

Moritz Strickert

Metadaten und kontrollierte Vokabulare

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 163–184
<https://doi.org/10.25364/978390337423210>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz, ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

Moritz Strickert, Humboldt-Universität zu Berlin, Universitätsbibliothek, moritz.strickert@ub.hu-berlin.de |
ORCID iD: 0000-0001-9626-5932

Zusammenfassung

Metadaten sind ein zentraler Bestandteil heutiger Wissensorganisation und werden bei der Bewältigung von permanent wachsenden Informationsmengen zusehends wichtiger. Der vorliegende Artikel arbeitet heraus, was unter dem Terminus „Metadaten“ zu verstehen ist, welche Standards existieren und wozu Metadaten(-standards) gebraucht werden. Des Weiteren liegt der Fokus auf der Erschließung, die auf eine strukturierte Beschreibung von analogen und digitalen Ressourcen abzielt. Um die Austauschbarkeit zwischen verschiedenen Institutionen bzw. Datenquellen sicherzustellen, bedarf es dabei gemeinsamer Standards. Wichtig ist in diesem Zusammenhang der Rückgriff auf kontrollierte Vokabulare und Normdateien, insbesondere die im deutschsprachigen Raum zentrale Gemeinsame Normdatei (GND).

Schlagwörter: Forschungsdaten; Metadaten; Dublin Core; Inhaltserschließung; Thesaurus; Normdatei

Abstract

Metadata and Controlled Vocabularies

Metadata are a central component of today's knowledge organization and are becoming increasingly important in the management of permanently growing amounts of information. This article elaborates on what is meant by the term metadata, which standards exist, and what metadata (standards) are used for. Furthermore, the focus is on indexing, which aims at a structured description of analog and digital resources. In order to ensure exchangeability between different institutions or data sources, common standards are required. In this context, it is important to have reference to controlled vocabularies and authority files, especially the Gemeinsame Normdatei (GND), which is key in the German-speaking world.

Keywords: Research data; metadata; Dublin Core; subject cataloging; thesaurus; authority file

1. Einleitung

Der vorliegende Beitrag zielt auf die Einführung und Diskussion der Bereiche Metadaten, Erschließung und kontrollierte Vokabulare sowie Normdaten ab. Er gliedert sich folgendermaßen: Zu Beginn erfolgt eine allgemeine Einführung in den Bereich der Metadaten und Metadatenschemata, die mit konkreten Beispielen aus der Praxis angereichert ist. Im Anschluss werden verschiedene Formen der Erschließung von Ressourcen dargestellt, die diese für Nutzer:innen auffindbar und nutzbar machen sollen. Hierzu werden die Prozesse der Formal- und Sacherschließung vorgestellt und die Unterschiede zwischen Stichwörtern, Schlagwörtern und kontrollierten Vokabularen herausgearbeitet. Darauf aufbauend wird das Konzept der Normdaten und deren Vorteile in der Informationsverarbeitung erörtert und gleichzeitig die im deutschsprachigen Raum wichtige Gemeinsame Normdatei (GND) vorgestellt. Eine Kritik an kontrollierten Vokabularen sowie ein Fazit samt Ausblick runden den Beitrag ab.

2. Metadaten und Metadatenschemata

Metadaten dienen der Beschreibung von Ressourcen, um diese auffindbar und weiterverwendbar zu machen. Diese Ressourcen beinhalten u. a. digitale und nicht-digitale Literatur in Bibliotheken, Forschungsdaten sowie Objektdaten in Museen. Richard Gartner unterscheidet drei Metadatentypen: deskriptive, administrative und strukturelle Metadaten.¹ Deskriptive Metadaten – zentraler Inhalt dieses Beitrags – beschreiben die Ressourcen, auf die sie sich beziehen, so dass sie gefunden und mit anderen verknüpft werden können.² Strukturelle Metadaten sind Hintergrundinformationen, die sicherstellen, dass Ressourcen gespeichert, aufbewahrt, angezeigt und abgerufen werden können. Sie enthalten Informationen bezüglich des Aufbaus der Ressource und stellen die Funktionalität dieser sicher.³ Ein Teilbereich struktureller Metadaten sind technische Metadaten, die mitunter auch als gesonderter Datentyp beschrieben werden⁴: Diese umfassen alles, was ein System über ein digitales Objekt, z. B. eine Bild- oder Textdatei, wissen muss, um es korrekt darzustellen und nutzbar zu erhalten, bspw. die Pixelzahl oder das Dateiformat. Ferner enthalten administrative Metadaten Informationen darüber, welche Eigentumsrechte an Daten bestehen und welche Arten der Nutzung gestattet werden, z. B. in Form von Lizenzen.

1 Gartner, R. (2016), S. 6-8.

2 Sugimoto, S.; Nagamori, M.; Mihara, T.; Honma, T. (2015), S. 105.

3 Joudrey, D.; Taylor, A. (2018), S. 139.

4 Dierkes, J. (2021), S. 314.

Die meisten Metadaten besitzen drei Hauptkomponenten, die zumeist durch Standards definiert werden. Die Semantik definiert die Bedeutungen der Metadatenelemente. Das Datenmodell gibt an, welche inhaltliche Struktur die Daten annehmen sollen. Die Syntax strukturiert die Art und Weise, in der die Metadaten dargestellt sind z. B. in Tabellenform.⁵ Durch Metadaten ist es möglich, dass die lokale Objekterschließung, die beispielsweise durch Bibliotheken geleistet wird, mittels übergreifender Datenverzeichnisse anderen zugute kommt und Fremddaten, die von anderen Institutionen (erschlossen) bereitgestellt werden, übernommen werden können. Um einen reibungslosen Austausch zwischen verschiedenen Datenbeständen sicherzustellen, sind Metadatenstandards notwendig. Diese Standards sind auf verschiedenen Ebenen angesiedelt und oft fachspezifisch. Es existieren konzeptionelle Modelle für einzelne Domänen, beispielsweise das IFLA Library Reference Model (LRM), das die logische Struktur bibliografischer Erschließung beschreibt, oder das CIDOC Conceptual Reference Model als Modell zum Austausch von Information für das Feld des kulturellen Erbes. Daneben existieren Strukturstandards wie Dublin Core, die einfach gehaltene und einheitliche Standards für Metadatenelemente zur Beschreibung von Ressourcen liefern. Des Weiteren gibt es Standards, die die strukturierte Erschließung von Inhalten, beispielsweise bezüglich Schreibkonventionen, regeln (Resource Description and Access, RDA; Regeln für die Schlagwortkatalogisierung, RSWK), Wertestandards (Gemeinsame Normdatei, GND; Open Researcher and Contributor iD, ORCID iD), sowie Standards, die den Datenaustausch zwischen Institutionen regulieren (Machine-Readable Cataloging, MARC bzw. MARC21; Lightweight Information Describing Objects, LIDO). Jens Dierkes nennt als Gütekriterien für die Metadatenqualität folgende Parameter: „Vollständigkeit, Genauigkeit, Provenienz, Erwartungskonformität, logische Konsistenz und Kohärenz, Aktualität und Zugänglichkeit.“⁶

Jedes Metadatenschema beinhaltet als standardisierte Beschreibungskonvention verschiedene Metadatenelemente (Kategorien oder Felder), die die einzelnen Teile einer Ressourcenbeschreibung (z. B. Titel, Erstellungsdatum) umfassen. Dafür ist es wichtig, standardisiert festzulegen, auf welche Art und Weise die Elemente erfasst werden (z. B. Schreibweisen von Daten).⁷ Zur Beschreibung von Dokumenten und anderen Ressourcen ist Dublin Core als Standard sehr weit verbreitet. Gartner bezeichnet diesen gar als „lingua franca“⁸ für deskriptive Metadaten im Internet.

5 Joudrey, D.; Taylor, A. (2018), Anm. 3, S. 133.

6 Dierkes, J. (2021), Anm. 4, S. 322.

7 Joudrey, D.; Taylor, A. (2018), Anm. 3, S. 132.

8 Gartner, R. (2016), Anm. 1, S. 33.

Die Entwicklung des Dublin-Core-Standards hatte das Ziel, einen Kern von Elementen zu identifizieren, der auf jedes physische oder digitale Objekt angewendet werden kann. Dublin Core besteht in seiner einfachsten Form (Simple Dublin Core) aus lediglich 15 Elementen, von denen jedes so weit definiert ist, dass es für alle Anwender:innen verständlich und umfassend nutzbar sein sollte.⁹ Mit diesem Standard – der ursprünglich aus dem Bereich der Bibliotheken kam – lassen sich nunmehr grundsätzlich unterschiedliche elektronische Dokumente beschreiben, recherchieren und austauschen. Problematisch ist jedoch, dass bei Simple Dublin Core solche Elemente wie Schöpfer:in („Creator“) sehr weit gefasst sind: Sie sind so vage definiert, dass bei Auffinden zweier Datensätze mit dem gleichen Element nicht sicher ist, dass diese Elemente auch das Gleiche bedeuten.¹⁰ Als Beispiel nennt Gartner den Film „Rebecca“, in welchem sowohl Alfred Hitchcock (der Regisseur) als auch David O. Selznick (der Produzent) als „Creator“ gelistet werden. Nutzende haben aber oftmals ein Interesse daran, genauer zu wissen, welche Rolle eine Person einnahm. Für diesen präziseren Informationsbedarf wurden verschiedene Lösungen entwickelt. Ein früher Lösungsansatz ist Qualified Dublin Core, dabei werden Dublin-Core-Elemente mit weiteren Tags versehen, um sie durch zusätzliche Informationsanreicherung präziser zu machen.¹¹

Um die semantische Bedeutung von Elementen eindeutig verstehen und einordnen zu können, ist es notwendig, klar zu identifizieren, auf welcher Metadatendefinition diese basieren. Das Mapping, das als Prozess der Beziehungen zwischen Metadatenbegriffen verschiedene Schemata erkennt, abgleicht und verknüpft, ist eine entscheidende Aufgabe, um die Metadaten interoperabel zu machen.¹² Natürlich-sprachige Termini wie beispielsweise „Titel“ sind zu unpräzise. Aus diesem Grund wird auf Uniform Resource Identifier (URI) zurückgegriffen. Diese bestehen aus Zeichenketten (Buchstaben, Zahlen, Satzzeichen) und können eindeutige Auskunft darüber geben, dass z. B. der Titel in der vorliegenden Ressource nach den Maßgaben von Dublin Core definiert ist.¹³ Da sich die Beschreibungselemente von Dublin Core in der Hauptsache auf Textdokumente beziehen, ist der Standard jedoch beispielsweise für Museen nur begrenzt geeignet, weil die Bedarfe für eine umfassende Objektdokumentation, z. B. in Hinblick auf relevante physische Objekteigenschaften, nicht erfüllt werden können.¹⁴

9 Gartner, R. (2016), Anm. 1, S. 31f.

10 Gartner, R. (2016), Anm. 1, S. 35.

11 Gartner, R. (2016), Anm. 1, S. 32-34.

12 Sugimoto, S.; Nagamori, M.; Mihara, T. et al. (2015), Anm. 2, S. 105.

13 Gartner, R. (2016), Anm. 1, S. 52.

14 Team MusIS im BSZ (2020), S. 3f.

Darüber hinaus werden für die Beschreibung von Forschungsdaten auch disziplinspezifische Metadatenstandards verwendet. Ein Beispiel aus der sozialwissenschaftlichen Forschung stellt der DDI-Standard (Data Documentation Initiative Standard) dar. Er dient zur Beschreibung z. B. von Umfragen, Fragebögen oder statistischen Datensätzen und ermöglicht, Forschungsdaten über ihren gesamten Lebenszyklus hinweg zu beschreiben sowie effizient und nachhaltig zu verwalten, einschließlich der Veröffentlichung und Nachnutzung der Daten. Der Standard ist interoperabel mit anderen Standards wie DataCite und Dublin Core. Durch die Verwendung des DDI-Standards ist es möglich, Forschungsdaten für Sekundärnutzende verständlich zu erschließen. Zugleich sollen die Maschinenlesbarkeit sowie die Auffindbarkeit und der Austausch zwischen verschiedenen Organisationen sichergestellt werden. Von Wissenschaftsseite wurde an diesem Schema jedoch kritisiert, dass die Erfassung von audio-visuellen Ressourcen weitestgehend unmöglich ist.¹⁵

Da bei der Vergabe von Metadaten zwischen minimalen und umfassenden Anforderungen eine große Spannweite existiert, ist es notwendig abzuwägen, ob im Einzelfall eher generische oder spezifische bzw. stark kontextgebundene Beschreibungen für Forschungsdaten gebraucht werden sollen. Metadaten schemata, welche stark auf die jeweilige Forschungscommunity bezogen sind, erleichtern die Aufnahme wichtiger fachspezifischer Bedarfe. Deren Übertragbarkeit auf andere Disziplinen ist jedoch oft problembehaftet. Die Interoperabilität von Metadaten(-Schemata) benötigt darüber hinaus Pflege, um ihre Konsistenz dauerhaft sicherzustellen.¹⁶

Als Beispiel dient die qualitative sozialwissenschaftliche Forschung: Hier ist oft das Verstehen von Sinnzusammenhängen essenziell. Da Sekundärnutzende von Forschungsdaten nicht an der ursprünglichen Forschungssituation beteiligt waren, sind einfache Metadaten oft nicht ausreichend und eine zusätzliche Kontextualisierung ist notwendig.¹⁷ Diese verschafft Klarheit bezüglich des Forschungsprozesses

15 Imeri, S.; Sterzer, W.; Harbeck, M. et al. (2019), S. 25.

16 Sugimoto, S.; Nagamori, M.; Mihara, T. et al. (2015), Anm. 2, S. 107.

17 Doris Bambey, Alexia Meyermann, Maike Porzelt und Marc Rittberger differenzieren bei der Kontextualisierung die Makro-, Mikro- und Objektebene. Die Makroebene gibt Auskunft über Studienhintergründe und den Erstellungskontext, die in der Forschung verwendeten Methoden und Forschungsprozesse sowie rechtliche Aspekte. Die Mikroebene beschreibt das Setting der Forschung, beispielsweise die Interviewsituation oder zusätzliche biografische Informationen der Beforschten. Die Objektebene gibt Auskunft über das Vorgehen bei der Datenaufbereitung (z. B. Transkriptionsregeln), ob verschiedene Versionen bestehen, den Umfang der einzelnen Datensätze und in welcher Beziehung diese zueinander stehen, siehe dazu Bambey, D.; Meyermann, A.; Porzelt, M. et al. (2018), S. 63. Grundsätzlich ist ebenfalls abzuwägen, wie viele Ressourcen für die Kontextualisierung eingesetzt werden sollen, um analytisches Potenzial sicherzustellen.

auf verschiedenen Ebenen: dem institutionellen Hintergrund, dem theoretisch-methodischen Studienansatz, der Art und Weise, wie Daten erhoben, aufbereitet und analysiert wurden, welche strategischen Entscheidungen durch die Primärforschenden getroffen wurden und schließlich welche Bedingungen bei einer Nutzung zu beachten sind.¹⁸ Neben der inhaltlich-thematischen Übereinstimmung sind auch „Aktualität, Zugangsmöglichkeiten, Versionierung, Datenqualität, Erhebungsmethoden, Provenienz und Untersuchungsbereich“¹⁹ wichtige Auswahlkriterien für Forschende, die diese Daten nachnutzen wollen. Nachhaltige und standardisierte Erschließung mittels Metadaten ist mit Blick auf Forschungsdaten unverzichtbar, die erst so auffindbar werden und für eine potenzielle Nachnutzung bereitstehen. Überdies können Forschungsdaten durch die Erschließung zusehends als eigenständige Publikationsform angesehen werden.²⁰ Die Anwendung von persistenten Identifikatoren (z. B. Digital Object Identifier, DOI) ermöglicht, dass Verknüpfungen zu Dateien, Dokumenten oder Webseiten, trotz sich mitunter wandelnder Speicherorte, dauerhaft sichergestellt werden können. Fremddaten, die aus externen Quellen bezogen werden, können den eigenen Datenbestand ergänzen und überdies zur Vermeidung doppelter Erschließungsarbeit beitragen. Darüber hinaus ist es sinnvoll, dass Forschungsdaten in fachrelevante Bibliothekskataloge, Portale und Nachweissysteme (z. B. da|ra, Registrierungsagentur für Sozial- und Wirtschaftsdaten) eingespist werden und auf diesem Wege zu finden sind.²¹

Grundsätzlich ist innerhalb des Erschließungsprozesses zwischen der Formalerschließung und der Sach- bzw. Inhalterschließung zu unterscheiden. Erstere greift nur auf Informationen zurück, die sich unmittelbar ermitteln lassen. Bei einem gedruckten Buch sind dies beispielsweise Autor:in, Titel, Verlag, Seitenanzahl. Diese Informationen sind als Metadaten durch Standardisierung, z. B. durch das Standardisierungsregelwerk Resource Description and Access (RDA), mit anderen Einrichtungen austauschbar. Jenseits der genuin bibliothekarischen Sphäre ist die Formalerschließung weniger stark mittels Regelwerken strukturiert. Bei Forschungsdaten ist die Erschließung deshalb oftmals abhängig von den Anforderungen der entsprechenden Repositorien.

Einen konkret ausformulierten Ansatz für die durchzuführende Kontextualisierung beschreibt das Forschungsdatenzentrum Qualiservice in ihrer Handreichung zur Kontextualisierung qualitativer Forschungsdaten mit besonderem Fokus auf den sogenannten Studienreport, siehe dazu Heuer, J.-O.; Kretzer S.; Mozygemba, K. et al. (2020).

18 Smioski, A. (2011), S. 228f.

19 Friedrich, T.; Recker, J. (2021), S. 414.

20 Queckbörner, B. (2019), S. 66.

21 Klump, J. (2010), S. 112.

Die Sacherschließung dient der Erschließung der Ressourcen nach inhaltlich-thematischen Kriterien. Dies kann durch die verbale Erschließung mittels Schlagwörtern geschehen, hauptsächlich auf Grundlage der Regeln für die Schlagwortkatalogisierung (RSWK) oder durch die Zuordnung in Klassifikationen wie der Regensburger Verbundklassifikation (RVK) oder der Dewey-Dezimalklassifikation (DDC). Die Vergabe von Schlagwörtern findet zumeist noch auf intellektueller Ebene statt, wobei derzeit bereits Verfahren (weiter-)entwickelt werden, um dies mittels (semi-)automatisierter Text- und Datenanalyse maschinell durchzuführen.²²

3. Recherche auf Basis von Stichwörtern, Schlagwörtern und kontrollierten Vokabularen

Der Erschließungsprozess umfasst bestenfalls eine detaillierte und einheitliche Beschreibung von analogen und digitalen Ressourcen auf Grundlage von strukturierten Metadaten und kontrollierten Vokabularen. Die Erschließung soll Nutzer:innen gesuchte Ressourcen und Informationen möglichst umfassend zur Verfügung stellen. Es sind hierbei grob fünf Ziele zu unterscheiden: Sie soll 1.) zuverlässiges Auffinden ermöglichen; 2.) helfen, Verschiedenes zu unterscheiden; 3.) zusammenführen, was zusammengehört; 4.) Gefundenes übersichtlich machen; 5.) das Ausgewählte zugänglich machen.²³ An mögliche Suchergebnisse sind zwei zentrale Anforderungen zu richten, die als „Precision“ (Genauigkeit) und „Recall“ (Vollständigkeit) beschrieben werden. Die Genauigkeit ist ein Indikator, wie viele der abgerufenen Dokumente relevant sind; die Vollständigkeit ist ein Indikator, wie viele der relevanten Dokumente in einem System tatsächlich abgerufen werden. Es gilt: Suchergebnisse sollen möglichst präzise sein und das Suchmaschinensystem soll nur die tatsächlich interessanten Treffer finden. Gleichzeitig sollen jedoch auch alle für die Nutzer:innen relevanten Dokumente aufgefunden werden.²⁴

Im Prozess der Recherche wird von Suchenden vielfach auf Stichwörter und Schlagwörter zurückgegriffen. Das Stichwort wird dem Dokument, beispielsweise aus dem Titel, entnommen, das Schlagwort der Ressource auf Grundlage seines Inhalts

22 So erschließt die Deutsche Nationalbibliothek seit 2010 in zunehmendem Maße den stark wachsenden Anteil von digitalen Ressourcen mittels maschineller Verfahren. Dieses Verfahren wurde 2017 auch für physische Medien erweitert, siehe dazu Mödden, E.; Schöning-Walter, C.; Uhlmann, S. et al. (2018), S. 30. Für die klassifikatorische Erschließung wird dabei auf maschinelle Lernverfahren zurückgegriffen; bei der Vergabe von Schlagwörtern finden linguistische Verfahren Anwendung, eine kompakte Verfahrensbeschreibung findet sich in Uhlmann, S. (2013). Die Ergebnisse dieser Abgleichsverfahren sind derzeit jedoch noch fehleranfällig und oft unpräzise, siehe dazu Wiesenmüller, H. (2018), S. 28. Zur grundsätzlichen Kritik an diesen Entwicklungen siehe Ceynowa, K. (2017).

23 Joudrey, D.; Taylor, A. (2018), Anm. 3, S. 207f.

24 Joudrey, D.; Taylor, A. (2018), Anm. 3, S. 121ff.

hinzugefügt. Die Recherchepraxis mit Stichwörtern, z. B. in Form einer Volltextsuche, ist problematisch, da diese in den sprachlichen Formulierungen der Autor:innen verhaftet bleibt, jedoch ist sie gleichzeitig weit verbreitet. Es werden mitunter Wörter verwendet, die wenig nützlich zur inhaltlichen Beschreibung der Ressourcen sind. Auf diese Weise liefern Suchanfragen einerseits nicht relevante Ergebnisse und andererseits werden Ressourcen, die ähnliche Inhalte haben, nicht gefunden, wenn diese von den Verfasser:innen nicht mit denselben Begriffen benannt worden sind. Inhalte sind nur auffindbar, wenn sie explizit als Begriff in der Ressource auftauchen. Das Vorkommen eines Begriffs bedeutet nicht notwendigerweise, dass ein Text den gesuchten Themenkomplex tatsächlich behandelt. Insbesondere bei der Volltextsuche im Internet treten Probleme bei der Suche nach abstrakten Themen wie etwa Ethik oder Gesundheit auf, da sie eine viel zu große und wenig fokussierte Ergebnismenge liefern.²⁵ Die Ergänzung durch Inhaltsverzeichnisse und andere Informationen z. B. in Bibliothekskatalogen erhöht zwar den Suchtrefferumfang, die Treffer sind aber oft wenig präzise.²⁶ Die Art der Suche kann das Vorkommen von Wörtern in den gesuchten begrifflichen Zusammenhängen nicht vom Auftauchen der gleichen Wörter in nicht gesuchten Zusammenhängen trennen. Es können nur Ergebnisse mit den Suchbegriffen erzielt werden, die die Suchenden kennen – verwandte oder unbekannte Begriffe fallen weg.²⁷

Eine Stichwortsuche auf Grundlage von Titeln ist ebenfalls problembehaftet: Viele Titel geben zu wenig Aufschluss über den Inhalt. Es fehlen vielfach Informationen darüber, welchen zeitlichen oder räumlichen Bezug der Text besitzt. Personennamen, aber auch Körperschaften und Geografika sind durch Mangel an Kontrolle bei dieser Form der Recherche nicht eindeutig voneinander abgrenzbar und somit auch nicht zuordenbar. Vergebene Titel nutzen zudem oft bildliche Sprache, deren Metaphorik nicht durch eine Stichwortsuche in ihrem Bedeutungsgehalt erkannt bzw. übersetzt werden kann.²⁸ Des Weiteren können Stichwörter die Probleme von Synonymen/Homonymen, abweichenden Ausdrücken und verschiedenen Sprachen, die für dieselben Themen verwendet werden, nicht lösen. Oftmals gibt es mehrere Arten, ein vorhandenes Konzept zu bezeichnen. Dies umfasst verschiedene Schreibweisen desselben Sachverhalts (z. B. BE/AE „harbour“ / „harbor“), Abkürzungen und Initialen, das Problem veralteter Termini, die heute für Sachverhalte nicht mehr gebräuchlich sind oder in unterschiedlichen Disziplinen unter-

25 Beall, J. (2008), S. 442.

26 Gross, T.; Taylor, A. G.; Joudrey, D. N. (2015), S. 17f.

27 Mann, T. (2008), S. 163-165.

28 Flachmann, H. (2004), S. 767.

schiedlich verwendet werden. Bei Homonymen, also Wörtern mit mehreren Bedeutungen (z. B. „Tau“), wird eine Stichwortsuche das Wort in allen seinen verschiedenen Bedeutungen zurückliefern, obgleich die Suche nur auf eine bestimmte Bedeutung abzielt. Ein ähnliches Problem besteht darüber hinaus, wenn ähnlich oder identisch geschriebene Wörter in unterschiedlichen Sprachen, jedoch mit verschiedenen Bedeutungen, existieren. Die Folge ist auch hier eine sehr große unpräzise Ergebnismenge.²⁹ Die Recherche via Freitextfeldern sowie Stichworten und ohne Rückgriff auf kontrollierte Termini liefert somit eher dürftige Resultate.

Für die beschriebenen Probleme der Stichwortsuche bieten (Sach-)Schlagwörter eine Lösung. Schlagwörter können sowohl für die inhaltliche Erschließung von Ressourcen als auch als Grundlage für Recherchen verwendet werden. Bei der inhaltlichen Erschließung mit Schlagwörtern muss grob zwischen freien Termini, beispielsweise in Form von Folksonomien (z. B. durch Social Tagging³⁰), und genormten Schlagwörtern unterschieden werden. Dieser Erschließungsprozess ist ressourcenintensiv, ermöglicht jedoch eine tiefergehende inhaltliche Recherche, indem auch Ressourcen ohne das Wissen um bibliografische oder formale Angaben (z. B. Autor:in, Verlag etc.) auffindbar gemacht werden. Manche Ressourcen nicht textueller Art wie Bilder oder Grafiken brauchen bislang eine weitergehende Beschreibung, damit sie überhaupt recherchierbar werden können.³¹ Auch nicht-deutschsprachige Titel können nach Verschlagwortung mit Rückgriff auf deutsche Schlagwörter gefunden werden. Dies ist deshalb wichtig, da Recherchekompetenzen in einer Fremdsprache beispielsweise in Hinblick auf die Kenntnis relevanter Suchtermini oftmals deutlich schlechter ausgeprägt sind als bei einer Suche in der Erstsprache.³² Gleichzeitig liefern Suchanfragen weniger Treffer, die für die Nutzenden nicht von Relevanz sind. Tina Gross et al. kommen in ihrer Metastudie zu folgendem Ergebnis:

29 Beall, J. (2008), Anm. 25, S. 439f.

30 Soziales oder kollaboratives Tagging umfasst die Verschlagwortung von Ressourcen durch die Nutzenden selbst, wobei keine umfassenden Regelwerke zur Anwendung kommen. Die Begriffssammlung, die im Zuge dessen erstellt wird, wird als „Folksonomie“ bezeichnet. Vorteile dieses Vorgehens bestehen darin, dass bislang wenig bekannte bzw. bearbeitete Ressourcen und Themengebiete, die zu speziell für die Aufnahme in kontrollierte Vokabulare sind, präzise beschrieben werden können. Das große Problem bei dieser Art von Beschreibung ist jedoch, dass sie ungeordnet ist. Beispielsweise werden weder Schreibweisen kontrolliert noch Synonyme miteinander verknüpft sowie die vergebenen Tags nicht auf ihre Sinnhaftigkeit hin überprüft. Dies führt dazu, dass die Suchergebnisse große Lücken aufweisen, da Hinweise auf relevante Ressourcen fehlen, die zwar einen ähnlichen Inhalt besitzen, aber mit einer alternativen Schreibweise oder Begrifflichkeit versehen sind, siehe dazu Gartner, R. (2016), S. 96ff.; Gross, T.; Taylor, A. G.; Joudrey, D. N. et al. (2015), S. 14.

31 Gross, T.; Taylor, A. G.; Joudrey, D. N. et al. (2015), Anm. 26, S. 11.

32 Flachmann, H. (2004), Anm. 28, S. 767.

The 2005 study of the effect of controlled vocabulary on the results of keyword searching found that an average of 35.9 % of hits in keyword searches would be lost if subject headings were to be removed from or no longer included in catalog records. The current study found that with the addition of tables of contents and summaries or abstracts, an average of 27 % of hits would be lost if the subject headings were not present in the records.³³

Für die intellektuelle Erschließung durch Menschen und für maschinelle Verfahren, die eine automatisierte Erschließung auf Grundlage von Abstracts, Inhaltsverzeichnissen oder Volltextanalysen ermöglichen sollen, ist der Rückgriff auf zeitgemäßen und kontrolliertes Vokabular unersetzlich. Für zukünftige Entwicklungen im Bereich des Text- und Data-Minings³⁴, wie es beispielsweise im Bereich der Digital Humanities bereits erprobt wird, ist dies ebenfalls bedeutsam. Ein solches Vokabular, z. B. in Form eines Thesaurus, soll sicherstellen, dass die Schlagwortvergabe konsistent und mit präzisen Termini erfolgt. Dies macht eine terminologische Kontrolle notwendig, die auf die Mehrdeutigkeit von Begriffen reagiert und klar ausweist, was auf welche Weise bezeichnet werden soll. Es werden begriffliche Beziehungen nachgewiesen, indem eine Vorzugsbenennung (Deskriptor) festgelegt wird, die im Anschluss zur Verschlagwortung verwendet werden kann. Gleichzeitig müssen Synonyme möglichst vollständig aufgenommen und mittels Verweisen mit der Benennung verknüpft werden. Des Weiteren müssen die Bedeutungen bei allen Homonymen (siehe oben), Polysemen – dies sind Wörter gleichen etymologischen Ursprungs (z. B. „Flügel“: Körperteil und Instrument) – eindeutig identifiziert werden. Diese Erfassung ermöglicht, dass bei einer späteren Recherche Synonyme als Sucheinstiege fungieren können, die wiederum auf die entsprechende Vorzugsbenennung verweisen. Dies verbessert die Recherche, da bei der Suche mit Synonymen ebenfalls die vergebenen Vorzugsbenennungen und weitere Synonyme gefunden werden können. Zugleich sind formale Konventionen notwendig, die unter anderem die Schreibweise sowie die Verwendung im Singular oder Plural reglementieren. Für den deutschsprachigen Raum basiert dieser Festlegungsprozess auf den Regeln der Schlagwortkatalogisierung (RSWK). Außerdem existiert eine sogenannte „Zerlegungskontrolle“, die sehr lange Wortkomposita vermeiden will. An ihre Stelle tritt eine Kombination aus bereits existenten Deskriptoren der Schlagwortsammlung, die den Bedeutungsinhalt des zusammengeführten Wortes umfasst.

33 Gross, T.; Taylor, A. G.; Joudrey, D. N. et al. (2015), Anm. 26, S. 31.

34 Diese Verfahren zielen darauf ab, mittels Algorithmen Bedeutungsstrukturen in umfangreichen, wenig bis gar nicht strukturierten Textmengen zu erkennen und für eine tiefere Analyse nutzbar zu machen.

Es lassen sich zwei Verschlagwortungsprinzipien unterscheiden: Das gleichordnende Prinzip sieht vor, dass Schlagwörter einzeln nebeneinander stehen und eine Kombination mittels Suchoperatoren erfolgt. Beim syntaktischen Indexieren werden Schlagwörter hingegen in eine sinnhafte Beziehung zueinander gesetzt, sodass eine Art Kurzabstract gebildet wird.

Ein Thesaurus unterscheidet sich von einer alphabetisch geordneten Schlagwortliste darin, dass dessen Termini gleichzeitig in eine Hierarchie eingefügt werden. Hierarchische Relationen verbinden somit spezifischere Unter- und umfassendere Oberbenennungen: Äquivalenzrelationen umfassen Synonyme, wohingegen Assoziationsrelationen verwandte Deskriptoren beinhalten können. Vielfach sind Thesauri polyhierarchisch aufgebaut, sodass sie eine Benennung über mehrere Über- bzw. Unterbenennungen besitzen. Wichtig ist, dass sowohl diejenigen, die die Verschlagwortung vornehmen, als auch die Nutzenden passende Deskriptoren einfach und zuverlässig finden können.³⁵

Normierte Sucheinstiege sind auch für die Suche nach Personen relevant, da hierbei der bevorzugte Name einer Person festgelegt wird. Gleichzeitig können Verweisungen zu alternativen Schreibweisen angelegt werden. Das Ziel ist die eindeutige Referenzierung und Auffindbarkeit von Personen. Dies kann beispielsweise sichergestellt werden, indem Personendatensätze durch biografische Daten ergänzt werden und der einzelnen Person im Anschluss ein Identifikator zugeordnet werden kann.³⁶

4. Normdaten

Eine wichtige Rolle kommt sogenannten Normdaten zu, die verwendet werden, wenn eine eindeutige Benennung von Entitäten gewünscht wird. In einer Normdatei (authority file) werden diese dann gesammelt. Eine Entität ist in diesem Zusammenhang eine Informationseinheit, die eindeutig identifizierbar und abgrenzbar

35 Eine gute Möglichkeit, um relevante Terminologien zu recherchieren, bietet das von der Universität Basel entwickelte Nachschlageinstrument Basic Register of Thesauri, Ontologies & Classifications (BARTOC).

36 Neben der weiter unten beschriebenen Gemeinsamen Normdatei (GND) spielt hier das Virtual International Authority File (VIAF) als ein gemeinschaftliches Projekt zahlreicher Nationalbibliotheken und Verbünde eine wichtige Rolle. Insgesamt werden dort die Bestände für Personendaten von 25 Normdateien zusammengebracht und regelmäßig aktualisiert. Die Datenbestände sind miteinander verlinkt und können online recherchiert und genutzt werden. Jeder Datensatz erhält dabei eine eindeutig identifizierbare Normdatennummer in Form eines Uniform Resource Identifiers (URI). Darüber hinaus ist die Orcid iD bedeutsam, die von der Open Researcher Contributor Identification Initiative (ORCID) initiiert wurde und einen nicht-proprietären Code umfasst, mit welchem sich wissenschaftliche Autor*innen ebenfalls eindeutig identifizieren lassen.

ist, wie beispielsweise eine Bezeichnung für eine Sache, eine Person oder ein Ort. Diese Entitäten können durch Begriffsdefinitionen und eindeutige Identifikationsnummern, sogenannte persistente Identifikatoren (z. B. Digital Object Identifier, DOI), Begriffsdefinitionen sowie Quellenangaben angereichert werden. Auf dieser Grundlage können Ressourcen weitere semantische Beziehungen zugewiesen werden, die sie wiederum mit anderen verknüpfbar machen und ein Netz von verbundenen Datensätzen (linked data) ermöglichen, das sich im Web vielfältig nutzen lässt.³⁷ Die Verknüpfung und eindeutige Referenzierung von Ressourcen ermöglichen damit eine erhöhte Sichtbarkeit, Verfügbarkeit und Zitierbarkeit in heterogenen und neuen Anwendungskontexten.³⁸

Kontrollierte Vokabulare schaffen ein höheres Maß an Konsistenz in Hinblick auf die Erschließung, indem sie auf eine einheitliche Benennung von Sachverhalten, Personen und Körperschaften zurückgreifen. Damit ermöglichen sie eine präzisere und qualitativ hochwertigere Metadatenbeschreibung. Metadaten können auf diese Weise harmonisiert und vernetzt werden. Zugleich kann auf individualisierte Personen oder andere Entitäten referenziert und so Informationen angereichert werden. Das liefert bei der Recherche verlässlichere Ergebnisse und erhöht die Auffindbarkeit von Publikationen und verteilten Datenbeständen. Im deutschsprachigen Raum verknüpft beispielsweise Kalliope als überregionales Nachweisinstrument für Nachlässe, Autographen und Verlagsarchive diese Bestände mit den Normdaten der Gemeinsamen Normdatei (GND). Ein Darstellungsformat, das sich für Anwendungen jenseits des reinen Lesens von elektronischen Texten eignet, ist die Extensible Markup Language (XML). Als Auszeichnungssprache kann sie hierarchisch strukturierte Daten darstellen, die auch maschinenlesbar sind. Die Vorteile von XML sind, dass sie nicht an eine spezielle Software gebunden ist, sich leicht zwischen verschiedenen Systemen austauschen lässt und sich gut für die Datenarchivierung eignet.³⁹

37 Das Semantic Web, als erweitertes Netz von Webressourcen und Datensätzen, ermöglicht die Verlinkung von verschiedenen Ressourcen aus unterschiedlichen Fachcommunities, die sich auch maschinell verarbeiten lassen. Menschen und Computer könnten auf diese Weise besser interagieren. Dazu müssen die Ressourcen mit eindeutigen Uniform Resource Identifiers (URIs) identifiziert und beschrieben werden, siehe dazu Joudrey, D.; Taylor, A. (2018), S. 32. Auf Grundlage dessen ermöglicht beispielsweise das Resource Description Framework (RDF), verschiedene Ressourcen als linked data mittels logischer und maschinenlesbarer Aussagen miteinander zu verbinden. Dabei sind die sogenannten Tripel wie ein einfacher Satz aufgebaut, der ein Subjekt, ein Prädikat und ein Objekt enthält, siehe dazu Haffner, A. (2012), S. 5. Auf diese Weise können Aussagen getroffen werden, z. B. „Berlin ist die Hauptstadt von Deutschland“. Mittels Abfragesprachen wie SPARQL ist es dann auf Grundlage der Datenverknüpfung möglich, umfangreiche und komplexe Suchanfragen zu stellen.

38 Lill, J. (2019), S. 18.

39 Gartner, R. (2016), Anm. 1, S. 56.

Mit kontrollierten Vokabularen erschlossene Ressourcen stellen somit grundlegende Qualitätskriterien für eine standardisierte Datenerfassung und einen standardisierten Datenaustausch sicher. Bezüglich des Ressourcenteilbereichs der Forschungsdaten befördert dies auch die in den FAIR-Prinzipien formulierten Ziele – „Findable, Accessible, Interoperable, Reusable“⁴⁰ –, die ein nachhaltiges Forschungsdatenmanagement, eine erhöhte Interoperabilität und umfassendere Zugänglichmachung der Ressourcen einfordern.

Es besteht jedoch Weiterentwicklungsbedarf auf der Ebene der Rechtersysteme, um die Möglichkeiten des Schlagwortnetzes⁴¹ umfassend als Filter und Verweisung nutzen zu können. Wiesenmüller weist darauf hin, dass es in vielen Online-Katalogen nicht problemlos möglich ist, sich bei der Schlagwortrecherche ebenfalls verwandte bzw. unter- und übergeordnete Bezeichnungstreffer darstellen zu lassen, obgleich diese Informationsrelationen in den Normdateien verzeichnet sind.⁴² Dies würde ermöglichen, dass die Nutzenden auf zusätzliche, ebenfalls relevante Ressourcen aufmerksam gemacht werden könnten. Die Entwicklung von „navigierbare[n] Visualisierungen von Fachgebieten (mit Wikipedia-Anbindung, Diensten zur direkten Übernahme von Definitionen und Ähnliches) und den damit assoziierten Beständen“⁴³ würde den weiteren Nutzungsmehrwert ebenfalls erhöhen.

5. Die Gemeinsame Normdatei (GND)

Im deutschsprachigen Raum ist insbesondere die GND ein zentraler Referenzpunkt für kontrolliertes Vokabular.⁴⁴ Sie ist ein Zusammenschluss aus verschiedenen, bereits zuvor existenten Normdateien. Deren Pflege wird durch ein Redaktionssystem

40 Wilkinson, M. D.; Dumontier, M. et al. (2016)

41 Dieses Schlagwortnetz entsteht durch die Verknüpfung verschiedener Begriffsdatensätze beispielsweise in Form von abstrakteren Oberbegriffen mit präziseren Unterbegriffen. Bei der Recherche könnte bei zu vielen Suchtreffern auf die detaillierteren Unterbegriffe zurückgegriffen werden. Gibt es wenig Suchtreffer, könnten die allgemeineren Oberbegriffe genutzt werden, um den Ergebnismenge zu erweitern.

42 Wiesenmüller, H. (2018), S. 29.

43 Kasprzik, A.; Kett, J. (2018), S. 138.

44 In der 2016 durchgeführten Befragung zum Thema Objekterschließung in Bibliotheken, Museen und Archiven antworteten 75 % der Befragten (insbesondere in Bibliotheken und Museen), dass sie auf kontrollierte (normierte) Vokabulare zurückgriffen. Folgende Vokabulare wurden verwendet: „Neben der Gemeinsamen Normdatei / GND (35) und einem eigenen Thesaurus (27) wurden folgende Thesauri genannt: Wikipedia (7), GeoNames (4), VIAF (3), Icon-Class (2), Library of Congress Authorities (1), RDA (1), Eine eigene Klassifikation (1), TU Systematik (1), Catalogue of Life (1), Index Kewensis (1), id.loc.gov (1), ethnologue.com (1), RVK (1), OBZ (1).“ Vgl. Marković, B.; Kmyta, O. et al. (2016), S. 418f.

Im englischsprachigen Raum nehmen die Library of Congress Subject Headings als präkombiniertes, kontrolliertes Vokabular eine zentrale Rolle ein. Für eine dahingehende Übersicht siehe Stone, A. T. (2000), S. 1-15.

übernommen, das sich aus verschiedenen deutschsprachigen Bibliotheksverbünden rekrutiert. In der GND selbst sind derzeit circa neun Millionen Datensätze angelegt. Diese umfassen normierte Datensätze für Geografika, Körperschaften, Kongresse, Personen, Sachschlagwörter sowie Werktitel. Bislang wurden diese hauptsächlich zur Medienkatalogisierung in Bibliotheken genutzt. Zusehends finden sie aber auch umfassendere Anwendung zur Kulturgutvernetzung in Archiven, Museen und verschiedenen Projekt- und Webkontexten.

Hier ist für den deutschsprachigen Raum insbesondere das von der Deutschen Forschungsgemeinschaft (DFG) geförderte Projekt GND für Kulturdaten (GND4C) zu nennen, das auf einen Ausbau und eine Vernetzung verschiedener Institutionstypen abzielt. Der Fokus liegt dabei bislang auf Geografika, Personen, Sachbegriffen und Werken (Bau- und Kunstwerke).⁴⁵ Die vier Hauptziele sind: 1. Nachhaltiger Aufbau einer sparten- und fächerübergreifenden Organisation, 2. Weiterentwicklung des Datenmodells und der Regeln im Hinblick auf nicht-bibliothekarische Anwendungskontexte, 3. Bereitstellung von Schnittstellen und Werkzeugen zur Unterstützung nicht-bibliothekarischer Anwendungskontexte, 4. Stärkung der Kommunikation mit den verschiedenen Interessengruppen über verschiedene Kommunikationskanäle und Sichtbarmachen des GND-Netzwerks.⁴⁶

Die Datensätze basierten lange auf dem literary warrant (gedruckten Publikationsaufkommen). Das heißt, ein Schlagwort wurde nur dann angelegt, wenn es nachweislich Literatur gab, die nicht mit den bereits vorhandenen Begrifflichkeiten beschrieben werden konnte und somit ein neues Schlagwort zur Beschreibung forderte. Heute können Datensätze sowohl proaktiv als auch auf Grundlage anderer Bedarfe aus unterschiedlichen Bereichen des GLAM-Sektors angelegt werden. Innerhalb der GND wird einzelnen Entitäten jeweils eine eindeutige Identifikationsnummer zugewiesen. Bei Personen werden neben der normierten Hauptbenennung auch abweichende Namensformen im Datensatz hinterlegt. Datensätze können mit zusätzlichen biografischen Informationen angereichert werden, was eine eindeutige Zuordnung von Personen und eine Vernetzung mit anderen Datensätzen möglich macht.⁴⁷ Dies umfasst Lebensdaten, Geburts- und Sterbeorte, Berufe, Affiliationen, aber auch die Verknüpfung mit publizierten Werken. Bei Körperschaften (z. B. Universitätsinstituten) können auch ehemalige Bezeichnungen, Nachfolge- und Vorgängereinrichtungen und die Einbettung in größere Verwaltungseinheiten

⁴⁵ Lill, J. (2019), Anm. 38, S. 19.

⁴⁶ Kett, J. (2019), S. 62.

⁴⁷ Ohne diese zusätzlichen Informationen ist es nicht möglich, Personen mit gleichem Namen zu unterscheiden; beispielsweise: Beck, Kurt; Ethnologe, 1952- | 134285603, Beck, Kurt; Politiker, Elektromechaniker, 1949- | 120301342, Beck, Kurt; Werkzeugmacher, Fotograf, 1909-1983 | 1027234143.

dargestellt werden. Vielfach besitzt die GND zudem Verknüpfungen mit anderen Normdateien wie die der Library of Congress, sodass auch viele Ressourcen, die mit englischsprachigen Schlagwörtern versehen worden sind, mittels GND-Begriffen auffindbar sind. Die GND basiert dabei auf dem Prinzip der Postkoordination: Auf das Zusammensetzen von langen, festen Terminiketten wird verzichtet. Es werden vielmehr mehrere Schlagworte vergeben, die den komplexen Sachverhalt darstellen. Nutzende können dies bei der Suche im Bibliothekskatalog durch die Verwendung von booleschen Operatoren anwenden. Die GND kann darüber hinaus auch als Nachschlagewerk genutzt werden, um zu eruieren, welche Vorzugsbenennungen für die Recherche und die eigene Verschlagwortung geeignet sein könnten.⁴⁸ Jens Lill weist darauf hin, dass für den deutschsprachigen Museumsbereich weiterhin verschiedenste Erfassungssysteme genutzt werden, die gängige Datenmodelle und Metadatenstandards häufig nicht ausreichend einbeziehen. Die GND ist, bedingt durch ihre hohe Verbreitung, grundsätzlich gut geeignet zur Normdatenanreicherung. Einschränkend ist jedoch zu beachten, dass zahlreiche für Museen relevante Begriffe noch nicht Teil der GND sind.⁴⁹

6. Kritik an kontrollierten Vokabularen

Der weiter oben beschriebene Rekurs auf das Publikationsaufkommen, Einschränkungen durch bibliothekarische Regelwerke und mangelnde Ressourcen für eine kontinuierliche Datenanreicherung und -pflege sowie die unterschiedliche Güte einzelner Datensätze via Fremddatenübernahme führen dazu, dass die GND für einige Anwendungsbereiche unzureichend ist. Es fehlen Personen – beispielsweise relevante Personen aus dem Archiv- oder Museumsbereich, die teilweise im bibliothekarischen Kontext bislang nicht erwähnt wurden – oder die dazu gehörigen Datensätze verfügen über nicht ausreichende Angaben. Außerdem mangelt es an umfassenden Verknüpfungen zwischen Begriffen, sodass keine durchgängige Thesaurusstruktur vorhanden ist.

Neben diesen Leerstellen existiert eine politisch-ethische Kritik an Normdaten.⁵⁰ Kritiker:innen weisen darauf hin, dass Klassifikationssysteme nach Möglichkeit

48 Es existieren verschiedene Recherchemöglichkeiten, die einen Online-Zugriff auf die GND ermöglichen, beispielsweise die Homepage der OGND, die vom Bibliotheksservice-Zentrum Baden-Württemberg betreut wird. Lobid.org, gehostet vom Hochschulbibliothekszentrum des Landes NRW, besitzt neben einer Rechercheoberfläche auch Schnittstellen (APIs), mit denen ein automatisierter Abgleich und eine Harmonisierung von Daten mit der GND möglich ist.

49 Lill, J. (2019), Anm. 38, S. 20.

50 Für eine weitergehende Betrachtung dieses Komplexes siehe: Strickert, M. (2021). Für den deutschsprachigen Raum wurden zudem viele dieser Fragestellungen im Rahmen des digitalen Denkla-

vieldeutig und flexibel sein sollten, kontrollierte Vokabulare hingegen zu statisch seien. Die angestrebte Universalsprache hätte nur eine Scheinneutralität, wohingegen Wissen und Kategorien notwendigerweise (historisch) situiert sind. Die verwendeten Ausdrucksweisen zur Beschreibung von Themen sind unweigerlich ideologisch gefärbt⁵¹ und privilegieren bestimmte Subjektpositionen und -perspektiven.⁵² Es steht dabei in Frage, inwiefern eine Normierung die historische und kulturelle Perspektivität von Wissen berücksichtigen und trotz universellen Anspruchs gleichzeitig Diversität und Multiperspektivität in sich aufnehmen kann. Zugleich wird die Vorstellung von einem bzw. einer „neutralen Durchschnittsleser:in“ kritisiert, an die sich die Erschließungsarbeit richten soll.⁵³

Die Untersuchung von Bias in kontrollierten Vokabularen findet in den USA bereits seit den späten 1960ern statt.⁵⁴ Ein Diskussionsansatz stellt die demokratische Teilhabe mittels Folksonomien und anderen kooperativen Praktiken, wie der Integration von Suchtermini von Nutzer:innen in (kontrollierte) Vokabulare, dar.⁵⁵ Einige Autor:innen weisen zudem darauf hin, dass eine schlichte Umbenennung die Spuren von Historizität und Ideologie unsichtbar machen würde und dass vielmehr die Brüche in der Benennung deutlich gemacht werden sollten.⁵⁶ Dies gilt umso mehr,

bors Critical Library Perspectives (<https://lab.sbb.berlin/events/critical-library-perspectives/#programm>) erörtert. Eine Dokumentation der Projektergebnisse findet sich unter: 027.7 Zeitschrift für Bibliothekskultur. Critical Library Perspectives 9 (4) (2022).

51 Gartner, R. (2016), Anm. 1, S. 42.

52 Adler, M. (2017), S. 2.

53 Knowlton, S. A. (2005), S. 124.

54 Sanford Berman veröffentlichte mit „Prejudices and Antipathies. A Tract on the LC Subject Heads Concerning People“ die erste umfassende Auseinandersetzung mit problematischen Begrifflichkeiten im Hinblick auf die Library of Congress Subject Headings. Neben Bermans jahrzehntelanger Arbeit in diesem Feld gibt es auch immer wieder Interventionen zu einzelnen Termini, siehe hier exemplarisch die Arbeit zum Terminus „East Indians“ von Biswas, P. (2018), S. 1-18. Oder der Streit um die Ersetzung der Benennung „Illegal Aliens“ durch die Bezeichnung „Undocumented Immigrants“, der eine politische Debatte auslöste, in die sich sogar US-Senatsmitglieder einbrachten, siehe dazu Aguilera, J. (2016); Lo, G. (2019), S. 170-196. Siehe für grundsätzliche (queer-)theoretische Erwägungen zu diesem Thema: Drabinski, E. (2013), S. 94.

Auch im deutschsprachigen Raum gibt es Kritik: So wird der Androzentrismus innerhalb der GND bemängelt, der nicht nur politisch problematisch sei, sondern gleichzeitig auch präzise Suchanfragen erschwere und zusätzlichen Rechercheaufwand schaffe (vgl. Aleksander, K. (2014), S. 15; Marković, B.; Kmyta, O. et al. (2016); Sparber, S. (2016)). Eine weitere Möglichkeit stellt die projektbezogene Anpassung von kontrollierten Vokabularen dar. Ein Beispiel ist die AMA MAIN-LCSH Working Group, die die Library of Congress Subject Headings mit Bezug auf indigene Gruppen für den kanadischen Archivkontext anpasste und dabei kulturell unsensible Begriffe durch solche ersetzte, die dem indigenen Sprachgebrauch mehr entsprechen, siehe dazu Bone, C.; Loughheed, B. (2018), S. 84.

55 Olson, H. A. (2002)

56 Drabinski, E. (2013), Anm. 54, S. 101.

zumal jede Kategorisierung notwendigerweise Ausschlüsse schafft und Änderungen immer Kontingenz aufweisen sowie zeitlich unabgeschlossen sind.⁵⁷

7. Fazit und Ausblick

Metadaten spielen eine zentrale Rolle bei der Archivierung, Erschließung und Auffindbarkeit von Ressourcen. Eine Recherche über verschiedene Bestände und Portale hinweg ist dabei auf eine homogene Erschließung basierend auf konsistenten und aktuell gehaltenen, kontrollierten Vokabularen angewiesen. Einzelnen Forscher:innen helfen kontrollierte Vokabulare und Normdateien insofern, als sie der Orientierung dienen können, was gebraucht wird, wenn diese z. B. selbst Schlagwörter vergeben müssen, sodass eine größere Qualität der Ressourcenbeschreibung möglich ist.⁵⁸ Gleichzeitig erhöht eine präzise Begriffssammlung die Beschreibungsqualität bei der inhaltlichen Erschließung. Dies schafft ergänzende Sucheinstiege und liefert passendere Rechercheergebnisse, die das Potenzial einer höheren Auffindbar- und Sichtbarkeit der Ressourcen bieten.

Kontrollierte Vokabulare und ein darauf basierender Austausch von Metadaten ermöglichen eine Stärkung der Kooperation zwischen verschiedenen Institutionen sowohl aus dem universitären als auch außeruniversitären Bereich (Museen, Archive, Spezialbibliotheken etc.). Dies gilt trotz der berechtigten Kritik an jenen und ihrer zu reflektierenden politisch-historischen Perspektivität und Eingebundenheit. In Hinblick auf zukünftige Entwicklungen liegt ein großes Innovationspotenzial von Normdaten darin, dass diese die Grundlage für ein internationales semantisches Netzwerk sein können, das bislang voneinander getrennte Ressourcen aus den verschiedensten Institutionen maschinenlesbar zusammenführt und für weitere Anwendungen nutzbar macht. Auf diese Weise können neue Verbindungen etabliert, Bestände umfassender zugänglich gemacht und schließlich neues Wissen gewonnen werden. Um dies zu realisieren, müssen Ressourcen jedoch tatsächlich auf (aktuelle) Normdaten zurückgreifen und mit diesen durchgängig erschlossen werden.

⁵⁷ Adler, M. (2017), Anm. 52, S. 157.

⁵⁸ Für eine Anleitung, die Interessierten zeigt, wie sie bei der Recherche nach Normdaten vorgehen können, die als Schlagwörter z. B. bei der Einreichung eigener Artikel oder der Datenabgabe an Repositorien, aber auch zur Recherche verwendet werden können, siehe: Strickert, M. (2023).

Bibliografie

- 027.7 Zeitschrift für Bibliothekskultur (2022), 9 (4): Critical Library Perspectives.
<https://doi.org/10.21428/1bfadeb6.2d09bc20>
- Adler, Melissa (2017): *Cruising the Library. Perversities in the Organization of Knowledge*. 1st edition. New York: Fordham University Press.
- Aguilera, Jasmine (2016): Another Word for “Illegal Alien” at the Library of Congress: Contentious. *The New York Times*, 23.07.2016. <https://www.nytimes.com/2016/07/23/us/another-word-for-illegal-alien-at-the-library-of-congress-contentious.html> (abgerufen am 17.05.2023)
- Aleksander, Karin (2014): Die Frau im Bibliothekskatalog. In: *LIBREAS* 25. <https://libreas.eu/ausgabe25/02alexander/> (abgerufen am 17.05.2023)
- Bambey, Doris; Meyermann, Alexia; Porzelt, Maike et al. (2018): Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung. Das Forschungsdatenzentrum (FDZ) Bildung am DIPF. In: Hollstein, Betina; Strübing, Jörg (Hg.): *Archivierung und Zugang zu qualitativen Daten*. 1. Aufl. Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD Working Paper 267/2018), S. 59-68.
- Beall, Jeffrey (2008): The Weaknesses of Full-Text Searching. In: *Journal of Academic Librarianship* 34 (5), pp. 438-444.
- Berman, Sanford (1971): *Prejudices and Antipathies. A Tract on the LC Subject Heads Concerning People*. Lanham: Scarecrow Press Books.
- Biswas, Paromita (2018): Rooted in the Past: Use of “East Indians” in Library of Congress Subject Headings. In: *Cataloging & Classification Quarterly* 56 (1), pp. 1-18.
- Bone, Christine; Loughheed, Brett (2018): Library of Congress Subject Headings Related to Indigenous Peoples. Changing LCSH for Use in a Canadian Archival Context. In: *Cataloging & Classification Quarterly* 56 (1), pp. 83-95.
<https://doi.org/10.1080/01639374.2017.1382641>
- Ceynowa, Klaus (2017): In Frankfurt lesen jetzt zuerst Maschinen. *Frankfurter Allgemeine Zeitung*, 31.07.2017. https://www.faz.net/aktuell/feuilleton/buecher/maschinen-lesen-buecher-deutsche-nationalbibliothek-setzt-auf-technik-15128954.html?printPagedArticle=true#pageIndex_4 (abgerufen am 17.05.2023)
- Dierkes, Jens (2021): Planung, Beschreibung und Dokumentation von Forschungsdaten. In: Putnings, Markus; Neuroth, Heike; Neumann, Janna (Hg.): *Praxishandbuch Forschungsdatenmanagement*. 1. Aufl. Berlin: De Gruyter Saur, S. 303-326.
- Drabinski, Emily (2013): Queering the Catalog. Queer Theory and the Politics of Correction. In: *Library Quarterly: Information, Community, Policy* 83 (2), pp. 94-111.
- Flachmann, Holger (2004): Erschließung. Zur Effizienz bibliothekarischer Inhaltserschließung. Allgemeine Probleme und die Regeln für den Schlagwortkatalog (RSWK). In: *Bibliotheksdienst* 38 (6), S. 745-791.
- Friedrich, Tanja; Recker, Jonas (2021): Auffindbarkeit und Nutzbarkeit von Daten. In: Putnings, Markus; Neuroth, Heike; Neumann, Janna (Hg.): *Praxishandbuch Forschungsdatenmanagement*. 1. Aufl. Berlin: De Gruyter Saur, S. 405-426.

- Gartner, Richard (2016): *Metadata. Shaping Knowledge from Antiquity to the Semantic Web*. 1st edition. Basel: Springer.
- Gross, Tina; Taylor, Arlene G.; Joudrey, Daniel N. (2015): Still a Lot to Lose. The Role of Controlled Vocabulary in Keyword Searching. In: *Cataloging & Classification Quarterly* 53 (1), pp. 1-39.
- Haffner, Alexander (2012): Internationalisierung der GND durch das Semantic Web. https://wiki.dnb.de/download/attachments/43523047/20120716_internationalisierung-DerGndDurchDasSemanticWeb.pdf (abgerufen am 17.05.2023)
- Heuer, Jan-Ocko; Kretzer, Susanne; Mozygemba, Kati et al. (2020): Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung. Eine Handreichung für Forschende zur Erstellung eines Studienreports. Bremen: Universität Bremen. (Qualiservice Working Papers 1).
- Imeri, Sabine; Sterzer, Wjatscheslaw; Harbeck, Matthias (2019): *Forschungsdatenmanagement in den ethnologischen Fächern*. Berlin: Zentraleinrichtung Universitätsbibliothek. (Schriftenreihe der Universitätsbibliothek der Humboldt-Universität zu Berlin 67).
- Joudrey, Daniel; Taylor, Arlene (2018): *The Organization of Information*. 1st edition. Santa Barbara, California: Libraries Unlimited. (Library and Information Science Text Series).
- Kasprzik, Anna; Kett, Jürgen (2018): Vorschläge für eine Weiterentwicklung der Sacherschließung und Schritte zur fortgesetzten strukturellen Aufwertung der GND. In: *o-bib. Das offene Bibliotheksjournal* 5 (4), S. 127-140.
- Kett, Jürgen; Balter, Detlev; Fischer, Barbara K. et al. (2019): Content kuratieren. Das Projekt „GND für Kulturdaten“ (GND4C). In: *o-bib. Das offene Bibliotheksjournal* 6 (4), S. 59-97.
- Klump, Jens (2010): Digitale Forschungsdaten. In: Neuroth, Heike; Oßwald, Achim; Schefel, Regine; Strathmann, Stefan; Huth, Karsten (Hg.): *nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, S. 104-115. http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf (abgerufen am 17.05.2023)
- Knowlton, Steven A. (2005): Three Decades since Prejudices and Antipathies. A Study of Changes in the Library of Congress Subject Headings. In: *Cataloging & Classification Quarterly* 40 (2), pp. 123-145.
- Lill, Jens (2019): Gemeinsam neu definiert. Das Projekt „GND für Kulturdaten (GND4C)“. In: *AKMB-news* 25 (1), S. 18-23.
- Lo, Grace (2019): “Aliens” vs. Catalogers. Bias in the Library of Congress Subject Heading. In: *Legal Reference Services Quarterly* 38 (4), pp. 170-196.
- Mann, Thomas (2008): Will Google’s Keyword Searching Eliminate the Need for LC Cataloging and Classification? In: *Journal of Library Metadata* 8 (2), pp. 159-168.
- Marković, Barbara; Kmyta, Olga; Sucker, Irina (2016): Objekterschließung an Bibliotheken, Museen und Archiven in Österreich. Ergebnisse einer Erhebung. In: *Mitteilungen der VÖB* 69 (3/4), S. 414-421.
- Mödden, Elisabeth; Schöning-Walter, Christ; Uhlmann, Sandro (2018): Maschinelle Inhaberschließung in der Deutschen Nationalbibliothek. Breiter Sammelauftrag stellt hohe

- Anforderungen an die Algorithmen zur statistischen und linguistischen Analyse. In: Bub Forum Bibliothek und Information 70 (1), S. 30-35.
- Olson, Hope A. (2002): *The Power to Name. Locating the Limits of Subject Representation in Libraries*. 1st edition. Dordrecht: Springer Netherlands.
- Queckbörner, Boris (2019): *Forschungsdaten und Forschungsdatenmanagement in der Geschichtswissenschaft. Gegenwärtige Praxis und Perspektiven am Beispiel ausgewählter Sonderforschungsbereiche*. 1. Aufl. Berlin: Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin. (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft 441).
- Smioski, Andrea (2011): Wegweiser qualitative Datenarchivierung. Infrastruktur, Datenakquise, Dokumentation und Weitergabe. In: SWS-Rundschau 51 (2), S. 219-238.
- Sparber, Sandra (2016): What's the frequency, Kenneth? - Eine (queer)feministische Kritik an Sexismen und Rassismen im Schlagwortkatalog. In: Mitteilungen der VÖB 69 (2), S. 236-243.
- Stone, Alva T. (2000): The LCSH Century: A Brief History of the Library of Congress Subject Headings, and Introduction to the Centennial Essays. In: *Cataloging & Classification Quarterly* 29 (1-2), pp. 1-15.
- Strickert, Moritz (2021): Zwischen Normierung und Offenheit. Potenziale und offene Fragen bezüglich kontrollierter Vokabulare und Normdateien. In: LIBREAS. Library Ideas 40. <https://doi.org/10.18452/23807>
- Strickert, Moritz (2023): Vom Suchen und Finden. Handreichung zur Arbeit mit kontrollierten Vokabularen und Normdateien. <https://doi.org/10.18452/26390>
- Sugimoto, Shigeo; Nagamori, Mitsuharu; Mihara, Tetsuya et al. (2015): *Metadata in Cultural Contexts. From Manga to Digital Archives in a Linked Open Data Environment*. In: Ruthven, Ian; Chowdhury, Gobinda (eds.): *Cultural Heritage Information. Access and Management*. 1st edition. London: Facet Publishing, pp. 89-112.
- Team MusIS im BSZ (2020): *Regelwerke, Thesauri, Klassifikationen, Systematiken und Begriffslisten*. <https://wiki.bsz-bw.de/display/MUSIS/Regelwerke%2C+Thesauri%2C+Klassifikationen%2C+Systematiken+und+Begriffslisten> (abgerufen am 17.05.2023)
- Uhlmann, Sandro (2013): Automatische Beschlagwortung von deutschsprachigen Netzpublikationen mit dem Vokabular der Gemeinsamen Normdatei. In: *Dialog mit Bibliotheken* 25 (2), S. 26-36.
- Wiesenmüller, Heidrun (2018): Maschinelle Inhaltserschließung in der Deutschen Nationalbibliothek. Breiter Sammelauftrag stellt hohe Anforderungen an die Algorithmen zur statistischen und linguistischen Analyse. In: Bub Forum Bibliothek und Information 70 (1), S. 26-29.
- Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan et al. (2016): *The FAIR Guiding Principles for Scientific Data Management and Stewardship*. In: *Scientific Data* 3, 160018.

Moritz Strickert ist Ethnologe, Soziologe und wissenschaftlicher Bibliothekar. Derzeit ist er Mitarbeiter des Fachinformationsdienstes Sozial- und Kulturanthropologie (FID SKA) an der Universitätsbibliothek der Humboldt-Universität zu Berlin. Er arbeitet in einem Projekt zur GND aus ethnologischer Perspektive und ist in der Arbeitsgruppe Thesauri des Netzwerks für nachhaltige Forschungsstrukturen in kolonialen Kontexten aktiv.