstanden (z.B. LCSH, RAMEAU, VIAF), das Ergebnis von Kooperationsvereinbarungen hinsichtlich gemeinsamer Datennutzung und Pflege (z.B. Normdatenkooperation, Datenlieferungen an die Verbünde) oder auch weitestgehend unabhängig über Aktivitäten Dritter erarbeitet worden (Wikipedia).

Grenzen der Nationalbibliografie

Es ist schwer, die Grenzen der Nationalbibliografie exakt zu umreißen. Leichter ist es, Obermengen zu finden, die sie enthalten (vgl. Abb. 1): Zunächst ist sie eine echte Teilmenge aller *beschreibenden Daten* im WWW. Weiterhin ist sie eine Teilmenge aller beschreibenden Daten im Kontext von Kultur. Dies beinhaltet beispielsweise auch Daten zu Personen, Objekten, Konzepten und Orten, die im Kontext von Kulturgütern eine Rolle spielen, und deren Beziehungen zueinander. Hinzu kommen in beliebiger Reihenfolge

⁶ also jene Entitäten, die in FRBR als „Entitäten der Gruppe 1“ bezeichnet werden (vgl. IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009)

⁷ Die Gemeinsame Normdatei (GND – Integrated Authority File) ist eine deutsche Normdatei, die alle Arten von Einheiten abdeckt. Vgl. http://www.dnb.de/DE/Home/home_node.html

⁸ RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) ist eine französische Normdatei der BnF vgl. <http://rameau.bnf.fr/>

⁹ Library of Congress Subject Headings (LCSH) werden von der Library of Congress zur Erschließung verwendet. vgl. <http://id.loc.gov/authorities/subjects>

noch die Beschränkungen auf den nationalen Raum sowie (vorgegeben durch konkrete Regelung des Sammelauftrags) auf einen bestimmten Typus von Kulturgütern (z.B. Bücher, Karten, Postkarten und Tonaufnahmen etc.) .

Ein weiterer Aspekt zur Abgrenzung der Nationalbibliografie ist die administrative Verantwortlichkeit. Gerade auch um den Anforderungen nach Verlässlichkeit (Beständigkeit und Datenqualität) gerecht zu werden, muss die verantwortliche Institution ausreichend Rückhalt vom Staat haben und entsprechend ausgestattet sein.

Für eine Nationalbibliothek oder eine Nationale Bibliografieagentur ist das das Herz ihrer Daseinsberechtigung und sollte Teil ihres gesetzlichen Auftrags sein, der häufig - aber nicht immer - durch ein Pflichtablieferungsgesetz unterstützt wird.¹⁰

Das bedeutet im Umkehrschluss allerdings nicht, dass alle Daten, die von der Deutschen Nationalbibliothek verwaltet werden, zur Nationalbibliografie zu zählen sind: Zahlreiche Datensätze aus Datensets unter der Verantwortung der Deutschen Nationalbibliothek haben keinen Bezug zu den Titeldaten der DNB. Auch die Metadaten zu Sondersammlungen passen trotz administrativer Zuständigkeit durch die Deutsche Nationalbibliothek inhaltlich nicht in die Nationalbibliografie.

Inhalt der Nationalbibliografie

Auf einem internen Workshop an der Nationalbibliografie hatten Kolleginnen und Kollegen die Aufgabe, zu versuchen, die Frage nach dem Gegenstand der Nationalbibliografie zu beantworten.

Es ergaben sich sehr unterschiedliche Antworten mit Varianten, die alle Datensets beinhalteten, bis hin zu einer Reduktion ausschließlich auf die Titeldaten der Nationalbibliografie ohne Einbeziehung der Sondersammelgebiete. Aber selbst die Gruppe mit der zuletzt genannten "Minimallösung" wollte nicht ganz auf die Normdaten verzichten.

Diese Reaktion ist nicht weiter überraschend: Der Ursprung dieser Daten ist letztlich der bibliografische Erschließungsprozess. Desweiteren enthalten diese Daten Informationen wie den Namen des Autors, die beispielsweise für die Erzeugung eines ISBD-Eintrags zu einem Titel unverzichtbar sind. Die Normdaten enthalten also teilweise Informationen, ohne die der Titelsatz allein (für die meisten Verwendungszwecke) nicht mehr aussagekräftig genug ist. Andererseits sind nicht alle Eigenschaften, die ein Normdatensatz beinhaltet, für die Erzeugung einer ISBD-Darstellung relevant.

Die Frage nach dem Gegenstand der Nationalbibliografie lässt sich nicht auf der Ebene von Entitäten beantworten, sondern erst auf der Ebene einzelner Eigenschaften. Nicht alle Eigenschaften, die zum Nachweis eines Kulturguts benötigt werden, sind seine eigenen direkten, sondern es sind teilweise auch die Eigenschaften einer verknüpften Entität. Gleiches gilt für das Rechercheangebot: Der Online-Katalog sollte möglichst alle Daten aus der gesamten Wissensbasis zur Optimierung der Suche verwendet. Er sollte soweit sinnvoll auch Datenelemente nutzen, die außerhalb des Bibliothekswesens und ohne jeden Gedanken an bibliografische Erschließung erstellt wurden. Als ein (beliebiges) Beispiel hierfür seien geografische Koordinaten aus Datasets wie geonames.org genannt, die in einer Suche basierend auf geografischer Lokalisierung und Nähe genutzt werden

¹⁰ Cf. Andersen (1974) p. 11.

können. Die Bereitstellung einer ISBD-Beschreibung eines Kulturguts wie auch des Online-Katalog sind zwei von vielen Diensten rund um die Nationalbibliografie. Die ISBD Anzeige und die Suche sind dabei nur zwei von vielen Diensten, die die Nationalbibliografie umgeben.

Nicht alle Daten, die für diese Dienste genutzt werden, sollten zur Nationalbibliografie gerechnet werden. Der Gegenstand der Nationalbibliografie lässt sich daher nicht aus den darauf aufbauenden bibliografischen Dienstleistungen (gedruckte Bibliografie, Online-Katalog, Datendienste, etc.) ableiten.

Wenn also nationalbibliografische Dienste ohnehin auf Daten basieren, die nicht allein der Nationalbibliografie zuzurechnen sind, dann ist es nur konsequent, die Abhängigkeit zwischen Dienst und Daten ganz zu lösen. Ein konsistentes und klares Bild ergibt sich, beschränkt man den Gegenstand der Nationalbibliografie auf seinen Kern: auf Veröffentlichungen – und damit auf den minimalen, für sich genommen wenig brauchbaren Rumpf bestehend aus wenigen textuellen Elementen und zahlreichen Referenzen. Mit Blick auf RDA¹¹ und FRBR¹² lässt sich die Aussage wie folgt weiter präzisieren: Die Nationalbibliografie verzeichnet im Kern „Manifestationen“ – also die „physische Verkörperung einer Expression eines Werkes“ (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009). Daten auf Werks-, Expressions- und Exemplarebene sind eigenen Datensets zuzuordnen. Die Datensätzen der Nationalbibliografie enthalten lediglich Referenzen auf anderen Ebenen.

Erstellung der Nationalbibliografie

Die Nationalbibliografie wird zunehmend Teil eines globalen Systems aus wechselwirkenden Datenproduzenten und Datennutzern. Je stärker die Nationalbibliografie mit institutions- und domänenübergreifenden Daten verknüpft, desto mehr Anwendungen werden von dieser erweiterten Wissensbasis Gebrauch machen. Auch in Diensten, die Bibliotheken selbst anbieten, werden zunehmend Daten aus externen Datenquellen zum Einsatz kommen. Die zunehmende Verfügbarkeit von ergänzenden Informationen zu Entitäten mit kulturellem Bezug (seien es Kulturgüter, Personen, Konzepte, etc.) bietet die Möglichkeit, bibliografische Dienstleistungen zu verbessern. Weiterhin könnte in Teilen auf das eigene Erfassen von Daten verzichtet und vollständig auf die Nachnutzung von Fremddaten umgestellt werden.

¹¹ siehe RDA-Homepage für weitere Informationen zu RDA: <http://www.rda-jsc.org/>

¹² siehe (IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009)

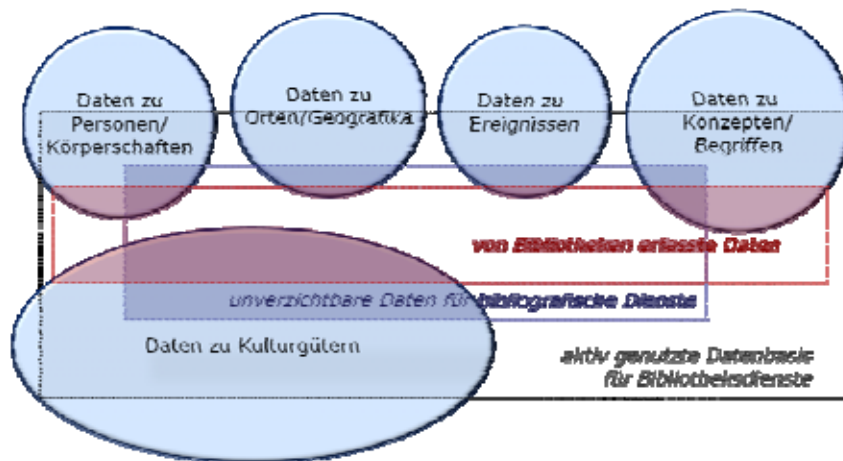


Abb. 2: Mögliche künftige Verteilung zwischen Datennutzung und Datenerfassung in Bibliotheken.

- 1.) Nicht alle von Bibliotheken erfassten Daten sind zwingend notwendig für bibliografische Dienste und dennoch unverzichtbar (z.B. Daten für Nutzungs- und Archivierungszwecke).
- 2.) Nicht alle für bibliografische Dienste benötigten Daten werden von Bibliotheken erstellt.
- 3.) Die von Bibliotheken genutzte Datenbasis ist weitaus größer als beide Mengen zusammengenommen.

Diese Richtung ist keineswegs neu: Bereits heute werden an der Deutschen Nationalbibliothek selbst im Kernbereich bibliografischer Dienste nicht mehr ausschließlich selbst erfasste Daten, sondern in Teilen auch Fremddaten (z.B. die Übernahme von Daten aus Verlagsmeldungen) und teilautomatisch gewonnenen Daten (z.B. Kataloganreicherung, automatische Erschließung) genutzt. Konsequenterweise wären die Ausweitung der Datennachnutzung und die kontrollierte Öffnung der Erschließung für Communities im WWW.

In erster Linie ist es für Bibliotheken eine Chance für eine klare Fokussierung und Besinnung auf die eigenen Stärken, wie die Erstellung von qualitativ hochwertigen Daten, Wissensorganisation und die Erstellung von Normdaten und so eine bessere Arbeitsteilung. Jeder Anbieter kann seine Eigenschaften und Stärken so einbringen und die Daten werden dann gemeinschaftlich genutzt. Der Anbieter einer Nationalbibliografie fühlt sich zuständig für alle Kulturgüter, die zum nationalen Schrifttum gezählt werden, für alle Referenzen, die diese Kulturgüter zu anderen Entitäten haben, und jene Eigenschaften dieser Entitäten, die für die nationalbibliografischen Dienstleistungen unverzichtbar sind und die sonst niemand anbietet. Andere Datenanbieter interessieren sich wiederum für andere Eigenschaften.

Die eigene Erarbeitung bzw. intellektuelle Prüfung von Metadaten wird sich angesichts der Mengenentwicklung zukünftig noch stärker auf die Bereiche beschränken müssen, die gegenüber den mit maschinellen Verfahren angereicherten Fremddaten noch einen Mehrwert bieten. Der Mehrwert ergibt sich auch hier nicht aus einer abstrakten Regelkonformität, sondern aus einer gesteigerten Verwertbarkeit der Daten für eigene Selektions- und Filterprozesse einerseits und für die Geschäftsmodelle der Bibliothekskunden andererseits.

Bestimmende Qualitätsmerkmale:

Damit allerdings die Qualität der Bibliotheksdienste unter dieser Entwicklung nicht leidet, sondern im Gegenteil zunimmt, muss die Nachnutzung von externen Datenquellen mit Sorgfalt bedacht sein. Gerade der Aspekt der Verlässlichkeit (Zitierbarkeit und Persistenz der Daten, Standardisierung, Qualitätssicherung) ist dabei entscheidend. Es muss jederzeit für jeden transparent sein, aus welcher Quelle Daten stammen und (falls diese

Information verfügbar ist) nach welchen Regeln sie erfasst wurden. Darüber hinaus müssen qualitätssichernde Prozesse etabliert werden. Insbesondere muss jede externe Datenquelle hinsichtlich ihrer langfristigen Perspektive eingestuft werden.

Die Nachnutzung von Datenbeständen in einer Nationalbibliografie kann zu einem Konflikt mit den existierenden Regelwerken führen. Den meisten Datenanbietern außerhalb der Bibliothekswelt sind die gängigen Regelwerke wie AACRII, RAK oder RDA nicht bekannt und wir können nicht erwarten, dass sie sich damit beschäftigen.

Ob Daten hinsichtlich der Nationalbibliografie qualitativ hochwertig sind, sollte zukünftig nur noch dann danach beurteilt werden, ob sie formal einem bibliografischen Regelwerk folgen, wenn die jeweilige Regel noch einem der durch die Nationalbibliografie angestrebten Zwecke dient.

Für die Übernahme von Metadaten anderer Aggregatoren sollten vertrauenswürdige Anbieter definiert werden. Für deren Bewertung sollte nach Vorbild des Kriterienkataloges der vertrauenswürdigen digitalen Langzeitarchive Kriterien erarbeitet werden. Dabei sollte ebenfalls in erster Linie darauf geachtet werden, inwieweit die Daten der Anbieter den veränderten Anforderungen an Interoperabilität gerecht werden.

Fremddaten, die nicht von solchen vertrauenswürdigen Anbietern stammen, sollten zwar verwendet, aber als solche gekennzeichnet werden.

Die verwendeten maschinellen Verfahren müssen in erster Linie zu nachvollziehbaren Ergebnissen führen. Je besser es gelingt, mit Hilfe solcher Verfahren homogen erschlossene Teilmengen der Nationalbibliografie zu generieren, desto größer wird die Akzeptanz bei den Nationalbibliografie-Nutzern sein. Dafür könnte es ebenfalls hilfreich sein, gemeinsam mit den Partnern einen Kriterienkatalog für vertrauenswürdige maschinelle Verfahren zu erarbeiten. Die regelmäßige Überprüfung der Verfahren sollte selbstverständlich sein.

Technische Umsetzung

Im vorherigen Abschnitt wurde erklärt, dass die Nationalbibliografie auf einem Minimalset von Metadaten für Manifestationen im Kombinationen mit Normdaten und verlinkten Datensets aufgebaut werden kann.

Die technische Grundphilosophie, die Linked-Data verfolgt, passt ideal auf die genannten Anforderungen: der Gedanke, die Datensätze selbst mit zugreifbaren Web-Identifiern (http-URIs) zu versehen und diese Identifier zur gegenseitigen Referenzierung einzusetzen. Auf diese Weise wird jeder einzelne Datensatz eine individuell benannte und zitierbare Ressource im WWW. Eine unmittelbare Integration der Nationalbibliografie ins WWW ist kaum denkbar.

Nach Tim Berners-Lee sind die Anforderungen für eine Implementierung von Linked Data ziemlich eindeutig:¹³

1. Use URIs as names for things

Jede Einheit in der Nationalbibliografie benötigt ihren eigenen eindeutigen Identifier. Das trifft mindestens auf Manifestationen, kontrolliertes Vokabular, Körperschaften und Personen zu und idealer Weise auch auf andere Normdaten wie Verlage.

¹³ Berners-Lee (2006)

2. Use HTTP URIs so that people can look up those names

Die von Bibliotheken vergebenen Identifier sollten sich an der Infrastruktur im Web orientieren.

3. When someone looks up a URI, provide useful information

Um die Nachnutzung der Bibliotheksdaten zu maximieren ist es zwingend notwendig konkrete Beschreibungen des Datenknotens, der durch die URI identifiziert wurde anzubieten. So können wir andere Datenanbieter dazu bewegen die Bibliotheksdaten nachzunutzen indem sie auf diese Normdaten verlinken.

4. Include links to other URIs so that they can discover more things

Wenn wir Normdaten von anderen Datenanbietern in unseren bibliografischen Daten verwenden, können wir zum einen einige Eigenschaften der verlinkten Entity direkt anzeigen, wie zum Beispiel den Autorennamen oder die Vorzugsbenennung eines Sachbegriffes und zum anderen sollten wir einen Link auf eine Seite für mehr Informationen wie zum Beispiel Geburtsdatum oder Beruf zu diesem Normdatensatz anbieten.

Hierbei ist zunächst weniger entscheidend, ob zur Datenmodellierung nun RDF eingesetzt wird oder ein anderes technisches Konzept. RDF und die darauf aufbauenden Technologien haben gewiss das Potential, die Semantik von Daten auf sehr generische Weise maschinenlesbar zu machen. Aber dieser Anspruch ist im Einzelfall mit einer hohen Modellierungskomplexität verbunden. Die Probleme, die beim praktischen Einsatz von RDF auftreten können, sind Anlass kontroverser Diskussionen¹⁴. Auch aus praktischer Erfahrung im Umgang mit RDF an der DNB kommen wir zu dem Schluss, dass RDF für einige Anwendungsfälle, die das volle Potential der Ausdrucksstärke dieser Sprache benötigen, (zumindest bis dato) noch nicht das passende Werkzeug ist. Solange man die genannten Linked-Data-Konventionen – Veröffentlichung der einzelnen Datenbankobjekte im WWW und das Anreichern dieser Objekte um Referenzen auf andere Objekte – konsequent durchführt, ist der Weg auch für die spätere Einführung ausdrucksstärkerer Beschreibungen geebnet.

Nutzung

Gerade auch kommerzielle Dienste wie Google, Bing und Facebook, die heute das Gesicht des Webs prägen und für die Masse der Nutzer der Einstiegspunkt darstellen, müssen in die Lage versetzt werden, bibliografische Daten kreativ in ihr Angebot zu integrieren, damit sie den Nutzern bei ihrer Recherche unmittelbar zugutekommen. So konnte die Europeana einen starken Anstieg von Zugriffen über Suchmaschinen verzeichnen, nachdem sie für jedes Objekt eine eigene Einstiegsseite und eine eindeutige URI vergeben hatte.¹⁵

Diese Dienste lösen Online-Kataloge und Portale als ersten Anlaufpunkt für bibliografische Recherchen ab. Der eigene Online-Katalog bzw. das Fachportal bleibt ein wichtiger Dienst, da er die beste Referenz für die Nutzung der eigenen Daten darstellt

¹⁴ <http://milicicvuk.com/blog/2011/07/19/ultimate-problem-of-rdf-and-semantic-web/>,
http://www.w3.org/2001/sw/wiki/index.php?title=RDF_Core_Work_Items&oldid=1980#Blank_Nodes

¹⁵ Clark et. al. (2011) p. 15-17

und die Bedürfnisse von spezifischen Nutzergruppen besser adressieren kann. Die Grenze zwischen dem Mehrwert, den Online-Kataloge gegenüber anderen Recherchediensten im Web bieten können, ist allerdings fließend. In Zukunft mag es Suchmaschinenanbieter geben, die sich mehr auf die Unterstützung bibliografischer Recherche konzentrieren, als es der aktuelle Markt tut. Anders gesagt: Es könnte sein, dass Anbieter für allgemeine Webdienstleistungen Online-Kataloge zunehmend obsolet machen. Für den Nutzer wäre dies, wegen des umfassenderen Angebots, in das bibliografische Recherchen dann fachgerecht eingebettet wären, eine gute Entwicklung.

Grundvoraussetzung für die einfache Nachnutzbarkeit der Bibliotheksdaten sind klare Lizenzbestimmungen für diese Daten. Angewendet speziell auf die Nationalbibliografie kann diese Forderung jedoch gut untermauert werden: Die Nationalbibliografie kann ihre Funktion als Nachweissystem für Publikationen nur über eine offene Nutzungslizenz voll entfalten. Nun kann man argumentieren, dass gerade Branchenriesen wie Google eigentlich genug Mittel zur Verfügung hätten, für einen derartigen Datendienst zu zahlen. Fraglich ist aber ob sie es tun würden. Dabei würde das eigentliche Potential der Datenvernetzung ignoriert: Die Daten erhalten erst dadurch den entscheidenden Mehrwert, dass sie von anderen Anbietern und Privatnutzern zitiert und genutzt werden. Und in den Aufwand des Zitierens wird im Allgemeinen nur dann investiert, wenn dadurch das eigene Angebot eine höhere Sichtbarkeit und Relevanz bekommt und dies gelingt nur, wenn die Daten in anderen Diensten mit einem breiterem bzw. anderem Publikum nachgenutzt werden. Dies beinhaltet auch offene Plattformen wie beispielsweise Wikipedia oder kleine bis mittelgroße Webangebote wie Online-Zeitungen oder Online-Händler. Der Nutzen der stärkeren Sichtbarkeit überwiegt definitiv gegenüber dem Verlust an Einnahmen in diesem Fall.

Aktivitäten der Deutschen Nationalbibliothek und Auswirkungen auf die Deutsche Nationalbibliografie

Die DNB ist (wie andere Nationalbibliotheken auch) bestrebt, die aufgezeigten notwendigen Entwicklungen durch eigene Aktivitäten voranzubringen:

Zum einen begann sie bereits 2008, die von ihr verwaltete Datenbasis als Linked-Data zu veröffentlichen. Ausgangspunkt war die Gemeinsame Normdatei. Es folgten weitere Datensets wie DDC-Deutsch. Nach 6 Monaten waren wir dann in der Lage einen Großteil der Titeldaten als Linked Data zu veröffentlichen, so dass mittlerweile 70% des gesamten Datenbestandes in Linked Data zur Verfügung steht. Abgesehen von den DDC-Daten, die unter einer CC BY-NC-ND zur Verfügung gestellt werden, sind alle anderen Daten unter einer Creative Commons Zero Lizenz nutzbar.

Desweiteren setzt die Deutsche Nationalbibliothek verstärkt auf automatisierte Verfahren, die die manuellen Erschließungsprozesse ergänzen und dem Endnutzer neue Sucheinstiege ermöglichen, wie z.B. die Anreicherung des Katalogs durch Inhaltsverzeichnisse und die verstärkte Übernahme der Fremddaten von Verlagen und Universitäten. So werden für Online Publikationen in der Reihe O der Nationalbibliografie ausschließlich beschreibende Informationen der Ablieferer nachgenutzt.¹⁶

Darüber hinaus wird in einem aktuellen Projekt evaluiert, ob sich Sachbegriffe automatisch für Bibliografische Daten vergeben lassen. Dieser Prozess baut auf einem automatischen Indexer auf. Wir analysieren dabei wie der Indexer mit verschiedenen

¹⁶ vgl. Gömpel and Svensson (2011)

Dokumenttypen umgeht und sich dies auf die Qualität der Suchergebnisse auswirkt. Erste Ergebnisse wird es in 2013 geben.

Die automatische Indexierung der Webressourcen wird dabei mit demselben kontrollierten Vokabular durchgeführt, das für die intellektuelle Erschließung verwendet wird. So garantieren wir eine möglichst homogene Suchumgebung und können verlässliche Normdaten anbieten.

Die heutigen Nutzer haben sehr unterschiedliche Erwartungen an Bibliografische Daten insbesondere was die Suche über Sachbegriffe betrifft. Diese Erwartungen beinhalten einen kompletten Überblick über die verfügbare Literatur, Zitierbarkeit oder der direkte Zugriff auf die Publikation.

An der Deutschen Nationalbibliothek folgen wir der IFLA Überzeugung, dass eine kohärente Indexierungspolitik und die Verwendung von kontrollierten Sucheinstiegen auch weiterhin wichtig sind, um konsistente Daten anzubieten.¹⁷ Alle unsere Nutzer profitieren davon, dass wir einheitliche und gut organisierte Zugriffe auf alle Arten von Publikationen anbieten können, unabhängig von Datenformaten oder Vertriebskanälen der Nationalbibliografie.

Die Klassifikationssysteme und Schlagwörter helfen den Nutzern, die relevanten Informationen zu erhalten bzw. zu bewerten.

Ein Großteil der Nutzer (sowohl Endnutzer wie auch professionelle Datennutzer) ist dabei nur an Teilbeständen und nicht an der gesamten Nationalbibliografie interessiert. Darüber hinaus wird wahrscheinlich ein Großteil der Nutzer gar nicht wissen, dass sie gerade die Nationalbibliografie nutzen, wenn sie in unserem Online-Katalog recherchieren. Eine immer größer werdende Zahl von Nutzern profitiert dabei von persönlichen Profilen in der Bibliografie oder abonniert Rss-Feeds auf die Suchergebnisse. Kataloganreicherung wie Inhaltsverzeichnisse oder Abstracts helfen den Nutzern darüber hinaus, die Informationen zu bewerten.

Aufgrund der immer stärker ansteigenden Zahl von veröffentlichten Informationen wird es wichtiger, die Informationen in verwaltbare Einheiten zu kategorisieren, die sich lesen, auswählen und gezielt durchsuchen lassen. Idealerweise wird eine komplette und detaillierte Indexierung auf alle Dokumente angewendet. Angesichts der enormen Mengen von gedruckten Veröffentlichungen und der stets wachsenden Zahl von veröffentlichten Webressourcen unter der .de-Domain, aber auch angesichts der finanziellen und personellen Ressourcen hat sich die Deutsche Nationalbibliothek zu einer graduellen Erschließung entschieden, die es uns erlaubt, den gesamten nationalen Output in unterschiedlichen Detailebenen zu erschließen. Diese modulare Vorgehensweise führt zu verschiedenen Sucheinstiegen für verschiedene Medienarten, ist aber trotzdem transparent und durch Qualitätskriterien kontrolliert.

Zum Minimalset der Erschließungsdaten gehört eine sehr grobe Klassifikationsnummer, die Sachgruppe, die für fast alle Dokumente angeboten wird. Die Struktur dieser Sachgruppe basiert auf den obersten Hierarchiegruppen der DDC Klassifikation.

Ein wichtiger Aspekt ist die Vernetzung der eigenen Daten mit externen Datentöpfen. Auch hier setzt die DNB auf einen Mix aus kooperativen, manuellen und automatischen Verfahren. Beispielsweise wurde die PND im Rahmen des VIAF-Projektes durch ein automatisiertes Verfahren international vernetzt. Das Projekt CrissCross verknüpfte die

¹⁷ IFLA (2012)

SWD über ein redaktionelles Verfahren mit RAMEAU und LCSH. Eine Kooperation mit der deutschen Wikipedia brachte die Vernetzung zwischen Wikipedia-Artikeln und der PND.

Fazit

Die Nationalbibliografie und auch bibliografische Daten im Allgemeinen sollten orientiert an Linked-Data-Prinzipien und unter freien Nutzungsbedingungen im World Wide Web veröffentlicht werden. Damit diese Daten einer möglichst breiten Masse von Endnutzern unmittelbar zugutekommen, sollten mit den großen Anbietern von Websuchen neue Kooperationsmöglichkeiten erarbeitet werden.

Die Optimierung der Erschließungspraxis auf die Bedürfnisse des WWW (auch unter Berücksichtigung absehbarer Entwicklungen) sollte der wichtigste Anspruch aktueller Regelwerksdiskussionen sein. Die Relevanz der erhobenen Daten kann nicht mehr in erster Linie an bibliografischen Regelwerken orientiert sein, weil diese sich selbst nach wie vor im Wesentlichen an den Notwendigkeiten von klassischen Katalogen bzw. Bibliografien orientieren. Dabei ignorieren die Katalogisierungsregeln die heutigen technischen Möglichkeiten: Beispielsweise gibt es noch zahlreiche Regeln, unter welchen Bedingungen Nebeneintragungen gemacht bzw. nicht gemacht werden sollen. Würde hier konsequent nicht von Nebeneintragungen, sondern von Sucheinstiegen gesprochen und würden diese unabhängig von nur intellektuell bestimmbaren Randbedingungen generell gefordert, so könnten diese weitgehend maschinell generiert werden.

Es sollte mehr Zeit für die Erhöhung des Vernetzungsgrades der Daten investiert werden: Textuelle Informationen sollten Referenzen weichen. Die Beschreibung des Kulturguts sollte mit allen Objekten, Personen, Körperschaften, Orten, Ereignissen und Konzepten, die in seinem Kontext eine Rolle spielen, verknüpft sein.

Damit dies besser gelingt, müssen die Normdaten ausgebaut werden: eine internationale, kooperative gepflegte Werkdatei wäre insbesondere wichtig, um Domänen- und Sprachgrenzen zu überwinden. Eine allgemeine Verleger-Normdatei wäre nicht nur für die Recherche relevant, sondern gleichzeitig ein guter Anknüpfungspunkt, um Kooperationen mit den Verlagen auszubauen. Ein eigener veröffentlichter Normdateneintrag als zentraler Anlaufpunkt im Web, der auf alle Publikationen des Verlags verweist, könnte zu einem nicht unerheblichen Wettbewerbsvorteil erwachsen, wenn Suchmaschinenanbieter damit beginnen, derartiges in ihr Angebot zu integrieren.

Die Nationalbibliografie wird ein Teil des Webs. Diese Entwicklung ist notwendig, um ihrer Funktion als Nachweissystem auch in Zukunft noch gerecht zu werden, denn das Web hat sich zum zentralen Ort des Informationsaustauschs entwickelt.

Im Web wird die Nationalbibliografie endgültig ihren strukturellen Wandel vom geschlossenen Register zum offenen Graphen vollziehen. Dadurch wird es möglich, sie enger mit den Daten aus anderen Domänen zu verweben. Gerade auf die Vernetzung von Datenbeständen sollten Bibliotheken ein besonderes Augenmerk legen. Es müssen Werkzeuge und Prozesse geschaffen werden, die die Vernetzung vorantreiben. Gemeinsam mit den Daten anderer Kultureinrichtungen könnte auf diese Weise ein globaler Graph kulturellen Wissens wachsen, den Suchmaschinendienstleister oder öffentliche Portale wie die Deutsche Digitale Bibliothek oder Europeana durch Recherchedienstleistungen mit Leben füllen könnten.

Die Überführung der Nationalbibliografie ins Web erfordert praktische Aktivitäten, die auch kurzfristig machbar sind, aber auch ein ganz grundsätzliches Umdenken. Selbst den

Wandel des Informationswesens der 90er Jahre hat das Bibliothekswesen noch nicht ganz verdaut: Noch immer sind Erschließungspraxis und Produkte nicht konsequent auf die Nutzungsszenarien des digitalen Zeitalters angepasst. Es ist nicht anzunehmen, dass der Wandel im Informationssektor an Fahrt verlieren wird. Weitere Veränderungen werden kommen. Das Bibliothekswesen sollte daher versuchen, selbst ein zentraler Teil des Wandels zu werden, also von den Antreibern neuer Entwicklungen mitgenommen zu werden und diese mitzugestalten. Dazu braucht es zu allererst mehr Flexibilität und Offenheit.

Literaturverzeichnis

- Anderson (1974) Anderson, Dorothy: Universal bibliographic control : a long term policy; a plan for action. Pullach 1974.
- Berners-Lee (2006) Berners-Lee, Tim: Linked Data. Available at <http://www.w3.org/DesignIssues/LinkedData.html>
- Clark et. al (2011) Clark, D. J.; Nicholas, D.; Rowlands, I.: D3.1.3 – Publishable report on best practice and how users are using the Europeana service. Available at http://www.europeanaconnect.eu/documents/D3.1.3_eConnect_LogAnalysisReport_v1.0.pdf
- Gömpel and Svensson (2011) Gömpel, Renate; Svensson, Lars G.: Managing Legal Deposit for Online Publications in Germany. 2011. urn:nbn:de:101-2011061609. Online available at <http://nbn-resolving.de/urn:nbn:de:101-2011061609>
- Heath and Bizer (2011) Heath, Tom; Bizer, Christian: Linked Data: Evolving the Web into a Global Data Space. San Rafael (Calif.) 2011. (Also online at <http://linkeddatabook.com/editions/1.0/>)
- IFLA (1998) Functional requirements for bibliographic records : final report. München 1998. All references in this paper are to the 2009 revised online version at http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf
- IFLA (2009) IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its Glossary. Ed. Barbara Tillett and Ana Lupe Christián. München 2009. For a list of online versions (incl. translations) see <http://www.ifla.org/publications/ifla-series-on-bibliographic-control-37>
- IFLA (2012) , Guidelines for Subject Access in National Bibliographies. Ed. Yvonne Jahns. Berlin: 2012.
- Miličić (2011) Miličić, Vuk: The Ultimate Problem of RDF and the Semantic Web. Blog post, available at <http://milicicvuk.com/blog/2011/07/19/ultimate-problem-of-rdf-and-semantic-web/>
- Purcell (2011) Purcell, Kristen: Search and email still top the list of most popular online activities: Two activities nearly universal among adult internet users. 2011. Available online at http://pewinternet.org/~media//Files/Reports/2011/PIP_Search-and-Email.pdf
- Schuster and Rappold (2006) Schuster, Michael; Rappold, Dieter: Social Semantic Software – was soziale Dynamik im Semantic Web auslöst. In: Semantic Web: Wege zur vernetzten Wissensgesellschaft. Ed. Tassilo Peregrini. Berlin 2006.
- Svensson and Jahns (2010) Svensson, Lars G.; Jahns, Yvonne: **PDF, CSV, RSS and other acronyms: redefining the bibliographic services in the German National Library.** 2010. Available online at <http://www.ifla.org/files/hq/papers/ifla76/91-svensson-en.pdf>