
Quo vadis Inhaltserschließung der Deutschen Nationalbibliothek?

Herausforderungen und Perspektiven¹

Ulrike Junger, Deutsche Nationalbibliothek

Zusammenfassung

Seit der Ausweitung des Pflichtexemplarrechts auf sog. unkörperliche Medienwerke im Jahr 2006 steht die Deutsche Nationalbibliothek (DNB) vor der Herausforderung, über neue Wege zur Erzeugung bibliografischer Metadaten nachzudenken. Die steigende Menge an Publikationen, die es zu bearbeiten gilt, aber auch deren zunehmende Diversität machen dies erforderlich. Für die Inhaltserschließung der DNB bedeutet das, dass andere Verfahren neben die intellektuelle Erschließung treten müssen. Die Entwicklung und Einführung automatischer Verfahren für Klassifikation und Beschlagnwortung gehören ebenso dazu wie die Nutzung von Fremddaten oder die Gewinnung inhaltserschließender Daten über Konkordanzen. Dies hat Auswirkungen auf Arbeitsabläufe, Erschließungsinstrumente und -standards sowie Datenformate. Die DNB strebt ein Konzept für die Inhaltserschließung an, das zum Ziel hat, so viele Publikationen als möglich angemessen zu erschließen, um thematische Recherchen erfolgreich zu unterstützen.

Summary

Since the extension of the legal deposit mandate to so-called non-physical publications in 2006, the Deutsche Nationalbibliothek (DNB) has to deal with the challenge of finding new ways to create bibliographic metadata. This is due to the growing amount of publications and their increasing diversity. Consequently, in the area of subject cataloguing in DNB, intellectual cataloguing must be complemented by additional methods. These include the development and implementation of automated procedures for classification and indexing as well as the use of third party data or the production of subject data via mappings. All this has effects on the workflows, tools and standards for subject cataloguing and data formats. DNB aims at a concept for subject cataloguing which will make it possible to catalogue as many publications as possible in an appropriate way in order to successfully support topical searches.

Zitierfähiger Link (DOI): <http://dx.doi.org/10.5282/o-bib/2015H1S15-26>

1. Ausgangslage

Die bibliothekarische Erschließung befindet sich in einer Umbruchphase. Darauf ist in den vergangenen Jahren oft hingewiesen worden, auch auf die Gründe dafür, für die nur zwei der wichtigsten Stichworte genannt werden sollen: Digitalisierung und Ressourcenprobleme. Auch und gerade die Deutsche Nationalbibliothek (DNB) steht vor entsprechenden Fragestellungen, mit denen sie sich

¹ Dieser Artikel fußt auf einem Vortrag, der auf dem 103. Deutschen Bibliothekartag 2014 in Bremen gehalten wurde.

für sich als Institution, aber auch mit den bibliothekarischen Partnern, auseinandersetzen muss. Während die Formalerschließung, unter anderem wegen des bevorstehenden Umstiegs auf das internationale Regelwerk Resource Description and Access (RDA), in der Regel im Fokus steht, muss auch das traditionell zweite Standbein bibliothekarischer Erschließung, die Inhaltserschließung, in den Blick genommen werden. Die DNB hat dies in den vergangenen Jahren immer wieder getan und an einem Konzept für eine zukunftsfähige Inhaltserschließung gearbeitet, dessen Grundzüge im Folgenden skizziert werden.

Es gibt eine Reihe von Faktoren und Rahmenbedingungen in der gegenwärtigen bibliothekarischen Landschaft, die eine Auswirkung auf die Inhaltserschließung in der DNB bzw. für das Thema insgesamt haben.

Für die DNB ist – wie so oft – zunächst der erweiterte Sammelauftrag zu nennen.

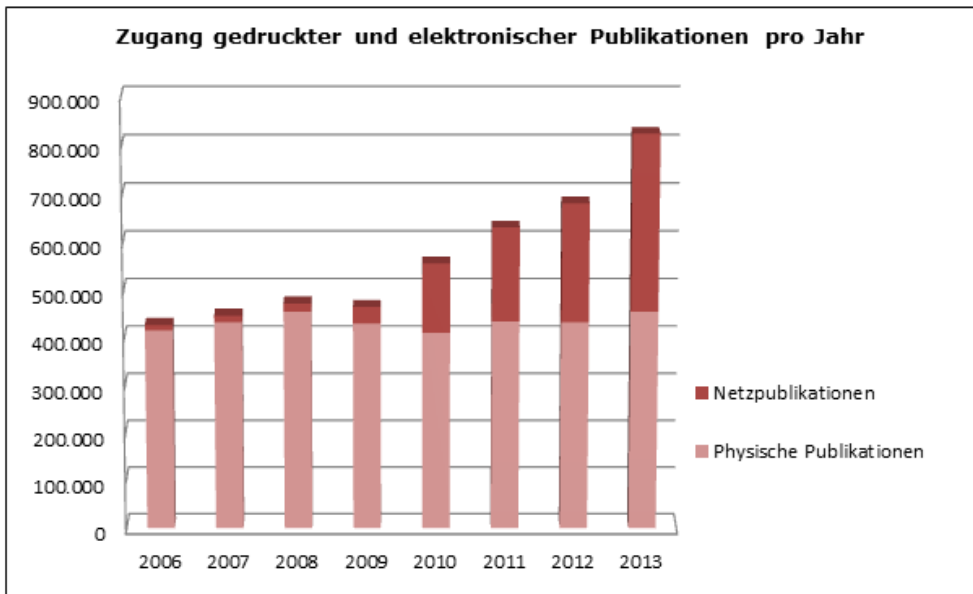


Abb. 1: Entwicklung des Zugangs in der Deutschen Nationalbibliothek

Acht Jahre sind seit der Novellierung des Gesetzes zur Deutschen Nationalbibliothek² vergangen, und die DNB unternimmt große Anstrengungen, den erweiterten Sammelauftrag umzusetzen, das heißt ihre Sammlung von sog. Netzpublikationen (alle Arten von nichtkörperlichen Medienwerken, einschließlich Webseiten) auf- und auszubauen. Über 1.000 digitale Objekte werden derzeit pro Tag eingesammelt und müssen verarbeitet werden. Darunter sind auch Netzpublikationen, die keine gedruckte Parallelausgabe mehr haben oder bei denen die gedruckte Ausgabe nicht mehr

2 Deutsche Nationalbibliothek: Rechtliche Grundlagen. http://www.dnb.de/DE/Wir/Recht/recht_node.html (14.12.2014).

gesammelt wird (bei Hochschulschriften, die nicht in Reihe A erscheinen). Zweierlei lässt sich feststellen: Zum einen ist bisher kein wesentlicher Rückgang bei den physischen Veröffentlichungen festzustellen, zum anderen hat die Sammlung digitaler Objekte noch lange nicht ihr Limit erreicht.

Die Erweiterung des gesetzlichen Auftrags hat jedoch nicht nur Folgen für die Sammlung, sondern auch für die anderen Aufgaben, die der DNB obliegen, so für die bibliografische Verzeichnung und Archivierung. Auch da stellen Netzpublikationen eine Herausforderung dar, nicht allein wegen der schieren Menge, sondern auch wegen der unterschiedlichen Formate, die zu verarbeiten sind, aufgrund der multiplen Ablieferungswege und nicht zuletzt natürlich, weil sie ganz neue Fragen hinsichtlich des Sammelprofils stellen. Freilich bieten Netzpublikationen für die Erschließung auch Möglichkeiten, die physische Medienwerke nicht haben: Sie liegen digital vor, bringen oft ihre eigenen Metadaten mit, und die Verfügbarkeit des elektronischen Volltextes macht sie einer maschinellen Erschließung zugänglich. Das sind Gründe, die die DNB veranlasst haben, die intellektuelle Erschließung von Netzpublikationen im Jahr 2010 weitgehend einzustellen (eine Ausnahme ist bisher die Erschließung von fortlaufenden elektronischen Sammelwerken in der Zeitschriftendatenbank (ZDB)). Die Nachweise für die Deutsche Nationalbibliografie fußen auf den abgelieferten Metadaten. Allerdings gehören inhaltserschließende Daten (bis auf Sachgruppen) i.d.R. nicht dazu.

Darüber hinaus gibt es aber auch andere Einflussfaktoren: Zu nennen ist die bevorstehende Einführung der RDA. Es zeichnet sich ab, dass RDA auch Aussagen zu einer zumindest weitgefassten Beschreibung der Inhalte von Werken enthalten wird (sog. high level subject relationship³). Auch die in der deutschsprachigen Erschließungstradition bisher starke Trennung zwischen Formal- und Inhaltserschließung, die sich in vielen Bibliotheken auch in einer entsprechenden Arbeitsorganisation widerspiegelt, wird durch RDA aufgeweicht. Der Anfang ist gemacht mit einheitlichen Regeln für die Normierung von Personen und Körperschaften, aber auch Werken, und der Erfassung von Beschreibungselementen wie *Inhaltstyp*, der zumindest einen Teil der bisherigen Formschlagwörter ersetzt, wie sie in der Anlage 6 der Regeln für den Schlagwortkatalog (RSWK⁴) enthalten sind. Auch die Ankunft neuer Webtechnologien wie Linked Data stellt Fragen an die Inhaltserschließung: Welche Art von Daten werden benötigt, um im Web-of-Data präsent und relevant zu sein?

Zuletzt soll es hier noch um ein Thema gehen, das sicher nicht nur die DNB beschäftigt: Bis 2016 werden etwa 10 % des jetzigen Personals der Abteilung Inhaltserschließung in den Ruhestand gehen; in den vergangenen drei Jahren waren es bereits 10 %. Bei insgesamt gleichbleibenden Personalressourcen, aber sich verändernden Anforderungen, sind Verlagerungen von Stellen und Umschichtungen in veränderte Aufgabenfelder innerhalb der Bibliothek erforderlich.

3 Siehe Joint Steering Committee for Development of RDA: Outcomes of the 2014 JSC meeting. <http://www.rda-jsc.org/2014JSCmeetingoutcomes.html> (14.12.2014).

4 Die elektronische Form der RSWK (3. Aufl., Stand: 7. Erg.-Lief., Mai 2010) ist abrufbar über <http://nbn-resolving.de/urn:nbn:de:101-2012053100>.

All dies ist Anlass und Grund, darüber nachzudenken, wie Inhaltserschließung zukunftsfähig gemacht werden kann. Denn die Lösung kann nicht sein, lediglich immer weniger Publikationen inhaltlich zu erschließen.

2. Derzeitige Erschließungspraxis in der DNB

Es sei kurz rekapituliert, welche inhaltserschließenden Daten derzeit in der DNB erzeugt werden.

Die Inhaltserschließung beruht seit der Einführung der Dewey-Dezimalklassifikation (DDC)⁵ im Jahr 2006 auf einem *Drei-Säulen-Modell*: DDC-Sachgruppen als klassifikatorisches Gliederungselement für die Deutsche Nationalbibliografie⁶ sowie DDC-Notationen und Schlagwortfolgen nach RSWK als präzisere Form der Inhaltserschließung. Hinzu kommen seit 2012 sog. Gattungsbegriffe⁷ für Belletristik und Kinder- und Jugendbücher.

Diese Erschließungsarten finden in unterschiedlicher Kombination Anwendung in der intellektuellen Inhaltserschließung.

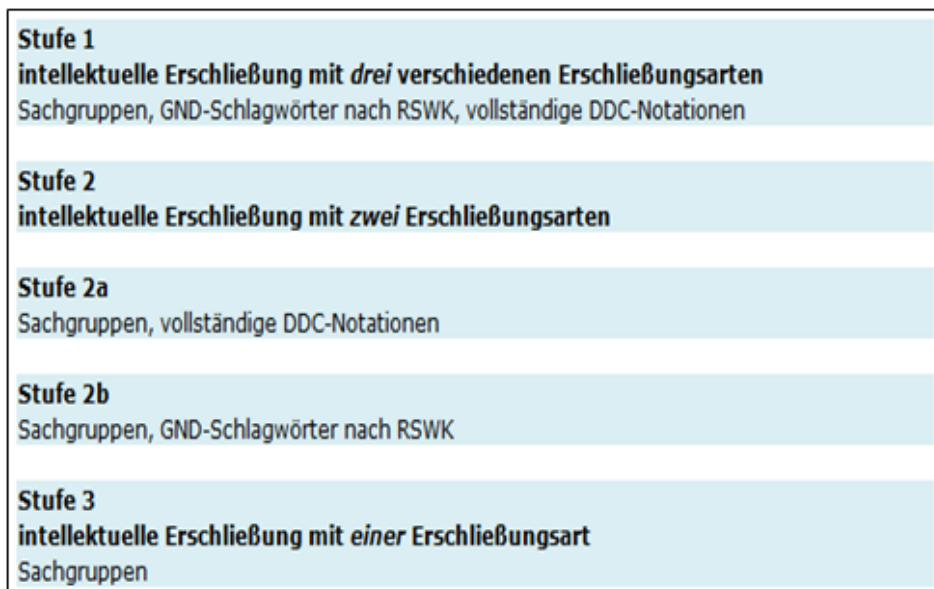


Abb. 2: Stufenschema der intellektuellen Inhaltserschließung der DNB

5 Dewey-Dezimalklassifikation. http://www.ddc-deutsch.de/Subsites/ddcdeutsch/DE/Home/home_node.html (14.12.2014).

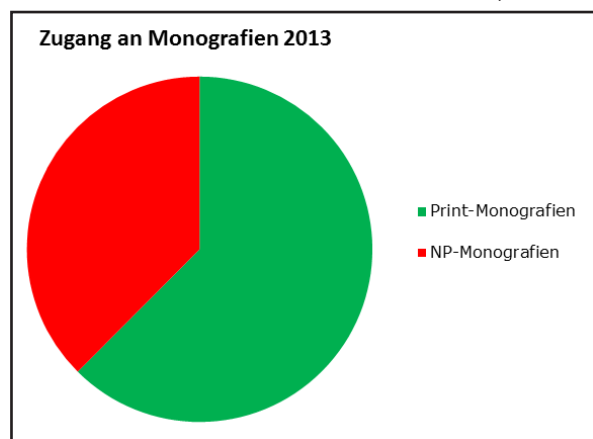
6 Gliederung der Deutschen Nationalbibliografie und des Neuerscheinungsdienstes. <http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/gliederungDnb.html> (14.12.2014).

7 Liste der in der DNB verwendeten Gattungsbegriffe. <http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/listeGattungsbegriffe.pdf> (14.12.2014).

Da die Deutsche Nationalbibliografie nach DDC-Sachgruppen gegliedert ist, müssen alle dort angezeigten Medienwerke eine Sachgruppe erhalten. DDC-Notationen und Schlagwortfolgen nach RSWK mit Vokabular der Gemeinsamen Normdatei (GND) werden in abgestufter Weise vergeben. Dabei werden nur Medienwerke erschlossen, die in den Reihen A, B und H angezeigt werden. In Stufe 1 finden sich z.B. die meisten Verlagspublikationen. Stufe 2a gilt für Zeitschriften, die meisten gedruckten Nichtverlagspublikationen sowie Dissertationen und Verlagspublikationen, die von eher nicht-wissenschaftlichem Interesse sind (z.B. Ratgeber, Hobbyliteratur, Erlebnisberichte). Nach Stufe 2b werden u.a. Kinderbücher, ursprünglich in deutscher Sprache verfasste Belletristik oder auch Comics und Cartoons erschlossen. Stufe 3 findet für eine Vielzahl an Publikationsgruppen Anwendung, so z.B. Musikalien und Tonträger, Schulbücher, fremdsprachige Literatur (Belletristik), auch in deutscher Übersetzung, Medienwerke unter 50 Seiten u.a.m.

Auf der Basis dieses Stufenschemas wurden im Jahr 2013 für 160.000 Publikationen Sachgruppen vergeben. Ca. 100.000 Medienwerke wurden darüber hinaus tiefer, klassifikatorisch und/oder verbal, erschlossen.

Wie oben erwähnt, besteht ein wachsender Anteil des Zugangs der DNB aus Netzpublikationen. Da diese nicht mehr intellektuell erschlossen werden, heißt das, dass ein wachsender Anteil der Sammlung keine intellektuell vergebenen Inhaltserschließenden Daten erhält, da die Medienwerke nicht mehr auf den Tisch bzw. den Bildschirm eines Inhaltserschließers kommen.



In Abbildung 3 ist der Zugang an gedruckten und elektronischen Monografien und Hochschulschriften für das Jahr 2013 zusammengefasst dargestellt. Die Printwerke sind alle in der einen oder anderen Weise durch die DNB inhaltlich intellektuell erschlossen, die Netzpublikationen nicht.

Das ist nicht befriedigend, weder für diejenigen, die im Portal der DNB nach Literatur recherchieren noch für die Institutionen, die die Nachweisdaten der DNB nachnutzen, oder für die Mitarbeiterinnen und Mitarbeiter. Dies gilt vor allem angesichts der Tatsache, dass der Bereich Netzpublikationen ein Wachstumsbereich ist, der nicht einfach ignoriert werden kann. Darüber hinaus haben Netzpublikationen für Forschung und Lehre eine immer größere Bedeutung gewonnen, wie beispielsweise die Verschiebung in den Erwerbungssetats zeigt.⁸

Abb. 3: Zugang an Monografien im Jahr 2013

Das ist nicht befriedigend, weder für diejenigen, die im Portal der DNB nach Literatur recherchieren noch für die Institutionen, die die Nachweisdaten der DNB nachnutzen, oder für die Mitarbeiterinnen und Mitarbeiter. Dies gilt vor allem angesichts der Tatsache, dass der Bereich Netzpublikationen ein Wachstumsbereich ist, der nicht einfach ignoriert werden kann. Darüber hinaus haben Netzpublikationen für Forschung und Lehre eine immer größere Bedeutung gewonnen, wie beispielsweise die Verschiebung in den Erwerbungssetats zeigt.⁸

8 Siehe dazu: Bibliotheken zählen. Berichtsjahr 2013. Auszüge aus der Deutschen Bibliotheksstatistik. http://www.hbz-nrw.de/dokumentencenter/produkte/dbs/aktuell/Datenposter/datenposter2013_web.pdf (14.12.2014).

Insgesamt hat sich die Inhaltserschließung in der DNB folgendes Ziel gesetzt: Es sollen so viele Publikationen wie möglich angemessen erschlossen werden.

Damit ist zweierlei impliziert: Es geht um Quantität (Erschließungsmenge) und Qualität (Angemessenheit der Erschließung). Beides hängt untrennbar zusammen und muss dennoch gegeneinander abgewogen werden. Um das quantitative Ziel zu erreichen, muss über veränderte Wege und Methoden nachgedacht werden. Die Frage der Angemessenheit berührt die Erschließungsstandards.

3. Weiterentwicklung des Erschließungskonzeptes

Die oben dargestellte Situation verlangt eine Weiterentwicklung des Erschließungskonzeptes der DNB. Das Drei-Säulen-Modell der intellektuellen Inhaltserschließung ist nicht mehr ausreichend, auch wenn das erwähnte Stufenkonzept der Inhaltserschließung bereits versucht, eine Antwort darauf zu geben, dass einerseits begrenzte Ressourcen zur Verfügung stehen, es andererseits aber auch nicht notwendig und sinnvoll ist, alle Publikationen dreifach inhaltlich zu erschließen.

Die Weiterentwicklung des Inhaltserschließungskonzeptes hält prinzipiell an den beiden Wegen – klassifikatorische und verbale Inhaltserschließung – fest. Klassifikationen und kontrollierte Vokabulare haben sich über viele Jahrzehnte als Mittel zur Wissensorganisation und Informationsverdichtung bewährt. Sie ermöglichen es auch im Kontext von Suchmaschinen und Discovery Systemen, die thematische Recherche zu unterstützen und sachlich Zusammengehöriges auffindbar zu machen. Indem beide Instrumente ins Zentrum gestellt werden, kann eine Einheitlichkeit der Sucheinstiege erreicht werden, unabhängig vom Verfahren, mit dem verbale oder klassifikatorische Daten erzeugt wurden. Darüber hinaus weisen DDC und GND Eigenschaften auf, die sie für die Anwendung im Web nützlich machen (z.B. Identifier, Strukturierung als Entitäten).

Hinsichtlich der Verfahren zur Erzeugung verbaler und klassifikatorischer Daten sollen neben der intellektuellen Erschließung in steigendem Umfang andere Formen der Datengewinnung eingesetzt werden. Das sind einerseits maschinelle oder automatische Erschließungsverfahren, andererseits das Datenrecycling, d.h. die verbesserte Nutzung bereits vorhandener Inhaltserschließungsdaten. Insbesondere intellektuell erzeugte Daten sind teuer und sollen möglichst gewinnbringend genutzt werden.

In den vergangenen Jahren wurde immer wieder über das Projekt PETRUS⁹ berichtet. Es wurde angestoßen, um Erschließungsvorgänge zu automatisieren, nicht nur für die Inhaltserschließung. Ziel ist es, solche Vorgänge zu Standardverfahren zu machen. In der Inhaltserschließung geht es dabei sowohl um die Klassifikation als auch um die Schlagwortvergabe.

9 Für einen Überblick siehe PETRUS - Prozessunterstützende Software für die digitale Deutsche Nationalbibliothek. <http://www.dnb.de/DE/Wir/Projekte/Abgeschlossen/petrus.html> (14.12.2014).

3.1 Maschinelle Sachgruppenvergabe

Wie erwähnt, sollen alle Publikationen, die in der Nationalbibliografie verzeichnet werden, eine DDC-Sachgruppe erhalten. Das ist auch für die Netzpublikationen der Fall. Beim Import einer Netzpublikation wird eine vom Ablieferer vergebene Sachgruppe übernommen oder eine Platzhaltersachgruppe erzeugt, beides ist jedoch oftmals inhaltlich nicht zutreffend. Das Verfahren zur maschinellen Sachgruppenvergabe, das zur Verbesserung dieser Situation entwickelt wurde, ist seit Anfang 2013 in Routine und wird weiterhin laufend optimiert und auf weitere Publikationstypen ausgeweitet.

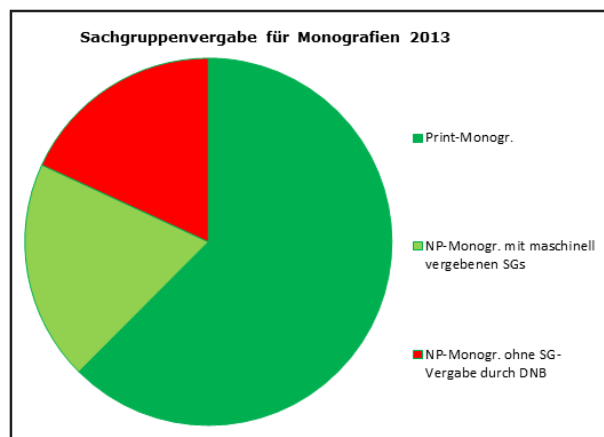


Abb. 4: Sachgruppenvergabe für Monografien im Jahr 2013

Wie Abbildung 4 zu entnehmen ist, wurden 2013 für ca. 95.000 Publikationen maschinell Sachgruppen vergeben und der Anteil der von der DNB erschlossenen Titel damit erheblich erhöht. Die Daten werden für die Nachnutzung ausgeliefert. Der Anteil der Publikationen, für die maschinell keine Sachgruppen vergeben werden konnten, resultiert daraus, dass hier das Verfahren noch nicht funktioniert, z.B. weil es Probleme mit den Formaten der abgelieferten Netzpublikationen gibt. Maschinelle Verarbeitungsketten müssen daran jeweils angepasst werden.

Darüber hinaus wird derzeit an einem Verfahren zur Vergabe medizinischer Kurznotationen für entsprechende Hochschulschriften gearbeitet, das allerdings erst im Jahr 2015 einem Produktivtest unterzogen werden kann.

3.2 Maschinelle Schlagwortvergabe

Die maschinelle Vergabe von Schlagwörtern aus der GND befindet sich derzeit in einem Produktivtest für elektronische Hochschulschriften. Für Einzelheiten zum Verfahren kann auf frühere Vorträge und Veröffentlichungen verwiesen werden.¹⁰ Etwa 20 bis 30 Werke werden derzeit täglich verarbeitet, mit zufriedenstellenden Ergebnissen. Eine spätere Auslieferung der Daten ist geplant, klar gekennzeichnet als maschinell erzeugt.

¹⁰ Z.B. Uhlmann, Sandro: Automatische Beschlagwortung von deutschsprachigen Netzpublikationen mit dem Vokabular der Gemeinsamen Normdatei (GND). In: *Dialog mit Bibliotheken* 2 (2013), S. 26-36; abrufbar unter <http://nbn-resolving.de/urn:nbn:de:101-20140305238>.

```

2242 |o|863706993
3000 |143943995|Hoge, Anne-Christin
3260 Hochbegabte Kinder und Jugendliche in der Verhaltens-Therapie
4000 Hochbegabte Kinder und Jugendliche in der (Verhaltens-)Therapie [[Elektronische Ressource]] : eine Analyse von Therapieverläufen / Anne-Christin Hoge
4022 #ont[Online-Ausg.]
4030 Münster : Verl.-Haus Monsenstein und Vannerdat
4048 Münster : Univ.- und Landesbibliothek
4060 258 S.
4061 Ill., graph. Darst.
4062 21 cm
4068 Online-Ressource
4083 =A $
4085 "HTTP"=q pdf=u http://miami.uni-muenster.de/servlets/DerivateServlet/Derivate-5942.html#x H=z LF
4180 #16#1010931636|Universität Münster: [Wissenschaftliche Schriften der WWU Münster / 6] ; Bd. 6
4190 MV Wissenschaft
4204 Zugl.: Münster (Westfalen), Univ., Diss., 2010
4233 "a"
4237 Online-Ausg.:
4700 |FE|Zei
5050 370
5051 $LS_20140304_dnbPS14_de
5540 [GND]104128223X|Hochbegabung$K0,084$D2014-03-18
5540 [GND]1040628744|Verhaltenstherapie$K0,044$D2014-03-18
5540 [GND]1040305503|Kind$K0,024$D2014-03-18
5540 [GND]1040597989|Therapie$K0,023$D2014-03-18
5540 [GND]1042825202|Verlauf$K0,012$D2014-03-18
[0101 ] leipzig dnb <101a>
7001 08-08-11 : x
7133 b
7800 212606018
    
```

Abb. 5: Beispiel für eine Online-Hochschulschrift mit maschinell erzeugten Schlagwörtern

Das Verfahren soll in den kommenden Jahren schrittweise auf weitere Publikationsgruppen ausgeweitet werden. Durch den Einsatz der maschinellen Erschließungsverfahren erhält die verbale Inhaltserschließung wieder ein stärkeres Gewicht, auch im Hinblick auf die Menge der mit Schlagwörtern erschlossenen Titel. In der intellektuellen Inhaltserschließung ist derzeit die klassifikatorische Erschließung mit der DDC das wichtigste Mittel, zumindest was die Anzahl der damit erschlossenen Titel betrifft.

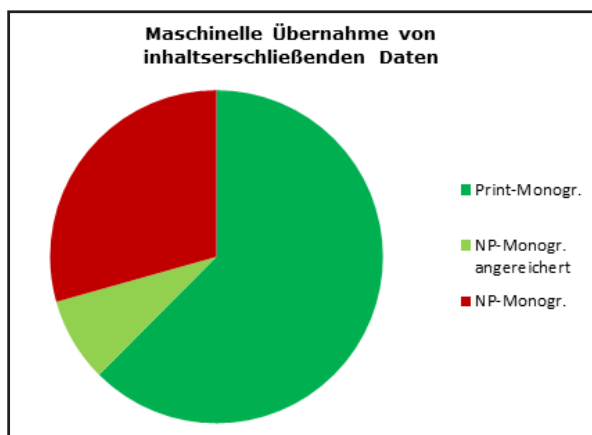
3.3 Wiedernutzung von Inhaltserschließungsdaten

Zum Thema Datenökonomie lässt sich sagen, dass die Nachnutzung von Erschließungsdaten, die von anderen erzeugt wurden, für die meisten anderen Bibliotheken schon lange eine Selbstverständlichkeit ist. Auch wenn die DNB ebenfalls z.B. Formalerschließungsdaten des Buchhandels oder von Ablieferern nutzt, so ist die systematische Nachnutzung von Daten in der Inhaltserschließung noch unterentwickelt.

Grundsätzlich lässt sich die Nutzung von Daten Dritter von der Mehrfach- oder Wiedernutzung selbst erzeugter Daten unterscheiden. Zu Ersterem gehört neben der direkten Übernahme von Daten Dritter wie z.B. anderer bibliothekarischer Einrichtungen auch die Gewinnung von Daten über Konkordanzen, zu Letzterem gehört die Übernahme von Daten aus Nachweisen für Parallelausgaben,

Vorauslagen etc. oder die Gewinnung von Daten ebenfalls über Konkordanzen oder die Verknüpfung unterschiedlicher Erschließungssysteme. Die DNB versucht, alle hier skizzierten Wege zu gehen. Ein Beispiel für die Übernahme von Schlagwörtern, die andere Bibliotheken erstellt haben, ist die Anreicherung von wirtschaftswissenschaftlichen Publikationen der Reihe B (die von der DNB nicht beschlagwortet werden) mit GND-Schlagwörtern. Diese werden über eine Konkordanz zum Standardthesaurus Wirtschaft der Zentralbibliothek für Wirtschaftswissenschaften in Kiel gewonnen.¹¹ Weiterhin sei ein Vorhaben genannt, das die DNB derzeit durchführt: Ziel ist es, die über Jahre im Projekt CrissCross¹² angelegten Verknüpfungen zwischen Schlagwörtern und DDC-Notationen für die Recherche fruchtbar zu machen, d.h. verbale Sucheinstiege zu klassifikatorisch erschlossenen Titeln zu gewinnen.

Daten des Buchhandels werden von der DNB ebenfalls schon sehr erfolgreich nachgenutzt. Seit einiger Zeit werden für Belletristik, Kinder- und Jugendbücher sog. Gattungsbegriffe vergeben. Diese werden zu ca. 90 % aus den Daten des Marketing- und Verlagsservice des Buchhandels (MVB) übernommen. Der MVB plant, in Zukunft ein differenziertes System namens Panthema zur thematischen Einordnung von Medienwerken einzusetzen. Es wird auf jeden Fall untersucht werden, ob sich daraus automatisch Inhaltserschließungsdaten gewinnen lassen.



Ein weiteres Beispiel dafür ist die maschinelle Übernahme von Inhaltserschließungsdaten aus Parallelausgaben, in der Regel von der intellektuell erschlossenen, gedruckten Ausgabe auf die Netzpublikation. Wie Abbildung 6 zeigt, konnten im Jahr 2013 immerhin rund 40.000 Titel auf diese Weise mit Schlagwortfolgen und/oder DDC-Notationen angereichert werden.

Abb. 6: Maschinelle Übernahme von Inhaltserschließungsdaten

3.4 Die nächste Version des gestuften Erschließungskonzeptes

Das oben beschriebene gestufte Erschließungskonzept konnte und musste um zwei Stufen erweitert werden.

11 Siehe Crosskonkordanz STW / SWD. <http://www.dnb.de/DE/Wir/Projekte/Abgeschlossen/crosskonkordanz.html> (14.12.2014) sowie Dolud, Lena; Kreis, Constanze: Die Crosskonkordanz Wirtschaft zwischen dem STW und der GND. In: Dialog mit Bibliotheken 2 (2012), S. 13-19; abrufbar unter <http://nbn-resolving.de/urn:nbn:de:101-20130822224>.

12 Nähere Informationen zum Projekt siehe CrissCross. <http://www.dnb.de/DE/Wir/Projekte/Abgeschlossen/crisscross.html> (14.12.2014).

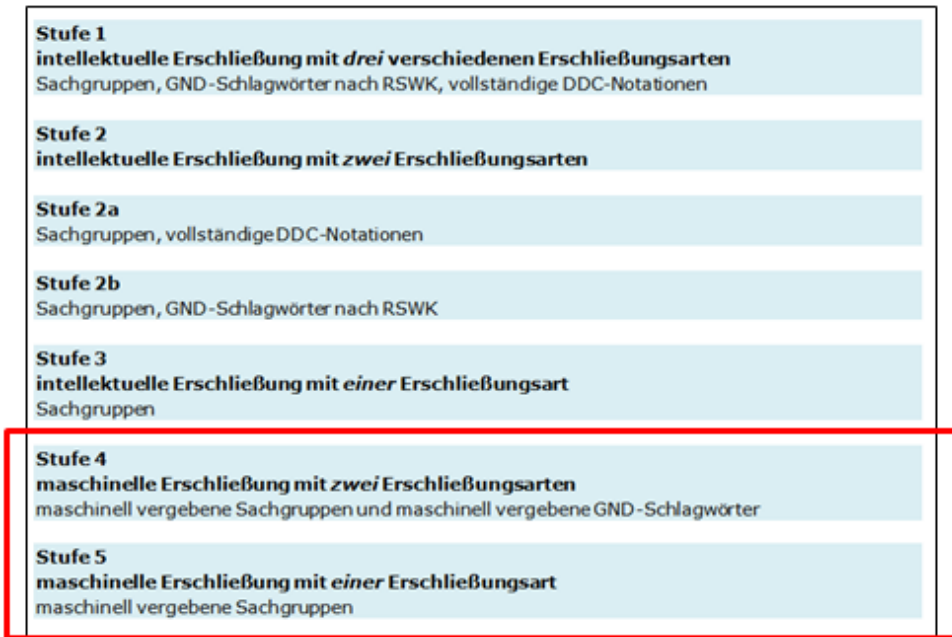


Abb. 7: Erweitertes Stufenschema der Inhaltserschließung der DNB

Abbildung 7 zeigt das derzeit gültige Erschließungskonzept für die Inhaltserschließung, wie es auch auf der Website der DNB veröffentlicht wurde, um Nutzer und Datenbezieher über die aktuelle Erschließungspolitik zu informieren.¹³

Dies wird aber sicherlich nicht die letzte Version des Konzeptes sein. Es wird ein wachsendes Reservoir weiterer inhaltserschließender Daten geben, aus weiteren Datenquellen und vor allem durch maschinelle Verfahren erzeugt, so dass sich für einzelne Medienwerke nicht mehr scharf fassen und beschreiben lässt, welcher Stufe sie zuzuordnen sind. So kann ein intellektuell nach DDC klassifizierter Titel der Reihe B künftig auch maschinell erzeugte Schlagwörter mit sich führen.

4. Schlussbemerkungen

Die DNB bleibt eine wichtige Lieferantin von Inhaltserschließung. Bei allen Veränderungen im Erschließungskonzept ist es das Ziel der DNB, auch künftig in großem Umfang inhaltserschließende Daten zu erstellen und zu liefern. Allerdings wird es vermehrt Daten zu diesen Daten geben müssen, d.h., dass ausgelieferte Nachweisdaten Kennzeichnungen enthalten, so dass Herkunft und „Güte“ nachvollzogen werden können. So hat die DNB eine Erweiterung im MARC21-Format angeregt, die

¹³ Siehe Inhaltliche Erschließung der Deutschen Nationalbibliothek. <http://www.dnb.de/SharedDocs/Downloads/DE/DNB/wir/inhaltserschliessung/nichterschliessenGesamt.pdf> (14.12.2014).

den Transport der Mitteilung erlaubt, dass die Inhaltserschließung maschinell erstellt wurde. Benötigt werden solche Informationen zu Auswertungs- und Steuerungszwecken, z.B. für die Nachnutzung solcher Daten in anderen Informationssystemen oder für die Anzeige in Katalogen und Portalen.

Die Weiterentwicklung der Inhaltserschließung und der dafür eingesetzten Werkzeuge und Methoden bringt noch eine Reihe weiterer Veränderungen und Herausforderungen mit sich:

Die Pflege der GND muss weiterentwickelt werden.

Die DNB ist davon überzeugt, dass der Einsatz eines normierten Vokabulars wie der GND einen Mehrwert hat, auch als verlässlicher Knoten im Web of Data. Die GND ist das gemeinsame Vokabular, das sowohl bei der intellektuellen wie bei der maschinellen Erschließung zum Einsatz kommt und so einheitliche Sucheinstiege ermöglicht. Allerdings erfordert der Einsatz namentlich maschineller Erschließungsverfahren neue Vorgehensweisen für die Pflege der Normdaten. Bisher entstehen neue Normdatensätze in der Regel, weil jemand ein Medienwerk erschließen möchte. Wenn einzelne Medienwerke nicht mehr auf den Tisch der Erschließenden kommen, sind neue Werkzeuge und Verfahren erforderlich, um das Vokabular aktuell zu halten; sie müssen allerdings erst entwickelt werden. Nicht auszuschließen ist dabei übrigens, dass es zu Wechselwirkungen zwischen der Normdatei und den maschinellen Verfahren kommt. So wäre es nach bisherigen Erfahrungen für die maschinelle Beschlagwortung nützlich, wenn die Entitäten in der GND von vornherein disambiguiert wären. Die RSWK sehen dies erst dann vor, wenn für eine zweite, gleichnamige Entität ein Normdatensatz angelegt werden soll.

Regelwerke und Datenmodelle sollten angepasst werden.

Regelwerke und Datenmodelle sollten ebenfalls auf die veränderte Situation reagieren. Die RSWK müssen zu einem Regelwerk zur verbalen Inhaltserschließung werden, das auch Aussagen und Empfehlungen z.B. zum Thema maschinelle Erschließung, aber auch zum Problem der Heterogenität von Daten enthalten sollte. Es muss sich darüber hinaus in den durch RDA vorgegebenen Rahmen einfügen. Die bisherigen Erfahrungen bei der Identifikation von Parallelausgaben eines Werkes zur automatischen Übernahme inhaltserschließender Daten zeigen, dass es ein erheblicher Vorteil wäre, wenn die Datenmodelle für bibliothekarische Erschließung den Functional Requirements for Bibliographic Records¹⁴ folgen würden. Dann könnte das Werk als geistige Schöpfung inhaltlich erschlossen werden, alle Manifestationen und Expressionen wären in identischer Weise thematisch zugreifbar.

Neue Aufgaben entstehen

Eine Veränderung der Methoden der Inhaltserschließung heißt mitnichten, dass damit Arbeit wegfällt – die Aufgaben verändern sich jedoch, es werden ggf. auch andere Kenntnisse und Fähigkeiten benötigt, z.B. ein Verständnis dafür, wie ein Wörterbuch für die maschinelle Beschlagwortung funktioniert und wie man es pflegt. Neue Aufgaben sind z.B. die Qualitätssicherung maschinell

¹⁴ IFLA Study Group on the Functional Requirements for Bibliographic Records: Functional requirements for bibliographic records. Final report, 1997, as amended and corrected through 2009.
http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf (14.12.2014).

erzeugter Daten oder die Überwachung und Optimierung von Prozessen. In der DNB wurde im Jahr 2013 eine neue Organisationseinheit gegründet, die sich mit dem Thema maschinelle Verfahren und Datengewinnung befasst. Dafür mussten Ressourcen für die intellektuelle Erschließung umgewidmet werden. Ein wesentlicher Grund für diese Entscheidung war die Einsicht, dass nur auf diesem Weg verhindert werden kann, dass der Anteil an Medienwerken in der Sammlung der DNB, der inhaltlich erschlossen wird, immer kleiner wird. Vor allem aber bieten maschinelle Verfahren die Chance, auch eine inhaltliche Erschließung z.B. von Aufsatzdaten ins Auge zu fassen.

Inhaltsererschließung bleibt für die DNB eine wichtige Aufgabe.

Literaturverzeichnis

- Beyer, Christian; Trunk, Daniela: Automatische Verfahren für die Formalerschließung im Projekt PETRUS. In: Dialog mit Bibliotheken 2 (2011), S. 5-10; abrufbar unter <http://nbn-resolving.de/urn:nbn:de:101-2012030831>
- Dolud, Lena; Kreis, Constanze: Die Crosskonkordanz Wirtschaft zwischen dem STW und der GND. In: Dialog mit Bibliotheken 2 (2012), S. 13-19; abrufbar unter <http://nbn-resolving.de/urn:nbn:de:101-20130822224>
- IFLA Study Group on the Functional Requirements for Bibliographic Records: Functional requirements for bibliographic records. Final report, 1997, as amended and corrected through 2009. http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf (14.12.2014)
- Junger, Ulrike: Can indexing be automated? The example of the Deutsche Nationalbibliothek. In: Cataloguing & Classification Quarterly 52, Nr. 1 (2014), S. 102-109.
- Schöning-Walter, Christa: Automatische Erschließungsverfahren für Netzpublikationen. In: Dialog mit Bibliotheken 1 (2011), S. 31-36; abrufbar unter <http://nbn-resolving.de/urn:nbn:de:101-2011101170>
- Uhlmann, Sandro: Automatische Beschlagwortung von deutschsprachigen Netzpublikationen mit dem Vokabular der Gemeinsamen Normdatei (GND). In: Dialog mit Bibliotheken 2 (2013), S. 26-36; abrufbar unter <http://nbn-resolving.de/urn:nbn:de:101-20140305238>