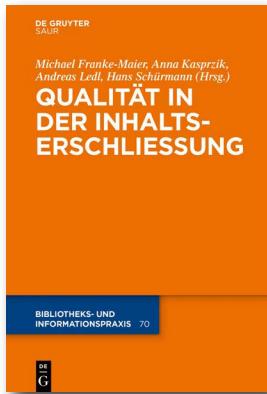


Rezension

<https://doi.org/10.1515/abitech-2022-0058>

Qualität in der Inhaltserschließung. Hrsg. von Michael Franke-Maier, Anna Kasprzik, Andreas Lendl, Hans Schürmann (Buch- und Informationspraxis; 70). Berlin/Boston: De Gruyter Saur, 2021. VI, 420 S. – ISBN: 9783110691498. € 72,95



Der Sammelband widmet sich Qualitätsaspekten in der Inhaltserschließung im Spannungsfeld zwischen etablierten intellektuellen Verfahren auf der einen und maschinellen Verfahren auf der anderen Seite. Die Publikation versteht sich, wie von den Herausgeberinnen und Herausgebern im Editorial beschrieben, als Fortsetzung der Diskussion, die spätestens 2017 mit der Veröffentlichung des angepassten Erschließungskonzeptes der Deutschen Nationalbibliothek (DNB) unter dem Titel *Grundzüge und erste Schritte der künftigen inhaltlichen Erschließung von Publikationen in der Deutschen Nationalbibliothek* aufgekommen ist.

Der Themenbereich des Verhältnisses von maschinellem versus intellektueller Erschließung im Kontext der Qualität der Inhaltserschließung geriet mit der Grundsatzentscheidung der DNB im Jahr 2010, die intellektuelle Erschließung für unkörperliche Medienwerke einzustellen und diese nunmehr auf der Grundlage von Metadaten maschinell zu erstellen, in den Fokus des bibliothekarischen Diskurses. Zu Diskussionen im bibliothekarischen Umfeld führte die im Jahr 2017 bereits angesprochene Veröffentlichung des angepassten Erschließungskonzeptes der DNB, das eine Ausweitung der maschinellen Erschließung

vorsieht.¹ Von der Aktualität des Sammelbandes zeugen nicht zuletzt die Beiträge und die Diskussion im Themenkreis 5 *Indexieren und Sortieren* auf dem Bibliothekskongress in Leipzig 2022.

Eingeleitet wird der Band mit einer kritischen Auseinandersetzung und Reflexion des Begriffs *Qualität*. Ausgehend von der ISO-Norm 9000:2015-1 beleuchten die Herausgeberinnen und Herausgeber ihn aus unterschiedlichen Perspektiven – einer idealistischen, einer rein fachlichen, einer praktischen und einer deskriptiven bzw. historischen. Sie verweisen auf die Multidimensionalität des Qualitätsbegriffs, der je nach Blickwinkel unterschiedlich betrachtet werden kann. Diese Vielschichtigkeit wird auch in den Beiträgen des Sammelbandes immer wieder deutlich und thematisiert.

Eine Fortsetzung der Diskussion des Qualitätsbegriffs im Editorial findet sich gleich im ersten Beitrag von Andreas Ledl zur „Qualität in der Inhaltserschließung – Ein Überblick aus 50 Jahren (1970–2020)“, der die Hauptlinien des Diskurses über Qualität in der Inhaltserschließung der Jahre 1970 bis 2020 zusammenfasst. Ledl zeigt auf, dass der Begriff der Qualität in der Inhaltserschließung in den 1970er Jahren aktuell wurde und verdeutlicht Veränderungen des Qualitätsdiskurses mit der Einführung von Neuerungen wie bspw. der OPAC-Einführung und dem Beginn des kooperativen, bibliothekarischen Arbeitens. In seinen Ausführungen spannt er den Bogen bis zur aktuellen Diskussion um das Verhältnis von automatisierten und intellektuellen Erschließungsverfahren.

Der Beitrag „Fit for Purpose – Standardisierung von inhaltserschließenden Informationen durch Richtlinien für Metadaten“ von Joachim Laczny bietet ebenfalls einen historischen Überblick, fokussiert sich jedoch auf die wissenschaftliche Diskussion zu Qualitätskriterien von Metadaten der letzten 20 Jahre. Der Autor geht der Frage nach, inwieweit Bibliotheken den Qualitätsanspruch an die Inhaltserschließung von Ressourcen durch die Formulierung und Veröffentlichung von bibliotheksspezifischen, übergeordneten Metadatenrichtlinie bzw. -policies beein-

¹ *Grundzüge und erste Schritte der künftigen inhaltlichen Erschließung von Publikationen in der Deutschen Nationalbibliothek*, Stand: Mai 2017; <https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/konzeptWeiterentwicklungInhaltserschliessung.pdf> (10.08.2022).

Rezensionen für ABI Technik: Sie haben Interesse, aktuelle Veröffentlichungen zu rezensieren? ABI Technik nimmt laufend Rezensionen von Neuerscheinungen entgegen. Wenn Sie Interesse haben, wenden Sie sich bitte an Peter Reuter, peter.reuter@bibs.uni-giessen.de. Sie erhalten dann ein Rezensionsexemplar.

flussen können. Im Beitrag werden mit Verweis auf die Initiative *Metadaten 2020* Prinzipien für qualitativ hochwertige Metadaten vorgestellt und Hinweise gegeben, wie diese im bibliothekarischen Kontext praktisch umgesetzt werden können.

Eine Zusammenfassung und kritische Betrachtung der Aktivitäten der DNB im Bereich der maschinellen Erschließung finden sich im Beitrag „Neue Wege und Qualitäten – Die Inhaltserschließung der Deutschen Nationalbibliothek“ von Ulrike Junger und Frank Scholze. Die Autorin und der Autor beschreiben die Grundlinien der Erschließungspolitik der Deutschen Nationalbibliothek und nennen Instrumente und Verfahren, die der Inhaltserschließung zugrunde liegen. Dabei gehen sie auf die Grundsatzentscheidung aus dem Jahr 2010 ein, die eine Einstellung der intellektuellen Erschließung für elektronische Werke vorsah, und fassen die sich daran anschließende Diskussion um die maschinelle Erschließung zusammen. Genannt werden sowohl Vorteile als auch Schwierigkeiten, die im Zusammenhang mit der maschinellen Inhaltserschließung entstehen. Eingegangen wird zudem auf das Erschließungskonzept der DNB aus dem Jahr 2017 sowie auf neueste Anpassungen des Konzeptes. Auch werden Maße und Kriterien beschrieben, die bei der Qualitätssicherung in der DNB verwendet werden.

Der Fokus der folgenden zwei Beiträge liegt auf Wissensbasen, auf die für die automatische Erschließung zurückgegriffen werden kann. In ihrem Beitrag „Wissensbasen für die automatische Erschließung und ihre Qualität am Beispiel von Wikidata“ befassen sich Lydia Pintscher, Peter Bourgonje, Julián Moreno Schneider, Malte Ostdorff und Georg Rehm mit den Eigenschaften von Wikidata und zeigen Werkzeuge und Prozesse zum Auffinden und Beheben von Qualitätsproblemen in Wikidata auf. Wünschenswerte zukünftige Entwicklungen im Bereich der Datenqualität bei Wikidata werden angesprochen. Der Beitrag schließt mit der Vorstellung der Ergebnisse einer Untersuchung des Deutschen Forschungszentrums für Künstliche Intelligenz, in der empirisch der Zusammenhang zwischen der Datenqualität in Wikidata und der Eigennamenerkennung (Named Entity Recognition) dargelegt wird.

Mit der Gemeinsamen Normdatei (GND) als Wissensbasis für die Inhaltserschließung beschäftigt sich Esther Scheven im Beitrag „Qualitätssicherung in der GND“. Die Autorin geht zunächst auf die Historie der GND ein und weist auf die Probleme hin, die durch die Zusammenführung von vier zuvor getrennt geführten Normdateien in der GND entstanden sind. Maßnahmen zur Reduzierung der Heterogenität der GND wie die Einrichtung des GND-Ausschusses oder der GND-Kooperative sowie Aspekte

und die Bedeutung der Qualitätssicherung in der GND werden genannt. Die Bedeutung der GND ergibt sich daraus, dass diese das zentrale Instrument für die verbale Inhaltserschließung in Deutschland darstellt und dadurch starken Einfluss auf die Qualität der verbalen Inhaltserschließung hat.

Der sich anschließende Beitrag „Qualitätskriterien und Qualitätssicherung in der inhaltlichen Erschließung – Thesenpapier des Expertenteams RDA-Anwendungsprofil für die verbale Inhaltserschließung (ET Ravi)“ greift das Thema der regelbasierten Erstellung von Normdaten der vorhergehenden Beiträge auf. Vorgestellt wird ein bislang unveröffentlichtes und für den vorliegenden Sammelband überarbeitetes Thesenpapier des Expertenteams, das 2019 im Auftrag des Standardisierungsausschusses erarbeitet und vorgelegt wurde. Das Expertenteam beklagt das derzeitige Nebeneinander verschiedener Verfahren und Erschließungssysteme und stellt verfahrensunabhängige Qualitätsdimensionen vor. Hervorgehoben wird abschließend die Bedeutung der Verknüpfung von Normdaten in der Inhaltserschließung, um logische Schlussfolgerungen aus den Daten sowie eine Navigation durch die Daten zu ermöglichen.

Wie Infrastrukturen zur Erstellung von Konkordanzen zur Unterstützung der Inhaltserschließung im Rahmen der Fachreferatsarbeit herangezogen werden können, zeigt der Beitrag „coli-conc – Eine Infrastruktur zur Nutzung und Erstellung von Konkordanzen“ von Uma Balakrishnan, Stefan Peters und Jakob Voß. Coli-conc ist ein webbasierter Dienst für die effektive Erstellung und Pflege von Konkordanzen zwischen bibliothekarischen Knowledge Organization Systems (KOS). Der Schwerpunkt des Dienstes liegt auf den im deutschsprachigen Raum verbreiteten Klassifikationen (u. a. DDC, RVK, GND). Beschrieben werden die Architektur von coli-conc, die verwendeten Formate und APIs, die die Interoperabilität zwischen Wissensorganisationssystem ermöglichen. Auch wird der Aspekt der Qualitätssicherung von Konkordanzen angesprochen. Abschließend wird ein Ausblick auf das Mapping-Verfahren mit Hilfe des Konzept-Hubs gegeben, das als Ergänzung zu bisherigen Verfahren prototypisch entwickelt wurde.

Einen Zusammenhang zwischen der OCR-Qualität und der Kuratierung von Daten und Metadaten im bibliothekarischen Kontext stellt der Beitrag „Methoden und Metriken zur Messung von OCR-Qualität für die Kuratierung von Daten und Metadaten“ her. Clemens Neudecker, Mike Gerber, Konstantin Baierer, Julián Moreno Schneider, Karolina Zaczynska und Georg Rehm beschreiben zunächst gängige Methoden und Metriken zur Qualitätskontrolle von OCR-Qualität von Digitalisaten. Bei der Diskussion dreier typischer Anwendungsfälle für die Nutzung von

OCR-Ergebnissen, der Eigennamenerkennung, der Verwendung von Digitalisaten in den Digital Humanities sowie im Bereich des Information Retrievals, wird die Relevanz der OCR-Qualität untermauert. Im bibliothekarischen Bereich kann eine Informationsextraktion aus digitalisierten Texten für die automatische Inhaltserschließung herangezogen werden.

Einen übergeordneten Blick auf Datenqualität als notwendige Grundlage für die Inhaltserschließung wirft der Beitrag „Datenqualität als Grundlage qualitativer Inhaltserschließung“ von Jakob Voß. Der Autor beschäftigt sich aus einem theoretischen Blickwinkel mit dem Themenbereich der Datenqualität und hebt dabei hervor, dass diese mehrere Dimensionen umfasst wie bspw. die Vollständigkeit, die Fehlerfreiheit, die Konsistenz und die Aktualität. Er geht zudem auf Faktoren ein, die die Datenqualität beeinflussen. Diese stammen aus den unabhängig voneinander bestehenden Bereichen der Standards und Regeln, der Anwendungsprogramme und der praktischen Faktoren. Aus der Betrachtung der Faktoren leitet Voß allgemeine Anforderungen an den Umgang mit Daten ab.

Im Gegensatz zu den vorherigen theoretischen Überlegungen zur Datenqualität hat der Beitrag „Bemerkungen zu der Qualitätsbewertung von MARC-21 Datensätzen“ von Rudolf Ungváry und Péter Király einen starken praktischen Bezug. Die Autoren stellen die Ergebnisse einer empirischen Untersuchung zur Qualität von MARC-21-Datensätzen vor und analysieren, in welchem Umfang MARC-21-Felder in den untersuchten Datensätzen genutzt werden, ob Stimmigkeit zwischen verwandten Unterfeldern vorherrscht und inwieweit die Möglichkeiten von MARC-21 ausgeschöpft werden. Untersuchungsgrundlage sind Katalogdaten aus Gesamtkatalogen von National- und Universitätsbibliotheken aus Ungarn, Deutschland und weiteren europäischen Ländern. Der Beitrag macht Probleme von MARC-21 deutlich und kommt zu dem Ergebnis, dass die Datenqualität der untersuchten Katalogdaten stark von der jeweils verwendeten Katalogisierungssoftware abhängig ist.

Auch der Beitrag „Named Entity Linking mit Wikidata und GND – Das Potenzial handkuratorierter und strukturierter Datenquellen für die semantische Anreicherung von Volltexten“ von Sina Menzel, Hannes Schnaitter, Josefine Zinck, Vivien Petras, Clemens Neudecker, Kai Labusch, Elena Leitner und Georg Rehm fasst Ergebnisse einer empirischen Studie zusammen und zeigt auf, wie die in vorherigen Beiträgen formulierten theoretischen Überlegungen zu Normdaten und Wissensbasen in die Praxis übertragen werden können. Nach einer theoretischen Einführung in das Verfahren der Eigennamenerkennung (Named Entity Linking) mit Hilfe von Normdaten aus Wis-

sensbasen und der Vorstellung von Evaluationsansätzen sowie Fehlertypen bei der Verlinkung von Entitäten folgt die Vorstellung der Studie *Interfaces for Data for Historical Social Network Analysis and Research (SoNAR)*. Die Studie verdeutlicht das Potenzial, das die Verlinkung von Eigennamen in historischen Texten mit Hilfe von Normdaten für die inhaltliche Erschließung bietet, und geht darüber hinaus auch auf Schwierigkeiten und Probleme beim Verfahren der Eigennamenerkennung ein.

Der Beitrag „Ein Protokoll für den Datenabgleich im Web am Beispiel von OpenRefine und der Gemeinsamen Normdatei“ von Fabian Steeg und Adrian Pohl beschäftigt sich ebenfalls mit dem Themenbereich des Named Entity Linking, allerdings aus einem anderen Blickwinkel. Betrachtet wird die Qualität von Datensätzen im Kontext anderer Datensätze sowie die Qualität von Tools und Algorithmen, die die Eigennamenerkennung und Verlinkung verbessern können. Am Beispiel der *lobid-gnd* stellen die Autoren die Möglichkeiten einer Reconciliation-API anhand ihrer Verwendung in OpenRefine zum Abgleich mit der GND vor. Diese ermöglicht das Identifizieren eindeutiger Entitäten einer Wissensbasis durch die Ermittlung von Identifikationen für Eigennamen. Daneben geben die Autoren einen Ausblick auf die Arbeit der Entity Reconciliation Community Group des World Wide Web Consortiums (W3C), die ein Protokoll zur Standardisierung der Reconciliation entwickelt.

Eine andere Perspektive auf Qualität in der Inhaltserschließung bietet Heidrun Wiesenmüller in ihrem Beitrag „Verbale Erschließung in Katalogen und Discovery-Systemen – Überlegungen zur Qualität“. Sie fokussiert sich bei ihren Überlegungen nicht auf den Input, sondern betrachtet das, was Recherchewerkzeuge aus Wissensorganisationssystemen und den mit ihnen generierten Metadaten machen. Wiesenmüller bewertet die Qualität des Outputs mit Blick auf die Bedürfnisse der Endnutzer. Der Beitrag ergänzt und vertieft die im Thesenpapier „RDA-Anwendungsprofil für die verbale Inhaltserschließung (ET RAVI)“ gegebenen Hinweise und stellt Qualitätsaspekte in Katalogen vor. Wiesenmüller formuliert zentrale Funktionen von inhaltlicher Erschließung mit Blick auf die Endnutzer (Zugangsfunktion, Orientierungsfunktion und Exploration) und betrachtet im Anschluss aktuelle Rechercheinstrumente in Hinblick auf die drei genannten Funktionen. Dabei kommt sie zu dem Ergebnis, dass Kataloge aktuell noch Defizite in dieser Hinsicht aufweisen.

Auch Jan Frederik Maas beschäftigt sich in seinem Beitrag „Inhaltserschließung für Discovery-Systeme gestalten“ mit dem Zusammenwirken von Inhaltserschließung und Discovery-Systemen. Er macht deutlich, wie und wann Merkmale der Inhaltserschließung in Discovery-

Systemen zur Anwendung kommen (u. a. Suche, Relevanzsortierung und Facettierung), und benennt Schwachstellen zwischen der Inhaltserschließung und dem, wie Discovery-Systeme diese nutzen. Aufgrund der Verschränkung von Discovery und Inhaltserschließung plädiert der Autor dafür, Discovery-Projekte nicht als reine IT-Projekte zu sehen, sondern die Bedeutung der Kommunikation zwischen IT und Bibliothek hervorzuheben, um bessere Ergebnisse für die Endnutzer zu generieren. Als Grundlage schlägt Maas ein gemeinsames Anforderungsmanagement der zwei Bereiche vor.

Auf die Interaktion zwischen Information Retrieval und Verschlagwortung gehen Christian Wartena und Koraljka Golub ein. Der Beitrag „Evaluierung von Verschlagwortung im Kontext des Information Retrievals“ fasst zunächst zusammen, welche Zwecke Verschlagwortung erfüllen soll und kann. Im Anschluss daran diskutieren der Autor und die Autorin, wie eine Evaluierung von Verschlagwortung im Kontext des Information Retrievals funktionieren kann. Der Fokus ihrer Betrachtung liegt dabei auf Next-Generation-OPACs und deren Retrieval-Ergebnissen. Nach einer Beschreibung möglicher Einflussfaktoren auf Retrieval-Ergebnisse wie Suchszenarien und -anfragen, Rankingalgorithmus oder Termgewichtung, ziehen Wartena und Golub das Fazit, dass es sich aufgrund der vielfältigen Einflussfaktoren schwierig gestaltet, Retrieval-Ergebnisse zur Bewertung der Verschlagwortung heranzuziehen.

Mit quantitativen Aspekten der Inhaltserschließung beschäftigen sich die nächsten beiden Beiträge. Eine Anreicherung mit Fremddaten ist vor allem bei E-Books aufgrund der häufig sehr großen Menge an Titeldaten, die für Bibliotheken bereitgestellt werden, unabdingbar. Im Beitrag „Die Qualität der Fremddatenanreicherung FRED“ prüft Cyrus Beck anhand einer Stichprobe einer ausgewählten Fächergruppe die Qualität der Fremddatenanreicherung FRED. Als Grundlage für die Bewertung der Qualität der Datenanreicherung wählt der Autor die Regeln für die Schlagwortkatalogisierung und die vier Grundaspekte von Erschließungsqualität, das sind: Vollständigkeit, Richtigkeit, Präzision und Themenbildung. Beck kommt zu dem Ergebnis, dass die Anforderungen des Regelwerkes für die untersuchte Fächergruppe überwiegend erfüllt werden und nur verhältnismäßig wenige falsche oder nicht präzise Schlagwörter im Rahmen der Fremddatenanreicherung importiert werden.

Der Beitrag „Quantität als Qualität – Was die Verbünde zur Verbesserung der Inhaltserschließung beitragen können“ von Rita Albrecht, Barbara Block, Mathias Kratzer und Peter Thiessen geht der Frage nach, inwieweit Quantität in der Inhaltserschließung selbst als Qualitäts-

kriterium gelten kann. Im Zentrum ihrer Analyse stehen Bibliotheken und Bibliotheksverbünde und ihre Sacherschließungssysteme und Sacherschließungstools für den Datenaustausch. Exemplarisch vorgestellt werden der *Digitale Assistent*, der die intellektuelle Sacherschließung im K10plus unterstützt, sowie das Beschlagwortungsprogramm *hebis-SET* des hebis-Verbunds, das ebenfalls die Nachnutzung von Sacherschließungsdaten ermöglicht. Die Autorinnen und Autoren benennen in ihrem Beitrag Herausforderungen und Potenziale der Nachnutzung von Sacherschließungsdaten und heben hervor, dass von der quantitativen Anreicherung der eigenen Daten mit Fremddaten sowohl Bibliotheken selbst als auch ihre Endnutzer profitieren. Während Bibliotheken von Effizienzaspekten profitieren, die bei der kollaborativen, automatisierten Anreicherung von Datensätzen entstehen, profitieren Endnutzer der Bibliothek bei ihrer Suche in Katalogen oder Discovery Systemen von möglichst umfangreich er-schlossenen Datensätzen.

Der Sammelband schließt mit dem Beitrag „Hybride Künstliche Intelligenz in der automatischen Inhaltserschließung“ von Harald Sack. Der Autor gibt zunächst einen Überblick zur Geschichte des maschinellen Lernens und nennt typische Einsatzgebiete in der Inhaltserschließung wie die automatisierte Verschlagwortung und Klassifizierung, das automatische Transkribieren historischer Dokumente und die automatisierte Erzeugung und Befüllung von Wissensgraphen. Im Bereich des Information Retrievals nennt Sack die semantische und explorative Suche, die Visualisierung von Suchergebnissen sowie Empfehlungssysteme. Sack zeigt sehr gut die Möglichkeiten auf, die neueste Deep-Learning-basierte Verfahren im Bereich der automatisierten Inhaltserschließung bieten können, kommt jedoch zu dem Ergebnis, dass Inhaltserschließung aktuell noch nicht vollautomatisch in allen Bereichen möglich sei. Er hebt abschließend die Vorteile semi-automatisierter Verfahren hervor, bei denen die intellektuelle Erschließung durch maschinelle Verfahren wie bspw. intelligente Vorschlagsmechanismen unterstützt wird.

Das von Sack im letzten Beitrag gezogene Fazit zum aktuellen Verhältnis maschineller und intellektueller Verfahren zieht sich gleichsam als roter Faden durch die 18 Beiträge des Sammelbandes. Anhand der Gesamtschau der in den Beiträgen behandelten Themenbereiche wird deutlich, dass maschinelle, automatisierte Verfahren zwar neue Möglichkeiten der Erschließung bieten und v. a. in quantitativer Hinsicht einen Quantensprung bei der inhaltlichen Erschließung darstellen, dass eine qualitativ hochwertige Inhaltserschließung jedoch nur durch eine Kombination beider Verfahren garantiert werden kann.

Die Beiträge des Sammelbandes zeigen nicht nur die im Editorial skizzierte Mehrdimensionalität des Qualitätsbegriffes auf, sondern auch das weite Spektrum an Bereichen und Aspekten, die Einfluss auf die Qualität der Inhaltserschließung haben und daher bei der Qualitätsbewertung berücksichtigt werden müssen.

In dieser inhaltlichen Ausrichtung passt der Sammelband *Qualität in der Inhaltserschließung* sehr gut in die Reihe *Bibliothek- und Informationspraxis*. Es werden nicht nur neue Themen und Fragestellungen aus dem Bereich der automatischen Erschließung genannt, sondern auch Erfahrungen aus der Praxis sowie aktuelle Forschungsarbeiten aus dem Themenbereich vorgestellt. In methodischer Hinsicht bietet der Sammelband wichtige Hinweise, die zur Bewertung der Qualität der Inhaltserschließung herangezogen werden können. Diese sind nicht nur für die bibliotheks- und informationswissenschaftliche Forschung relevant, sondern können auch für Praktiker in Informationseinrichtungen als Orientierung zur Entwick-

lung von Qualitätssicherungsmaßnahmen bei der Arbeit mit eigenen Daten herangezogen werden.

Im Sinne der Lesefreundlichkeit und besseren Erreichbarkeit des Sammelbandes wäre es wünschenswert gewesen, die Beiträge in verschiedenen thematischen Abschnitten zu gruppieren. Bei der Lektüre des gesamten Bandes sowie beim Blick ins Editorial wird zwar deutlich, dass die Beiträge nicht willkürlich angeordnet sind, diese Struktur hätte aber sichtbarer herausgestellt werden können. Aus Sicht der interessierten Lesenden wäre zudem eine einheitlichere Gestaltung der einzelnen Beiträge, z. B. in Form von vorangestellten Abstracts und abschließenden Zusammenfassungen, sinnvoll gewesen. Auch dies würde die Orientierung für die Leserinnen und Leser erheblich erleichtern und helfen, die für sie interessanten und relevanten Inhalte schnell zu finden.

Dr. Helena Meyer
UB Gießen