

Die Nationalbibliografie im Zeitalter des Internets

Die traditionelle Definition einer Nationalbibliografie als vollständige Verzeichnung aller Publikationen aus einem national definierten Raum ist nicht ohne Weiteres auf das Internetzeitalter übertragbar. Denn durch das Internet bekam der Begriff Publikation eine neue Bedeutung und der Akt des Publizierens und Recherchierens hat sich grundlegend verändert: Publikationen sind leicht zu veröffentlichen, in zunehmendem Maße dynamisch und als indexierbare Volltexte für Suchmaschinen verfügbar. Dieser Artikel beschäftigt sich mit den Fragen, welchen Mehrwert die traditionelle Erschließung noch bieten kann und welche Ausrichtung eine künftige Nationalbibliografie vor diesem Hintergrund haben müsste. Aufbauend auf einer Gegenüberstellung der traditionellen und heutigen Anforderungen wird die Idee einer künftigen Nationalbibliografie als Graph im World Wide Web entworfen. Der Artikel schließt mit einer Betrachtung von Aktivitäten der Deutschen Nationalbibliothek, die dabei helfen sollen, diese Idee in die Tat umzusetzen.

The traditional definition of a national bibliography as the complete index of all publications within a national, pre-defined region no longer applies as such in the Internet age. The concept of publication has acquired a new meaning and the acts of publication and research have changed fundamentally. Writings are easy to publish, they are increasingly dynamic in nature and, hence, already fully indexable for search machines. This article deals with two questions: what additional value can be offered by traditional cataloguing and what orientation is appropriate for a national bibliography in the future. Starting with a comparison between traditional and present-day needs, this article proposes the idea of a national bibliography as a graph within the World Wide Web. In closing, the article gives an overview of steps which the German National Library could take to put such an idea into action.

EINLEITUNG

Die Aufgabe einer Nationalbibliografie lässt sich definieren als die vollständige Verzeichnung aller Publikationen aus einem national definierten Raum. 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. Das weltweite Veröffentlichen von Informationen ist nun technisch auch ohne Verlage möglich.

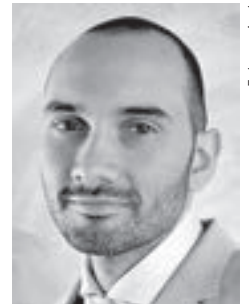
Auch hat das Web eine eigene Antwort auf die Verzeichnung seiner Inhalte gefunden: Suchmaschinen-dienste sorgen für eine grundsätzliche Auffindbarkeit der Informationen. Die Suche im Volltext hat die Nutzung von Katalogisaten und Klassifikationen an den Rand gedrängt. Grund dafür ist aber nicht unbedingt ein funktionaler Vorteil 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. Wissensplattformen wie Wikipedia, die OpenLibrary¹ und andere Web2.0-»Mitmach-Systeme« erzeugen Artikel und Metadaten zu Publikationen und deren Kontext und bilden damit eine perfekte Ergänzung zur Volltextsuche. Mit dem Einstieg von Google in das Digitalisierungsgeschäft fiel zudem noch eine weitere Hürde: 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 (vgl. Google Inc., 2011).

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 Form und der Inhalt der Nationalbibliografie waren und sind in erster Linie bestimmt durch die Anforderungen, die an die darauf basierenden Produkte und Dienste gestellt werden. Die Verbindung ist so eng, dass nicht immer zwischen dem Produkt und der Bibliografie als abstraktem Konzept unterschieden wird. Beim Ursprungsprodukt, der gedruckten Nationalbibliografie, war das auch nicht weiter verwunderlich, denn es war lange Zeit die einzige Repräsentanz der Nationalbibliografie. Doch die fortschreitende Informatisierung der Gesellschaft brachte neue Nutzungsszenarien. Es ist schon lange nicht mehr allein die gedruckte Nationalbibliografie, die bestimmend dafür ist, welche Daten erfasst werden bzw. auf welche Weise sie erfasst werden. Aus Sicht der gedruckten Nationalbibliografie gesprochen, haben sich hinter ihrem Rücken – in der Wissensbasis, aus der sie sich speist –



Jürgen Kett

Foto: privat



Mathias Manecke

Foto: privat



Sarah Beyer

Foto: Veit Stöfel

Produkt versus Konzept

weitaus komplexere Strukturen als für sie nötig entwickelt. Diese Strukturen bedienen andere Dienste und Prozesse wie den *Online-Katalog* und die *kooperative Datenpflege*: die Nationalbibliografie wandelte ihre Form von einem Zettelregister zu einer lokalen Datenbank. Die Strukturen dieser Datenbank wurden komplexer, Titelbeschreibungen setzen sich aus mehreren Entitäten (hierarchische Titelsätze, Normsätze etc.) zusammen, Normdaten entwickelten eine zunehmende Eigenständigkeit und der Datenaustausch zwischen Bibliotheken und damit die Außenverknüpfungen zu anderen Datenbanken intensivierten sich.

Um zu verstehen, was die Nationalbibliografie in Zukunft auszeichnet, welche Struktur und welchen Inhalt sie haben muss, ist es notwendig, die Gesamtheit aller Anforderungen, die sich aus den verschiedenen Sichten ergeben, zu verstehen. Aus der traditionellen Sicht der gedruckten Nationalbibliografie spielt das Kriterium Regelwerkskonformität eine maßgebliche Rolle. Im heutigen Tagesgeschäft sind Anforderungen hinzutreten, die aus diversen Diensten resultieren. So erfordert der Online-Katalog besondere Strukturen, um eine optimale Suche zu ermöglichen, z. B. weitestgehende Verlinkung zwischen den Daten und eine größtmögliche Granularität. Die kooperative Datenpflege wiederum verlangt Stabilität, Nachvollziehbarkeit und eine klare Verantwortlichkeit und Verlässlichkeit. Zu diesen Anforderungen tritt noch die Sicht auf die zukünftigen Erwartungen, Möglichkeiten und die sich abzeichnenden Perspektiven für eine zukunftsorientierte Nationalbibliografie.

Traditionelle Anforderungen

Etwas mehr als hundert Jahre nach der Erfindung des Buchdrucks mit beweglichen Lettern begann mit den Messkatalogen der großen deutschen Buchmessen (in Frankfurt am Main 1564 und in Leipzig 1594) in Deutschland die allgemeine und möglichst vollständige Verzeichnung der gesamten Buchproduktion des jeweiligen Messejahres. Während diese Aufgabe in den anderen europäischen Nationalstaaten von den sich meist aus königlichen Bibliotheken entwickelnden großen Nationalbibliotheken übernommen wurde, blieb sie in Deutschland über mehrere Jahrhunderte verschiedenen Buchhandelsunternehmungen vorbehalten. Die erwähnten Messkataloge wurden in Frankfurt am Main bis 1749 und in Leipzig bis 1860 fortgeführt und dann durch verschiedene Verlagsunternehmen ersetzt. Die wichtigsten waren Hinrichs Bücher-Catalog (1742 bis 1913), das allgemeine Bücherlexikon von Wilhelm Heinsius (1812 bis 1894) und Christian Gottlob Kayser's Verzeichnis (1834 bis 1911) (Grimmer, et al., 2002).

Die Entstehung der Deutschen Nationalbibliografie ging wiederum auf die Initiative des Börsenvereins der Deutschen Buchhändler zu Leipzig zurück. Sie stand in direktem Zusammenhang mit der enormen Steigerung der Buchproduktion am Beginn des 20. Jahrhunderts. Die Verzeichnung in der Deutschen Nationalbibliografie verlieh einer Publikation und nicht zuletzt auch ihrem festen Ladenpreis offiziellen Charakter. Das war sowohl aus kommerzieller als auch aus wissenschaftspolitischer Sicht bedeutsam (vgl. Kummer, 1995). 1931 erschien die Deutsche Nationalbibliografie erstmals unter dem Titel »Deutsche Nationalbibliographie: wöchentliches Verzeichnis der ... Neuigkeiten des Deutschen Buchhandels« von Beginn an mit den Reihen A und B. Auch die verschiedenen Kumulationen zu Halbjahres-, Jahres- und Fünfjahresverzeichnissen übernahm von nun an die Deutsche Bücherei in Leipzig. Für die zurückliegenden Jahrhunderte war die zusammenfassende Verzeichnung der nationalen Verlagsproduktion auf Grund des Fehlens einer einheitlichen Nationalbibliographie ein schwieriges Unterfangen. So entstanden die großen Projekte zu einer retrospektiven Nationalbibliographie: Gesamtverzeichnis des deutschsprachigen Schrifttums (GV 1700 bis 1965, teilweise bis 1980 im Verlag K.G. Saur), die Verzeichnisse des 16. und 17. Jahrhunderts (VD 16 und VD 17) und der Gesamtkatalog der Wiegendrucke (GW). VD 16 und 17 haben sich inzwischen zu leistungsfähigen Portalen entwickelt,² ein VD 18 ist in Arbeit.³ Für die Drucke der Inkunabel-Zeit hat sich mit dem INKA-Portal ebenfalls ein deutsches WEB-Portal etabliert.⁴

Die Deutsche Nationalbibliografie war traditionell für im Wesentlichen drei Nutzergruppen bedeutsam:

- **Buchhandel / Verlagswesen**; u. a. für:
 - den Titelnachweis der eigenen Publikationen,
 - Marktanalysen sowohl der Verlage als auch des Buchhandels,
 - Buchhandels- / Verlagsstatistik.
- **Bibliotheken**; u. a. für:
 - die direkte Übernahme der Erschließungsleistung (Erwerb der bibliografischen Daten) für eigene Kataloge,
 - die Nachnutzung der Normierungsleistungen (Normdaten, Regelwerkspflege),
 - den eigenen Bestandsaufbau,
 - die eigene Auskunftstätigkeit.
- **Endnutzer**; u. a. für:
 - Fachwissenschaft: Diese erwartet vor allem einen Überblick über relevante Literatur zum Fachgebiet und die Sicherung der Zitierbarkeit der eigenen Publikationen,
 - Medien-, Buch-, Verlags-, Literaturwissenschaft:

Diese erwarten vor allem einen Überblick über die Verlagsproduktion und den Literaturbetrieb einer bestimmten Zeit, zu einem bestimmten Thema, einer bestimmten Region, einer bestimmten Branche, eines bestimmten Verlages.

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)

In doppelter Hinsicht waren Vollständigkeit und Verlässlichkeit die zentralen Qualitätskriterien der traditionellen Nationalbibliografie.

Zum einen galt es, das Paradigma der »vollständigen Verzeichnung des nationalen Schrifttums« zu erfüllen. Für Buchhandel, Bibliothekswesen und Fachwissenschaften war es gleichermaßen von zentraler Bedeutung, dass diese Vollständigkeit weder durch politische noch durch inhaltlich wertende Einschränkungen geschmälert wurde. Einziges Auswahlkriterium sollte und konnte die formale Tatsache sein, dass es sich bei dem zu verzeichnenden Objekt um ein selbstständig erschienenes Werk des nationalen Schrifttums handelte. Für die Definition dessen, was unter einer selbstständigen Veröffentlichung und was unter nationalem Schrifttum zu verstehen ist, gab es durchaus Spielräume und dementsprechend auch im Laufe der Geschichte Wandlungen, aber im Wesentlichen konnte diese Definition immer klar formuliert und damit die Aufgabe immer klar umrissen werden.

Zum anderen war auch die Vollständigkeit der bibliografischen Daten zu einem einzelnen Werk zwingende Voraussetzung für die Nutzbarkeit einer Nationalbibliografie. Bemerkenswert ist in diesem Zusammenhang, dass in den europäischen Nationalstaaten die Übernahme der Funktion des zentralen bibliografischen Nachweises als staatliche Aufgabe eng mit der Durchsetzung des nationalen Pflichtexemplars verbunden war (vgl. Blum, 1990). Nur mit Hilfe der auf das Pflichtexemplar gründenden Autopsie war es hinreichend verlässlich möglich, vollständige bibliografische Daten regelwerksgetreu in einer Nationalbibliografie zu publizieren. Dabei spielte die Regelwerkskonformität eine große Rolle: Brüche hatten unmittelbare Konsequenzen auf die Nutzbarkeit und Interoperabilität der Daten.

2. Aktualität

Ein weiteres zentrales Qualitätskriterium der traditionellen Nationalbibliografie war die 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 (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.

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. Jahrhunderts 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.

Heutige Anforderungen:

Willkommen im World Wide Web

Die Frage nach den heutigen Anforderungen stellte sich auch die »IFLA Working Group on Guidelines for National Bibliographies«. In ihrem Zwischenbericht kommt die Arbeitsgruppe im Großen und Ganzen auf dieselben Nutzergruppen wie Buchhandel und Verlagswesen, Bibliotheken und Endnutzer (vgl. IFLA Working Group on Guidelines for National Bibliographies, 2008). Es werden aber auch einige neue Anforderungen, die sich aus dem elektronischen Zeitalter ergeben, genannt – so beispielsweise die technische Nutzbarkeit von Datensätzen in Computeranwendungen.

Das WWW, das sich in seinem Ursprung aus Webseiten, später dann auch aus Webanwendungen für Endnutzer zusammensetzte, wird zunehmend zu einem Ort des offenen Datenaustauschs. Schon gegen Anfang des neuen Jahrtausends wurden mit der Definition von Standards für die Maschine-Maschine-Kommunikation im Web, den so genannten *Webser-*

Grundeigenschaften der Nationalbibliografie

WWW: Ort des offenen Datenaustauschs

vices⁵, die Weichen für diese Entwicklungen gestellt. Im Laufe der 2000er Jahre haben sich Webservices und deren Weiterentwicklungen durchgesetzt, und mit ihnen sind verteilte Web-basierte Anwendungen zu einem selbstverständlichen Bestandteil von Dienstleistungen geworden (vgl. Schuster, et al., 2006).

Auch im Bibliothekswesen hat diese Technologie mit Schnittstellen wie OAI und SRW/U und dem Einsatz von XML-Austauschformaten Einzug gehalten. Webservices waren von Beginn an als Schnittstellen zur Domänen-übergreifenden Vernetzung gedacht. Allerdings hatte man dabei (wie der Name schon sagt), eher höherwertige, rechenintensive oder anderweitig komplexe Dienstleistungen im Auge – Dienstleistungen, die interne Datenbanken als Basis für ihre Berechnungen verwenden. Mit dem Erfolg von *Social Networks* und Community-Projekten wie *Wikipedia* und der davon inspirierten *Linking-Open-Data*-Bewegung verschob sich der Fokus auf die nackten Daten selbst. Dahinter steht die Philosophie: Stelle deine Daten zur Verfügung, und die Web-Gemeinde wird daraus mit ihrem kreativen Potential völlig neue Dienste erschaffen (vgl. W3C, 2011). Seither werden im Web Daten aus unterschiedlichsten Quellen (den sogenannten *Data sets*) über offene Schnittstellen angeboten und miteinander verknüpft – Daten aus Social Networks, geografische Daten, wissenschaftliche Primärdaten und zunehmend auch Daten aus dem kulturellen Sektor.⁶

Es sind Datennetze wie die Linked Open Data Cloud, in die Nationalbibliografien integriert werden müssen, um den Erwartungen des World Wide Web gerecht und somit auch in Zukunft noch als relevant wahrgenommen zu werden. Es gibt keine bessere Plattform für das Bibliothekswesen, um den Wert seiner Erschließungsarbeit unter Beweis zu stellen. Denn durch die *Linking-Open-Data*-Bewegung steht zwar inzwischen eine beachtliche Datenbasis zur Verfügung, aber bezüglich Qualität und Beständigkeit lässt sich nur vereinzelt eine Aussage machen. Die noch junge Geschichte der Suchmaschinen hat gezeigt, wie zweischneidig das Geschäft mit der Aufmerksamkeit ist: Unternehmen bewegen erhebliche Summen⁷, um bei der Recherche im Netz möglichst häufig gefunden zu werden. Dabei spielt nicht selten die Korrektheit der Aussagen eine völlig untergeordnete Rolle. 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. Das Netz kann nur wachsen, wenn die Informationen, über die Aussagen gemacht werden, auch eine Woche später noch existieren. Sie müssen zitierbar sein (vgl. Schuster, et al., 2006).

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:

1. Vollständigkeit, Verlässlichkeit, Interoperabilität

Schon immer gab es hinsichtlich der Vollständigkeit der Nationalbibliografie eine Schere zwischen Anspruch und Realität. Während diese aber bei der traditionellen Nationalbibliografie erstens relativ klar definiert und messbar war und damit den potenziellen Benutzerinnen und Benutzern der Nationalbibliografie transparent gemacht werden konnte und zweitens durch geeignete formale Selektionskriterien (nur selbstständige Publikationen, nur Publikationen mit einem bestimmten Mindestumfang ...) gering gehalten werden konnte, ist es nach dem Hinzutreten von Netzpublikationen zum Sammelauftrag der Deutschen Nationalbibliothek (DNB) nicht mehr möglich, einigermaßen verlässliche Aussagen zum Umfang dessen, was Bibliografie-relevant wäre, zu treffen. Man erreicht nicht nur keine Vollständigkeit mehr, man kann nicht einmal mehr annähernd verlässlich sagen, wie weit man in der Realität von diesem Anspruch entfernt ist. Und dennoch sollte dieser Anspruch auch im Online-Zeitalter nicht aufgegeben werden. Es muss nur klar sein, dass es völlig neuer Definitionen bedarf, für welchen Teil der Veröffentlichungen Vollständigkeit erreicht werden soll.

Im Netz gibt es zum einen Zusammenstellungen von Texten und Bildern, die sehr wohl mit den traditionellen selbstständigen Werken vergleichbar sind (Verlagsproduktionen, Hochschulschriften, [wissenschaftliche] Netzzeitschriften), auch wenn nicht absehbar ist, wie lange solche Publikationsarten noch aufrechterhalten werden. Ohne deren Einbeziehung wäre eine Nationalbibliografie völlig unzureichend und damit auf Dauer wertlos. Für diese Art von Netzpublikationen sollte die Vollständigkeit definierbar und damit messbar sein.

Das größere Problem bereitet auf den ersten Blick das Thema Web-Harvesting von Publikationen, die mit traditionellen selbstständigen Werken nicht vergleichbar sind: Wie oft müssen diese geharvestet werden, um hier dem Begriff der Vollständigkeit gerecht zu werden?

Dieses Problem entschärft sich mit der Erkenntnis: Das Web ist nicht archivierbar. Es ist ein dynamisches, interaktives Gesamtkunstwerk. Geharvestete Web-

nackte Daten als
Eigenwert

WWW: eine ideale
Plattform für das
Bibliothekswesen

sites sind nur Sammlungen von Schnappschüssen, die in einem möglichst (so weit technische Rahmenbedingungen dies zulassen) engen Zeitraum entstanden sind. Die Snapshots sind nicht exakte Kopien der jeweiligen Websites, sondern geben lediglich – ähnlich der Szenen-Fotografie einer Theateraufführung – einen Eindruck wieder, was zum Zeitpunkt der Archivierung auf den Websites zu sehen war. Die Snapshots sind neue, eigenständige Objekte, die im Moment der Verlinkung in einem Online-Katalog zu einer neuen Publikation werden. Die Nationalbibliografie enthält nur die Beschreibung dieser Schnappschüsse – nicht etwa die Beschreibung der Websites – und ist in dieser Hinsicht vollständig. Bibliotheken und Archive können nicht das Web, sondern lediglich die von ihnen selbst erstellten Snapshots verzeichnen und archivieren.

Auch die Verlässlichkeit hinsichtlich der Datenqualität bleibt eine zentrale Anforderung an die Nationalbibliografie. Fehler wie inkorrekte Aussagen können sich in einer hochgradig vernetzten Welt schnell potenzieren. Allerdings kann die Datenqualität nicht mehr allein am Maßstab der Regelwerkskonformität gemessen werden, sondern auch daran, welchen Nutzen die Daten für die verschiedenen Dienstleistungen bringen. 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.

Für die Interoperabilität der Daten ist nach wie vor die Regelwerkskonformität von großer Bedeutung. Die Forderung nach Regelwerkskonformität bezieht sich aber nicht mehr nur und im immer geringeren Maße auf bibliografische Regelwerke, sondern zunehmend (auch) auf technische Anforderungen an Datenstrukturen, Strukturierungs- und Verknüpfungsregeln. In diesen Bereichen stellt das World Wide Web völlig andere Anforderungen als das traditionelle Bibliothekswesen. Und auch die internationalen Rahmenbedingungen unterscheiden sich stark von jenen des deutschsprachigen Raumes. Die Anforderungen werden sich durch den Umstieg auf MARC21 zwar stärker annähern, aber dennoch sind auch die verbleibenden Unterschiede unübersehbar. Die Nationalbibliografie muss mit dieser Heterogenität umgehen, sich an bestehenden Standards und »Common Practices« orientieren, den technischen Anforderungen der verschiedenen Kundengruppen gerecht werden, muss aber auch selbst gestaltend aktiv werden, insbesondere dort,

wo es gilt, Widersprüche zwischen den verschiedenen Kundengruppen zu überwinden.

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.

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 Löschregularien, Timestamps usw. eine wesentliche Rolle.

DIE NATIONALBIBLIOGRAFIE: EIN GRAPH IM GRAPHEN DES GRAPHEN DER GRAPHEN

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

Die Nationalbibliografie wandelte ihre Form von einem Zettelregister zu einer lokalen Datenbank mit immer komplexeren Strukturen 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.

Struktur der Nationalbibliografie:

Ein Graph im WWW

Strukturell muss man die Nationalbibliografie als

**Anforderungen an die
Nationalbibliografie im
Zeitalter des WWW**

**Nationalbibliografie darf
nicht neben dem WWW
stehen**

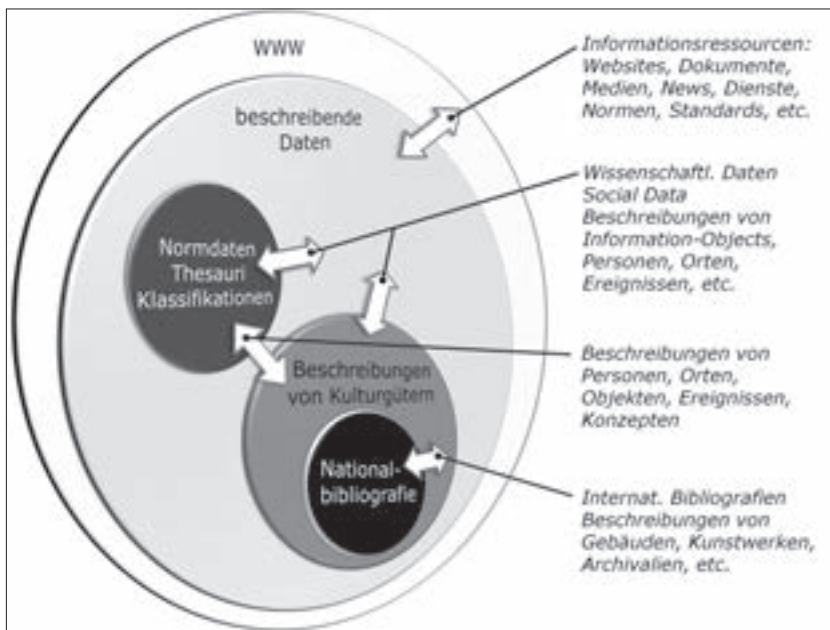


Abb. 1: Die Nationalbibliografie der Zukunft ist ein Graph im WWW. Die Ellipsen stellen Teilgraphen dar. Die Pfeile zwischen den Teilgraphen zeigen beispielhaft mögliche Querverbindungen zwischen den Knoten der Teilgraphen

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.

kein geschlossenes System

Die Nationalbibliografie 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 Deutschen Nationalbibliothek wäre ein verlustbehaftetes Derivat.

Jeder Knoten dieses Graphen repräsentiert ein Produkt intellektueller bzw. künstlerischer Anstrengungen,⁸ 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 (vgl. Abb. 2): Es sind nach vorgegebenen Regeln erfasste Daten. Der Hauptsachtitel einer Publikation ist weiterhin dieselbe Zeichenkette. 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 Gemeinsamen Normdatei – GND). 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 Bibliothekswesens gepflegt werden.

So gesehen, ist die Deutsche Nationalbibliografie bereits heute hochgradig vernetzt in einem Graphen, den wir die *Datenbasis der DNB* 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 Deutsche Nationalbibliothek die redaktionelle Verantwortung trägt, sondern auch alle Datenbestände, die mit diesen assoziiert sind. Die Erweiterung der Wissensbasis um Fremddaten wird mit der Öffnung unserer Datensets 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 Kooperationsverein-

URL	http://d-nb.info/1014919797/about	← Bezeichner und URL des bibliogr. Eintrags
Titel	Der Staat (Politeia)	
Autor	Platon	← Referenz auf die Personen in der GND
Verlagsort : Verleger	Hamburg : tredition	← Referenzen auf den Ort und Verlag in der GND
Erscheinungsjahr	2011	← Referenz auf das Konzept »Jahr 2011« bei dbpedia.org
Persistent Identifier	urn:nbn:de:101:1-201109033292	← Referenz auf die Publikation im WWW
ISBN	978-3-8424-2088-5	
Sachgruppe(n)	800 Literatur, Rhetorik, Literaturwissenschaft	← Referenz auf das Konzept »Sachgruppe 800«

Abb. 2: Beispielhafter (und unvollständiger) Eintrag der Nationalbibliografie mit Link-Labels. Referenzen auf andere WWW-Knoten werden durch mit Referenzen hinterlegte Bezeichnungen (Link-Labels) dargestellt. In der aktuellen Erschließungspraxis werden nicht wie in diesem Beispiel so viele Referenzen wie möglich eingesetzt (Verlagsort und Verleger beispielsweise werden bis dato weitgehend nicht mit Normdaten verknüpft)

barungen 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:

Ein Teil des Graphen der Kultur

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 noch die Beschränkungen auf den nationalen Raum sowie – vorgegeben durch konkrete Regelung des Sammelauftrags der Deutschen Nationalbibliothek (vgl. Deutsche Nationalbibliothek, 2009) – auf einen bestimmten Typus von Kulturgütern. (Wie bereits im vorherigen Kapitel diskutiert, ist eine exakte Regelung des Sammelauftrags im Zeitalter des WWW nicht mehr möglich. Derartige Regelungen können lediglich Annäherungen sein.)

Ein weiterer Aspekt zur Abgrenzung der Nationalbibliografie ist die administrative Verantwortlichkeit. Die Daten der Deutschen Nationalbibliografie sollten nach heutigem Stand ohne Zweifel vollständig von der Deutschen Nationalbibliothek⁹ verwaltet werden. Gerade auch um den Anforderungen nach Verlässlichkeit (Beständigkeit und Datenqualität) gerecht zu werden. Ein Dataset wie geonames.org würde nach diesem Kriterium klar außerhalb liegen, selbst wenn die darin enthaltenen Daten von der Deutschen Nationalbibliothek genutzt würden. 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 DNB-Verantwortung haben keinen Bezug zu den Titeldaten der Deutschen Nationalbibliothek. Auch die Metadaten zu ihren Sondersammlungen passen trotz administrativer Zuständigkeit durch die Deutsche Nationalbibliothek inhaltlich nicht in die Nationalbibliografie.

Gegenstand der Nationalbibliografie:

Manifestationen

Wie Abbildung 1 zu entnehmen, zählen die Normdaten, Klassifikationen und Thesauri nicht zur Nationalbibliografie. Diese Entscheidung war ein Diskussions-

punkt: Auf einem DNB-internen Workshop 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 der Deutschen Nationalbibliothek beinhalteten, bis hin zu einer Reduktion ausschließlich auf die Titeldaten der Deutschen Nationalbibliothek ohne Einbeziehung ihrer Sondersammlungen. 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 (vgl. Abb. 3). Andererseits sind nicht alle Eigenschaften, die ein Normdatensatz beinhaltet, für die Erzeugung einer ISBD-Darstellung relevant.

Teilmenge aller beschreibenden Daten im Kontext von Kultur

URL	http://d-nb.info/1014919797/about
Titel	Der Staat (Politeia)
Autor	http://d-nb.info/gnd/118594893
Verlagsort	http://d-nb.info/gnd/2013138-0
Verleger	http://d-nb.info/gnd/16092129-6
Erscheinungsjahr	http://dbpedia.org/page/2011
Persistent Identifier	urn:nbn:de:101:1-201109033292
ISBN	978-3-8424-2088-5
Sachgruppe(n)	http://d-nb.info/ddc-sg/800

Abb. 3: Beispielhafter Eintrag der Nationalbibliografie mit bloßen Referenzen. Die bloßen Referenzen (URIs) haben im Allgemeinen keinen inhaltlichen Informationsgehalt. Um die damit verknüpften Entitäten beurteilen zu können, müssen die Referenzen aufgelöst werden

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 verwenden. 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 Datensets wie geonames.org genannt, die in einer Suche basierend auf geografischer

Normdaten kein Bestandteil der Nationalbibliografie

Lokalisierung und Nähe genutzt werden können. Die Bereitstellung einer ISBD-Beschreibung eines Kulturguts wie auch des Online-Katalogs sind zwei von vielen Diensten rund um die Nationalbibliografie. 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 in Abbildung 3 skizzierten minimalen, für sich genommen wenig brauchbaren Rumpf, bestehend aus wenigen textuellen Elementen und zahlreichen Referenzen. Mit Blick auf RDA¹⁰ und FRBR¹¹ 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 diese anderen Ebenen.

Erstellung der Nationalbibliografie:

Prozess in einem globalen »Datenökosystem«

Die Nationalbibliografie wird zunehmend Teil eines

globalen Systems aus wechselwirkenden Datenproduzenten und Datennutzern. Je stärker die Nationalbibliografie mit institutions- und domänenübergreifenden Daten verknüpft ist, 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 (vgl. in Abb. 4 die »aktiv genutzte Datenbasis« im Verhältnis zu den in »von Bibliotheken erfassten Daten«). 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.

Diese Richtung ist keineswegs neu: Bereits heute werden 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 gewonnen Daten (z.B. Kataloganreicherung, automatische Erschließung) genutzt. Konsequenter wären die Ausweitung der Datennachnutzung und die kontrollierte Öffnung der Erschließung für Communities im WWW.

In erster Linie ist es eine Chance für eine klare Fokussierung, eine Besinnung auf die eigenen Stärken und eine bessere Arbeitsteilung: jeder Anbieter kann seine Eigenschaften und Stärken einbringen, 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.

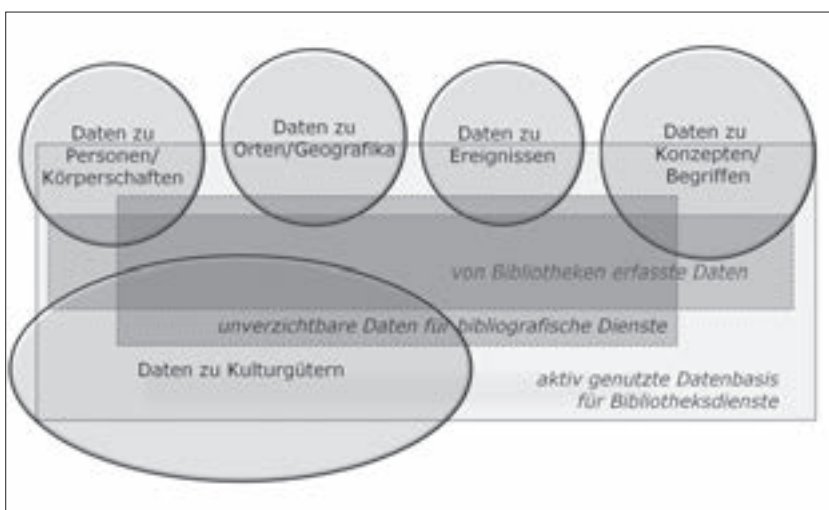


Abb. 4: Mögliche künftige Verteilung zwischen Datennutzung und Datenerfassung in Bibliotheken. 1.) Nicht alle von Bibliotheken erfassten Daten sind zwingend notwendig für bibliografische Dienste und dennoch unverzichtbar (z.B. Daten für Nutzungs- und Archivierungszwecke). 2.) Nicht alle für bibliografische Dienste nötigen Daten werden von Bibliotheken erstellt. 3.) Die von Bibliotheken genutzte Datenbasis ist weitaus größer als beide Mengen zusammengenommen

Bestimmende Qualitätsmerkmale:

Persistenz, Transparenz, Nachnutzbarkeit,

Vernetzungsgrad

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. Wie bereits im letzten Abschnitt erläutert wurde, ist gerade der Aspekt der Verlässlichkeit (Zitierbarkeit und Persistenz der Daten, Standardisierung, Qualitätssicherung) 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.

Einzuschätzen sind Fragen wie: Ist die Quelle vertrauenswürdig und fachlich kompetent? Gibt es bei der Quelle funktionierende qualitätssichernde Mechanismen? Welche der Daten, die die Quelle erstellt, benötigt sie selbst für ihr Kernanliegen? (Denn gerade für diese Daten darf man annehmen, dass sie sorgsam erstellt und auch intensiv gepflegt werden).

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.

Es kann nicht davon ausgegangen werden, dass die Anbieter Willens und/oder in der Lage sind, sich in ausreichendem Maße Kenntnisse über die Strukturierungs-, Ansatzungs- und sonstigen Regeln der Bibliothekswelt zu erarbeiten. Die bibliothekarischen Katalogisierungsregeln müssen entschlackt und auf die im WWW üblichen Strukturierungsregeln angepasst werden (konkrete Ansätze hierzu folgen im nächsten Abschnitt).

Für die Übernahme von Metadaten anderer Aggregatoren sollten vertrauenswürdige Anbieter definiert werden. Für deren Bewertung sollte nach Vorbild des Kriterienkataloges vertrauenswürdiger digitaler Langzeitarchive (Nestor AG, 2008) 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:

Orientierung an Linked-Data-Konzepten

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 (vgl. Berners-Lee, 2009). Auf diese Weise wird jeder einzelne Datensatz eine individuell benannte und zitierbare Ressource im WWW. Eine unmittelbarere Integration der Nationalbibliografie ins WWW ist kaum denkbar.

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 Diskussionen.¹² Aus der praktischen Erfahrung im Umgang mit RDF an der Deutschen Nationalbibliothek resultiert der Schluss, dass RDF für einige Anwendungsfälle, die das volle Potential der Ausdruckstärke 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:

bibliografische Recherche = Recherche im Web

Gerade auch kommerzielle Dienste wie Google, Bing und Facebook, die heute das Gesicht des Webs prägen und für die Masse der Nutzerinnen und Nutzer den Einstiegspunkt darstellen, müssen in die Lage versetzt werden, bibliografische Daten kreativ in ihr Angebot zu integrieren, damit sie den Nutzerinnen und Nutzern bei ihrer Recherche unmittelbar zugutkommen. Diese Dienste lösen Online-Kataloge und Portale als

qualitätssichernde
Prozesse etablieren

Linked Data als Vehikel
ins WWW

Datensatz muss den
Nutzer finden und nicht
umgekehrt

Was muss sich ändern?

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 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 die Benutzung wäre dies, wegen des umfassenderen Angebots, in das bibliografische Recherchen dann fachgerecht eingebettet wären, eine gute Entwicklung.

Nutzungsbedingungen: Offene Lizenz

Gerade in der Linked-Data-Community wird die Veröffentlichung von Daten unter einer offenen Lizenz (vgl. Open Knowledge Foundation, 2011) häufig als Grundvoraussetzung für deren Nachnutzbarkeit angeführt. Eine grundsätzliche Diskussion dieses Themas steht nicht im Fokus dieses Artikels. 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. 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 breiteren bzw. anderen Publikum nachgenutzt werden. Dies beinhaltet auch offene Plattformen wie beispielsweise Wikipedia oder kleine bis mittelgroße Webangebote wie Online-Zeitungen oder Online-Händler. Das Web ist in diesem Sinne ein eigenes Ökosystem, dessen Grundvoraussetzung die Verlinkung und die freie Nutzung dieser Links ist. Das volle Potential kann die Nationalbibliografie also nur entfalten, wenn die Daten auf allen Ebenen dieses Ökosystems ohne Einschränkung eingesetzt werden können.

Kreativität braucht Freiheit

neue Wege in der Erschließungspraxis

KONSEQUENZEN UND OFFENE FRAGEN

Öffnung der Nutzungsmöglichkeiten

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 Produktvielfalt (Online-Katalog, PDF-Versionen der Nationalbibliografie-Hefte, verschiedene Datendienste) rund um die Nationalbibliografie muss weiter ausgebaut werden und individueller an die eigenen Bedürfnisse angepasst werden können. Daraus ergibt sich ein Paradigmenwechsel: Aufgabe des Prinzips der Präkoordination in Nationalbibliografie-Heften zugunsten vielfältiger Möglichkeiten der Postkoordination der Daten in variablen Datensets.

Erschließung flexibilisieren und öffnen

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 Nebeneinträgen gemacht bzw. nicht gemacht werden sollen. Würde hier konsequent nicht von Nebeneinträgen, sondern von Sucheinstiegen gesprochen und würden diese unabhängig von nur intellektuell bestimmbar Randbedingungen generell gefordert, so könnten diese weitgehend maschinell generiert werden. Desweiteren sind Sortierregeln ein fester Bestandteil der RAK, obwohl sie in Online-Katalogen kaum Anwendung finden. Sortierregeln sollten zu Softwareanforderungen umformuliert, aus dem bibliothekarischen Regelwerk allerdings entfernt werden. Auf textuelle Datenelemente, die weder als Sucheinstieg geeignet sind noch für die Bestandserhaltung wichtig sind, sollte weitestgehend verzichtet werden. Regelungen für Beschreibungselemente (Aufgliederung von Fußnoten in unterschiedliche Felder, Regeln für die Schreibweise nicht suchbarer Datenelemente wie z. B. Seitenzahlen) werden entbehrlich.

Der RDA-Standard hat sich die Modernisierung der Erschließungsregeln zum Ziel gesetzt: »RDA is being developed as a new standard for resource description and access designed for the digital world« (Joint Steering Committee for Development of RDA, 2007). Auch wenn zu Recht kritisiert wird, dass auch die RDA in mancher Hinsicht noch immer zu stark in traditionellen Kategorien denkt und nicht weit genug reicht (vgl. Koyle, et al., 2007), bieten sie doch hinsichtlich ihrer Flexibilität einen deutlichen Fortschritt gegenüber älteren Regelwerken. Die Modernisierung der Erschließungsregeln ist allerdings nur eine Facette der notwendigen Veränderungen. Viel entscheidender ist die Fähigkeit, mit Daten, die außerhalb der eigenen Regelwelt erzeugt wurden, zu interagieren. Es sollten verstärkt Fremddaten genutzt und die Erschließung auch für Communities außerhalb des Bibliothekswesens geöffnet werden. Selbst der Werbetext eines Verlages kann einen sinnvollen Sucheinstieg darstellen. Voraussetzung dafür ist allerdings die allgemeine Nachvollziehbarkeit des Ursprungs von Daten sowie die Einführung von (automatischen und intellektuellen) Qualitätssicherungsmaßnahmen.

Zitierbarkeit als Qualitätskriterium etablieren

Die Zitierbarkeit von Daten sollte als Qualitätskriterium etabliert werden. Als »zitierbar« gekennzeichnete Datensätze werden niemals gelöscht und haben eine vollständige Änderungshistorie und belegen die Urheberschaft bzw. Quelle jedes einzelnen Datenelements. Die Prozesse rund um die Datenpflege müssen so angepasst werden, dass sie dies gewährleisten. (Nicht alle veröffentlichten Daten müssen dieses Kriterium erfüllen. Das wäre nicht praktikabel, denn die Konservierung von Datensätzen ist mit Aufwand verbunden.) Zur Verlässlichkeit gehört weiterhin die Transparenz darüber, aus welcher Quelle Daten stammen und nach welchen Regeln mit welchen Verfahren sie erzeugt wurden. Der Nachweis der Datenquelle ist insbesondere für den Ausbau der Fremddatennutzung und Öffnung der Erschließung für Communities außerhalb des Bibliothekswesens wichtig. Auf diese Weise bleibt transparent, woher die Daten stammen bzw. welche der Daten maschinell erzeugt wurden: eine entscheidende Voraussetzung für Qualitätssicherungsmaßnahmen.

Vernetzungsgrad erhöhen

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 Konzep-

ten, die in seinem Kontext eine Rolle spielen, verknüpft sein.

Damit dies besser gelingt, müssen die Normdaten ausgebaut werden: eine internationale, kooperativ 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, derartige Einträge in ihr Angebot zu integrieren.

AKTIVITÄTEN DER DEUTSCHEN NATIONALBIBLIOTHEK

Die Deutsche Nationalbibliothek ist (wie andere Nationalbibliotheken auch¹³) 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 (vgl. Deutsche Nationalbibliothek, 2011). Ausgangspunkt waren die Vorarbeiten zur Gemeinsamen Normdatei.

Es folgten weitere Datensets wie DDC-Deutsch und zuletzt im Januar 2012 eine erste Veröffentlichung der Nationalbibliografie als Linked-Data. Bis auf das Dataset DDC-Deutsch, für das spezielle Lizenzvereinbarungen mit OCLC gelten, stehen alle Datensets unter einer offenen Lizenz (CCo) zur Verfügung. Eine schrittweise Ausweitung dieser Regelung auf alle anderen Repräsentationsformen der Deutschen Nationalbibliografie (zum Beispiel MARC21) ist in Arbeit: in einem ersten Schritt werden ältere Daten kostenlos zur Verfügung gestellt, mittelfristig alle Daten der Nationalbibliografie.

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.

Die Erhöhung des Vernetzungsgrades wird in verschiedenen Bereichen vorangetrieben. Als guter Startpunkt haben sich Optimierungsarbeiten am eigenen Datenbestand erwiesen: Die Umwandlung textueller Beschreibungen in Referenzen ist in einigen Fällen größtenteils automatisierbar. Einen noch experimentellen Status hat der Ansatz, über Named-Entity-Recognition-Verfahren auch Nennungen von Entitäten im

RDA als flexibles Regelwerk

Linked Open Data Service

Datenvernetzung und Semantic Web

Text automatisch zu erkennen und diese mit Normdateneinträgen zu verknüpfen (vgl. CONTENTUS Project, 2011). Ein wichtiger Aspekt ist die Vernetzung der eigenen Daten mit externen Datentöpfen. Auch hier setzt die Deutsche Nationalbibliothek 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 SWD über ein redaktionelles Verfahren mit RAMEAU und LCSH. Eine Kooperation mit der deutschen Wikipedia brachte die Vernetzung zwischen Wikipedia-Artikeln und der PND. Desweiteren wird im Rahmen des Projektes CONTENTUS ein Verfahren zur Vernetzung der Gebietskörperschaften und Schlagworte mit den entsprechenden Wikipedia-Artikeln und dem Datenset geonames.org entwickelt.

Um derartige Vernetzungsaktivitäten nachhaltig voranzutreiben, wurde die Plattform culturegraph.org gegründet. Die Idee dahinter ist der Aufbau einer technischen und organisatorischen Infrastruktur, die gezielt die Vernetzung bibliografischer Daten, Thesauri, Klassifikationen und Normdaten unterstützt und die gewonnen Verknüpfungen allgemein (u. a. als Linked Open Data) zugänglich macht. Eine Kerneigenschaft der Plattform ist die Unterstützung von Zitierbarkeit (im Sinne des letzten Abschnitts). Einer der nächsten Schritte ist die Öffnung der Plattform für andere Domänen (im Kontext der Deutschen Digitalen Bibliothek¹⁴) und interessierte Web-Communities. Die Plattform startete Ende 2010 mit einem von der Arbeitsgemeinschaft der Verbundsysteme beauftragten Projekt: Ziel ist der Aufbau eines »Resolving- und Lookup-Dienstes für bibliografische Identifier«¹⁵. Hierzu werden die Daten aller Verbundpartner zu Veröffentlichungen seit 1945 über Ähnlichkeitsberechnungen geclustert. Auf diese Weise sollen Datensätze, die die gleiche Manifestation beschreiben, zusammengeführt und mit einem gemeinsamen Identifier versehen werden. Desweiteren können auch andere Cluster für allgemeinere Ähnlichkeitsebenen (z. B. Werksebene) gebildet werden. Nach Projektende gilt es zu analysieren, ob sich auf dieser Basis eine nationale Werksdatei bilden lässt.

Alle nationalen Aktivitäten müssen sich allerdings in einen internationalen Kontext einordnen. Die daher vielleicht wichtigste Aktivität bleibt das Engagement in internationalen Projekten und Gremien, die die Zukunft der Erschließungspraxis und des Datenaustauschs bestimmen.

FAZIT

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 1990er Jahre hat das Bibliothekswesen noch nicht ganz verdaut: Noch immer sind Erschließungspraxis und Produkte nicht konsequent an die Nutzungsszenarien des digitalen Zeitalters angepasst. Und 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 zuallererst mehr Flexibilität und Offenheit.

LITERATUR

Berners-Lee, Tim: Linked Data – Design Issues. [Online] 2009. www.w3.org/DesignIssues/LinkedData.html [Zitat vom: 01.10.2011].

Blum, Rudolf: Nationalbibliographie und Nationalbibliothek. In: Archiv für Geschichte des Buchwesens. 1990, Bd. 35, S. 1–294.

CONTENTUS Project. CONTENTUS Fact Sheet Collection. [Online] 2011. www.contentus-projekt.de/fileadmin/media/factsheets/CONTENTUS-FactSheet-Collection2011_v1-01-web.pdf

Deutsche Nationalbibliothek. Der Linked Data Service der Deutschen Nationalbibliothek – Version 3.1. 2011. http://files.d-nb.de/pdf/linked_data.pdf

Deutsche Nationalbibliothek. Sammelrichtlinien – Stand: 1. Juni 2009. [Online] 2009. <http://nbn-resolving.de/urn:nbn:de:101-2009033003>

DNB; hbz. 2011. culturegraph.org. [Online] Deutsche Nationalbibliothek, 2011. www.culturegraph.org/ [Zitat vom: 20.10.2011].

Google Inc. 2011. Über die Google Buchsuche. [Online] 2011. <http://books.google.com/intl/de/googlebooks/about.html> [Zitat vom: 20.10.2011].

Grimmer, Roland; Junker, Christa: Geschichte der gedruckten Deutschen Nationalbibliographie. In: Dialog mit Bibliotheken 14 (2002) 3, S. 10–20.

IFLA Study Group on the Functional Requirements for Bibliographic Records. 2009. Funktionale Anforderungen an bibliografische Datensätze. [Online] 2009. <http://nbn-resolving.de/urn:nbn:de:101-2009022600>. [urn:nbn:de:101-2009022600](http://nbn-resolving.de/urn:nbn:de:101-2009022600)

IFLA Working Group on Guidelines for National Bibliographies. 2008. Guidelines for National Bibliographies in the Electronic Age – Draft. [Online] 2008. <http://archive.ifla.org/VII/s12/guidelines-national-bibliographies-electronic-age.pdf>

Joint Steering Committee for Development of RDA. 2007. Strategic Plan for RDA 2005–2009. [Online] November 2007. www.rda-jsc.org/stratplan.html

Koyle, Karen; Hillman, Diane: Resource Description and Access (RDA) – Cataloging Rules for the 20th Century. In: D-Lib Magazine. Februar 2007, Volume 13, Number 1/2.

Kummer, Dietmar: Bibliothekarisches Studium in Vergangenheit und Gegenwart : Festschrift aus Anlass des 80jährigen Bestehens der bibliothekarischen Ausbildung in Leipzig im Oktober 1994. [Hrsg.] Engelbert Plassmann und Dietmar Kummer. Zeitschrift für Bibliothekswesen und Bibliographie : Sonderhefte. 1995, Bd. 62, S. 122.

Nestor. Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. [online] 2008. <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>

Open Knowledge Foundation. 2011. Open Definition – Version 1.1. [Online] 2011. <http://opendefinition.org/okd/> [Zitat vom: 20.10.2011].

Schuster, Michael; Rappold, Dieter: Social Semantic Software – was soziale Dynamik im Semantic Web auslöst. [Hrsg.] Tassilo Pellegrini. Semantic Web: Wege zur vernetzten Wissensgesellschaft. 2006.

W3C. 2011. SWEO Community Project: Linking Open Data on the Semantic Web. [Online] 2011. <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData> [Zitat vom: 13.10.2011].

¹ Homepage: <http://openlibrary.org/>

² www.vd16.de/ und www.vd17.de/

³ <http://vd18-proto.bibliothek.uni-halle.de/>

⁴ www.inka.uni-tuebingen.de/

⁵ Mehr Informationen zu Webservices finden sich unter www.w3.org/2002/ws/Activity und www.w3.org/TR/soap12-part1

⁶ Einen Überblick über die aktuellen Stand der LOD-Cloud bietet www4.wiwiss.fu-berlin.de/locloud/state/

⁷ Siehe Web-Angebote wie www.profi-ranking.de und www.iseo-gmbh.de/ u.a.

⁸ 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).

⁹ Bzw. von den für die retrospektive Nationalbibliographie verantwortlichen Portalen (INKA, VD16, VD17 ...).

¹⁰ Siehe RDA-Homepage für weitere Informationen zu RDA: www.rda-jsc.org/

¹¹ Siehe IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009.

¹² <http://milicivuk.com/blog/2011/07/19/ultimate-problem-of-rdf-and-semantic-web/> und www.w3.org/2001/sw/wiki/index.php?title=RDF_Core_Work_Items&oldid=1980#Blank_Nodes

¹³ Übersicht bibliografischer Linked-Data: <http://thedatahub.org/group/bibliographic/>

¹⁴ Siehe Projekt-Homepage: www.ddb.de

¹⁵ Siehe Homepage: <http://culturegraph.org>

DIE VERFASSER

Jürgen Kett ist Stellvertretender Abteilungsleiter der Abteilung Informationstechnik, Deutsche Nationalbibliothek, Adickesallee 1, 60322 Frankfurt am Main, E-Mail: j.kett@dnb.de

Mathias Manecke, Deutsche Nationalbibliothek, Deutsches Buch- und Schriftmuseum, 04103 Leipzig, E-Mail: m.manecke@dnb.de

Sarah Beyer, Deutsche Nationalbibliothek, Abteilung Digitale Dienste, Adickesallee 1, 60322 Frankfurt am Main, E-Mail: s.beyer@dnb.de