

Ulrike Junger und Frank Scholze

Neue Wege und Qualitäten – Die Inhaltserschließungspolitik der Deutschen Nationalbibliothek

1 Einleitung

Es kommt nicht oft vor, dass ein bibliothekfachliches Thema Gegenstand eines ganzseitigen Artikels im Feuilleton einer der wichtigsten überregionalen Zeitungen in Deutschland wird. Am 31. Juli 2017 war dies der Fall: Die *Frankfurter Allgemeine Zeitung* veröffentlichte einen Artikel des Generaldirektors der Bayerischen Staatsbibliothek, Klaus Ceynowa, in dem dieser sich kritisch mit einem Konzept zur inhaltlichen Erschließung auseinandersetzte,¹ das die *Deutsche Nationalbibliothek* (DNB) zuvor für die deutschsprachige bibliothekarische Community veröffentlicht hatte.

Hatten bereits zuvor die Bemühungen der DNB, Verfahren zur maschinellen Dokumenterschließung zu entwickeln und einzusetzen, zu kontroversen Reaktionen im Bibliothekswesen geführt, so sorgte dieser Artikel noch einmal in besonderer Weise für Aufmerksamkeit und Diskussionen zu einem Thema, das vielen als eher verstaubt und unattraktiv galt: die Inhaltserschließung.

Der folgende Beitrag zeichnet einige Grundlinien der Erschließungspolitik der DNB seit 2010 nach und beschreibt, welche Instrumente und Verfahren bei der Inhaltserschließung zum Einsatz kommen, welche konzeptionellen Entscheidungen ihr zugrunde liegen, wie versucht wird, Qualität zu erfassen und welche Entwicklungs- und Handlungsfelder für die Zukunft gesehen werden.

2 Verfahren der Inhaltserschließung in der DNB

In der bibliothekarischen Sach- oder Inhaltserschließung gibt es zwei klassische Herangehensweisen: Einerseits die verbale Inhaltserschließung, bei der Dokumenten Schlagwörter aus einem kontrollierten Vokabular zugewiesen werden und andererseits die klassifikatorische Erschließung, die Dokumente in ein meist hierarchisch gegliedertes System der Wissensorganisation einordnet. Im

¹ Siehe <https://www.faz.net/aktuell/feuilleton/buecher/maschinen-lesen-buecher-deutsche-nationalbibliothek-setzt-auf-technik-15128954.html> (4.12.2020).

Laufe der Geschichte der DNB kamen unterschiedliche Systeme und Methoden sachlicher Erschließung zum Einsatz. Derzeit bedient sich die DNB sowohl eines verbalen als auch eines klassifikatorischen Systems.

Als Herausgeberin der *Deutschen Nationalbibliografie* spielt die DNB traditionell eine wichtige Rolle für andere Bibliotheken. Dienstleistungen für und Kooperation mit anderen bibliothekarischen Einrichtungen gehören zu ihren gesetzlich verankerten Aufgaben.² Gemeinsame Regeln und Standards erleichtern den Austausch und die gemeinsame Nutzung von Katalogdaten. Dies ist einer der Gründe, warum die DNB auch bei den Verfahren und Systemen, die für die Inhaltsererschließung eingesetzt werden, auf kooperative Vorhaben setzt.

So trägt die DNB seit den 1980er Jahren zur Entwicklung des *Regelwerks für die Schlagwortkatalogisierung* (RSWK) bei. Dieses Regelwerk, das seither mehrere Überarbeitungen erfahren hat, bildet bis heute die Grundlage für die Schlagwortvergabe in der DNB. Auf der Basis der RSWK wurde ein kooperativ erstelltes und gepflegtes kontrolliertes Schlagwortvokabular erstellt – die *Schlagwortnormdatei* (SWD). Diese hat sich als das am häufigsten verwendete Schlagwortvokabular im deutschsprachigen Bibliothekswesen durchgesetzt.

Die SWD wurde 2012 mit den Normdateien, die in der Formalerschließung Verwendung fanden,³ zur *Gemeinsamen Normdatei* (GND) zusammengeführt.⁴

Die GND, die mittlerweile auch über das Bibliothekswesen hinaus Anwendung findet, ist das Reservoir an kontrollierten Begriffen, das sowohl für die intellektuelle als auch für die maschinelle Erschließung in der DNB verwendet wird.

Während die verschiedenen Reihen der Deutschen Nationalbibliografie traditionell nach Fachgebieten gegliedert waren, setzten weder die *Deutsche Bücherei Leipzig* noch die *Deutsche Bibliothek Frankfurt* eine Feinklassifikation zur inhaltlichen Erschließung von Medienwerken ein. In den 1990er Jahren gab es entsprechende Forderungen an die DNB, die schließlich zum Einsatz der *Dewey-Dezimalklassifikation* (DDC) in Deutschland führten.⁵

² Siehe das Gesetz über die Deutsche Nationalbibliothek, §2, Abs. 3, online abrufbar unter <https://www.gesetze-im-internet.de/dnbg/index.html> (4.12.2020).

³ Es handelt sich um die Personennormdatei (PND), die Gemeinsame Körperschaftsdatei (GKD) sowie die Einheitssachtitel-Datei des Deutschen Musikarchivs (EST-DMA).

⁴ Für weitergehende Informationen zur GND siehe den Beitrag von Esther Scheven, Qualitätssicherung in der GND, in diesem Band.

⁵ Einen Überblick über die DDC, ihre Einführung und Verwendung in der DNB gibt Heidrun Alex, Die Dewey-Dezimalklassifikation (DDC) in: *Klassifikationen in Bibliotheken*, S. 65 ff.

Es gibt aktuell drei Varianten der Nutzung der DDC in der DNB:⁶

- Die sogenannten DDC-Sachgruppen dienen seit 2004 dazu, die Deutsche Nationalbibliografie thematisch zu gliedern. Jedes in der Deutschen Nationalbibliografie verzeichnete Medienwerk erhält mindestens eine der rund 100 DDC-Sachgruppen.
- Seit 2007 werden vollständige DDC-Notationen vergeben, dabei wird das gesamte System der Haupt- und Hilfstafeln der DDC genutzt.
- Im Zusammenhang mit der Einführung maschineller Erschließungsverfahren werden sogenannte DDC-Kurznotationen entwickelt. Dabei handelt es sich um fachlich ausgewählte kürzere DDC-Notationen, die feingranularer sind als die Notationen der DDC-Sachgruppen, aber kürzer als vollständige DDC-Notationen.

In der digitalen Welt bieten jedoch nicht nur Schlagwörter und Notationen einen thematischen Zugang zu Medienwerken, sondern z. B. auch digital vorliegende Inhaltsverzeichnisse. Seit 2008 digitalisiert die DNB die Inhaltsverzeichnisse des laufenden Zugangs an körperlichen Medienwerken, aber auch retrospektiv für ältere Bestände. Diese Inhaltsverzeichnisse sind als durchsuchbarer Volltext in den Katalog der DNB eingebunden und unterstützen thematische Recherchen.⁷

3 Die Inhaltserschließungspolitik der DNB seit 2010

Die Novellierung des Gesetzes über die Deutsche Nationalbibliothek (DNBG) im Jahr 2006 stellte eine Zäsur für die DNB dar. Die Ausweitung des gesetzlichen Sammelauftrages auf sogenannte unkörperliche Medienwerke,⁸ d. h. im Internet zugängliche Veröffentlichungen, hatte auch erhebliche Auswirkungen auf die anderen Aufgaben, die das DNBG der DNB vorgibt: So sind etwa für die Archivierung, Verfügbarmachung und Bereitstellung unkörperlicher Medienwerke neue, andersartige Verfahren und Infrastrukturen erforderlich. Aber auch die Erschließung wurde vor neue Herausforderungen gestellt. Die Aufgabe der

⁶ Siehe auch https://www.dnb.de/DE/Professionell/DDC-Deutsch/DDCinDNB/ddcindnb_node.html (4.12.2020).

⁷ Für nähere Informationen siehe <https://www.dnb.de/DE/Professionell/Metadatendienste/Metadaten/Kataloganreicherung/kataloganreicherung.html> (4.12.2020).

⁸ Vgl. DNBG § 3, Abs. 3.

Erschließung und bibliografischen Verzeichnung als solche hat sich mit der Novellierung des DNBG nicht geändert.

Grundsätzlich wäre es denkbar gewesen, auch unkörperliche Medienwerke in herkömmlicher Weise formal und inhaltlich zu erschließen, d. h. durch die intellektuelle Erstellung entsprechender Katalogdaten. Im Wesentlichen gab es drei Gründe, warum die DNB einen anderen Weg eingeschlagen hat:

- Zwar gab es im Vorfeld der Novellierung des DNBG Modellrechnungen, wie sich der digitale Publikationsmarkt entwickeln würde. Diese Projektionen erwiesen sich jedoch schnell als unzureichend. Obwohl die DNB in den Jahren ab 2006 zunächst erst einmal grundlegende Konzepte und eine Infrastruktur zur Sammlung und Archivierung von unkörperlichen Medienwerken aufbauen und erstellen musste, so zeigte sich schon bald, dass die Menge der zu sammelnden Medienwerke alle Projektionen drastisch überstieg (Abb. 1). Auch wenn der DNB im Zuge der Novellierung des DNBG Personalressourcen für die Bewältigung der neuen Aufgaben zugewiesen worden waren, so war absehbar, dass die für die Erschließung verfügbaren Kapazitäten bei weitem nicht ausreichen würden, um die Menge an unkörperlichen Medienwerken auf herkömmliche Weise zu erschließen.
- Die zu erschließenden digitalen Objekte selbst bieten die Grundlage für den Einsatz maschineller Verfahren.
- Mit den rasch wachsenden Mengen an digitalen Objekten ist die steigende Verfügbarkeit beschreibender Metadaten und ergänzender Informationen wie Inhaltsverzeichnisse verknüpft.

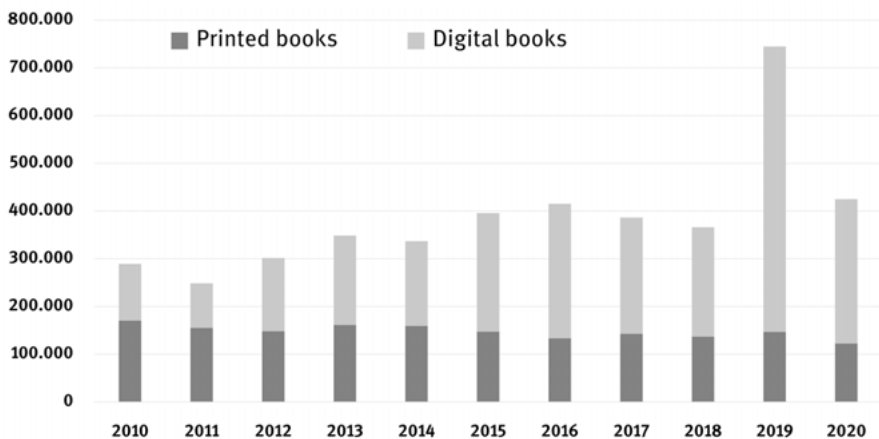


Abb. 1: Entwicklung des Zugangs monografischer Print- und Netzpublikationen der DNB

Im Zusammenhang mit der Einführung der neuen Reihe O der Deutschen Nationalbibliografie, die der Verzeichnung der von der DNB gesammelten unkörperlichen Medienwerke dient, traf die DNB im Jahr 2010 die Grundsatzentscheidung, die intellektuelle Erschließung für unkörperliche Medienwerke einzustellen.⁹ Katalogisate für unkörperliche Medienwerke sollten auf der Grundlage von Metadaten erstellt werden, die zusammen mit dem digitalen Objekt abzuliefern sind.¹⁰ Diese Daten werden über entsprechende maschinelle Prozesse in die Katalogisierungsdatenbank der DNB integriert und weiterverarbeitet. Die verpflichtend abzuliefernden Metadaten umfassen vor allem solche, die sonst in der Formalerschließung erstellt worden wären. Inhaltserschließende Angaben sind fakultativ.

Nun hätte die Entscheidung zum Verzicht auf die intellektuelle Erschließung unkörperlicher Medienwerke bedeuten können, dass auf eine weitergehende inhaltliche Erschließung von unkörperlichen Medienwerken insgesamt verzichtet wird. Dies stand jedoch nie zur Diskussion. Dass die Möglichkeiten eines thematischen oder inhaltlichen Zugriffs auch in digitalen Kontexten wichtig sind, zeigt der Blick z. B. auf kommerzielle Verkaufsportale: Inhalte nach Themen, Facetten und Kategorien suchen und strukturieren zu können, ist wichtiger Bestandteil eines nutzungsfreundlichen Service.

Es war klar, dass eine über Sachgruppen hinausgehende inhaltliche Erschließung notwendig ist, um gerade die rasant wachsenden Mengen an unkörperlichen Medienwerken sinnvoll thematisch zugreifbar zu machen.

Die dafür eingesetzten Systeme der inhaltlichen Erschließung sollten dieselben sein wie in der intellektuellen Inhaltserschließung, um einheitliche thematische Zugriffe auf den Gesamtbestand der Medienwerke in den Sammlungen der DNB zu erlauben, unabhängig von der Materialart. In Bezug auf die Art und Weise, wie mit diesen Systemen inhaltserschließende Daten erzeugt werden, sollten hingegen neuartige Wege beschritten werden. Konkret ging es darum, maschinelle Verfahren zu entwickeln und einzusetzen, die auf der Basis der digital vorliegenden Texte inhaltstragende Informationen extrahieren, diese auf die GND und die eingesetzten Klassifikationen abbilden und mit den Ergebnissen die Nachweise für die Medienwerke anreichern.

Maschinelle Erschließungsverfahren haben einige grundsätzliche Vorzüge:

- Sie erlauben die Erschließung sehr großer Dokumentmengen, die nur unter unverhältnismäßig hohem Ressourceneinsatz intellektuell erschlossen werden könnten.

⁹ Vgl. Gömpel u. a. 2010.

¹⁰ Siehe https://www.dnb.de/DE/Professionell/Sammeln/Unkoerperliche_Medienwerke/unkoerperliche_medienwerke_node.html#doc210120bodyText3 (4.12.2020).

- Es können auch Dokumentgruppen erschlossen werden, die traditionell nicht im Fokus bibliothekarischer Erschließung stehen, z. B. Zeitschriftenartikel.
- Bei einer Veränderung der Verfahren können bereits erschlossene Dokumente mit überschaubarem Aufwand erneut einer Erschließung unterzogen werden.
- Sie lassen sich auch retrospektiv auf bis dahin nicht erschlossene Bestände anwenden, sofern ein digitaler Text oder ein digitales Substrat vorliegt.

Im Zusammenhang mit der Entscheidung, unkörperliche Medienwerke nicht mehr intellektuell zu erschließen, wurde das Projekt PETRUS¹¹ auf den Weg gebracht, in dem Lösungen für die automatisierte Erschließung von Netzpublikationen, Digitalisaten und Print-Publikationen erarbeitet werden sollten. Nach einigen Jahren wurde eine neue Organisationseinheit in der DNB etabliert, die sich dauerhaft um die Pflege und Fortentwicklung maschineller Erschließungsverfahren kümmert.

Innovationsprojekte benötigen Ressourcen. Um das Vorhaben zur Entwicklung maschineller Erschließungsverfahren mit entsprechendem Personal ausstatten zu können, war es erforderlich, Ressourcen umzulenken. Dies betraf auch die intellektuelle Inhaltserschließung körperlicher Medienwerke. Weniger Ressourcen und die Erwartung, langfristig intellektuelle Erschließung auf breiter Fläche auch für körperliche Medienwerke durch maschinell erzeugte Daten ersetzen zu können, mündeten in eine stufenweise Reduktion und weitere Differenzierung der intellektuellen Inhaltserschließung in der DNB, die bezogen auf verschiedene Gruppen von Medienwerken bereits unterschiedlich gehandhabt wurde.¹² Die Einschränkung der intellektuellen Inhaltserschließung für körperliche Medienwerke erfolgte aufgrund des benötigten Vorlaufs asynchron und asymmetrisch zur Entwicklung und Bereitstellung maschinell erzeugter Daten für unkörperliche Medienwerke.

Die Einschränkungen bei der intellektuellen Inhaltserschließung – insbesondere der Schlagwortvergabe – orientierten sich an folgenden Annahmen:

- Haben Medienwerke aussagekräftige Titel, so ist der Mehrwert einer Beschlagnwortung für das Retrieval gering.

¹¹ Vgl. Schöning-Walter 2020.

¹² Z. B. wurden Publikationen der Reihe B seit der Einführung der Dewey-Dezimalklassifikation in der DNB nicht mehr beschlagwortet, andere Gruppen von Medienwerken wurden nur in Auswahl erschlossen.

- Intellektuelle Inhaltserschließung sollte auf solche Medienwerke konzentriert werden, die im Bestand vieler Bibliotheken vorhanden sind, d. h. bei denen die Metadaten der DNB stark nachgenutzt werden.
- Es gibt Unterschiede in der Publikationskultur verschiedener Wissenschaftsfächer. In Fächern, in denen die wissenschaftliche Kommunikation vornehmlich über Zeitschriften erfolgt, erscheint eine Einschränkung der Inhaltserschließung von Büchern vertretbar.
- Die intellektuelle Inhaltserschließung war bereits zuvor nur bei ausgewählten Medienwerken vorgenommen worden, z. B. die Beschlagwortung von Belletristik.

Ein erster Schritt wurde im Jahr 2012 vollzogen. Für Kochbücher, Reiseführer, Bastelbücher, Sprachwörterbücher, Ratgeber u. ä. sowie Zeitschriften wurde die Beschlagwortung eingestellt, ein inhaltlicher Zugriff auf diese Medienwerke über Titelstichwörter und die DDC-Erschließung wurde als ausreichend erachtet.

Zudem wurde die ohnehin nur in ausgewählten Fällen durchgeführte Beschlagwortung von Belletristik auf ursprünglich deutschsprachige Werke beschränkt. Zugleich wurden jedoch die über die Metadaten des Marketing- und Verlagservice des Buchhandels (MVB) verfügbaren Gattungsbegriffe für Belletristik sowie Kinder- und Jugendliteratur übernommen.¹³ Es wurden Normsätze im Katalogisierungssystem der DNB angelegt, die Verweisungen und zusätzliche Informationen enthalten und auf die die MVB-Daten bei der Einspielung abgebildet und verknüpft werden. Sind für ein Medienwerk keine MVB-Daten vorhanden, so werden die Gattungsbegriffe intellektuell vergeben. Im Ergebnis sind alle Werke der Belletristik und Kinder- und Jugendliteratur entsprechend erschlossen.

Die nächste Veränderung hing mit der Einführung des Standards *Resource Description and Access* (RDA) im Jahr 2015 zusammen: Das im deutschsprachigen Bibliothekswesen verabredete sogenannte Standardelemente-Set¹⁴ sieht die Angabe von Art des Inhalts und Zielgruppe eines Medienwerks vor. Diese Angaben waren aus Sicht der DNB ausreichend, um Schul- und Berufsschulbücher zu erschließen, weshalb die Schlagwortvergabe für Medienwerke in der DDC-Sachgruppe S (Schulbücher) eingestellt wurde.

Das Jahr 2017 brachte eine weitere Veränderung in der Inhaltserschließungspolitik der DNB mit sich. Nachdem seit 2012 mit der Inbetriebnahme der

¹³ Die Liste der Gattungsbegriffe kann unter <https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/listeGattungsbegriffe.pdf> (4.12.2020) abgerufen werden.

¹⁴ Siehe <https://wiki.dnb.de/pages/viewpage.action?pageId=114430616> (4.12.2020).

maschinellen DDC-Sachgruppenvergabe für deutsch- und englischsprachige Netzpublikationen ein erster Meilenstein in der Nutzung maschineller Verfahren erreicht werden konnte, folgte 2014 die maschinelle Schlagwortvergabe mit der GND für deutschsprachige Netzpublikationen und 2015 erstmals die maschinelle Vergabe von DDC-Kurznotationen für medizinische Publikationen der Reihe O. Letztere konnte inzwischen erfolgreich¹⁵ auf andere Fächer ausgeweitet werden. Seit 2018 werden englischsprachige Publikationen der Reihe O maschinell mit Sachbegriffen der *Library of Congress Subject Headings* (LCSH) versehen kombiniert mit Schlagwörtern der GND für Individualbegriffe.

Die Erfahrungen und Ergebnisse der maschinellen Inhaltserschließung in der DNB ließen es denkbar erscheinen, auch körperliche Medienwerke mit digitalisierten Inhaltsverzeichnissen mit maschinell erzeugten Schlagwörtern und Notationen zu versehen.

Das in der Folge entwickelte Konzept¹⁶ sieht vor allem zwei Maßnahmen vor:

- Die intellektuelle Beschlagwortung von Kinder- und Jugendliteratur sowie belletristischer Werke wird generell aufgegeben.
- Die Klassierung von Publikationen der Reihen B und H mit vollständigen DDC-Notationen wird zugunsten der maschinellen Vergabe von Schlagwörtern aus der GND eingestellt.

Es enthält außerdem die Aussage, dass Verfahren zur maschinellen Inhaltserschließung perspektivisch flächendeckend anstelle intellektueller Vergabe von Schlagwörtern und Notationen zum Einsatz kommen sollen.

Über dieses Konzept informierte die DNB kooperierende Einrichtungen und Kund:innen im Frühsommer 2017, angewandt wurde es ab September 2017.

4 Reaktionen auf das Erschließungskonzept 2017 und aktuelle Anpassungen

Diese Maßnahmen fanden in der bibliothekarischen Öffentlichkeit ein gemischtes Echo und führten u. a. zu dem in der Einleitung erwähnten Artikel von Klaus Ceynowa in der FAZ. Kritik wurde insbesondere daran geübt, dass die Ergebnis-

¹⁵ Die Inbetriebnahme der maschinellen Vergabe von DDC-Kurznotationen für ein Fachgebiet ist abhängig davon, dass der Algorithmus anhand einer ausreichenden Menge an Dokumenten so trainiert werden kann, dass ein F-Wert von 0,65 erreicht wird.

¹⁶ Das Konzeptpapier ist abrufbar unter <https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/konzeptWeiterentwicklungInhaltserschliessung.pdf> (4.12.2020).

se, die über maschinelle Verfahren erzielt werden, unzureichend seien und damit die Ablösung der intellektuellen Inhalterschließung durch maschinelle Verfahren insbesondere für körperliche Medienwerke der Reihe A nicht vertretbar sei.

Die DNB reagierte einerseits mit einem Papier, das die getroffenen Entscheidungen nochmals erläuterte,¹⁷ andererseits wurde vom Standardisierungsausschuss eine Initiativgruppe *Kooperative Erschließung und Informationsversorgung im DACH-Raum am Beispiel der Inhalterschließung* eingerichtet. Diese Arbeitsgruppe erarbeitete zusammen mit der DNB ein sogenanntes Übergangskonzept zur Inhalterschließung. Es sieht vor, dass die DNB zwar am Primat der maschinellen Verfahren festhält, jedoch zusichert, dass für Publikationen der Reihe A eine intellektuelle Inhalterschließung (Schlagwörter, DDC-Vollnotationen, Sachgruppen) vorgenommen wird, solange maschinelle Verfahren keinen adäquaten Ersatz liefern. Diese Garantie bezieht sich auf wissenschaftliche Literatur in sogenannten buchaffinen Fächern; das sind im Wesentlichen die Geistes- und Sozialwissenschaften sowie das Ingenieurwesen.¹⁸ Am grundsätzlichen Ziel des Einsatzes maschineller Erschließungsverfahren auch für körperliche Medienwerke wird festgehalten.

Auch die Öffentlichen Bibliotheken meldeten sich zu Wort. Bereits die Einstellung der Schlagwortvergabe für Ratgeber, Reiseführer und Hobbyliteratur sowie Schulbücher hatte ihren Widerspruch hervorgerufen. Die Nicht-Erschließung von Belletristik und insbesondere von Kinder- und Jugendliteratur war aus Sicht dieser Gruppe von Bibliotheken nicht akzeptabel, da diese Gattung bei ihren Nutzer:innen besonders nachgefragt wird und eine gute inhaltliche Erschließung aus Sicht der Öffentlichen Bibliotheken für eine qualifizierte Beratung und Auskunftserteilung erforderlich ist.

Die DNB bot daraufhin an, zusammen mit Vertretungen der Öffentlichen Bibliotheken in einer gemeinsamen Arbeitsgruppe an Alternativen zu arbeiten. Im Fokus stand dabei die Nachnutzung von Inhalterschließungsdaten des Buchhandels und ihre Umwandlung in GND-Schlagwörter. Seit Herbst 2020 läuft ein entsprechendes Verfahren für Kinder- und Jugendliteratur im Produktivbetrieb, weitere Gruppen von Medienwerken sollen folgen.¹⁹

Mit dem Übergangskonzept im Juli 2019 wurde auch die maschinelle Schlagwortvergabe für Publikationen der Reihe B wieder aufgegeben, da der hohe Anteil sozialwissenschaftlicher Publikationen mit wenig spezifischem

¹⁷ Vgl. Junger, Schwens 2017.

¹⁸ Das Konzept ist abrufbar unter <https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/veraenderungenInhalterschliessungDnbJuli2019.pdf> (4.12.2020).

¹⁹ Ausführliche Informationen finden sich bei Mödden u. a. 2020.

Vokabular und ein hoher Anteil an inhaltstragenden Individualbegriffen wie Körperschaften bei nichtwissenschaftlichen Publikationen wie z. B. Ortsführern zu einer hohen Quote an Erschließungsfehlern führte. Zudem wären Verbesserungen nur durch fortlaufende, aufwendige Modifikationen des eingesetzten Wörterbuches, das für die maschinelle Schlagwortvergabe verwendet wird, zu erreichen gewesen. Aufrechterhalten wird jedoch die maschinelle Schlagwortvergabe für Dissertationen der Reihe H, da hier die Ergebnisse akzeptabel sind. Dies ist vermutlich auch darauf zurückzuführen, dass die Inhaltsverzeichnisse wissenschaftlicher Schriften oft ein detailliertes spezifisches Vokabular aufweisen, das den maschinellen Verfahren eine gute Ausgangsbasis bietet.

Reihe H (2017) Verteilung der Bewertungen

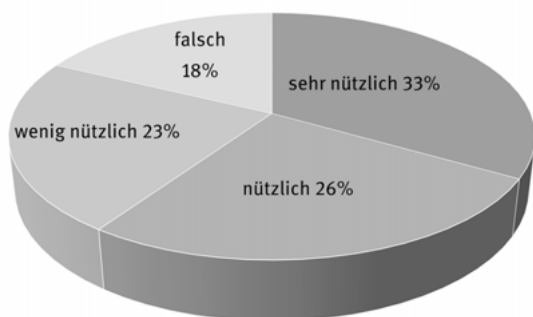


Abb. 2: Auswertungsergebnisse für die maschinelle Schlagwortvergabe für Publikationen der Reihe H

5 Zur Frage der Qualität maschineller Inhaltserschließung

Die Auseinandersetzung um die Frage, was Qualität ist und wie man sie misst, begleitet die Entwicklung und den Einsatz maschineller Erschließungsverfahren von Beginn an. Es ist zu konstatieren, dass es keinen einfachen Qualitätsbegriff etwa in Form einer allgemein akzeptierten Maßzahl für die Bewertung der Ergebnisse (maschineller) inhaltlicher Erschließung gibt, auf den zurückgegriffen werden kann.

Im Folgenden wird beschrieben, welche Maße und Kriterien bei der Entwicklung und Qualitätssicherung der in der DNB eingesetzten maschinellen

Erschließungsverfahren angewandt werden. Diese stellen die Grundlage für Aussagen zur Qualität dar:

1. Vergleiche von Erschließungsergebnissen, die auf unterschiedlichem Wege erzeugt wurden.
2. Bewertungen von Erschließungsergebnissen durch Personen.
3. Gütemaße, die im Information Retrieval Anwendung finden, konkret *recall* und *precision* sowie das daraus abgeleitete *F-Maß*.
4. Maße aus der Wahrscheinlichkeitsrechnung in Form von Konfidenzwerten, die im Prozess der Generierung maschineller Erschließungsdaten gebildet werden.

Eine vergleichende statistische Bewertung von Erschließungsergebnissen wird von der DNB in folgenden Kontexten durchgeführt: Bei der Klassierung werden zum einen intellektuell erschlossene Dokumente und deren maschinell bearbeitete digitale Parallelausgaben betrachtet, wobei maschinell vergebene Sachgruppen und Kurznotationen mit den intellektuell vergebenen abgeglichen werden. Zum anderen werden maschinell vergebene Sachgruppen und Kurznotationen im Zuge des laufenden stichprobenbasierten Qualitätsmanagements intellektuell bewertet. Intellektuell vergebene Sachgruppen und Kurznotationen fungieren dabei als Goldstandard, deren Richtigkeit als gesetzt betrachtet wird. Der Abgleich ergibt prozentuale Werte richtig und falsch zugewiesener Notationen. Bei der Schlagwortvergabe werden solche Abgleiche zwischen intellektuell und maschinell erschlossenen Parallelausgaben gleichfalls herangezogen.

Ein Hauptzweck bibliothekarischer Erschließung ist, Dokumente in Nachweissystemen zu finden. Bei einer weiteren Form der Qualitätskontrolle stehen daher Retrievalaspekte im Zentrum. Zur Beurteilung maschinell verbogener Schlagwörter wird eine intellektuelle Bewertung der Nützlichkeit eines Schlagworts für das Wiederauffinden eines Dokuments vorgenommen. Dabei trifft eine Person eine subjektive Aussage darüber, ob sie ein maschinell vergebenes Schlagwort für geeignet hält, ein Dokument im Rahmen einer thematischen Recherche finden zu können. Die dafür verwendete vierstufige Skala wurde von der DNB selbst festgelegt. Sie umfasst folgende Stufen (vgl. Abb. 2):

- sehr nützlich: Das einzelne Schlagwort beschreibt einen wichtigen Aspekt des Textes ausreichend und trifft absolut zu.
- nützlich: Das einzelne Schlagwort beschreibt einen wichtigen Aspekt des Textes aus einer weiteren (oder auch engeren) Perspektive und trifft zu.
- wenig nützlich: Das einzelne Schlagwort beschreibt einen wichtigen Aspekt des Textes nicht ausreichend, ist aber auch nicht völlig unzutreffend oder falsch.

- falsch: Das einzelne Schlagwort beschreibt keinen wichtigen Aspekt des Textes und ist falsch.

Beim stichprobenbasierten Qualitätsmanagement der DNB werden nicht nur einzelne maschinell erzeugte Schlagwörter entsprechend bewertet, sondern auch die Gesamtheit aller Schlagwörter zu einem Dokument.

Zu beachten ist dabei jedoch, dass das Ziel maschineller Schlagwortvergabe nicht die Erstellung eines Kurz-Abstracts des Inhalts eines Dokumentes ist, wie es etwa die verbale Inhaltserschließung nach den Regeln für die Schlagwortkatalogisierung leisten soll.²⁰ Vielmehr geht es um die Anreicherung mit thematischen Zugriffspunkten, die einen sachlich korrekten Bezug zum Textinhalt und das Auffinden eines Dokuments im Zuge einer thematischen Recherche ermöglichen sollen.

Die zusammenfassende Bewertung maschineller Erschließungsergebnisse basiert auf gängigen Maßen für die Güte des Information Retrieval, und zwar auf *recall* (d. h. der Anteil der vorhandenen Dokumente aus einem Korpus, die mit Hilfe eines Suchterms tatsächlich gefunden wird) und *precision* (d. h. der Anteil korrekter Dokumente im Suchergebnis).

Die DNB hat sich entschieden, das Mittel aus *recall* und *precision* als Gütemaß zu nutzen, das sogenannte *F-Maß*. F-Maße werden sowohl für maschinell vergebene Schlagwörter wie Notationen berechnet, auf der Grundlage größerer Dokumentmengen, etwa den Publikationen eines Jahrgangs der Deutschen Nationalbibliografie.

Im Zuge der maschinellen Bearbeitung von Dokumenten wird für alle maschinell erzeugten Erschließungsdaten ein Konfidenzwert ausgegeben, der sich zwischen 0 und 1 bewegt. Dieser intrinsische Wert kann aber nicht unmittelbar als Qualitätsmaß interpretiert werden, denn er gibt wieder, wie „sicher“ der maschinelle Algorithmus ist, ein zutreffendes Ergebnis produziert zu haben. Eine intellektuelle Bewertung derselben Daten fällt möglicherweise anders aus. Indirekt kann der Konfidenzwert jedoch als Qualitätsmaß genutzt werden, indem nur solche maschinell erzeugten Daten in das Katalogisat übernommen werden, die einen festgelegten Schwellenwert überschreiten.

Alle Maße und Kriterien dienen einer laufenden Überwachung der Prozesse der maschinellen Inhaltserschließung und werden zudem genutzt, um Veränderungen bzw. Fortschritte bei der Weiterentwicklung der Verfahren einschätzen zu können.

²⁰ Vgl. dazu RSWK §13,1.

Für eine Diskussion um die Qualität (maschineller) Inhaltserschließung sollte neben der Frage, welches die richtigen Maße und Kriterien sind, vor allem das Verhältnis von *precision* und *recall* in den Blick genommen werden.

Maschinelle Verfahren ermöglichen es, auch Dokumente zu erschließen, die traditionell nicht im Fokus bibliothekarischer Erschließung stehen. So werden in der DNB auch digitale Zeitschriftenartikel maschinell erschlossen und damit die Menge an Dokumenten mit inhaltserschließenden Daten, die im Katalog der DNB recherchiert werden kann, erheblich ausgeweitet.

Bei maschineller Beschlagwortung kann etwa eine höhere *precision* bedeuten, dass zwar weniger falsche Schlagwörter vergeben werden, dafür aber auch weniger relevante; umgekehrt kann die Entscheidung für einen höheren *recall* zur Folge haben, dass eine höhere Ausbeute bei thematischen Recherchen erzielt werden kann, jedoch möglicherweise mehr falsche Schlagwörter in Kauf genommen werden müssen. Wie gehen Nutzer:innen bibliothekarischer Suchsysteme damit um? Kommt es nicht vor allem darauf an, dass Suchsysteme auf intelligente Weise inhaltserschließende Metadaten bündeln und thematische Pfade in einen Bestand legen?

Die DNB ist interessiert daran mitzuwirken, einschlägige und zielgerichtete Forschung in diesem Bereich zu unterstützen, beispielsweise bei Retrievalstudien mit Standardkorpora.

6 Was bringt die Zukunft – eine Skizze

Auch wenn die Umsetzung der Erschließungspolitik der DNB in den vergangenen zehn Jahren immer wieder unterschiedlich diskutiert wurde, so steht die DNB zu der grundsätzlichen Entscheidung, wo immer möglich maschinelle Erschließungsverfahren einzusetzen.

Eines der fünf strategischen Handlungsfelder der DNB²¹ in den kommenden Jahren lautet dementsprechend *Digitale Erschließungsverfahren ausbauen* und umfasst die folgenden Ziele:

- Die Erschließungsprozesse basieren auf dem digitalen Objekt.
- Die Erschließungsprozesse sind für Kooperationen geöffnet.
- Eine differenzierte und anforderungsorientierte Erschließungspolitik ist gestaltet und kommuniziert.

²¹ Die Strategischen Prioritäten 2021–2024 sind abrufbar unter https://www.dnb.de/DE/ueber-uns/Strategie/strategie_node.html#doc210304bodyText2 (11.1.2021).

- Es ist ein modulares System maschineller Erschließungsverfahren vorhanden.

Die Etablierung maschineller Erschließungsverfahren in der DNB war und ist kein einmaliger, abgeschlossener Vorgang. Vielmehr sind die laufende Fortentwicklung und Pflege maschineller Erschließungsverfahren zu neuen Daueraufgaben geworden. Wie alle Verfahren bibliothekarischen Arbeitens müssen sie regelmäßig auf den Prüfstand gestellt werden, auch, um neue (wissenschaftliche) Erkenntnisse aufzunehmen und umzusetzen. Aus diesem Grund führt die DNB derzeit ein größeres Vorhaben durch, um die maschinelle Inhaltserschließung auf eine veränderte methodische und technische Basis zu stellen und den Einsatz neuer Werkzeuge vorzubereiten. Das Schlagwortvokabular der GND und Notationen der DDC bleiben weiterhin Zielsysteme maschineller Inhaltserschließung, um Nutzer:innen einheitliche thematische Wege zum Bestand der DNB zu ermöglichen.

Eine inhaltliche Herausforderung kommender Jahre wird es sein, neben den Prozessen der Formalerschließung auch die Normdatenarbeit stärker durch maschinelle Verfahren zu unterstützen. Dazu gehört etwa das Etablieren eines Vorschlagssystems für neue Schlagwörter auf der Basis maschineller Erschließungsverfahren.

Die DNB verfolgt das Ziel, Hauptakteurin eines Kompetenznetzwerks für maschinelle Erschließungsverfahren im Bibliotheks- und Informationswesen zu werden. Die Kooperation mit anderen Einrichtungen und Personen aus dem Kultursektor sowie in Wissenschaft und Forschung, aber auch einschlägigen Unternehmen der Wirtschaft sind ein zentraler Baustein dafür. Der Aufbau eines Netzwerks *Maschinelle Erschließungsverfahren* soll u. a. durch jährliche, von der DNB organisierte Tagungen gefördert werden.

Kooperationen im Feld (Inhalts-)Erschließung waren und bleiben ein zentrales Thema für die DNB. Der Aufbau von Trainingsdatenbanken, die Neu- oder Weiterentwicklung von Tools, der Aufbau (verteilter) Testumgebungen für die Evaluierung der Werkzeuge und vieles andere mehr lässt sich nur in Kooperationen dauerhaft gut bewältigen. Transparenz und Offenheit sind essenzielle Voraussetzungen, um Vertrauen und Arbeitsteilung zu erreichen. Als Beispiel für einen offenen Datenaustausch seien hier die frei zugänglichen Datensammlungen des 2020 neu eingerichteten *DNBLabs* genannt.²² Beispiele für bestehende oder entstehende Kooperationen sind u. a. die Fortsetzung der Zusammen-

²² Nähere Informationen sind zu finden unter <https://www.dnb.de/DE/Professionell/SerVICES/WissenschaftundForschung/DNBLab/dnblab.html?nn=578704> (4.12.2020).

arbeit mit den Öffentlichen Bibliotheken und der MVB GmbH²³, Absprachen mit der Gruppe der TU9-Bibliotheken²⁴ zu einer verteilten Inhaltserschließung von Publikationen aus Naturwissenschaften und Technik, aber auch die Evaluation des Konzeptes zur Inhaltserschließung mit der Initiativgruppe *Kooperative Erschließung und Informationsversorgung im DACH-Raum am Beispiel der Inhaltserschließung* des Standardisierungsausschusses.

Die DNB zieht nach einem Jahrzehnt maschineller Inhaltserschließung eine insgesamt positive Bilanz. Maschinelle Verfahren können erfolgreich für die Erschließung bibliothekarischer Sammlungen eingesetzt werden. Die DNB ist zuversichtlich, dass Fortschritte in den Feldern Künstliche Intelligenz und Text- und Data-Mining zunehmend für die Anwendung in Bibliotheken fruchtbar gemacht werden können und solche Entwicklungen dazu beitragen, dass Bibliotheken ihre Dienstleistungen verbessern und sich grundlegend enger mit Wissenschafts- und Kultureinrichtungen vernetzen und verbinden können.

7 Literaturverzeichnis

- Alex, Heidrun: Die Dewey-Dezimalklassifikation (DDC). In: *Klassifikationen in Bibliotheken*. Hrsg. von Heidrun Alex, Guido Bee und Ulrike Junger (=Bibliotheks- und Informationspraxis Bd. 53). Berlin/Boston: De Gruyter Saur 2018. S. 65–110.
- Deutsche Nationalbibliothek, Arbeitsstelle für Standardisierung (Hrsg.): *Regeln für die Schlagwortkatalogisierung*. 4. vollst. überarb. Aufl. Frankfurt a. M.: Deutsche Nationalbibliothek 2017. <https://nbn-resolving.org/urn:nbn:de:101-2017011305> (4.12.2020).
- Gömpel, Renate, Ulrike Junger und Elisabeth Niggemann: *Veränderungen im Erschließungskonzept der Deutschen Nationalbibliothek*. In: *Dialog mit Bibliotheken* (2010) H. 1. S. 19–22. <https://nbn-resolving.org/urn:nbn:de:101-2011012858> (14.6.2021).
- Junger, Ulrike und Ute Schwens: *Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft*. In: *Dialog mit Bibliotheken* (2017) H. 1. S. 4–7. <https://d-nb.info/1140661434/34> (4.12.2020).
- Mödden, Elisabeth: *Inhaltserschließung im Zeitalter von Suchmaschinen und Volltextsuche*. In: *b.i.t.online* (2018) Bd. 21 Nr. 1. S. 47–51. <https://www.b-i-t-online.de/heft/2018-01-interview-moedden.pdf> (4.12.2020).
- Mödden, Elisabeth: *Weiterentwicklung der Inhaltserschließung*, In: *Dialog mit Bibliotheken* (2020) H.1. S. 35–37. <https://nbn-resolving.org/urn:nbn:de:101-20200326148> (4.12.2020).

²³ Mödden u. a. 2020.

²⁴ Siehe https://www.ub.uni-stuttgart.de/ubs/kooperationen/Flyer_TU9-Bibliotheken_de.pdf (4.12.2020).

- Mödden, Elisabeth, Christa Schöning-Walter und Sandro Uhlmann: Maschinelle Inhaltser-schließung in der Deutschen Nationalbibliothek. In: BuB (2018) H. 1. S. 30–35. <https://b-u-b.de/wp-content/uploads/2018-01.pdf> (4.12.2020).
- Mödden, Elisabeth und Katrin Tomanek: Maschinelle Sachgruppenvergabe für Netzpublikationen Vom Projekt PETRUS in die Praxis. In: Dialog mit Bibliotheken (2012) H. 1. S. 17–24. <https://nbn-resolving.org/urn:nbn:de:101-2012100834> (4.12.2020).
- Mödden, Elisabeth, Anne Dreger, Klaus Peter Hommes, Letitia Mölck, Loredana Pinna und Daniela Sitte-Zöllner: Der Weg zur Gründung der AG Erschließung ÖB-DNB und die Entwicklung eines maschinellen Verfahrens zur Verschlagwortung der Kinder- und Jugendliteratur mit GND-Vokabular. In: o-bib. Das offene Bibliotheksjournal (2020) Bd. 7 Nr. 4. S. 1–9. <https://doi.org/10.5282/o-bib/5637>.
- Schöning-Walter, Christa: PETRUS: Prozessunterstützende Software für die digitale Deutsche Nationalbibliothek. In: Dialog mit Bibliotheken (2020) H. 1. S. 15–18. <https://nbn-resolving.org/urn:nbn:de:101-2011012844> (4.12.2020).
- Uhlmann, Sandro: Automatische Beschlagwortung von deutschsprachigen Netzpublikationen mit dem Vokabular der Gemeinsamen Normdatei (GND). In: Dialog mit Bibliotheken (2013) H. 2. S. 26–36. <https://nbn-resolving.org/urn:nbn:de:101-20140305238> (4.12.2020).