

Rezensionen

Paul Gooding, Melissa Terras, Sarah Ames (Hrsg.): Library Catalogues as Data – Research, Practice and Usage. London: Facet Publishing, 2025. 228 S.

Besprochen von **Vivien Petras**: Institut für Bibliotheks- und Informationswissenschaft, Humboldt-Universität zu Berlin, Dorotheenstr. 26, D-10099 Berlin, E-Mail: vivien.petras@ibi.hu-berlin.de

<https://doi.org/10.1515/bfp-2025-0037>

Bibliothekskataloge dienten seit jeher der Dokumentation der Bestände ihrer haltenden Institutionen und der Recherche bzw. dem Zugang der Bibliotheksnutzer*innen. Die bibliothekarischen Katalogmetadaten u. a. Daten (z. B. Ausleih- und Nutzungsdaten) werden ebenfalls für das Bestandsmanagement und die Bestandserhaltung benutzt.

Der Sammelband zeigt nicht nur, wie bibliothekarische Daten für das institutionelle Management und die Bestandsbeschreibung genutzt werden, sondern auch, wie sie einer interdisziplinären Forschungsgemeinschaft Einblicke in die Epistemologien und den Zeitgeist bibliothekarischer und nutzender Communities eröffnen.

Thomas Padilla eröffnet in seinem Vorwort einen weiten, beinahe lyrischen Rahmen für die unterschiedlichen Forschungsinteressen: „A library catalogue is an accretion, a fossil record of subjective decisions that evidence what an organisation operating in a particular sociocultural context believes is important and, consequently, what is not important. The library catalogue channels contested collection descriptions that directly influence how communities encounter and assess knowledge, circulation data complicates any sense of certainty with respect to the zeitgeist, and in the aggregate, library catalogues provide a dynamic map that characterises the waxing and waning of knowledge production globally.“ (S. xxv).

Die Herausgeber*innen (von der schottischen Nationalbibliothek sowie den Universitäten in Edinburgh und Glasgow) verstehen Bibliothekskataloge als Teil eines Datenökosystems, in dem unterschiedliche Datenquellen und -ressourcen (z. B. Lizenzdaten, Suchsysteme, digitalisierte Materialien, Webressourcen oder Verlagsinformationen) auf verschiedene Bibliotheksfunktionen und -aktivitäten treffen (Beschreibung und Erschließung, Bestandsmanagement, Nutzungsaktivitäten) und deren Interaktionen auch wieder Daten hervorrufen. In ihrer prägnanten Einführung führen sie den Leser vom traditionellen bestandsfokussierten Bibliothekskatalog über das Bibliotheksmanagementsystem (integrated library system) zum Datenökosystem bi-

bliothekarischer Daten. Erst in der „Konversation“ (S. xxxi) aller Ressourcen miteinander entsteht ein komplexes und facettenreiches Bild der modernen Bibliothek und ihrer Aktivitäten.

Sie beschreiben das Ziel des Sammelbandes als „imagine and define a field of study which seeks to bring together the various siloes of data contained within the library system in order to enhance our understanding of library collections, practices, users, and more“ (S. xxxi), also der Definition einer gemeinsamen Forschungsagenda bzw. eines Forschungsfeldes zu Bibliotheksdaten. Dieses muss zwangsläufig interdisziplinär betrachtet und aufgrund des Volumens der Daten mit automatisierten und maschinellen Verfahren angegangen werden. Die Herausgeber*innen formulieren nicht nur eine Reihe von Forschungsfragen, sondern beschreiben auch die Herausforderungen – technisch, rechtlich und menschliche Komplexität, die in jeder Analyse berücksichtigt werden müssen. Mit dem Aspekt der menschlichen Komplexität erkennen sie an, dass bibliothekarische Daten Wissensstrukturen darstellen, die von Menschen und ihren Intentionen, Perspektiven und Kulturen geprägt sind. So sollen Analysen des bibliothekarischen Datenökosystems eben nicht nur sammlungsbasiert und deskriptiv bleiben, sondern können – als Reflektion menschlicher Praktiken und Interaktionen – zugleich gesellschaftliche Strukturen und Praktiken widerspiegeln.

Im Sammelband werden zehn Aufsätze von einem internationalen Autor*innenkollektiv (Bulgarien, Dänemark, Deutschland, Finnland, Niederlande, Polen, Schweiz, Spanien, Tschechien, UK, Ungarn, USA) präsentiert, die diese unterschiedlichen Analyseansätze beispielhaft verdeutlichen sollen. Nicht nur die Internationalität, sondern auch die disziplinäre und institutionelle Herkunft der Autor*innen (Archivar*innen, Bibliothekar*innen, Forscher*innen, Kurator*innen aus Bibliotheks- und Informationswissenschaft und bibliothekarischen Praxis, Digital Humanities, Geschichtswissenschaften, Informatik, Literaturwissenschaft, Sozialwissenschaft u. a.) gewährleistet ein breites Spektrum an Perspektiven, methodischen Ansätzen und Forschungsfragen.

Die Herausgeber*innen teilen den Band in vier Sektionen auf, die (1) Methoden zur Analyse bibliothekarischer Daten (Kapitel 1–3), (2) die Politik- und Machtstrukturen von Katalogen und Bibliotheksdaten (Kapitel 4–5), (3) die interdisziplinären Potenziale (Kapitel 6–7) sowie (4) praktische Nutzungen und Anwendungen, insbes. im Zeitalter der KI (Kapitel 8–10) von Katalogen als Daten aufzeigen.

Wie bei einer wissenschaftlichen Anthologie zu erwarten, sind die Beiträge durch unterschiedliche Abstraktionsgrade, wissenschaftliche Methodiken und Schreibstile geprägt, die, genau wie von den Herausgeber*innen intendiert, ein vielschichtiges Bild der Forschungslandschaft zu Katalogen als Daten darbieten. Gemein ist allen Beiträgen eine konsequente und mit hilfreichen Zwischenüberschriften, Grafiken und Tabellen untermalte Darstellung und zahlreiche Quellen, die nicht nur zur weiteren Lektüre einladen, sondern ein umfangreiches Quellenwerk zu weiteren Studien und Perspektiven im bibliothekarischen Datenökosystem darbieten.

In einem interessanten Auftaktkapitel verdeutlicht Brian Lavoie von OCLC anhand mehrerer Beispiele mit OCLC-Verbunddaten (WorldCat), wie kollektive Sammlungsdaten (collective collections), also Verbunddaten, für strategische Entscheidungen im Bestandsmanagement benutzt werden können. So zeigten z. B. verschiedene Analysen für die GoogleBook-Digitalisierung oder die Bestände US-amerikanischer und kanadischer Bibliotheken in WorldCat die Überlappungen und einzigartigen Bestände dieser Bibliotheken. Dieses Kapitel zeigt nicht nur die Macht großer Datenbestände und der damit möglichen Analysen auf, sondern plädiert – wenn auch nicht ausdrücklich – für die Entwicklung weiterer Verbunddatensammlungen.

Das zweite Kapitel ist eine aktualisierte und gekürzte Version des DARIAH-ERIC Konsortium White Papers zur Analyse der bibliografischen Datenlandschaft in den Geisteswissenschaften. Hier wird eine abstrakte Perspektive eingenommen und die verschiedenen Stakeholder der „Bibliodata“ Landschaft textlich und sehr schön visuell kartiert. Dies ist ebenfalls eine sehr gute Einführung in die Herausforderungen, die in dieser integrierten Landschaft für Infrastrukturen und andere Akteure entstehen.

Das dritte Kapitel in der Sektion Methoden beschreibt ausschnittsweise quantitative Indikatoren für die Datenqualität in Bibliothekskatalogen und veranschaulicht diese anhand von drei Szenarien zur Anreicherung von Katalogdaten, Diversität von Katalogen sowie Shape Expressions zur Validierung von RDF-Daten.

Kapitel 4 beschreibt eine äußerst interessante Fallstudie zur maschinellen Identifikation mittels maschineller Sprachverarbeitung (NLP) von Geschlechterbias in Archivkatalogdaten, z. B. die Auslassung von weiblichen Akteurinnen in Beschreibungen. Konkrete Beispiele verdeutlichen die Problematik.

Kapitel 5 – ebenfalls in der Sektion zu Machtstrukturen – geht mit der Brille der 2023 Cyberattacke auf die British Library der Frage nach, wie umfassend und sicher die sog. Dark Archives, die im Katastrophenfall bibliothekarische u. a. Daten wiederherstellen sollen, diese Daten wie-

derherstellen können. Die Ergebnisse dieser empirischen Studie sind eher niederschmetternd und ein wichtiger Hinweis, auch die digitale Bestandssicherung nicht aus den Augen zu verlieren.

Die nächsten zwei Kapitel sind wiederum wunderbare Fallstudien, die bibliothekarische Daten für die Digital Humanities auswerten können. Kapitel 6 zur Entwicklung von Buchformaten, Druckpraktiken und Lesegewohnheiten im Europa des 18. Jahrhunderts postuliert, dass mit der quantitativ nachweisbaren Entwicklung zu kleineren Printformaten ein individuelles Lesen (anstatt Vorlesen oder Predigen von großformatigen Werken), eine größere Verbreitung der Werke und die Entwicklung von Privatbibliotheken einherging.

Kapitel 7 beschreibt die Ergebnisse eines Projekts zur Erforschung von Ausleihdaten in 18 schottischen Bibliotheken von 1750–1830 und dokumentiert nicht nur Datengrundlage und Methodik der Analyse, sondern auch Einsichten in das zeitgenössische Leseverhalten, z. B. dass heute als kanonisch geltende Autor*innen nicht zwangsläufig die meistgelesenen waren, oder dass weibliche Autoren auch ausgeliehen wurden.

Die letzte Sektion des Buchs zu praktischen Anwendungen schaut hauptsächlich auf KI-Anwendungen.

Kapitel 8 fällt dabei etwas aus dem Rahmen, da es nicht explizit auf bibliothekarische Daten, sondern auf bibliometrische Daten eingeht, die zum großen Teil von kommerziellen Produzenten extrahiert und verbreitet werden. Die Autoren beschreiben mithilfe von illustrativen Beispielen, wie ChatGPT in unterschiedlichen Phasen bibliometrischer Analysen und der anschließenden Schreibprozesse eingesetzt werden kann und natürlich auch, welche Hürden und Einschränkungen dabei zu beachten sind.

Kapitel 9 untersucht, wie digitalisierte und OCR-te Katalogdaten mithilfe verschiedener großer Sprachmodelle (Claude, GPT4-Turbo, Gemini Advanced) strukturiert für weitere Analysen aufbereitet werden können. Die Fallstudie zeigt mit den konkreten Prompts, wie vier kleine historische Bibliothekskataloge mithilfe der LLMs in feldstrukturierte Daten überführt werden und diese weiter bereinigt und angereichert werden konnten. Jeder Schritt wird genau dokumentiert und evaluiert. Die Studie verdeutlicht, dass selbst bei kleinen Datensammlungen die automatisierte Strukturierung die automatisierte Strukturierung nur bedingt funktioniert und weiterhin erheblicher Korrekturaufwand erforderlich ist.

Auch das letzte Kapitel ist eine interessante korpuslinguistische Fallstudie der Katalogisierungspraxis und -terminologie des Inkunabelkatalogs der British Library (Catalogue of Books Printed in the 15th Century now at the British Museum). Zu Beginn des Kapitels werden weitere Analyse-

projekte der British Library anschaulich vorgestellt, bevor die Methodik und Ergebnisse für die vorliegende Studie beschrieben werden. Die Studie zeigt z. B. die häufigsten Token und Adverbien, die zur Beschreibung der Bücher verwendet wurden, und zeigt eine erstaunlich sprachliche Vielfalt.

Die Kapitel verdeutlichen auf anschauliche Weise, dass Bibliothekskataloge nicht nur Instrumente der Dokumentation und des Zugriffs, sondern Teil eines komplexen Daten-

ökosystems sind, die auch gesellschaftliche, technische und kulturelle Entwicklungen reflektieren können. Je nach Forschungsinteresse werden unterschiedliche Leser*innen sicherlich unterschiedliche Kapitel aufschlussreicher finden. Wie die Herausgeber*innen selbst schreiben: „This landscape will only lead to further insight and questions about the human voices, computational processes and effort behind the data“ (S. xxxvii).