

Rita Albrecht, Barbara Block, Mathias Kratzer, Peter Thiessen

Quantität als Qualität – Was die Verbände zur Verbesserung der Inhaltserschließung beitragen können

1 Begriffsdefinitionen

Im Standardwerk *Bibliothekarisches Grundwissen*¹ verwendet der Autor für die „inhaltliche Beschreibung und Erschließung“ von Ressourcen den klassischen Begriff *Sacherschließung* und subsumiert darunter die Schlagwortvergabe, die klassifikatorische Sacherschließung sowie weitere Verfahren der Sacherschließung wie maschinelle Verfahren oder von Benutzer:innen vergebene Tags.²

Daneben hat sich in den vergangenen zehn Jahren für den gleichen Sachverhalt auch der Begriff *Inhaltserschließung* herausgebildet. Die Deutsche Nationalbibliothek (DNB) nutzt ihn konsequent für alle von ihr eingesetzten Anwendungen und Verfahren für die inhaltliche Beschreibung ihrer Medien und nennt auch die dafür zuständige organisatorische Einheit „Abteilung Inhaltserschließung“.³

Im nachfolgenden Text werden die beiden Begriffe *Sacherschließung* und *Inhaltserschließung* deshalb ebenfalls synonym verwendet.

2 Quantität als Qualitätskriterium der Inhaltserschließung

Als die DNB im Mai 2017 ankündigte, einen Teil der inhaltlichen Erschließung ihrer Medienwerke noch im Verlaufe des Jahres auf maschinelle Verfahren umzustellen,⁴ ging ein Schrei der Entrüstung durch die deutschsprachige bibliothekarische Fachwelt. Die darauf folgende Diskussion schaffte es durch einen

1 Gantert, Klaus: *Bibliothekarisches Grundwissen*. 9., vollständig aktualisierte und erweiterte Auflage. Berlin, Boston: De Gruyter Saur 2016. <https://doi.org/10.1515/9783110321500>.

2 Gantert 2016, S. 197.

3 <https://www.dnb.de/DE/Ueber-uns/Organisation/organisation.html> (19.11.2020).

4 Grundzüge und erste Schritte der künftigen inhaltlichen Erschließung von Publikationen in der Deutschen Nationalbibliothek. Mai 2017. https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/konzeptWeiterentwicklungInhaltserschliessung.pdf?__blob=publicationFile&v=4 (19.11.2020).

Gastbeitrag des Generaldirektors der Bayerischen Staatsbibliothek, Klaus Ceynowa, sogar bis ins Feuilleton der Frankfurter Allgemeinen Zeitung.⁵ Darin prognostiziert Ceynowa, dass die DNB durch den Einstieg in maschinelle Erschließungsverfahren ihre „Kernaufgabe einer hochqualitativen Inhaltserschließung“ nicht mehr erfüllen werde. Er beklagt, dass Bibliotheken wie die Bayerische Staatsbibliothek ihre Personalkapazitäten im Bereich der Inhaltserschließung deutlich erhöhen müssten, wollten sie die Erschließung der deutschsprachigen Medienzugänge auf dem bisherigen Niveau halten und maschinell erzeugte Daten der DNB nachbearbeiten und korrigieren.

Traditionell nutzen Bibliotheken die Inhaltserschließung der Deutschen Nationalbibliothek nach, indem sie deren Metadaten beziehen und in ihre Datenbestände einspielen. Für die Bibliotheken, die Mitglied in einem Verbundsystem sind, erledigt das im Allgemeinen die zuständige Verbundzentrale durch regelmäßige Importe in die jeweilige Verbunddatenbank. Auf diese Weise werden die in der DNB nach den Regeln für die Schlagwortkatalogisierung (RSWK)⁶ vergebenen Schlagwörter und die Notationen der Dewey Decimal Classification (DDC)⁷ für einen breiten Anwenderkreis verfügbar gemacht. Ob eine Bibliothek diese Daten unbesehen nachnutzt oder gezielt nachbearbeitet, bleibt ihr allein überlassen. Sie entscheidet nach ihren individuellen internen Rahmenbedingungen und Anforderungen, ob Daten aus anderen Sacherschließungssystemen wie z. B. der im deutschsprachigen Raum weitverbreiteten Regensburger Verbundklassifikation (RVK)⁸ ergänzt werden. Ebenso ist es in ihr Ermessen gestellt, ob sie Medien, für die keine Metadaten der DNB zur Verfügung stehen, vollständig selbst erschließt oder auf eine Inhaltserschließung ganz verzichtet, z. B. weil dafür keine Personalkapazitäten vorhanden sind.

Erschließungsdaten können jedoch nicht nur von der DNB bezogen und übernommen werden: Als weitere Quellen bieten sich die Daten der deutschsprachigen Verbundsysteme,⁹ des OCLC WorldCat oder sonstiger bibliografischer Datenbanken an. In der Regel stehen einer Bibliothek entsprechende Re-

5 Ceynowa, Klaus: In Frankfurt lesen jetzt zuerst Maschinen. In: Frankfurter Allgemeine Zeitung, 31.7.2020. <https://www.faz.net/-gr0-909kq> (19.11.2020).

6 Regeln für die Schlagwortkatalogisierung. RSWK. Erarbeitet von der Expertengruppe Sacherschließung im Auftrag des Standardisierungsausschusses. 4., vollständig überarbeitete Auflage, Stand März 2017. Leipzig, Frankfurt a. M.: Deutsche Nationalbibliothek 2017. <https://d-nb.info/1126513032/34> (19.11.2020).

7 https://www.dnb.de/DE/Professionell/DDC-Deutsch/ddc-deutsch_node.html (19.11.2020).

8 <https://rvk.uni-regensburg.de/regensburger-verbundklassifikation-online> (19.11.2020).

9 Bibliotheksverbund Bayern (BVB), Gemeinsamer Bibliotheksverbund (GBV), Hessisches Bibliotheksinformationssystem (hebis), Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz), Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV), Österreichischer Bibliotheksverbund (OBV), Südwestdeutscher Bibliotheksverbund (SWB).

cherche- und Übernahmetools auf Basis von Z39.50 oder SRU zur Verfügung. Recherchen in den deutschsprachigen Verbundsystemen haben in der Regel gute Aussichten auf Treffer mit Elementen der Inhaltserschließung wie RSWK-Schlagwörter oder RVK-Notationen, die zusammen mit den formalen bibliografischen Daten übernommen werden können.

Für Bibliothekar:innen kann somit resümiert werden: Sie sind umso zufriedener, je mehr regelkonforme Vorleistungen für die eigenen Aufgaben der Inhaltserschließung zur Verfügung stehen.

Aus Sicht der Nutzer:innen von Bibliotheken ist Quantität ebenso ein wichtiges Qualitätsmerkmal: Sie müssen davon ausgehen können, dass eine inhaltliche Recherche nach Schlagwörtern oder Klassifikationen den vollständigen Bestand an Medien bzw. Daten in der Bedienoberfläche abdeckt. Tatsächlich ist es aber längst Realität, dass nur noch Teile der Bibliotheksbestände mit Schlagwörtern oder Notationen erschlossen sind. Das resultiert einerseits aus einer selektiven aktiven Inhaltserschließung für Neuzugänge und andererseits aus den Mengen an nicht oder nach anderen Systemen inhaltlich erschlossenen Altbeständen und elektronischen Ressourcen.¹⁰

Tab. 1: Anteil an Beständen mit Inhaltserschließung in ausgewählten Verbunddatenbanken

Verbund-Datenbank	Titel mit Beständen	mit RSWK und RVK	mit RSWK	mit RVK	mit BK	ohne RSWK, RVK, BK ^I
B3Kat ^{II}	26 652 028	4 945 599 (18,56 %)	9 274 005 (34,80 %)	7 396 135 (27,75 %)	— ^{III}	14 927 487 (56,01 %)
hebis ^{IV}	14 091 535	2 497 814 (17,73 %)	4 387 329 (31,13 %)	3 959 162 (28,09 %)	— ^V	8 242 858 (58,50 %)
K10plus ^{VI}	72 799 945	k. A.	12 156 240 (16,70 %)	6 419 446 (8,82 %)	7 315 654 (10,05 %)	57 531 368 (79,03 %)

I Ggf. können die Datensätze aber Elemente der Inhaltserschließung aus anderen Systemen enthalten (z. B. von Verlagen).

II Stand: 1.10.2020.

III Die Basisklassifikation wird im B3Kat nicht angewandt.

IV Stand: 19.7.2019.

V Die Basisklassifikation wird im hebis-Verbund nicht angewandt.

VI Stand 1.1.2020; K10plus klassifikatorisch erschlossene Titel gesamt 20 939 779; verbal erschlossene Titel gesamt 21 470 563. Im GBV wird RVK oft nur im Exemplarbereich erfasst.

¹⁰ Ob die Sacherschließungsleistungen von Verlagen weiterhin pauschal als qualitativ unzureichend betrachtet werden und deshalb ungenutzt bleiben können, wenn der Anteil von *nur noch* elektronisch erscheinenden Ressourcen ein gewisses Limit überschritten hat, bleibt abzuwarten.

Recherchen nach inhaltlichen Kriterien können deshalb stets nur eine Teilmenge der zutreffenden Bestände offenlegen. Leider wird diese Tatsache von den Bibliotheken nicht immer transparent kommuniziert. Auf die Zufriedenheit der Benutzer:innen mit den Bibliotheksdienstleistungen dürfte die Unvollständigkeit der Treffermengen somit kaum Auswirkungen haben. Allerdings müssen sich die Verantwortlichen in den Bibliotheken selbst fragen, ob ein solches „Verbergen“ von Beständen ihrem Selbstverständnis als Dienstleistungseinrichtung entspricht. Ein Ausweg aus dem Dilemma könnte darin bestehen, noch mehr Sacherschließungsdaten aus anderen Quellen zu übernehmen und die Abdeckungsquote damit zu erhöhen. Dabei kommt sowohl eine manuell-intellektuelle Übernahme wie auch eine maschinell-automatisierte in Frage, eine Übernahme von Elementen gemäß bibliothekarischer Standards oder auch die Übernahme von inhaltserschließenden Informationen aus dem nicht-bibliothekarischen Umfeld. Dies erfordert jedoch die Bereitschaft der Bibliotheken, gegebenenfalls Unschärfen und eine gewisse Fehlertoleranz in Kauf zu nehmen, beispielsweise, wenn die Daten nicht den üblichen bibliothekarischen Regeln entsprechen oder der zur maschinellen Übernahme genutzte Algorithmus nicht in allen Fällen eine eindeutige und richtige Zuordnung vornehmen kann.

Eine wichtige Aufgabe der Bibliotheksverbände als Dienstleistungseinrichtungen für die in ihnen zusammengeschlossenen Bibliotheken ist somit die Verfügbarmachung von Sacherschließungsdaten in größerem Mengenumfang – als Vorleistung für Bibliothekar:innen oder zur bewussten Ergänzung von Beständen, die nicht sachlich erschlossen werden können. Dafür können die bewährten Mechanismen des Datenaustauschs zwischen den Verbänden genutzt werden, für den sich aber aufgrund der heterogenen Anforderungen an die verwendeten Sacherschließungssysteme besondere Herausforderungen ergeben. Ergänzend dazu müssen auch neue Wege beschritten werden, mit denen sich der Abdeckungsgrad an Sacherschließungsdaten signifikant steigern lässt, ohne zu große inhaltlich-qualitative Abstriche machen zu müssen.

3 Aktuelle Praxis und Herausforderungen

3.1 Praxis

3.1.1 Sacherschließungssysteme

Die Praxis der Inhaltserschließung ist von Heterogenität geprägt – sowohl innerhalb eines Verbundes als auch über Verbundgrenzen hinweg. So autorisieren die Verbünde bestimmte Systeme der Inhaltserschließung als Verbundstandard und den Verbundbibliotheken wiederum steht es frei, welche(n) dieser Standards sie nutzen oder bedienen wollen. Für die als Standard definierten Systeme werden oft Fremddatenübernahmen garantiert und Supportmaßnahmen gewährleistet, beispielsweise durch die zentrale Bereitstellung und Pflege unterstützender Tools für die aktive Erschließungsarbeit oder indem regionale Spiegel überregionaler Normdateien (RVK, GND) zur Verfügung gestellt werden.

Einheitlich über sämtliche Verbundgrenzen hinweg wird lediglich die verbale Inhaltserschließung nach den *Regeln für die Schlagwortkatalogisierung* als Standard definiert. Alle Verbünde führen dazu unterstützende Spiegel der Gemeinsamen Normdatei (GND) in ihren Verbunddatenbanken mit, die den Thesaurus für die Schlagwortvergabe bildet. Allerdings beteiligen sich nicht alle Bibliotheken in den Verbänden an der Schlagwortvergabe nach RSWK; insbesondere Spezialbibliotheken mit bestimmten fachlichen Ausrichtungen bevorzugen oft andere Thesaurus-Systeme zur inhaltlichen Erschließung ihrer Bestände.

Innerhalb der letzten beiden Jahre wurden in den Verbänden BVB, GBV, hebis, KOBV und SWB Umfragen zur Praxis der Sacherschließung in den teilnehmenden Bibliotheken durchgeführt.

Zusammengefasst ergab sich folgendes Bild:

hebis:

- Von 60 angefragten Bibliotheken/Bibliothekssystemen haben 53 Daten bereitgestellt.
- Von den 53 Bibliotheken, die ihre Daten bereitgestellt haben, betreiben 46 aktiv Sacherschließung.
- Die drei häufigsten Erschließungsverfahren sind (Mehrfachnennungen möglich):
 - Gemeinsame Normdatei (GND): 28 Bibliotheken
 - Regensburger Verbundklassifikation (RVK): 10 Bibliotheken
 - Hauseigene Erschließungssysteme: 5 Bibliotheken

BVB/KOBV:

- Von 237 angefragten Bibliotheken/Bibliothekssystemen haben 127 Daten bereitgestellt.
- Von den 127 Bibliotheken, die ihre Daten bereitgestellt haben, betreiben 127 aktiv Sacherschließung.
- Die drei häufigsten Erschließungsverfahren sind (Mehrfachnennungen möglich):
 - Gemeinsame Normdatei (GND): 83 Bibliotheken (Einzelschlagwörter oder Schlagwortfolgen)
 - Hauseigene Erschließungssysteme: 61 Bibliotheken
 - Regensburger Verbundklassifikation (RVK): 49 Bibliotheken

GBV/SWB:

- Es wurde eine Online-Befragung der Bibliotheken in GBV und SWB durchgeführt, geantwortet haben 129 Bibliotheken (41 aus dem GBV, 85 aus dem SWB).
- Von diesen Bibliotheken betreiben aktiv 90 Bibliotheken verbale und 107 Bibliotheken eine klassifikatorische Erschließung ihrer Bestände.
- Die drei häufigsten Erschließungsverfahren sind (Mehrfachnennungen möglich):
 - Gemeinsame Normdatei (GND) in 76 Bibliotheken
 - Hauseigene Erschließungssysteme (Haus-Systematiken in 61 Bibliotheken, Haus-Thesauri in 19 Bibliotheken)
 - Regensburger Verbundklassifikation (RVK): in 41 Bibliotheken

3.1.2 Sacherschließungstools

In den Verbänden und Bibliotheken stehen verschiedene Tools zur Verfügung, die die Bibliothekar:innen bei der praktischen Sacherschließungsarbeit unterstützen und zumindest eine halbautomatisierte Arbeitsweise ermöglichen. Wir stellen zwei Beispiele vor.

3.1.2.1 Digitaler Assistent (GBV und SWB)

Der Digitale Assistent¹¹ ist ein Tool, das die intellektuelle Sacherschließung im K10plus unterstützt. In einem Gemeinschaftsprojekt der Bibliotheksverbände

¹¹ Beckmann, Regine, Imma Hinrichs, Melania Janßen, Gerard Milmeister und Peter Schäuble: Der Digitale Assistent DA-3. In: o-bib (2019) Bd. 6 Nr. 3, S. 1–20. <https://doi.org/10.5282/o-bib/2019H3S1-20>.

GBV und SWB mit der Firma Eurospider wird der im SWB eingesetzte DA2 zum DA3 weiterentwickelt, um die Möglichkeiten zu erweitern und um das Handling zu verbessern.

Der DA3 ist webbasiert und unterstützt Bibliotheken bei der verbalen und klassifikatorischen Sacherschließung. Die Bearbeiter:innen starten mit einer Recherche über SRU im K10plus. Der gesuchte Titelsatz wird in das Tool geladen, wobei wesentliche Felder der Formalerschließung und die bereits vorhandene Sacherschließung in den DA3 übernommen und angezeigt werden. Ausgehend vom zu erschließenden Titel werden dann durch das Tool Fremddaten (wie der B3Kat und die Kataloge von OBV, DNB und Library of Congress – LoC) auf vorhandene Erschließungselemente durchsucht und Vorschläge für Schlagwörter und Notationen zur Verfügung gestellt. Es folgt die intellektuelle Prüfung der Vorschläge und ggf. deren Modifizierung.

Können keine Vorschläge generiert werden oder werden die Vorschläge abgelehnt, besteht die Möglichkeit, den Titel intellektuell zu erschließen. Hierfür kann im DA3 auf verschiedene Normdateien – wie GND, RVK und Basisklassifikation; im Testbetrieb auch den Standard-Thesaurus Wirtschaft (STW) – zugegriffen werden. Danach werden die Daten in den K10plus zurückgeschrieben.

Im Projekt liegt die fachliche Koordination bei den Projektbibliotheken (für den SWB die Universitätsbibliothek Stuttgart und für den GBV die Staatsbibliothek zu Berlin). Die Firma Eurospider ist verantwortlich für die Systementwicklung und die Verbundzentralen von SWB und GBV haben die Bereitstellung der Schnittstellen für die Katalogsuche und für den Re-Import der um Sacherschließungselemente angereicherten Titeldatensätze übernommen. Die Daten werden vom DA3 in MARXML zur Verfügung gestellt. Für den Import in den K10plus¹² werden sie alle 70 Sekunden abgefragt und in das Pica-Internformat umgesetzt. Für die Ergänzung der Sacherschließung in die bestehende Titelaufnahme waren besondere Anpassungen erforderlich, da die im K10plus vorhandene Sacherschließung auf bibliografischer Ebene nicht überschrieben werden sollte. Für das Merging auf Lokal- und Exemplarebene wurden bibliotheksspezifische Anpassungen (zum Beispiel für die SBB Berlin, die ZBW – Leibniz-Informationszentrum Wirtschaft oder den Index Theologicus) implementiert. Die über den DA3 generierte Sacherschließung wird beim Import in den K10plus durch den Code DA-3 gekennzeichnet und ist so in der Katalogisierungsdatenbank erkennbar.

¹² Kühn, Armin und Andreas Krausz: Technik des DA-3. Anschluss an K10plus. 4. Workshop Computerunterstützte Inhaltserschließung am 11. und 12. November 2020. <https://wiki.dnb.de/x/ew-VCg> (19.11.2020).

3.1.2.2 Beschlagwortungsprogramm/hebis-SET

Im hebis-Verbund wird bereits seit vielen Jahren ein Sacherschließungstool eingesetzt, das an der Universitätsbibliothek Frankfurt entwickelt wurde. Der Name Beschlagwortungsprogramm macht seine ursprüngliche Bestimmung deutlich: Konzipiert wurde es zunächst ausschließlich für die Vergabe von Schlagwörtern nach RSWK. Nachdem jedoch in den 2000er Jahren mehr und mehr hebis-Bibliotheken mit (Teil-)Aufstellungen ihrer Bestände nach der Regensburger Verbundklassifikation (RVK) begonnen hatten, wurde auch die Vergabe von RVK-Notationen und die Generierung von Signaturen daraus als weitere Funktion in das Tool integriert.

Das Beschlagwortungsprogramm war zunächst ein eigenes Anwendungsprogramm. Inzwischen wurde es jedoch vollständig in die hebis-Version der Client-Software WinIBW integriert; die Funktionen werden standardmäßig mit den Installationsdateien ausgeliefert. Die Vergabe und Bearbeitung von Schlagwörtern bzw. von RVK-Notationen erfolgt jeweils in einem eigenen Fenster. Fremdleistungen vorausgewählter Verbünde können über Z39.50-Recherchen gesucht und übernommen werden. Bei der Übernahme werden Schlagwortfolgen automatisch in Einzelschlagwörter zerlegt, da der hebis-Verbund diese Praxis verbundintern bereits Ende der 1990er Jahre eingeführt hat.

Stehen keine Fremdleistungen zur Verfügung, können aus dem Bearbeitungsfenster heraus Suchen in den GND-Normdaten sowie in der RVK-Online ausgelöst und die passenden Begriffe bzw. Notationen ausgewählt werden.

Für die Schlagwortsuche wird im Hintergrund der in die Verbunddatenbank integrierte Spiegel der GND benutzt. Wenn die Bearbeitung des Titelsatzes von den Sacherschließenden abgeschlossen wird, werden die neu vergebenen Schlagwörter automatisch dem Titeldatensatz zugefügt und mit dem entsprechenden Normdatensatz in der hebis-Datenbank verknüpft.

Bibliotheken, die nach der RVK aufstellen, können aus den Notationen, die für einen Titel verfügbar sind oder neu vergeben wurden, im Tool direkt eine gültige Individualsignatur erzeugen. Dazu wird eine Notation ausgewählt und es werden die erforderlichen weiteren Angaben gemacht. Anschließend steht eine Validationsprüfung zur Verfügung, mit der sich die Vollständigkeit der Signatur prüfen lässt. Ist die Signatur korrekt gebildet, kann sie direkt in das dafür vorgesehene Exemplar in der Verbunddatenbank übernommen werden.

Inzwischen hat die hebis-Verbundzentrale damit begonnen, die Funktionen des Beschlagwortungsprogrammes in ein neues, webbasiertes Tool zu überführen. Damit begegnet sie zum einen der ungewissen Lebensdauer der WinIBW und bereitet zum anderen eine Systemmigration vor, die in den kommenden Jahren ansteht. Die erste Version des neuen Sacherschließungstools hebis-SET steht bereit und enthält die Funktionen für die Vergabe von RVK-Notationen

und -Signaturen. Der Funktionsumfang der WinIBW-Version wurde bei der Übertragung voll erhalten. Das Roll-out in die hebis-Bibliotheken hat begonnen.

3.1.3 Nachnutzung von Sacherschließungsdaten

Nicht erst seit den Evaluationen durch die Deutsche Forschungsgemeinschaft (DFG) und den Wissenschaftsrat¹³ und deren Empfehlungen sowie ihrer Kritik an tatsächlichen und vermeintlichen Ineffizienzen in der bibliothekarischen Verbundlandschaft werden technische Möglichkeiten des Datenaustauschs über die reine Fremddatenübernahme hinaus genutzt, um auch nachträglich Sacherschließungselemente in Datensätzen anzureichern, die – prinzipiell erst einmal autark – in ihren Verbunddatenbanken für sich existieren. Die Grundsatzdiskussion über Pro und Contra eines föderalen Flickenteppichs an Verbundkatalogen gegenüber einer wünschenswerten zentralen Katalogisierungsplattform (mag diese nationaler oder internationaler Natur sein) soll hier weniger interessieren als die Tatsache, dass auch bei verteilter Datenhaltung Synergieeffekte im Bereich Sacherschließung durch maschinelle und z. T. völlig automatisiert ablaufende Prozesse möglich sind.

Beispielhaft ist hier die maschinelle Schlagwortnachführung mit dem Merging von RSWK-Schlagwörtern in den eigenen Titelaufnahmen, die jeder Verbund durchführt, zu nennen. Hierbei werden durchaus unterschiedliche Strategien verfolgt bzw. Verfahren angewandt. So werden für den B3Kat im Rahmen von Datenübernahmen aus anderen Verbänden Anreicherungen nur für diejenigen Sacherschließungselemente (RVK-Notationen, Formangaben oder Schlagwortfolgen bzw. Einzelschlagwörter) vorgenommen, die in den eigenen Sätzen noch nicht vorhanden sind. Abweichend dazu werden allerdings die Schlagwortnachführungen aus DNB-Daten behandelt, die additiv zu evtl. bereits vorhandenen RSWK-Elementen erfolgt. Dies ist darauf zurückzuführen, dass einerseits viele Bibliotheken bei Titeln, die in das Sammel- und Erschließungsspektrum der DNB fallen, bewusst keine Inhaltserschließung vornehmen und explizit auf Nachnutzung der DNB-Daten setzen. Andererseits wünschen Bibliotheken daneben die Möglichkeit, auch eigene Sacherschließung einbringen zu können, die im Detail evtl. von der DNB-Vergabe abweicht oder diese ergänzt.

¹³ Gemeinsame Erklärung der Deutschen Forschungsgemeinschaft und des Wissenschaftsrats zur Zukunft der Bibliotheksverbände als Teil einer überregionalen Informationsinfrastruktur in Deutschland. <http://www.wissenschaftsrat.de/download/archiv/1003-11.pdf> (19.11.2020).

Eine Grundkonstellation bzw. -problematik, der man sich bei Anreicherungsprozeduren und Fremddatenübernahmen jedoch immer bewusst sein muss, ist die Tatsache, dass es die verschiedenen „Sprachen“ der Intern- und Austauschformate gibt (siehe dazu auch Abschnitt 3.2.1). Diese machen beim Import und Export von Datensätzen „Übersetzungen“ notwendig – mit nicht immer vermeidbaren Kollateralschäden. So geht beispielsweise die syntaktische Struktur bei Schlagwortfolgen verloren, wenn damit ausgestattete Datensätze durch Verbünde importiert werden, die nur Einzelschlagwörter in ihrer Verbunddatenbank halten. Ebenso ist ein Informationsverlust zu konstatieren, wenn die analytische Struktur von synthetischen DDC-Notationen, d. h. ihre Zerlegung in einzelne Notationsbestandteile, in einem MAB2-orientierten Internformat nicht exakt analog zum Austauschformat MARC 21 abbildbar ist. Solche Einschränkungen und Fallstricke lassen sich nur teilweise auf technische Rahmenbedingungen zurückführen. Vielmehr sind sie von der Auswahl zwischen möglichen „Übersetzungsvarianten“ (wie wird importiert) und Entscheidungen darüber, was überhaupt „übersetzt“ (importiert und exportiert) werden soll, sowie von den Sacherschließungsleitlinien der Bibliotheken bzw. der jeweiligen Verbundgremien abhängig.

Oft können Sacherschließungselemente bereits im Rahmen der Formaler-schließung bei Fremddatenübernahmen in der Verbunddatenbank mitgespeichert werden. Abhängig vom Importweg (manuelle Einzelübernahme via Online-Schnittstelle oder Batchimport von größeren Datenmengen wie bspw. E-Book-Paketen) und Quell- bzw. Liefersystem (Verbünde, DNB, Library of Congress oder OCLC WorldCat) kommen hierbei z. T. individuell entwickelte Konversionsroutinen zum Einsatz, um den jeweiligen Dialekten des Lieferformats gerecht zu werden. Abgesehen von regelmäßigen Formaterweiterungen und -änderungen, die laufende Anpassungen der Konversionsroutinen erforderlich machen, stellen diejenigen Sacherschließungselemente ein besonderes Problem dar, bei denen Transport- und Internformat stark divergieren.

Zu Zeiten von MAB2 war dies z. B. für die Library of Congress Subject Headings (LCSH) der Fall. Diese werden von OCLC oder LoC in MARC 21 wie folgt ausgeliefert:

```
650#0 $a Representative government and representation
      $z Latin America.
650#0 $a Political participation $z Latin America.
650#0 $a Public opinion $z Latin America.
650#0 $a Presidents $z Latin America.
650#0 $a Executive power $z Latin America.
650#0 $a Political planning $z Latin America.
650#0 $a Policy sciences $z Latin America.
```

Erkennbar ist die granulare Unterfeldstruktur für die mehrgliedrigen Begriffe und die Identifikation des Thesaurus (LCSH) über den zweiten Indikator 0.

Solange eine ähnliche Feldstruktur in der Verbunddatenbank für LCSH noch nicht vorhanden war bzw. man sich mit Fokus auf deutschsprachige Sacherschließungsbegriffe für LCSH nur sekundär interessierte, wurden die englischsprachigen Begriffe (beim Import) mit Granularitätsverlust in einem Unterfeld des Internformats zusammengeführt, was Stichwortsuchen weiterhin uneingeschränkt zulässt. Folglich können sie heute in Altdaten oder für Daten, bei denen noch alte Konversionsroutinen im Einsatz sind, nur mit entsprechendem Qualitätsverlust wieder nach MARC 21 rückübersetzt und ausgeliefert werden. Für den B3Kat ergibt sich beim Export beispielsweise:¹⁴

```
650#4 $aRepresentative government and representation / Latin America
650#4 $aPolitical participation / Latin America
650#4 $aPublic opinion / Latin America
650#4 $aPresidents / Latin America
650#4 $aExecutive power / Latin America
650#4 $aPolitical planning / Latin America
650#4 $aPolicy sciences / Latin America
```

Erkennbar ist außerdem die fehlende Kennzeichnung des Thesaurus: nun 4, d. h. *Source not specified* als zweiter Indikator. Bei einer weiteren Nachnutzung der B3Kat-Daten können die verloren gegangenen Informationen also kaum mehr rekonstruiert werden.

3.2 Herausforderungen

3.2.1 Babylonische Formatverwirrung

Für den Datenaustausch haben sich die Verbünde im deutschsprachigen Raum auf ein einheitliches MARC-21-Austauschformat verständigt. Um dies bedienen zu können, haben sie ihre Export- und Importschnittstellen entsprechend angepasst, denn in den einzelnen Verbänden kommen unterschiedliche Bibliothekssysteme mit unterschiedlichen Internformaten zum Einsatz.

Die Pica-Verbünde setzen eine Software von OCLC mit proprietärem Format ein. Dieses kann anwenderseitig auf die individuellen Bedarfe angepasst werden und hat sich in den ca. 30 Jahren seit Einführung jeweils stark verändert. BVB und KOBV haben als Aleph-Verbünde das Format Aleph Sequential (ASEQ)

¹⁴ Ähnliches (mit Semikolon statt Schrägstrich als Trennzeichen) gilt für den Export von Altdaten aus dem K10plus.

der Firma Ex Libris, das im deutschsprachigen Markt MAB2-basiert ist. Dieses ist derzeit auch beim hzb im Einsatz, wird aber mit der Einführung von Alma (wie Aleph ebenfalls ein Produkt der Firma Ex Libris) durch ein MARC-basiertes Format abgelöst werden.

Die Formate der Pica-Verbünde wurden für bestimmte Anforderungen an die Sacherschließung in den jeweiligen Verbänden unterschiedlich erweitert.

Im hebis-Format gibt es u. a. Felder für:

- Klassifikation der Hessischen Bibliographie
- Klassifikation der BDSL (Bibliographie der deutschen Sprach- und Literaturwissenschaft)
- Klassifikation der BLL (Bibliography of linguistic literature)
- Schlagwörter der BDSL (Bibliographie der deutschen Sprach- und Literaturwissenschaft)
- Schlagwörter für Schulprogramme

Im K10plus-Format (GBV und SWB) gibt es u. a. Felder für:

- verschiedene Systematiken von Öffentlichen Bibliotheken
- Schlagwörter des STW
- Erschließung von Musikalien nach Besetzung und Form bzw. Gattung

Für den B3Kat (BVB und KOBV) sind ergänzend dazu dedizierte Felder für bspw. die Systematiken der Bayerischen Bibliographie oder Haussystematiken bayerischer Bibliotheken zu nennen.

Die Erweiterung der Formate hat dazu geführt, dass Sacherschließungsfelder beim Datenausch der Verbünde bislang nicht alle in gleicher Weise berücksichtigt werden. Es liegt in der Hoheit des empfangenden Verbundes zu selektieren, welche Daten übernommen werden.

Letztlich gibt es aber nur einen kleinen Kanon von inhaltserschließenden Angaben, die alle Verbünde ausliefern und übernehmen. Diese sind:

- RVK-Notationen
- DDC-Notationen
- Schlagwortfolgen bzw. Einzelschlagwörter nach RSWK

Während die Verwendung der GND bei der Sacherschließung schon lange in allen Verbänden etabliert ist, gilt dies für die RVK-Normdatei, die erst seit 2016 mit eindeutig referenzierbaren IDs ausgeliefert wird, noch nicht. Inzwischen ist die RVK aber sowohl im B3Kat als auch in hebis und im K10plus als Normdatei verfügbar und es können beim Datenexport IDs mitausgeliefert werden. Diese machen eine Datenübernahme bzw. ein Update sehr viel einfacher.

Das Format für den Austausch von RSWK-Schlagwortfolgen wurde beim MARC-Umstieg ausführlich diskutiert und festgelegt. Um die Folgen austauschen zu können, wurde das anwenderspezifische MARC-21-Feld 689 definiert.¹⁵

Für den Austausch der LCSH stehen im MARC-21-Format die Felder 600, 610, 611, 630, 647, 648, 650 und 651 zur Verfügung. Diese Felder wurden u. a. im K10plus-Format nachgebildet, damit die LCSH möglichst unverändert übernommen werden können. Im GBV und SWB gab es zunächst eine sehr viel größere Feldstruktur im Internformat. Dies hat zur Folge, dass die LCSH aus Altdaten nicht mit dem vollständigen MARC-21-Unterfeldspektrum ausgeliefert werden können. Im B3Kat-Format ist das Feld für die LCSH am MAB2-Format orientiert, bietet mittlerweile jedoch die Möglichkeit, das gesamte MARC-21-Unterfeldspektrum korrekt ablegen zu können: Der Entitätstyp wird dazu, anders als in MARC 21 nicht über die Kategoriennummer, sondern den Indikator gesteuert. Diese Gegebenheiten haben insgesamt zur Folge, dass die LCSH zwischen den Verbänden nur in unterschiedlicher Granularität ausgetauscht werden können.

Bei einer Datenübernahme aus dem OCLC WorldCat sind für die Sacherschließung hauptsächlich DDC, LCC (Library of Congress Classification) und LCSH zu erwarten, jedoch keine Schlagwortfolgen nach RSWK (MARC 689) – selbst dann nicht, wenn man Datensätze der DNB oder der deutschen Verbände auswählt. Grund dafür ist, dass OCLC anwenderspezifische Felder als lokale bibliografische Informationen speichert und nur an die liefernde Bibliothek wieder exportiert.

Um die Datenübernahmen zwischen den deutschsprachigen Verbänden besser nachvollziehen zu können, wurde eine Provenienzkennzeichnung vorgeschlagen. Hierbei wäre zu bedenken, dass diese Datenübernahme dezentral und gelegentlich über mehrere Stationen, somit nicht notwendigerweise direkt vom ersterfassenden Verbund erfolgt. Auch innerhalb der Verbände arbeiten über die Zeit hinweg in der Regel stets mehrere verschiedene Bibliotheken kooperativ an einem Datensatz. Allein die (jeweils letzte) liefernde Institution des gesamten Datensatzes festzuhalten, ist daher nicht ausreichend. Um ein kohärentes Hilfsmittel zur Qualitätssicherung darzustellen, müsste eine Provenienzkennzeichnung vielmehr feldbezogen sein und die erfassende Institution, jede überarbeitende Institution sowie das zur Erzeugung bzw. Überarbeitung des Feldinhalts eingesetzte Verfahren dokumentieren. Diese im Allgemeinen längeren Ketten von Provenienzen würden zusammen jedoch einen eigenen *Meta-Metadatenatz* bilden, der sich keinesfalls allein über multiple MARC-21-Felder

¹⁵ <https://wiki.dnb.de/display/MARC21AE/Anwenderebene+Titeldaten+Feld+689> (19.11.2020).

883 adäquat abbilden lässt. Letztere sind aber zurzeit beim Austausch über MARC 21 die einzige feldbezogene Kennzeichnungsoption für Datenprovenienzen.

Eine Provenienzkennzeichnung ist bisher beim Datenausch unter den Verbundsystemen und der DNB nur zur Kennzeichnung der Datenübernahme im Rahmen von Culturegraph¹⁶ verpflichtend festgelegt worden. Sie dient der Unterscheidung zwischen ggf. intellektuell neu hinzugefügten und lediglich von mehreren Verbänden maschinell aus Culturegraph-Bündeln übernommenen Elementen.

3.2.2 Probleme durch bzw. bei maschineller Übernahme von Sacherschließung

Anhand welcher Kriterien ein passender Quellsatz für die Übernahme von Sacherschließungselementen ermittelt wird, ist nur ein Faktor, der Einfluss auf die Qualität des Ergebnisses nimmt. Mindestens ebenso wichtig ist aber auch die gewählte Strategie. Selbst die denkbar einfachste Variante *Anreichern, wo der eigene Datensatz noch nichts aufweist* hat bereits ihre Tücken: Sie bedarf zwar keines aufwendigen Abgleichs, ob ein bestimmtes Schlagwort bzw. eine bestimmte Notation schon im Zielsatz vorhanden ist, und erhöht in Masse vollzogen (wie etwa beim sogenannten Pfeffer-Verfahren¹⁷ oder bei Übernahmen aus Culturegraph-Bündeln) schlagartig den Abdeckungsgrad mit Sacherschließung. Man übernimmt damit aber natürlich auch unbesehen jeden Erschließungsfehler, was problematisch ist, da dieser sich mit keinem Übernahmeverfahren nach derselben Strategie mehr korrigieren lässt. Stattdessen müsste hierzu ein erheblich aufwendigeres Verfahren implementiert werden, das Updates in Abhängigkeit von der Provenienz der übernommenen Daten steuern kann.

Für alle Verbünde, deren Verbundsystem die angeschlossenen Lokalsysteme noch mit einer Versorgungsschnittstelle bibliografisch auf dem Laufenden hält, können Massendatenmanipulationen grundsätzlich auch immer ein Durchsatzproblem aufwerfen. Typischerweise drängen sich solche Aktionen latent an den Wochenenden, doch, wenn wie bei einem periodischen Verfahren zur Schlagwortnachführung a priori nicht bekannt ist, wie viele Datensätze es

¹⁶ https://www.dnb.de/DE/Professionell/Standardisierung/AGV/_content/culturegraph_akk.html (19.11.2020).

¹⁷ Pfeffer, Magnus: Automatische Vergabe von RVK-Notationen mittels fallbasiertem Schließen. In: Wissen bewegen – Bibliotheken in der Informationsgesellschaft. 97. Deutscher Bibliothekartag in Mannheim 2008. Hrsg. von Ulrich Hohoff. Frankfurt a. M.: Klostermann, 2009. S. 245–254.

im nächsten Lauf betroffen wird, ist schlichtweg nicht mehr planbar, ob der unvermeidliche Rückstau auf der Versorgungsschnittstelle sich bis zum Montag nach dem Wochenende wieder aufgelöst haben wird.

3.2.3 Maschinelle Prüfung von „Qualität“

Die Frage, ob der Inhalt eines Feldes oder Unterfeldes zur Felddefinition passt, mag zwar auf den ersten Blick formaler Natur erscheinen, ist allerdings bei etwas genauerer Betrachtung in vielen Fällen eher intellektuelle Auslegungssache. Eine Eigenschaft, die sich maschinell in bestimmten Feldern oder Unterfeldern sehr effizient und verlässlich überprüfen lässt, ist die *syntaktische* Korrektheit, also zum Beispiel, ob in dem ausschließlich für RVK-Notationen vorgesehenen Feld eine Zeichenkette steht, die auch wie eine RVK-Notation aufgebaut ist.

Erheblich aufwendiger wird es, wenn man bei fehlender Normdateiverknüpfung prüfen möchte, ob die fragliche Zeichenkette auch wirklich eine gültige RVK-Notation ist, weil man sie dazu prinzipiell gegen alle aktuell(!) gültigen Systemstellen vergleichen muss. Eine analoge Herausforderung stellt die Prüfung eines unverknüpften Schlagworts auf Zugehörigkeit zu einem bestimmten Vokabular (RSWK, LCSH, Medical Subject Headings – MeSH usw.) dar. Beide Probleme sind bereits in nicht ganz unerheblichem Maß *semantischer* Natur und deshalb nur entsprechend aufwendig maschinell zu lösen – auch wenn sie noch mindestens eine Komplexitätsstufe unterhalb der klassischen Qualitätsfragen im Kontext der Diskussion um maschinell generierte Sacherschließung anzusiedeln sind:

- Wie *zutreffend* ist ein vorliegendes Sacherschließungselement für eine gegebene Ressource?
- Wie *präzise* ist ein vorliegendes Sacherschließungselement für eine gegebene Ressource?
- Kategorisiert ein vorliegendes Sacherschließungselement eine Ressource *besser* als ein anderes?

Eine maschinelle Beantwortung dieser Fragen mit gemeinhin akzeptabler Konfidenz ist derzeit Zukunftsmusik und bleibt erst noch zu entwickelnden künstlichen Intelligenzen vorbehalten. Richten wir unseren Blick deshalb im Folgenden wieder zurück auf das, was technisch heute schon im Bereich des Machbaren liegt.

Ein oft beklagter Qualitätsmangel, der sich beim Austausch bzw. der Nachführung von Sacherschließung mit bzw. aus anderen Verbunddatenbanken

ohne entsprechende Vorkehrungen schnell einschleichen kann, ist die Duplizität von Schlagwortung oder Notationen. Auch hier ist die Forderung wieder weitaus einfacher formuliert als im Detail umgesetzt:

Während eine in der Titelaufnahme bereits vergebene Notation durch den schlichten paarweisen Vergleich von Zeichenketten treffsicher ignoriert werden kann, lassen sich zwei gegebene Schlagwortfolgen maschinell im Allgemeinen nur dann zuverlässig als *gleich* (meint hier in der Regel: identisch bis auf äquivalente Permutation) erkennen, wenn vorab alle Folgenglieder derselben Entität alphabetisch sortiert werden.

Als Königsdisziplin darf schließlich gelten, konkordante Schlagwörter aus unterschiedlichen Vokabularen gegeneinander zu deduplizieren oder – der umgekehrten Strategie folgend – als Zeichenkette identische Schlagwörter koexistieren zu lassen, sofern sie aus unterschiedlichen Vokabularen stammen.

Die angeführten Beispiele zeigen: Je komplexer der Sachverhalt ist, der *Gleichheit* ausmacht, desto aufwendiger gestaltet sich nicht nur die Implementierung seiner maschinellen Überprüfung, sondern auch die Formulierung insgesamt konsistenter Vorgaben dafür.

4 Potenziale

Die aktuellen Dienstleistungen der Verbände für die Inhaltsschließung decken die seitens ihrer Verbundbibliotheken gestellten Anforderungen ab. Der Aufwand, der in den Verbundzentralen für Entwicklung und Pflege solcher Dienstleistungen erbracht wird, steht im Gesamtkontext sämtlicher ihrer Dienstleistungen und spiegelt daher implizit den Stellenwert wider, welcher der Sacherschließung innerhalb des jeweiligen Verbundes eingeräumt wird. Es ist jedoch durchaus vorstellbar, dass die Services für Sacherschließung intensiviert und ausgebaut werden, wenn in den Verbänden der strategische Wille dazu besteht. Im Folgenden werden einige Optionen dafür dargestellt.

4.1 Erstkatalogisierungs-ID als Anker für vermehrte und gezielte Datenübernahmen

Im Jahr 2006 verständigten sich die deutschsprachigen Verbände und die DNB in der AG Kooperative Verbundanwendungen (AGKVA) auf die Einführung eines verbundübergreifenden Identifikators für bibliografisch nachgewiesene Ressourcen: den Erstkatalogisierungs-Identifizier (EKI). Er sollte eine zuverlässige

und fortwährende maschinelle Nachführung der Sacherschließung zwischen den Verbänden ermöglichen. Wird eine Titelaufnahme samt EKI durch eine Institution in Verbund A aus der Titeldatenbank von Verbund B übernommen, so können zu jedem beliebigen späteren Zeitpunkt allein anhand des übereinstimmenden EKI sowohl aus Verbund B nach Verbund A als auch umgekehrt (und sogar zwischen zwei beliebigen Verbänden, die die Titelaufnahme aus Verbund B nachgenutzt haben!) nachträglich hinzugefügte Sacherschließungselemente maschinell übernommen werden. Es lohnt sich also nicht nur für den eigenen Verbund, die anderen Verbunddatenbanken gründlich nach einer exakt auf die vorliegende Manifestation passenden Titelaufnahme abzusuchen. Leider werden jedoch in der Praxis gerade für Folgeauflagen oder andere Ausgaben Kopien bereits vorhandener Katalogisate hergestellt und angepasst. Das mag bezogen auf den Einzelfall Recherche- und Bearbeitungszeit sparen, bedeutet aber in einer überregionalen Sichtweise eine Mehrarbeit bei der Sacherschließung. Das liegt daran, dass Abgleiche nach anderen Identifiern als dem EKI oder gar nach Ansetzungen von Titeln und Autor:innen grundsätzlich mit weitaus mehr Aufwand verbunden und zudem teilweise mit inhärenten Unschärfen behaftet sind.

Tab. 2: EKI-Überschneidungsmengen und -quoten zwischen den Verbunddatenbanken bezogen auf die jeweilige Gesamtanzahl von EKIs^{*}

mit Überschnei- dung von	B3Kat	hbz	hebis	K10plus	Nachnutzung- spotenzial**
B3Kat		1 498 657 (5,11 %)	1 006 237 (3,43 %)	2 223 239 (7,59 %)	< 16,13 %
hbz	1 498 657 (7,49 %)		1 187 321 (5,93 %)	3 189 695 (15,93 %)	< 29,35 %
hebis	1 006 237 (5,41 %)	1 187 321 (6,39 %)		2 481 715 (13,35 %)	< 25,16 %
K10plus	2 223 239 (3,27 %)	3 189 695 (3,66 %)	2 481 715 (4,70 %)		< 11,63 %

^{*} Quelle: Gemeinsamer Verbände-Index (GVI), Stand: 11.11.2020. Die Autor:innen danken Stefan Lohrum von der KOBV-Zentrale herzlich für die Bereitstellung der dieser Auswertung zugrunde liegenden EKI-Listen.

^{**} Das tatsächliche *Nachnutzungspotenzial* ist gegeben durch die Menge der EKIs, die in mindestens einer anderen Verbunddatenbank ebenfalls vorkommen. Die zeilenweise gebildete Summe über die paarweisen Überschneidungsquoten ist dafür lediglich eine obere Schranke, weil etwa der B3Kat und die hbz-Datenbank z. T. dieselben EKIs gemeinsam haben wie der B3Kat und der K10plus.

Weshalb fallen die Überschneidungsquoten nach wie vor eher niedrig aus? Bei einer Bewertung ist zunächst einmal zu berücksichtigen, dass der EKI erst ab dem Jahr 2007 seine Wirkung entfalten konnte. Selbst wenn Verbünde für zu diesem Zeitpunkt bereits vorhandene Titelaufnahmen rückwirkend EKIs erzeugt haben, wurden diese außer bei Retrokonversionsprojekten allenfalls punktuell in andere Verbunddatenbanken übernommen. Für die in ihrer Zahl stark zunehmenden E-Book-Aufnahmen werden außerdem zwar Metadatenlieferungen der Verlage teilweise von einer Verbundzentrale für alle anderen aufbereitet, dann aber von jedem Verbund mit eigenen EKIs eingespielt, weil die Bibliotheken allenthalben eine möglichst schnelle Bereitstellung der Titelnachweise erwarten.

Ein weiterer Grund für die niedrigen Überschneidungsquoten ist schließlich die Tatsache, dass der OCLC WorldCat als weltweit größte bibliografische Nachweisdatenbank eine Quelle für Fremddatenübernahmen ist, die sich in den letzten Jahren in einigen Verbänden stetig steigender Beliebtheit erfreut. So kamen laut aktueller Übernahmestatistik des B3Kat im Zeitraum Januar bis August 2020 fast 47% aller per Z39.50 in den B3Kat übernommenen Datensätze aus dem WorldCat. Klar ist aber auch: Wenn in allen Verbänden als erstes im WorldCat nach einer passenden Aufnahme gesucht wird, kommt es vermehrt zu parallelen Übernahmen derselben Titelaufnahme, und als Folge davon erhält diese übernommene Titelaufnahme in allen Verbänden einen jeweils anderen EKI – für die gegenseitige Nachführung von hinzugefügter Sacherschließung ist das eher hinderlich.

Anhand des EKI ließe sich Sacherschließung in der Theorie besonders einfach und zuverlässig zu jedem beliebigen Zeitpunkt aus anderen Verbunddatenbanken übernehmen. Ob die Verbundzentralen die hierfür erforderliche Maschinerie nie vollständig ausgebaut haben, weil die Überschneidungsquoten zu dürftig geblieben sind, oder die Überschneidungsquoten zu dürftig sind, weil die Verbundzentralen nie ein permanent laufendes Verfahren zum automatischen Austausch implementiert haben, sei dahingestellt.

4.2 Anreicherung aus den Werkclustern von Culturegraph

Deutliches Potenzial zur Weiterentwicklung des Datenaustauschs zwischen den Verbänden und der Anreicherung von Sacherschließungselementen aus Fremddaten und damit eine Effizienzsteigerung der Sacherschließung im Sinne überregionaler und verbundübergreifender Zusammenarbeit bietet seit neuestem das Projekt Culturegraph der DNB. Im Rahmen dieses Projektes werden inzwischen regelmäßig aus sämtlichen Titeldaten der deutschsprachigen Verbände maschinell sogenannte Werkcluster gebildet, d. h. Bündelungen von Datensätzen

zen, die gemäß algorithmischer Entscheidung – trotz abweichender Erschließungstiefe und v. a. abweichender Formalerschließungselemente – alle dasselbe Werk beschreiben.

So werden beispielsweise von Lehrbüchern sämtliche Auflagen, für die regelkonform in den Verbunddatenbanken jeweils eigene Titeldatensätze erstellt werden, in einem Bündel zusammengefasst. Ebenso können in den Bündeln Aufnahmen von Werken und sogar Manifestationen zusammenfinden, für die in den einzelnen Titelaufnahmen z. T. keine eindeutigen Identifier (wie ISBN oder EKI) erfasst sind und die deshalb bei den bisherigen Abgleichverfahren unberücksichtigt bleiben.

Culturegraph als zentrale Plattform erleichtert durch sein Werkclustering und das von den Verbänden gemeinsam mit der DNB entwickelte, XML-basierte Bündelformat die maschinelle Übernahme von Sacherschließung schon rein organisatorisch: Statt auf Verdacht nacheinander bilateral aktuelle Änderungen aus allen anderen Verbunddatenbanken (wie oben in Abschnitt 3.1.3 beschrieben) zu harvesten und anhand des EKI gegen den entsprechenden Datensatz aus der eigenen Verbunddatenbank abzugleichen, stehen im Werkcluster bereits alle grundsätzlich interessanten Angaben samt Provenienz zur Übernahme bereit.

Durch vertiefte Betrachtung der in den Bündeln enthaltenen Datensätze sowie detaillierte Analyse deren jeweiliger Erschließungselemente können passgenaue Übernahmevorgaben entwickelt und entweder im Rahmen einer Einmalkorrektur für den gesamten Datenbestand eines Verbundes oder prinzipiell als fortlaufender regelmäßiger Prozess von Datenupdates implementiert werden. Um ein solches Übernahmeverfahren technisch implementieren zu können, sollte eine API für Culturegraph bereitstehen.

Auch hier wird sich zeigen müssen, wie weitreichend Effizienz und automatisierte verbundübergreifende Zusammenarbeit auf Basis dieser technischen Möglichkeiten zur Ergänzung von Sacherschließung zugelassen werden. Indem man Abwägungen im Spektrum von Qualität *und* Quantität zu einer rein ausschließenden Gegenüberstellung macht, läuft man Gefahr, das große Potenzial des Verfahrens stark zu beschneiden, nur um der Sicherheit den Vorrang einzuräumen, ausschließlich hochwertige Sacherschließungselemente aus wenigen ausgewählten Quellen zu erhalten.

4.3 Technische Unterstützung arbeitsteiliger Verfahren

Bibliotheken versuchen seit einigen Jahrzehnten, die (teure) Aufgabe der Erschließung ihrer Medien durch geeignete Maßnahmen effizienter und res-

sourcenschonender zu gestalten. Dieser Anspruch hat nicht zuletzt die Unterstützung durch elektronische Datenverarbeitung und die Gründung von Bibliotheksverbänden begünstigt. Damit verbunden war eine hohe Erwartungshaltung an die Schaffung von Wertschöpfungsketten – im Idealfall muss eine Ressource nur einmal erschlossen werden und jeder kann diese Daten anschließend bei eigenem Bedarf nachnutzen. Services für die Nutzung von Fremddaten sind daher Kerngeschäft jeder Verbundzentrale.

Trotzdem ist es aktuell in der bibliothekarischen Realität eher so, dass die gleiche Ressource mehrfach in den einzelnen Verbänden erschlossen wird. Die Gründe dafür mögen vielfältiger Natur sein – suboptimale Workflows in den Katalogisierungsumgebungen, überzogene Erwartungen an die Datenqualität, manchmal auch mangelnde Kenntnis größerer Zusammenhänge – entscheidend ist das Ergebnis und dass kaum jemand ernsthaft über eine Optimierung nachdenkt. Dabei ist für den Bereich der Formalerschließung noch am ehesten nachvollziehbar, dass man ein neu eingegangenes Medium schnellstmöglich für die Bibliotheksbenutzer:innen sicht- und damit nutzbar machen möchte.¹⁸ In der Folge wird die Titelaufnahme dann lieber selbst erstellt als sich auf die eventuell zeitaufwendigere Suche nach passenden Datensätzen in den anderen Verbunddatenbanken zu machen.

Für den Bereich der Inhaltserschließung stellt sich die Sachlage jedoch anders dar. Die Sucheinstiege, die die Inhaltserschließung einem Katalogisat zugefügt, ergänzen die der Formalerschließung und können deshalb auch zu einem späteren Zeitpunkt erfasst werden, ohne dass dies für Benutzer:innen schädlich ist.

Inhaltserschließung ist aufwendig, da man sich dafür mit verschiedenen Wissenschaftsdisziplinen, ihren Teilgebieten und ihrer Terminologie auseinandersetzen muss. Und längst stehen nicht mehr in jeder größeren Bibliothek einschlägig ausgebildete Wissenschaftler:innen für das jeweilige Fachreferat zur Verfügung; diese finden sich heute eher, wenn eine Bibliothek grundsätzlich spezialisiert ist (z. B. Spezialbibliotheken, zentrale Fachbibliotheken etc.) oder fachwissenschaftliche Projekte betreut (z. B. Fachinformationsdienste). Somit böte es sich durchaus an, durch Expert:innen erstellte Inhaltserschließung gezielt nachzunutzen. Um ein hohes Maß an Vollständigkeit und Verlässlichkeit zu erzielen, wäre dafür ein überregional arbeitsteiliges System denkbar: Bibliothek A ist zuständig für eine Disziplin, Bibliothek B für eine andere, Bibliothek C

18 Ähnliches gilt für Bestellkatalogisate: Sie werden eher ohne umfangreichere Vorabrecherchen selbst erstellt, weil die Zeit drängt und oft auch erst wenige bibliografische Informationen vorliegen. Einige Bibliotheken nutzen zudem Systeme des Buchhandels zur Bestellautomatisierung, die zu jedem bestellten Titel umgehend ein Bestellkatalogisat zurückliefern.

für eine weitere usw. Wichtig hierfür sind klare und transparente Vereinbarungen der Zuständigkeiten für die einzelnen Bibliotheken und deren Bereitschaft, eine solche Verantwortung dauerhaft zu übernehmen. Das schüfe einerseits eine erhebliche Entlastung in den Bibliotheken, die sich heute noch selbst um eine mehr oder weniger vollständige Inhaberschließung ihrer Bestände kümmern, und andererseits einen hohen Qualitätsstandard für die Inhaberschließung.

Die Aufgaben der Verbände bzw. der Verbundzentralen in einem solchen arbeitsteiligen Modell sind Aufbau und Pflege der notwendigen technischen Infrastruktur zum regelmäßigen Bezug von Inhaberschließung für die Bestände der eigenen Bibliotheken. Dazu müssen Online-Abfrageschnittstellen zwischen den Verbänden aufgebaut und darüber neu gefundene Erschließungsdaten (wo immer möglich unter Nutzung des EKI) kontinuierlich übernommen werden. Das ist über die Grenzen mehrerer Verbände hinweg eine komplexe und keinesfalls triviale Aufgabe, für die entsprechende Ressourcen zur Verfügung stehen müssen. Technisch weitaus einfacher ist es, eine API auf Culturegraph bereitzustellen. Sie würde es erlauben, aus den Verbänden heraus jeweils nur *eine* Schnittstelle abfragen zu müssen, um von dort Inhaberschließungsdaten aus allen deutschsprachigen Verbänden übernehmen zu können.

4.4 Effizienzsteigerung durch Entitäten-basierte Erschließung

1998 veröffentlichte die IFLA die *Functional Requirements for Bibliographic Records* (FRBR)¹⁹ als einen neuen Blick auf das bibliografische Universum. Kern des Konzepts ist ein Entity-Relationship-Modell, wie es in der Informationswissenschaft und insbesondere bei der Modellierung von Datenbanken üblich ist. Das Modell definiert Entitäten, erläutert die individuellen Merkmale zur Beschreibung dieser Entitäten und listet alle Beziehungen auf, die zwischen den Entitäten bestehen können. Das 2017 erschienene *IFLA Library Reference Model*²⁰ (IFLA LRM) schreibt das konzeptionelle Modell der FRBR fort und deklariert es noch konsequenter durch als sein Vorgänger.

19 Functional requirements for bibliographic records. Final Report. IFLA Study Group on the Functional Requirements for Bibliographic Records. München: K. G. Saur, 1998. (UBCIM Publications, New Series 19). Auch verfügbar: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records> (22.6.2021).

20 Riva, Pat, Patrick Le Bœuf und Maja Žumer: IFLA Library Reference Model. A conceptual model for bibliographic information. Consolidation Editorial Group of the IFLA FRBR Review Group. As amended and corrected through December 2017. Den Haag 2017. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf (19.11.2020).

Beide Modelle ordnen die Inhaltserschließung der Entität Werk bzw. *work* zu. Ein Werk hat demnach eine Beziehung zu Themen-Entitäten, die durch Nomen wie einem Sachbegriff oder einer Notation aus einer Klassifikation repräsentiert werden. Da das Werk zudem in Beziehung zu allen seinen Expressionen und Manifestationen steht, ergibt sich für diese automatisch ebenfalls eine Beziehung zu den Themen-Entitäten des Werks.

Für die praktische Erschließung eröffnet dieses Konzept die Möglichkeit, durch einmalige Erfassung von Elementen der Inhaltserschließung für ein Werk diese automatisch auf die damit verknüpften Manifestationen zu „vererben“, ohne dass für jede Manifestation eine eigene Erfassung notwendig wird. Insbesondere für Werke, die in vielen verschiedenen Ausgaben und Auflagen erscheinen, wäre dieses praktische Vorgehen eine deutliche Verbesserung der Effizienz bei der Erschließung. Bedauerlicherweise stehen aktuell keine Bibliothekssysteme zur Verfügung, die das hinreichend unterstützen würden. So fehlt es an entsprechenden Indexierungsoptionen, die das Vererben von Daten über Relationen hinweg unterstützen, und auch an Formaten, die die Erfassung und den Austausch von Sacherschließung in Werksätzen erlauben. MARC 21 als internationales Austauschformat setzt uns hier entsprechende Grenzen.

Diese Grenzen können nur durch ein konsequentes Umdenken und eine Neuorientierung aufgebrochen werden. Eine adäquate und möglichst vollständige Umsetzung des IFLA LRM kann nur dann erfolgen, wenn die Erschließung in einem System erfolgt, das auf Entitäten und deren Beschreibungen basiert und diese gemäß LRM zueinander in Beziehungen setzt. Ein solches Entitäten-basiertes Katalogisieren braucht auch ein Format, in dem solche Daten erfasst und transportiert werden können. Ein möglicher Kandidat dafür könnte BIBFRAME sein, das als potenzielles Nachfolgeformat von MARC 21 entwickelt wurde. BIBFRAME basiert auf dem Linked-Data-Prinzip und bietet somit eine ideale Voraussetzung für die Abbildung von Entitäten und Beziehungen.

Allerdings hat auch BIBFRAME einen Nachteil, da in seinem Datenmodell derzeit nicht alle in IFLA LRM vorgesehenen Entitäten abgebildet werden können; es fehlt z. B. die Entität Expression. Im Modell vorhanden sind jedoch die Entitäten *work* (inhaltlich nicht vollständig mit LRM in Übereinstimmung) und *subject*, die miteinander in Beziehung gesetzt werden. Die Entität *instance*, dem Pendant zur Manifestation, ist über ihre Beziehung zum *work* ebenfalls mit *subject* verknüpft. Inwieweit das BIBFRAME-Modell tragfähig ist, untersuchen derzeit mehrere konkrete Anwendungsprojekte.

Bibliothekssysteme haben einen langen Lebenszyklus von mehreren Jahrzehnten. Im anglo-amerikanischen Raum sind sie zudem weitestgehend untrennbar mit dem MARC-21-Format verbunden. Grundlegende Änderungen an den Systemen benötigen deshalb viel Zeit. Die deutschsprachigen Verbände be-

finden sich derzeit in einer Transformationsphase und lösen mindestens einen Teil ihrer bisherigen Systeme durch neue ab. Dabei sind sie jedoch von dem abhängig, was die Systemhersteller derzeit anbieten, und das ist vom Entitätenbasierten Erschließen noch weit entfernt. Immerhin hat Ex Libris für sein System Alma angekündigt, dass darin in naher Zukunft auch BIBFRAME als Erfassungsformat zur Verfügung stehen wird, und OCLC arbeitet seit Anfang 2020 in einem von der Mellon Foundation geförderten Projekt *Shared entity management infrastructure* (SEMI)²¹ insbesondere an der Darstellung der Entitäten Person und Werk und deren Vernetzung über die Bibliotheksgemeinschaft hinaus.

5 Fazit

Die Autorinnen und Autoren dieses Beitrags sind von der Grundthese ausgegangen, dass ein bedeutender Aspekt der *Qualität* von Sacherschließung darin besteht, diese auch *quantitativ* auf eine solide Basis zu stellen. Im Vorangegangenen wurden deshalb diejenigen Aspekte des verbundübergreifenden Datenaustauschs erörtert, die wesentlich dazu beitragen, die Qualität von Sacherschließung nicht nur sicherzustellen, sondern – im Sinne eines möglichst hohen Abdeckungsgrades zur Maximierung der Trefferquote (engl.: *recall*) – auch signifikant zu erhöhen. Es ist dabei deutlich geworden, dass entsprechende Handlungsoptionen sowohl im Einflussgebiet der Verbundzentralen liegen als auch komplementär von anderen Agierenden wahrgenommen werden müssen.

So ist ausgeführt worden, dass die Grundlagen für eine optimale Nachnutzungsmöglichkeit der Sacherschließungsdaten bereits in der Phase der Formalererschließung gelegt werden sollten und durchaus noch verbesserungswürdig sind (Abschnitt 4.1). Wenn begleitend dazu die intellektuelle Sacherschließung noch arbeitsteiliger organisiert werden kann (Abschnitt 4.3), bilden diese Maßnahmen zusammen mit weiterzuentwickelnden Möglichkeiten des technischen Datenaustauschs eine wirksame Strategie, das Ziel einer durch Qualität *und* Quantität geprägten Sacherschließungspraxis zu verwirklichen.

Zweifellos besteht noch die Notwendigkeit einer weitergehenden Abstimmung im Datenaustausch zwischen den Verbänden und mit den Expert:innen in den Sacherschließungsgremien (Abschnitt 3.2). Allerdings wird eine Abstimmung über formal-inhaltliche Kriterien (Abschnitte 3.2.1 und 3.2.2) alleine nicht ausreichen, um den Qualitätsaspekt *Quantität* signifikant zu verbessern.

²¹ <https://www.oclc.org/en/worldcat/oclc-and-linked-data/shared-entity-management-infrastructure.html> (19.11.2020).

Als wesentliche Maßnahmen zur Erhöhung der Quantität von Sacherschließungsdaten wurden neben der intensivierten Nutzung des EKI-basierten Verfahrens (Abschnitt 4.1) die Implementierung eines auf Culturegraph aufbauenden kontinuierlichen Übernahmeprozesses identifiziert (Abschnitt 4.2). Culturegraph eröffnet eine schon in naher Zukunft umsetzbare Möglichkeit, um von einem vollständigen Datenflussnetzwerk hin zu einem sternförmigen zu gelangen und so einen effizienteren Ablauf im Datenaustausch zu gewährleisten. Bei Batchverfahren, die auf Culturegraph aufsetzen, steht zunächst vor allem die Erhöhung der Quantität im Vordergrund, wobei die massive Anreicherung mit RSWK-Schlagwörtern auch andere Qualitätsaspekte berührt. Bei einzelfallbezogenen Übernahmen mittels Culturegraph-API sind der gesteigerte Komfort und die maschinelle Unterstützung der intellektuellen Sacherschließung hervorzuheben – insbesondere, wenn dieses Feature auch in Tools wie den DA-3 integriert wird.

6 Literaturverzeichnis

- Beckmann, Regine, Imma Hinrichs, Melania Janßen, Gerard Milmeister und Peter Schäuble: Der Digitale Assistent DA-3. In: o-bib (2019) Bd. 6 Nr. 3. S. 1–20. <https://doi.org/10.5282/o-bib/2019H3S1-20>.
- Ceynowa, Klaus: In Frankfurt lesen jetzt zuerst Maschinen. In: Frankfurter Allgemeine Zeitung, 31.7.2020. <https://www.faz.net/-gr0-909kq> (19.11.2020).
- Deutsche Nationalbibliothek: Grundzüge und erste Schritte der künftigen inhaltlichen Erschließung von Publikationen in der Deutschen Nationalbibliothek. Mai 2017. https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/konzeptWeiterentwicklungInhaltserschliessung.pdf?__blob=publicationFile&v=4 (19.11.2020).
- Expertengruppe Sacherschließung: Regeln für die Schlagwortkatalogisierung. RSWK. Erarbeitet von der Expertengruppe Sacherschließung im Auftrag des Standardisierungsausschusses. 4., vollständig überarbeitete Auflage, Stand März 2017. Leipzig, Frankfurt a. M.: Deutsche Nationalbibliothek 2017. <http://nbn-resolving.de/urn:nbn:de:101-2017011305>.
- Gantert, Klaus: Bibliothekarisches Grundwissen. 9., vollständig aktualisierte und erweiterte Auflage. Berlin, Boston: De Gruyter Saur 2016. <https://doi.org/10.1515/9783110321500>.
- IFLA Study Group on the Functional requirements for bibliographic records: Functional requirements for bibliographic records. Final Report. München: K. G. Saur, 1998. (UBCIM Publications, New Series 19). Auch verfügbar: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records> (22.6.2021).
- Kühn, Armin und Andreas Krausz: Technik des DA-3. Anschluss an K10plus. 4. Workshop Computerunterstützte Inhalterschließung am 11. und 12. November 2020. <https://wiki.dnb.de/x/ew-VCg> (19.11.2020).
- Pfeffer, Magnus: Automatische Vergabe von RVK-Notationen mittels fallbasiertem Schließen. In: Wissen bewegen – Bibliotheken in der Informationsgesellschaft. 97. Deutscher Biblio-

- thekartag in Mannheim 2008. Hrsg. von Ulrich Hohoff. Frankfurt a. M.: Klostermann 2009. S. 245–254.
- Riva, Pat, Patrick Le Bœuf und Maja Žumer: IFLA Library Reference Model. A conceptual model for bibliographic information. Consolidation Editorial Group of the IFLA FRBR Review Group. As amended and corrected through December 2017. Den Haag 2017. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf (19.11.2020).
- Wissenschaftsrat und Deutsche Forschungsgemeinschaft: Zur Zukunft der Bibliotheksverbände als Teil einer überregionalen Informationsinfrastruktur in Deutschland. <http://www.wissenschaftsrat.de/download/archiv/1003-11.pdf> (19.11.2020).

