

Bibliotheksdaten als Forschungsdaten in Digital Humanities-Kollaborationen

Librarydata as researchdata in digital humanities collaborations

Martin Krickl

Zusammenfassung

Kooperationen zwischen Bibliotheken und Digital Humanities (DH)-Forschung entwickeln sich zunehmend, auch wenn noch viel Potenzial für Bibliotheken vor allem jenseits einer servicierenden Rolle besteht. Das Modell von embedded-librarianship bietet sich für gleichberechtigte Kooperationen an. Werden Bibliotheksdaten genutzt, sollte für einen gleichberechtigten Austausch der Datenfluss nicht nur unidirektional sein. Katalogdaten können qualitativ hochwertige Forschungsdaten für DH-Projekte sein. Deren Bearbeitung und Bereitstellung sollte nach Schwerpunkten und unter Einbezug des Domänenwissens von Forschungspartnern umgesetzt werden. Damit dies erfolgen kann, ist ein verändertes Verständnis von Katalogen und Katalogdaten nötig.

Schlagwörter: *Digitale Geisteswissenschaften; Bibliothekskatalog; Embedded Librarian; Forschungsdaten; Metadaten*

Abstract

Joint collaborations between libraries and Digital Humanities Researches are expanding. Still there is a lot of potential for libraries especially when leaving the service-approach behind. The concept of embedded librarianship offers a valuable model for such partnerships. Whenever library-data

Martin Krickl, Österreichische Nationalbibliothek | E-Mail: martin.krickl@onb.ac.at

ORCID iD: <https://orcid.org/0000-0002-8569-0340>

Künstliche Intelligenz in Bibliotheken, Hg. v. Köstner-Pemsel, Stadler, Stumpf, 2020, S. 113–128

<https://doi.org/10.25364/guv.2020.voeb5.10>

Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/), ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

becomes research-data, the direction of the data-flow is unfortunately often one-way. Metadata from library catalogs can be research-data of distinct quality. To ensure quality and relevance, libraries should integrate domain-expertise. Nonetheless a different understanding of library catalogs is needed as well.

Keywords: *digital humanities; library catalog; research data; embedded librarian; metadata*

1 Kollaborationen zwischen Bibliotheken und DH-Forschung

Kollaborationen zwischen Bibliotheken und Forschungsprojekten aus dem Bereich der Digital Humanities sind naheliegend, zumal beide als „natural allies“ angesehen werden können.¹ Für eine beiderseits fruchtbare Zusammenarbeit gibt es bereits zahlreiche Studien, Modelle, Guidelines und Best-practice-Beispiele.² Die Skala reicht dabei von „low barrier“ bis

-
- 1 Alex H. Poole: „Natural allies“. Librarians, archivists, and big data in international digital humanities project work. In: *Journal of Documentation* 74 (2018), 4, S. 804–826.
 - 2 Chris Alen Sula: Digital Humanities and Libraries: A Conceptual Model. In: *Journal of Library Administration* 53 (2013), 1, S. 10–26; Micah Vandegrift, Stewart Varner: Evolving in Common: Creating Mutually Supportive Relationships Between Libraries and the Digital Humanities. In: *Ebd.*, S. 67–78; Alix Keener: The Arrival Fallacy: Collaborative Research Relationships in the Digital Humanities. In: *Digital Humanities Quarterly* 9 (2015), S. 2 <http://www.digitalhumanities.org/dhq/vol/9/2/000213/000213.html> (abgerufen am 15.02.2020); Tibor Koltay: Are you ready? Tasks and roles for academic libraries in supporting Research 2.0. In: *New Library World* (2016), 117.1/2, S. 94–104; Samantha Guss: A Studio Model for Academic Data Science. In: *Databrarianship: the academic data librarian in theory and practice*. Hg. von Lynda M. Kellam und Kristi Thompson. Chicago: Association of College and Research Libraries 2016, S. 9–24; Poole: „Natural allies“ (Anm. 1); Miriam Posner: No Half Measures: Overcoming Common Challenges to Doing Digital Humanities in the Library. In: *Journal of Library Administration* 53 (2013), H.1, 43–52 <https://scholarship.org/uc/item/6q2625np> (abgerufen am 25.05.2020). Für den deutschsprachigen Raum siehe v. a.: Petra Maier: Digital Humanities und Bibliothek als Kooperationspartner. *DARIAH-DE Working Papers* (2016), Nr. 19; Frédéric Döhl: Digital Humanities und Bibliotheken. Über technisch-organisatorische Infrastruktur hinausgedacht. In: *ZfBB* 66 (2019), 1, S. 4–18.

„resource intensive“,³ wobei als Ressourcen sämtliche materiellen und immateriellen Ressourcen berücksichtigt werden müssen. Unter den „low barrier“-Kooperationen dominieren Ad-hoc-Dienstleistungen wie technischer Support, Schulungen und Beratungen. „Resource intensive“ werden Kooperationen dann, wenn beispielsweise aufwendige Systeme wie Repositorien für Forschungsdaten eingebracht werden, vor allem aber durch den Einsatz von Personalressourcen. Neben vorhandenem oder fehlendem Know-how bestimmen Entscheidungen über Verfügbarkeit von Personalressourcen wesentlich die Bereitschaft zur Kooperation. Repräsentative Umfragen unter amerikanischen und europäischen wissenschaftlichen Bibliotheken haben gezeigt, dass die Angebote seit 2011 signifikant gewachsen sind.⁴ Unter den europäischen Bibliotheken boten 2014 die Mehrheit der Befragten keine „Research Data Services“ an oder beschränkten sich weitgehend auf Infrastruktur und servicierende Angebote. Die Bereitschaft für konsultative Services sei zwar hoch und die Kollaborationen mit Universitäten stark etabliert (circa 60% Prozent der Befragten), dennoch sind aktive Kooperationen („direct participation with researchers on a project“) noch selten.⁵ Leider beschränken sich viele Bibliotheken auf den „service-and-support approach“ oder werden von ihren Partnern darauf beschränkt.⁶

1.1 Embedded options

Unter den aktiven Kooperationen mit Personalressourcen sind „embedded options“⁷ vielversprechend, die – wenn auch in unterschiedlichem Ausmaß und unter Referenz auf unterschiedliche Begrifflichkeit – auf den

3 Cynthia Hudson-Vitale: Embedded Options: A Common Framework. In: *Data-brarianship* (Anm. 2), S. 25–34, hier S. 33.

4 Carol Tenopir: Research Data Services in Academic Libraries. In: *Journal of eScience Librarianship* 4 (2015), 2, S. 1–21; Dies. u. a.: Research Data Services in European Academic Research Libraries. In: *LIBER* 27 (2017), S. 23–44.

5 Siehe auch Poole: „Natural allies“ (Anm. 1), S. 810.

6 Posner: No Half Measures (Anm. 2), S. 46; Poole: „Natural allies“ (Anm. 1), S. 806f. und 811.

7 Hudson-Vitale: Embedded Options (Anm. 3).

Konzepten des „embedded librarian“ bzw. „blended librarian“ beruhen.⁸ Zentral ist dabei die Forderung, dass embedded librarians aus ihrem originären Kontext in das domänenspezifische Feld der Forschungspartner treten.⁹ „Embedded librarianship is another attempt to do just that: engage“ – wie Stephanie J. Schulte es prägnant formuliert.¹⁰ Als embedded librarian zu fungieren heißt nicht nur, physisch in den Institutionen und bei den Teams der Partner präsent zu sein, sondern auch Präsenz zu zeigen. Eine aktive Rolle einzunehmen, heißt für Bibliothekar*innen nicht ausschließlich in der servicingierenden Rolle präsent zu sein, sondern sich als gleichwertiger Partner*innen in das Forschungsprojekt, in dessen Konzeption, Antragstellung genauso wie in die Generierung und Umsetzung von Forschungsfragen einzubringen.¹¹ Dafür braucht es risikobereite, experimentierfreudige, offene, flexible und selbstbewusste Bibliotheken.¹² Widerstand kommt oftmals aus den etablierten domänenspezifischen Selbstverständnissen der Forschenden, Bibliothekar*innen als gleichwertige Partner*innen im Forschungsprozess wahrzunehmen und damit aus

8 David Shumaker, Mary Talley: Models of Embedded Librarianship: A Research Summary. In: *Information Outlook* 14 (2010), 1, S. 27–35; Jake Carlson, Ruth Kneale: Embedded librarianship in the research context: navigating new waters. In: *College & Research Libraries News* 72 (2011), 3, S. 167–170; Clare McCluskey: Being an embedded research librarian: supporting research by being a researcher. In: *Journal of Information Literacy* 7 (2013), Nr. 2, S. 4–14; Sheila Corral: Educating the academic librarian as a blended professional: a review and case study. In: *Library Management* 31 (2010), 8/9, S. 567–593.

9 Fabian Cremer, Claudia Engelhardt und Heike Neuroth: Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven, Potentiale. In: *Bibliothek Forschung und Praxis* 39 (2015), 1, S. 13–31.

10 Stephanie J. Schulte: Embedded Academic Librarianship. A Review of the Literature. In: *Evidence Based Library and Information Practice* 7 (2012), Nr. 4, S. 122–138, hier S. 134.

11 So auch die Take-away-message bei Posner: No Half Measures (Anm. 2); Poole: „Natural allies“ (Anm. 1); Keener: Arrival fallacy (Anm. 2).

12 Poole: „Natural allies“ (Anm. 1), S. 808. Zur Bibliothek als creative place vgl. u. a. Vandegrift, Varner: *Evolving in Common* (Anm. 2), S. 72f.; Jennifer Vinopal, Monica McCormick: Supporting digital scholarship in research libraries: scalability and sustainability. In: *Journal of Library Administration* 53 (2013), 1, S. 67–78, hier S. 40.

einem ungünstigen Gefälle in der Rollenverteilung.¹³ Zudem wirkt das ab dem 19. Jahrhundert im Zuge der Professionalisierung des Berufsstandes etablierte Selbstverständnis nach, demgemäß Bibliothekar*innen Informationsdienstleister*innen seien, die außerhalb ihres Feldes keine Forschung zu betreiben hätten.¹⁴ Forscher*innen produzieren in dieser traditionellen Rollenverteilung Wissensbestände, Bibliothekar*innen verwalten diese. Ein Problem für „effective cross-functional collaboration“¹⁵ ist somit, „dass viele Forschende Bibliotheken derzeit gar nicht als Partner innerhalb des Forschungsprozesses wahrnehmen“.¹⁶ Demgegenüber sind die Potenziale von „embedded options“ durch „supporting research by being a researcher“ zu stärken.¹⁷ Sich auf Kooperationen mit DH-Forschung einzulassen, heißt für Bibliotheken Selbstverständnisse aufzubrechen, sich einer Forschungskultur auszusetzen, die „unpredictable, highly customized, and prone to failure“ operiert, und zu akzeptieren, dass „business as usual“ nicht reicht.¹⁸

1.2 Travelogues: ein kooperatives Projekt

Als Beispiel sei hier das internationale interdisziplinäre Projekt „Travelogues: Perceptions of the Other 1500–1876“ (FWF I 3792, DFG 398697847) vorgestellt, in welchem ein großer Korpus von über 3.000 Reiseberichten aus dem digitalisierten Bestand der Österreichischen Nationalbibliothek (ÖNB) mit avancierten Methoden und Tools der Digital

13 Vandegrift, Varner: *Evolving in Common* (Anm. 2), S. 76 spricht gar von einem „academic inferiority complex“. Döhl: *Digital Humanities und Bibliotheken* (Anm. 2), S. 10.

14 Uwe Jochum: *Die Idole der Bibliothekare*. Würzburg: Königshausen & Neumann 1995, S. 21–29.

15 Katie Gibson et al.: *Traversing the Gap. Subject Specialists Connecting Humanities Research and Digital Scholarship Centers*. In: *Datalibrarianship* (Anm. 2), S. 3–17, hier S. 16.

16 Inka Tappenbeck: *Welche Dienstleistungen braucht die „digitale Wissenschaft“?*. In: *ZfBB* 64 (2017), H. 5, S. 223–233, hier S. 230. Die Annahme deckt sich mit Ergebnissen der Studie von Poole: „Natural allies“ (Anm. 1), S. 816f.

17 So die Forderung von McCluskey: *Embedded research librarian* (Anm. 8).

18 Posner: *No Half Measures* (Anm. 2), S. 50f.

Humanities analysiert wird.¹⁹ Neben dem Institut für Neuzeitforschung der ÖAW (Project lead), dem Austrian Institute of Technology und dem Institut L3S aus Hannover ist die ÖNB gleichwertiger Partner, der die Antragsstellung und die Konzeption der Forschungsfrage aktiv mitgetragen hat und dessen „embedded librarian“ (Verfasser dieses Artikels) gleichwertig in sämtliche Prozesse des Forschungsprojektes eingebunden ist. Disziplinenübergreifend wird an Lösungen – wie OCR-Postcorrection, einem Genre-Classifer oder späterhin Analyse-tools – gearbeitet, welche auch der Bibliothek als Input zugute kommen. Ein derart „enge[r] Austausch in der Know-how-Bildung“ ermöglicht einen maximalen Forschungs-Output für sämtliche Beteiligten.²⁰

1.3 Chance und Barrieren

Bibliothekar*innen – zumal als embedded librarians – können und sollen in gelingenden kooperativen Projekten eine zentrale vermittelnde Rolle einnehmen. In datenintensiven Projekten bringen sie Know-how zu Ressourcen, Datenmanagement, Datenformaten und Datenrichtlinien ein, behalten den Überblick über Datensets (Kollektionen, Korpora), gewährleisten standardisierte Formate und Datenqualität. Sie bringen sich aktiv in die Datenanalyse und in die Interpretation der Ergebnisse ein. Sie erbringen anerkannte Forschungsleistung und publizieren diese. Zudem eignen sie sich als Übersetzer und Vermittler zwischen Domänen der Humanities und der Informationswissenschaften.²¹

Es gibt zahlreiche Momente, die neben einem Mangel an zur Verfügung stehenden Ressourcen Barrieren für gelingende Kooperationen zwischen Bibliotheken und DH-Projekten sein können und die es als Risiken aktiv zu adressieren gilt. Domänenspezifisches Wissen und diver-

19 Travelogues-project.info. Doris Gruber, Elisabeth Lobenwein und Arno Strohmeier: Travelogues – Perceptions of the Other 1500–1876. A Computerized Analysis. In: Achtzehntes Jahrhundert digital: zentraleuropäische Perspektiven. Hg. von Thomas Wallnig, Marion Romberg und Joëlle Weis. Wien: Böhlau 2019, S. 129–131.

20 Döhl: Digital Humanities und Bibliotheken (Anm. 2), S. 13.

21 Siehe auch Kathleen Kasten-Mutkus, Laura Costello und Darren Chase: Raising Visibility in the DH Landscape: Academic Engagement and the Question of the Library's Role. In: Digital Humanities Quarterly 13 (2019), Nr. 2.

gente Communities of practice wirken oftmals erschwerend in interdisziplinärer Projektarbeit oder provozieren Missverständnisse.²² Für Bibliotheken hinderlich sind – wie bereits angedeutet – oftmals ein überkommenes Selbstverständnis und beschränkende tradierte Muster.²³ Divergente Infrastrukturen, divergente Standards und Datenformate sowie dafür fehlendes Know-how können genauso problematisch sein wie unklare Lizenzen. Die unklare bzw. nicht vorhandene Kennzeichnung von Urheberschaft (Autorschaft) bzw. Arbeitsleistung ist ein weiteres gewichtiges Thema,²⁴ wobei – wie ich weiter unten an Katalogdaten als Forschungsdaten zeige – Bibliotheksdaten in der Regel als gemeinfrei gelten ohne expliziter Kennzeichnung von Autorschaft. Insofern haben Forschende oftmals wenig Interesse, ihre Forschungsleistung dort zu investieren, wo keine Kennzeichnung gegeben wird. Sollen auf Forschung basierende Bibliotheksdaten „FAIR“²⁵ sein, sollte auf dieses Dilemma Rücksicht genommen werden.

1.4 Das Einbahn-Szenario

Die Beschränkung auf eine servicerende Rolle als Lieferant von Daten und Dienstleistungen hat leider allzu oft zur Konsequenz, dass Bibliotheksdaten extrahiert und verarbeitet werden, nicht aber wieder in die Datensysteme der Bibliothek zurückfließen. Für das Beispiel Katalogdaten wäre ein typisches Szenario, dass „Rohdaten“ aus dem Katalogsystem extrahiert und unter Einsatz von Forschungsleistung in Projekten angereichert werden, folglich in eine Stand-off-Datenbank oder eine Printpublikation münden, das in den Katalogdaten präsentierte Informa-

22 Alex H. Poole, Deborah A. Garwood: Interdisciplinary scholarly collaboration in data-intensive, public-funded, international digital humanities project work. In: *Library and Information Science Research* 40 (2018), S. 184–193, v. a. S. 187–189.

23 Posner: No half Measures (Anm. 2), S. 44–48; Vanedgrift, Varner: Evolving in Common (Anm. 2), S. 74.

24 Tarez Samra Graban: Introduction. Questioning Collaboration, Labor and Visibility in DH-Research. In: *Digital Humanities Quarterly* 23 (2019), Nr. 2.

25 FAIR steht für findable, accessible, interoperable and re-usable. Siehe dazu: Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg u.a.: The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific Data* 3 (2016), Nr. 1 (<https://doi.org/10.1038/sdata.2016.18>).

tionsniveau jedoch fehlerhaft oder unzureichend bleibt. Für solche Szenarien ohne Reintegration von Wissen könnte das Bild der Einbahn als Veranschaulichung dienen.

Ein weitaus bereichernderes Szenario für Kooperationen zwischen Bibliotheken und DH-Forschung scheint sich durch einen Austausch und einen multidirektionalen Datenfluss zu ergeben. Darin werden Daten extrahiert, verarbeitet, reintegriert, erneut verarbeitet oder es wird wie im Falle des kooperativen Projektes *Travelogue* möglichst viel Datenanreicherung im Bibliotheksverwaltungssystem direkt vorgenommen. Die Strategien und Workflows mögen unterschiedlich ausgeprägt und je Projekt anders gewichtet sein, gemeinsam sollte ihnen aber ein reintegrierender Austausch sein.

2 Katalogdaten als Forschungsdaten

2.1 Bibliothekskataloge im digitalen Zeitalter? – „A dynamic age for catalogues“

Über Grenzen und Möglichkeiten von Katalogdaten in DH-Projekten nachzudenken, heißt zuerst sich Leitfragen zuzuwenden, wie sie 2014 in Rom unter dem Motto „Faster, Smarter and Richer. Reshaping the Library Catalogue“ diskutiert wurden:²⁶ Was sind Bibliothekskataloge und welche Funktion(en) können und sollen sie Anfang des 21. Jahrhunderts erfüllen? Die lange Tradition des Konzeptes „Bibliothekskatalog“ hat zur Etablierung persistenter Muster von Formatierungen, Funktionen und Erwartungen geführt, die eine gewisse Schwerfälligkeit suggerieren und eine Last gegen Innovation zu sein scheinen. Rachel Ivy Clarke zeigt am Beispiel der persistenten Formatierung des „records“ vom handschriftlichen Zettel bis zum Datensatz in der Bibliotheksverwaltungssoftware, wie schwer es fällt, sich von solchen selbstverständlichen Formatierungen zu lösen.²⁷

26 Eine Auswahl an Beiträgen erschien im Jahrgang 52.3 (2015) der Zeitschrift *Cataloging & Classification Quarterly* unter dem Titel: Reshaping the Library Catalogue: Selected Papers from the International Conference FSR2014 (Rome, February 27th–28th, 2014).

27 Rachel Ivy Clarke: Breaking Records: The History of Bibliographical Records and Their Influence in Conceptualizing Bibliographical Data. In: *Cataloging & Classification Quarterly* 52 (2015), 3, S. 286–302.

Lange Zeit etablierte Selbstverständnisse lösen sich auf, was als Chance oder Bedrohung gewertet werden kann. Aus dem Datensatz (record), der nach klaren Regeln erstellt bzw. abgearbeitet wird, wird ein Netz von Entitäten (Linked entities).²⁸ Aus der seriellen Folge von Informationseinheiten der Ergebnisliste wird ein Ergebnisnetz (Graph), und aus dem Datensilo wird ein offenes, erweiterbares Datennetzwerk. Als Konsens wurde nicht nur in Rom unterstrichen, dass angesichts der Entwicklungen von Semantic Web, Linked Data und der Teilnahme gewichtiger privatwirtschaftlicher Stakeholder wie Google sich Informationskompetenzen von und Erwartungshaltungen gegenüber Bibliotheken radikal verändern (werden) und ein „dynamic age for catalogues“²⁹ bzw. ein fundamentaler „turning point“³⁰ erreicht sei.

Traditionalisten beharren auf der Sonderstellung von Bibliothekskatalogen, wollen deren Funktionen dezidiert abgegrenzt sehen und unterstreichen dabei den Bonus der Reliabilität und Persistenz gegenüber der Gefahr ephemerer, unsauberer und unstrukturierter Daten im Semantic Web.³¹ Andere arbeiten euphorisch daran, die Datensilos der Bibliotheken zu öffnen und zu dynamisieren, diagnostizieren eine fatale Versäumnis und sehen solche Ambitionen oftmals gehemmt durch traditionelle Denkmuster und Zuschreibungen. Das Feld scheint weit offen zu stehen und die Angst dominierend, sich in divergierende Richtungen zu verlaufen. Dabei müssen sich die hier bewusst überspitzt gezeichneten Positionen nicht ausschließen. In dieser „neuen Offenheit“ gilt es vielmehr durch Best-Practice-Beispiele gangbare Wege überzeugend darzulegen – experimentieren und scheitern inklusive. Insofern sehe ich im „dynamic age for catalogues“ auch eine Zeit für Experimente mit der Flexibilität des

28 Ebd., S. 299.

29 So Klaus Kempf in seinem Vortrag „A silent revolution, or the transformation of the catalogue from a local data silo to a multiple net based access and meta data system“. Abstract unter <https://www.aib.it/attivita/congressi/fsr-2014/fsr2014-abstract/2014/40480-fsr-session2a-20140227/> (abgerufen am 15.02.2020).

30 Carlo Bianchini, Mauro Guerrini: A Turning Point for Catalogs: Ranganathan's Possible Point of View. In: *Cataloging & Classification Quarterly* 52 (2015), 3, S. 341–351.

31 Martha M. Yee: Cataloging Compared to Descriptive Bibliography, Abstracting and Indexing Services, and Metadata. In: *Cataloging & Classification Quarterly* 44 (2007), 3–4, S. 307–327.

Konzeptes Bibliothekskatalog, dessen Innovationspotenzial meines Erachtens unterschätzt wird. Dafür gilt es neben theoretischen Überlegungen pragmatische Lösungen zu finden. Im Projekt *Travelogue* wird ein möglichst umfassender Einsatz des Bibliothekskataloges als persistente, leistungsstarke Infrastruktur für Metadaten erprobt.

2.2 Vom Index zum Data hub – „smart data“ und Datenqualität

Über Jahrhunderte schien das Verständnis der Funktion eines Bibliothekskataloges als Instrument des Nachweises und der eindeutigen Beschreibung von Objekten in Aggregation zum Zweck der Identifikation und Adressierung von Objekten in Aggregationen (Sammlungen) selbstverständlich zu sein.³² Etabliert hat sich eine funktionale Differenzierung zwischen Bibliothekskatalogen, Fachbibliografien, Indizes und webbasierten Findeinstrumenten. Der Katalog als „guide to a particular collection“ hat Verweisfunktion.³³ Wird er benützt, wird in der Regel nur durch ihn hindurchgesehen auf etwas anderes, nämlich Bestände, die dann benützt, gelesen, durchforscht werden. Diese indexikalische Funktion wird bei Zugriffen derjenigen „researchers looking for heavy gold through data mining“ sekundär.³⁴ In solchen (zumindest teilweise) auf quantitativen Methoden basierenden DH-Projekten werden Bibliothekskataloge als Aggregate extrahierbarer reliabler Informationen zunehmend interessant und damit ein „material for research in its own right“.³⁵ Die Nuggets, nach denen gesucht wird, sind die Informationseinheiten in den Bibliothekskatalogen selbst, die wie ein „trading good“ gehandelt werden, deren Wert sich nach Glaubwürdigkeit und Zugänglichkeit bemesse. Je höher die markierte Glaubwürdigkeit, je extensiver und intensiver die Informationsgabe und je stärker die Verlinkungen zu anderen Informationselementen, desto höher und stabiler ist der Wert. Die Menge der Daten, Validität bzw.

32 Siehe etwa die Darstellung der Geschichte und Funktionen von Bibliothekskatalogen bei Lucia Sardo: *La Catalogizzazione: Storia, Tendenze, Problemi aperti*. Milano: Editrice Bibliografica 2017.

33 Yee: *Cataloging Compared* (Anm. 31), hier S. 314.

34 Raphaëlle Lapôtre: *Library Metadata on the web: the example of data.bnf.fr*. In: *JLIS.it* 8 (2017), 3.

35 Z. B. Sandra Tuppen, Stephen Rose, Loukia Drosopoulou: *Library Catalogue Records as a Research Resource*. In: *Fontes Artis Musicae* 63 (2016), 2, S. 67–88.

Qualität, Zugänglichkeit und Referenzierbarkeit sind somit die entscheidenden Parameter, nach denen Katalogdaten für Forschungen selektiert werden. Dabei ist – Lapôtre konsequent weiterdenkend – davon auszugehen, dass bereits ein Markt mit unterschiedlich extensiven und unterschiedlich qualitativen Angeboten sich formiert hat, der auch Raum für Nischen lässt. So wird sich auch die Qualität innerhalb lokaler, vernetzter Kataloge (Verbundkataloge) und Aggregationen (Europeana u. a.) stärker diversifizieren. Das aus dem Modell des konsistenten Zettelkataloges tradierte Ideal eines homogenen Katalogs verschwindet sukzessive durch qualitativ unterschiedlich kuratierte Subkollektionen. Einzelne Bibliotheken und Sammlungen werden gemäß dieser Vorstellung Schwerpunkte setzen und für diese mehr Ressourcen und Know-how in die Qualitätssteigerung von Katalogdaten investieren.

Für qualitative Daten als Produkt kuratorischer Pflege wurde in der Forschung der Begriff „smart data“ vorgeschlagen, wobei Bibliotheken zu deren stärksten Produzenten und Distribuenten gezählt werden.³⁶ Reliabilität und Qualität durch kuratorische Prozesse sind die starke Währung der Bibliotheken, auch wenn sie aktiv gegen deren Deflation arbeiten müssen. Qualitätssicherung für Metadaten (bzw. Katalogdaten) wird eine permanente Herausforderung bleiben, zumal sich die Bemessungsgrundlage von „Qualität“ zunehmend relativiert. Die Geschichte der Katalogisierung begleitet die Frage nach der Qualität von Katalogen und Katalogeinträgen. Wann ist ein Bibliothekskatalog ein guter Katalog, wann ist ein Katalogeintrag qualitativ ausreichend? Zumal Katalogisieren ein stark regelgeleiteter Prozess ist, schien die Frage nach der Qualität nur eine Frage nach Konsens innerhalb einer „community of practice“ zu sein.³⁷ Die Ergebnisse fallen dennoch heterogen aus, wie man sich durch die Angabe von Parametern wie „completeness, accuracy, conformance to expectations, logical consistency, timeliness“ nur im Kreis dreht – von einem

36 Christof Schöch: Big? Smart? Clean? Messy? Data in the Humanities. In: *Journal of Digital Humanities* 2 (2013), 3; Michael J. Giarlo: Academic Libraries as Data Quality Hubs. In: *JLSC* 1 (2013), 1. Für Giarlo sind Qualitätskriterien von Bibliotheksdaten „trust“, „authenticity“, „understandibility“, „usability“ und „integrity“.

37 Karen Snow: Defining, Assessing, and Rethinking Quality Cataloging. In: *Cataloging & Classification Quarterly* 55 (2017), 7/8, S. 438–455, v. a. S. 448f.

relativen Parameter zum nächsten.³⁸ Letztlich bleibt die Erkenntnis der Relativität von Qualität angesichts spezifischer Zwecke und spezifischer Nutzergruppen.³⁹ Ein „One-size-fits-all“-Standard ist eine Illusion. Zumal nicht auf den Erwartungshorizont jedes Nutzers, jeder Nutzerin individuell adaptiert werden kann, lohnt es sich vor allem dort, wo Schwerpunkte in der Qualitätssicherung von Katalogdaten gesetzt werden, potenzielle Nutzergruppen und Domänenexperten in ihren Ansprüchen und ihrer Expertise abzuholen. Wird etwa im Projekt *Travelogues* eine kuratierte Kollektion von Reiseberichten aus dem Bestand der ÖNB erstellt, so werden durch die Zusammenarbeit des embedded librarian mit Domänenexperten (Historiker*innen) deren Anforderungen berücksichtigt. Angesichts von Usancen im Verbund, Regelwerk und den spezifischen Domänen-Anforderungen wurde ein möglichst pragmatisches Application Profile für die Datenedition bzw. Datenkorrektur verhandelt. Kriterien für „accuracy“, „completeness“ und „extent“ sind damit je Schwerpunktsetzung, je Kooperation und je Domänennutzer unterschiedlich auszuhandeln und als Kompromisse umzusetzen.

2.3 Erschließungstiefe und bibliografischer Anspruch

Sollte in der indexikalischen Funktion des Kataloges das Informationsvolumen möglichst reduziert gehalten werden und jegliche Informationsgabe, welche nicht dem Zweck der Identifikation und Adressierung von Ressourcen dient, möglichst ferngehalten werden, so verändert sich auch die Frage nach „extent“ und „completeness“ angesichts der neueren Funktion des Kataloges als Datenpool. Pragmatische Kompromisse sollten bei der Erweiterung der Kernelemente (RDA) um optionale Elemente und bei der Extension der Anmerkungen je nach Anwendung und Nutzungszusammenhang erarbeitet werden. Für das Projekt *Travelogues* etwa sind detaillierte Angaben zu bibliografischen Referenzen und zum illustrierenden Inhalt genauso zweckdienlich wie Angaben zu den beschriebe-

38 Thomas R. Bruce, Diane I. Hillmann: The Continuum of Metadata Quality: Defining, Expressing, Exploiting. In: *Metadata in Practice*. Hg. von Diane I. Hillmann und Elaine L. Westbrook. Chicago: American Library Association 2004, S. 238–256, hier S. 243.

39 Snow: *Quality Cataloging* (Anm. 37), S. 451.

nen Reisen, die über Beschlagwortung abgebildet werden. Für mehrere Nutzergruppen sinnvoll wäre eine konsequente Implementierung des FRBR-Modells, die derzeit im Österreichischen Verbundkatalog nur unzureichend umgesetzt ist. Für eine Vielzahl an quantitativ, bibliometrisch ausgerichteten Zugriffen aus dem Feld der Digital Humanities ist etwa die Auswertung von Relationen zwischen Werken und Manifestationen erheblich. Pragmatische Kompromisse sind, wie am Beispiel des FRBR-Modells ersichtlich, vor dem Status quo zu entwickeln, nicht ohne zugleich auf nötige Veränderungen zu drängen. Für manche Projekte mögen auch spezifische Informationen in lokalen, indizierten Feldern eine entsprechende Lösung sein. Für Datensätze des *Travelogues*-Korpus wurde etwa lokal im Katalog der ÖNB ein Marker in Kombination mit Geographika verwendet, der im OPAC zweckdienlich als Suchbegriff fungiert (z. B. *TravelogueD19 Europa Orient* für einen Reisebericht des 19. Jahrhunderts zu den Regionen Europa und Orient – zumal auf Letzterer der Fokus des Projektes liegt). Die funktionale Differenzierung von Bibliothekskatalog und Bibliografie wird dabei willentlich verschliffen.⁴⁰ In deren traditionellen Distinktion könnten sich Bibliothekar*innen zwar auf Ergebnisse der (analytischen) Bibliografie stützen; dennoch besteht in dieser Distinktion eine Arbeitsteilung zwischen Forschungsleistung und Katalogisierung,⁴¹ die zum oben ausgeführten servicerenden Selbstverständnis von Bibliothekar*innen passt. Im Rahmen von projektbasierten Schwerpunkten besteht hingegen die Chance, Katalogdaten um Forschungsergebnisse zu bereichern, Wissensstände zu aktualisieren und auf Objekte oder Kollektionen bezogene bibliografische Qualität zu offerieren, um so die Attraktivität von Katalogdaten als Forschungsdaten zu erhöhen.

2.4 Referentialität

Meines Erachtens dafür noch nicht ausreichend geklärt ist die Referentialität von Bibliotheksdaten. Das betrifft zum einen die Frage nach dis-

40 Neben dem *Travelogues*-Korpus nenne ich als Beispiel die Alfred Döblin Personalbibliografie im Kallias-Katalog des Deutschen Literaturarchivs Marbach. <https://www.dla-marbach.de/bibliothek/bibliografien/alfred-doeblin-personalbibliografie/> (abgerufen am 15.02.2020).

41 Yee: *Cataloging Compared* (Anm. 31), S. 315.

tinkten Identifiern wie sie für FAIR-Research-Data unabdingbar sind, zum anderen die Frage nach der Sichtbarkeit und damit Referenzierbarkeit von intellektueller (kuratorischer) Forschungsleistung. Für die oben geforderte zirkuläre Re-Integration von Forschungsergebnissen in die Bibliotheksdaten (Katalogdaten) ist der Mangel an Markierung erbrachter intellektueller Forschungsleistung oftmals ein vehementes Hindernis. Während Forscher*innen auf die Referenzierbarkeit ihrer Forschungsleistung angewiesen sind und dies zunehmend auch die zugrundeliegenden Forschungsdaten betrifft, gelten Katalogdaten weitgehend als gemeinfrei (CC-Common) und deren Ersteller als Kompilatoren ohne Recht auf Autorschaft. Es ist nicht ungewöhnlich, dass die Präsentation von Forschungsergebnissen zu den jeweiligen Ressourcen oder Elementen in Bibliothekskatalogen aufgrund divergierender Lizenzen ausbleibt. Zuträglich wäre hier eine Einigung im Bibliothekenverbund auf eine entsprechende Markierung, um somit von Einzelnen oder im Rahmen eines Projektes kollektiv erbrachte Forschungsleistung auch an den Daten selbst sichtbar und referenzierbar zu machen. AC-Nummern sind zwar innerhalb des Bibliothekenverbundes ausreichende Identifikatoren für Datensätze, im Semantic Web als Identifikatoren jedoch nicht ausreichend.

2.5 Datenbereinigung

Die Bereinigung von historisch gewachsenen Altdaten, Korruptionen aus Migrationsprozessen und von unzähligen Dubletten aus Aggregationsprozessen sind unabdingbare, wenn auch äußerst zeitintensive Maßnahmen zur Qualitätssicherung.⁴² Für die Attraktivität von spezifischen Katalogdaten für DH-Projekte und den Erfolg von Kooperationen ist die Frage nach erfolgter Qualitätssicherung beziehungsweise die Abschätzung von Ressourcen dafür nicht unerheblich. Entsprechend gilt es bei der Konzeption von Kooperationsprojekten die Datenqualität zu evaluieren, die Ressourcen für „Datapreparation“ (bzw. Datacleaning) abzuschätzen und gegebenenfalls ausreichend Personalressourcen dafür einzuplanen. Im Projekt *Travelogues* hat sich deutlich gezeigt, dass der Aufwand für Datacleaning (in diesem Fall die Bereinigung von Dubletten, die Erstellung

⁴² Alberto Petrucciani: Quality of Library Catalogs and Value of (Good) Catalogs. In: *Cataloging & Classification Quarterly* 53 (2015), 3/4, S. 303–313.

von Bandaufnahmen) und Metadataenrichment (Tiefenerschließung der Ressourcen) nicht zu unterschätzen sind. Auch hierfür hat sich das Modell der „embedded option“ im Rahmen von Projekten als zweckdienlich erwiesen.

2.6 Zugänglichkeit

Um die Sicherstellung von Zugänglichkeit (accessibility) als weiteres Qualitätsmerkmal bemühen sich diverse Linked Open Data Initiativen unterschiedlicher Ausrichtung und Extension.⁴³ Prinzipielles Ziel dieser Initiativen ist die Öffnung der „Datensilos“ und die darin eingelagerten Daten aus dem toten Winkel des Semantic Web zu holen.

Ist im Österreichischen Bibliothekenverbund das Edieren von Linked Data in Formaten wie RDF oder BIBFRAME noch Zukunftsszenario, so wird zur Gewährleistung von Zugänglichkeit konsequent am Angebot von Schnittstellen gearbeitet. An der ÖNB werden im Rahmen der Strategiegruppe Linked Open Data über die Library Labs Katalogdaten via SRU (Search and Retrieve via URL) oder OAI-PMH-Sets (Open Archives Initiative Protocol for Metadata Harvesting) ausgegeben.⁴⁴ Das Projekt *Travelogues* kann diese weitreichende und nachhaltige Infrastruktur nutzen und so etwa Daten in verschiedenen Formaten extrahieren.

Die Funktionen von Bibliothekskatalogen werden sukzessive über deren primären Zweck der Identifizierung und Adressierung von Ressourcen erweitert in Richtung eines offenen „Datenpools“ bzw. „Datennetzwerks“, um zwischen diesen und darüber hinaus „in andere Kataloge oder nichtbibliothekarische Zusammenhänge zu Navigieren“.⁴⁵ Wird zudem nicht gescheut, pragmatische Kompromisse zwischen Regelkonformität

43 Vgl. z. B. den Überblick bei Irfan Ali, Nosheen Fatima Warraich: Linked data initiatives in libraries and information centres: a systematic review. In: *The Electronic Library* 36 (2018), 5, S. 925–937, und R. Hastings: Linked Data in libraries: Status and future directions. In: *Computers in Libraries* 35 (2015), 9, S. 12–28.

44 <https://labs.onb.ac.at/en/dataset/catalogue/> (abgerufen am 15.02.2020).

45 IFLA Cataloguing Section und dem IFLA Meetings of Experts on an International Cataloguing Code (Hg.), Erklärung zu den internationalen Katalogisierungsprinzipien (ICP). Ausgabe 2016 mit geringfügigen Überarbeitungen 2017, hier S. 13. https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-de.pdf (abgerufen am 15.02.2020). Ich danke der Redaktion für den Hinweis.

und Nutzeranforderungen zu erarbeiten, erweist sich der Katalog als flexibler und qualitativer Datenpool für eine Fülle an unterschiedlichsten Forschungsprojekten. Wenn es gelingt, möglichst viele Parameter der Qualitätssicherung im Katalogsystem entsprechend den Anforderungen für die betroffenen Datenkollektionen umzusetzen, erweist sich die Infrastruktur von Bibliothekskatalogen als tragfähig für eine nachhaltige, qualitätsvolle, offene und zugängliche Sicherstellung von Katalogdaten als Forschungsdaten.