



Das offene
Bibliotheksjournal



3

2022



Verein Deutscher
Bibliothekarinnen
und Bibliothekare

o-bib – das offene Bibliotheksjournal ist eine Open-Access-Publikation und zugleich die Mitgliederzeitschrift des VDB. Sie erscheint jährlich mit vier Heften und ist über folgende Website zugänglich:

<https://www.o-bib.de>
ISSN 2363-9814

VDB – Verein Deutscher Bibliothekarinnen und Bibliothekare e.V.
Konstanze Söllner, 1. Vorsitzende
Universitätsbibliothek Erlangen-Nürnberg
91051 Erlangen

Geschäftsführende Herausgebende:
Heidrun Wiesenmüller, Hochschule der Medien Stuttgart
Achim Oßwald, Technische Hochschule Köln

Herausgeberkreis:
Klaus-Rainer Brintzinger, Ulrich Hohoff, Thomas Stäcker,
Helge Steenweg, Achim Oßwald, Heidrun Wiesenmüller

Zuständige für den Verbandsteil:
Nina Frank

Chefredakteurin:
Susanne Maier

Inhaltsverzeichnis

Aufsätze

Volltexte für die Forschung: OCR partizipativ, iterativ und on Demand

Anke Hertling, Leibniz-Institut für Bildungsmedien, Georg-Eckert-Institut
Sebastian Klaes, Leibniz-Institut für Bildungsmedien, Georg-Eckert-Institut

Konferenzbeiträge strategisch publizieren

Automatisierte Workflows zur individuellen Veröffentlichung von Konferenzbeiträgen am Beispiel des Verbands Digital Humanities im deutschsprachigen Raum e.V.

Patrick Helling, Data Center for the Humanities (DCH), Universität zu Köln
Anke Debbeler, Data Center for the Humanities (DCH), Universität zu Köln
Rebekka Borges, Rheinische Friedrich-Wilhelms-Universität Bonn

Infrastrukturen und Services für die wissenschaftliche Nutzung von Webarchiven

Ein Überblick

Tobias Beinert, Bayerische Staatsbibliothek, München
Katharina Schmid, Bayerische Staatsbibliothek, München
Konstanze Weimer, web all in One, München

Open Access zwischen Community-Arbeit und Sammlungsmanagement

Ein Praxisbericht

Tabea Lurk, Hochschule für Gestaltung und Kunst FHNW Basel

Tagungsberichte

LIBER Architecture Group Seminar "Designing for Learning and Scholarship: a challenge for librarians, architects and all" in Luxemburg

Alice Rabeler, Universitäts- und Landesbibliothek Bonn

Berichte und Mitteilungen

Der Zertifikatskurs Forschungsdatenmanagement in NRW

Eine modular aufgebaute Weiterqualifikation für das professionelle Datenmanagement

Benjamin Slowig, Landesinitiative fdm.nrw, Duisburg-Essen
Mirjam Blümm, Technische Hochschule Köln
Konrad U. Förstner, ZB MED - Informationszentrum Lebenswissenschaften und Technische Hochschule Köln
Marvin Lanczek, ZBIW, Technische Hochschule Köln
Birte Lindstädt, ZB MED - Informationszentrum Lebenswissenschaften, Köln
Rabea Müller, ZB MED - Informationszentrum Lebenswissenschaften, Köln
Ulrike Nickenig, Landesinitiative fdm.nrw, Duisburg-Essen
Stephanie Rehwald, Landesinitiative fdm.nrw, Duisburg-Essen
Lioba Schreyer, Landesinitiative fdm.nrw, Duisburg-Essen

Multiplikator*innen für eine offene Wissenschaft

Das Konzept der Train-the-Trainer-Workshops von open-access.network

Sabrina Stockhusen, Universität Bielefeld

Linda Martin, Open-Access-Büro Berlin, Freie Universität Berlin

Bericht aus der 82. Sitzung der Arbeitsgemeinschaft der Verbundsysteme am 28. April 2022

Edith Röschlau, Deutsche Nationalbibliothek

Aus der Deutschen Forschungsgemeinschaft

Ulrike Hintze, Deutsche Forschungsgemeinschaft

Gruppe ‚Wissenschaftliche Literaturversorgungs- und Informationssysteme‘ (LIS)

Ansichten – Einsichten – Diskussionen

Zuschrift von OCLC zum Beitrag „Das Lesen der Anderen“ von Renke Siems

Stellungnahme der Herausgebenden

Heidrun Wiesenmüller und Achim Oßwald

geschäftsführende Herausgebende von o-bib

Facetten eines Missverständnisses

Ein Debattenbeitrag zum Terminus „Diamond Open Access“

S. Dellmann, X. van Edig, J. Rücknagel, S. Schmeja, Technische Informationsbibliothek (TIB), Hannover

Rezensionen

Qualität in der Inhaltserschließung / herausgegeben von Michael Franke-Maier, Anna Kasprzik, Andreas Ledl und Hans Schürmann. – Berlin, Boston: De Gruyter Saur, 2021. – VI, 420 Seiten : Illustrationen. – (Bibliotheks- und Informationspraxis ; Band 70). – ISBN 978-3-11-069149-8 : EUR 69.95 (auch als E-Book im Open Access verfügbar unter <https://doi.org/10.1515/9783110691597>)

Martin Völkl, Universitätsbibliothek Augsburg

Bedrohte Bücher : eine Geschichte der Zerstörung und Bewahrung des Wissens / Richard Ovenden ; aus dem Englischen von Ulrike Bischoff. – Erste Auflage. – Berlin: Suhrkamp, 2021. – 416 Seiten : Illustrationen. – Originaltitel: Burning the books. – ISBN 978-3-518-43007-1 : EUR 28.00 (auch als E-Book verfügbar)

Achim Oßwald, Technische Hochschule Köln

Kommissionen

Gemeinsame Kommission Informationskompetenz von VDB und dbv

„Alles IK oder was? Ideen und Beispiele für die Vernetzung von Literacies in Bibliotheken gesucht!“

Best-Practice-Wettbewerb Informationskompetenz 2022

Christina Schmitz, Staatsbibliothek zu Berlin

Erik Senst, Leuphana Universität Lüneburg

Timo Steyer, UB Braunschweig

Volltexte für die Forschung: OCR partizipativ, iterativ und on Demand

Anke Hertling, Leibniz-Institut für Bildungsmedien, Georg-Eckert-Institut

Sebastian Klaes, Leibniz-Institut für Bildungsmedien, Georg-Eckert-Institut

Zusammenfassung

Für die Forschungsarbeit mit digitalisierten Quellen stellt die Leistung der Volltexterkennung, also die Genauigkeit der Optical Character Recognition (OCR), eine wesentliche Grundlage dar. Die Volltexterkennung avanciert damit zu einem Qualitätskriterium von digitalen Sammlungen und Bibliotheken müssen als zentrale Digitalisierungsakteure ihrer Verantwortung im Hinblick auf die Evidenz von auf Volltexten basierenden wissenschaftlichen Ergebnissen gerecht werden. Ausgehend von einer Digitalisierung, die explizit an der Zielgruppe der digitalen Forschung ausgerichtet ist, greift der folgende Beitrag Formate und Workflows zur Organisation der Volltexterkennung als partizipativen und iterativen Prozess in Zusammenarbeit mit der Forschung auf. Vor dem Hintergrund der aktuellen OCR-D-Förderphase wird ein on-Demand-Ansatz, bei dem OCR-Prozesse nach spezifischen Bedarfen durchgeführt werden, vorgestellt.

Abstract

For working with digitized sources in research, the quality of full-text recognition, i.e. the accuracy of Optical Character Recognition (OCR), is essential. Full-text recognition is thus advancing to become a quality criterion of digital collections, and libraries – as central actors in digitization – must live up to their responsibility regarding the evidence of scientific results based on full text. Starting from a digitization process that is explicitly oriented towards digital research, the paper discusses formats and workflows for organizing full-text recognition as an iterative and participatory process in collaboration with researchers. Against the background of the current OCR-D funding phase, the paper also presents an on-demand approach for OCR processes according to specific requirements.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5832>

Schlagwörter: Digitalisierung; Volltext; OCR; Optische Zeichenerkennung

Autorenidentifikation:

Hertling, Anke: GND: [1033153737](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0011-9); ORCID: <https://orcid.org/0000-0002-3163-2233>;

Klaes, Sebastian: ORCID: <https://orcid.org/0000-0003-3597-7017>

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Einleitung

Mit der wachsenden Bedeutung einer Forschung, die wie die Digital Humanities (DH) mit digitalen Methoden und Werkzeugen arbeitet, verändern sich die Anforderungen an von Bibliotheken bereitgestellte digitale Sammlungen und somit auch an die Digitalisierung von Quellen. Die

Auseinandersetzung mit der Volltexterkennung stellt dabei neben der Bereitstellung verschiedener Daten- und Exportformate¹ einen, so formuliert es die vom Verband „Digital Humanities im deutschsprachigen Raum“ (DHd) 2019 gegründete Arbeitsgruppe DHd-AG OCR, „Schlüssel für die Umsetzung der Forderungen der DH“² dar. Insbesondere bei historischen Quellen, bei denen Schriftarten, Layout, Sprache und Orthographie vielfach variieren, ist die Volltexterkennung mit ihren Prozessen der Vorverarbeitung in Form von Bildoptimierung und Binarisierung (Preprocessing), der Layoutsegmentierung (Region Segmentation), der Zeichenerkennung (Character Recognition) und der Nachbearbeitung im Sinne einer Fehlerkorrektur nach wie vor eine große Herausforderung. 2021 startete die Deutsche Forschungsgemeinschaft (DFG) deshalb in die dritte Phase ihrer OCR-D Förderinitiative und unterstützt die OCR-Weiterentwicklung, um die Volltexttransformation der im deutschen Sprachraum erschienenen Drucke des 16. bis 18. Jahrhunderts zu ermöglichen und so zu optimieren, dass die Quellen nicht nur durchsuch-, sondern verstärkt digital analysierbar sind.³ Ziel der aktuellen DFG-Förderphase ist es, den seit 2015 entwickelten OCR-D-Softwareprototyp in Workflow- und Digitalisierungssysteme zu integrieren und die Erzeugung qualitativ hochwertiger Volltexte in den bibliothekarischen Regelbetrieb zu überführen.⁴ Koordiniert wird die Förderphase von der Herzog August Bibliothek in Wolfenbüttel, die mit ihren Partnerinstitutionen einen regelmäßigen Austausch zwischen den Projektbeteiligten, darunter die Universitätsbibliotheken Braunschweig und Mannheim, die SLUB Dresden, die SUB Göttingen und die Gesellschaft für wissenschaftliche Datenverarbeitung in Göttingen (GWDG), die ULB Sachsen-Anhalt, die Johannes Gutenberg-Universität Mainz und die FAU Erlangen-Nürnberg, organisiert. Die Forschungsbibliothek des Leibniz-Instituts für Bildungsmedien | Georg-Eckert-Institut (GEI) arbeitet gemeinsam mit dem Institut für Mensch-Computer-Medien (HCI) und dem Zentrum für Philologie und Digitalität „Kallimachos“ (ZPD) der Universität Würzburg ebenfalls im Projektverbund und bringt dabei besonders ihre enge Zusammenarbeit mit der Digital-Humanities-Forschung ein. Bereits seit Beginn der Digitalisierung erfolgte der Auf- und Ausbau der digitalen Sammlung am GEI forschungsbasiert. Die langjährige gemeinsame Arbeit mit der Fachcommunity wird im Folgenden dargelegt und entsprechende Instrumente sowie Erfahrungen aus der engen Kooperation mit der Forschung aufgezeigt. Besonders vor dem Hintergrund der Anforderungen aus den Digital Humanities wird ein Verständnis von Volltexterkennung als partizipativer und iterativer Prozess vorgestellt, der neue Workflows und Organisationsstrukturen bei allen Akteuren erfordert. Damit optimierte und nachhaltige OCR-Ergebnisse für die Forschung generiert werden können, wird ein on-Demand-Ansatz entwickelt, bei dem die Volltexterkennung entsprechend spezifischer Forschungs- und Materialbedarfe auf Korpusebene durchgeführt wird.

- 1 U.a. Gasser, Sonja: Das Digitalisat als Objekt der Begierde. Anforderungen an digitale Sammlungen für Forschung in der Digitalen Kunstgeschichte, in: Andraschke, Udo; Wagner, Sarah (Hg.): Objekte im Netz. Wissenschaftliche Sammlungen im digitalen Wandel, Bielefeld 2020, S. 261–276, hier S. 267 ff. Online: <<https://doi.org/10.14361/9783839455715>>.
- 2 Vgl. die auf ihrer Webseite formulierten Arbeitsschwerpunkte der DHd-AG OCR: <<https://dig-hum.de/ag-ocr/>>, Stand: 26.04.2022.
- 3 Koordinierte Förderinitiative zur Weiterentwicklung von Verfahren der Optical Character Recognition. Das OCR-D-Projekt. Online: <<https://ocr-d.de/de/about>>, Stand: 26.04.2022; sowie Engl, Elisabeth: OCR-D kompakt. Ergebnisse und Stand der Forschung in der Förderinitiative, in: Bibliothek – Forschung und Praxis 44 (2), 2020, S. 218–230. Online: <<https://doi.org/10.1515/bfp-2020-0024>>. Stand: 26.04.2022>.
- 4 DFG: Implementierung der OCR-D-Software zur Volltextdigitalisierung. Information für die Wissenschaft Nr. 15 | 27. Februar 2020. Online: <https://www.dfg.de/foerderung/info_wissenschaft/2020/info_wissenschaft_20_15/index.html>, Stand: 26.04.2022.

2. Sondierung der Interessen beim Aufbau der digitalen Schulbuchbibliothek GEI-Digital

Auch in ihrem aktuellen Förderprogramm „Digitalisierung und Erschließung“ erklärt die DFG die Wissenschaft zur primären Zielgruppe und die „nachdrückliche Stimulierung und Stärkung wissenschaftlicher Forschung“⁵ als eine Voraussetzung für die Förderung von Digitalisierungsprojekten. Eine wissenschaftliche Beteiligung bei der Digitalisierung kann zum Beispiel in Form eines Digitalisierungsbeirates gewährleistet werden. So haben am GEI Vertreter*innen u.a. aus der Geschichts- und Erziehungswissenschaft gemeinsam mit der Forschungsbibliothek einen Digitalisierungsplan mit der Zielsetzung entwickelt, die gesamte deutsche historische Schulbuchsammlung des Instituts, die die Fächer Geschichte, Geographie, Politik, Realien, Werteerziehung/Religion sowie den (Erst-) Leseunterricht umfasst, zu digitalisieren und zugänglich zu machen. Im Fokus der Zusammenarbeit mit dem Digitalisierungsbeirat zum Aufbau der digitalen Schulbuchbibliothek GEI-Digital standen zunächst fachwissenschaftliche Kriterien zur Erstellung von Korpora, die als Grundlage für einen vom GEI 2007 in der Förderlinie „Digitalisierung der DFG-Sondersammelgebiete“ eingereichten DFG-Antrag fungierten. Gemeinsame Abstimmungsprozesse waren auch deshalb notwendig, weil Schulbücher als Gebrauchsmedien kaum und selbst nach Gründung der Deutschen Bücherei zumeist nicht systematisch gesammelt wurden und bis heute keine Bibliografie vorliegt, die alle deutschsprachigen Schulbücher verzeichnet. Spätestens ab 1871 gab es zahlreiche Auflagen sowie eine Vielzahl an regional ausdifferenzierten Ausgaben, deren vollständige Digitalisierung Finanz- und Zeitrahmen, wie sie in DFG-Projekten möglich wären, erheblich überschritten hätten. So wurde u.a. mit dem Digitalisierungsbeirat festgelegt, dass mehrere Auflagen nur dann digitalisiert werden, wenn zwischen der frühestmöglichen und der spätesten Auflage signifikante Veränderungen zum Beispiel in Form von Umfangssteigerungen festzustellen sind.⁶

Obwohl die Anfänge der Schulbuchproduktion im 17. Jahrhundert liegen, waren die Entwicklungen des Schulbuchs im 19. Jahrhundert für die Entscheidung des Digitalisierungsbeirates maßgeblich, Bestände aus dieser Zeit prioritär zu digitalisieren. Erst im 19. Jahrhundert setzten sich Schulbücher, wie sie heute gebräuchlich sind, durch. Schulbücher zählen dabei zu den ersten modernen Massenmedien. Mit der Einführung der allgemeinen Schulpflicht Anfang des 19. Jahrhunderts avancierten sie zu einem staatlichen Steuerungsinstrument und wurden zum festen Bestandteil des Unterrichts. Große Verlage wie Ferdinand Hirt in Breslau oder Velhagen & Klasing in Bielefeld richteten ihr Kerngeschäft auf die Konzeption und Produktion von Schulbüchern verschiedener Schulfächer aus.⁷ Insbesondere vor dem Hintergrund der Ausdifferenzierung im höheren Schulwesen stieg die Zahl der im Deutschen Kaiserreich publizierten Schulbücher in Bezug auf Auflage und Exemplare stark an und erreichte die

5 DFG: Merkblatt und ergänzender Leitfaden – Digitalisierung und Erschließung, DFG Vordruck 12.15 – 09/21, S. 5. <https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/digitalisierung_erschliessung/formulare_merkblaetter/index.jsp>, Stand: 26.04.2022.

6 Hertling, Anke; Klaes, Sebastian: Historische Schulbücher als digitales Korpus für die Forschung. Auswahl und Aufbau einer digitalen Schulbuchbibliothek, in: Nieländer, Maret; De Luca, Ernesto William (Hg.): Digital Humanities in der internationalen Schulbuchforschung. (Eckert, Expertise 9), Göttingen 2018, S. 22–44. Online: <<https://repository.gei.de/handle/11428/296>> (DOI 10.14220/9783737009539), Stand: 26.04.2022.

7 Jäger, Georg: Der Schulbuchverlag, in: Ders. et al. (Hg.): Geschichte des deutschen Buchhandels im 19. und 20. Jahrhundert. Bd. 1: Das Kaiserreich 1870-1918, Teil 2, Frankfurt am Main 2003, S. 62–102.

Schulbuchproduktion zwischen 1871 und 1918 einen quantitativen Höhepunkt. So wurden rund 40% aller Geschichtsschulbücher, die im Zeitraum zwischen 1700 und 1945 in den zum deutschen Staatsgebiet zählenden Territorien erschienen sind, in dieser Zeit veröffentlicht.⁸

Angesichts der forcierten Schulbuchproduktion im Deutschen Kaiserreich sowie der Übernahme des Bildungsmonopols durch den Staat plädierte der am GEI eingesetzte Digitalisierungsbeirat, dass Schulbücher aus dieser Epoche von besonderer wissenschaftlicher Relevanz sind und diese zuerst digitalisiert und möglichst als Volltexte zugänglich gemacht werden sollten. Problematisiert wurde dabei, dass den älteren und zumeist unikalen Quellen aus Sicht der Bestandserhaltung eine höhere Digitalisierungspriorität hätte eingeräumt werden müssen. So wäre die sehr wertvolle historische Fibel-Sammlung aus Sicht der GEI Forschungsbibliothek bevorzugt zu digitalisieren. Die Community der Fibel-Forschung ist jedoch sehr überschaubar und für die Fibeln mit ihren vielen Abbildungen und unterschiedlichen Schriftarten war zu dieser Zeit eine Volltexterkennung nicht zielführend. Notwendig war demnach ein genaues Abwägen und ein gemeinsamer Verständigungsprozess, bei dem das GEI vor allem davon profitierte, dass die Volltexttransformation der digitalen Quellen gleich zu Beginn seiner Digitalisierungsplanungen mit einkalkuliert worden war. Der Digitalisierungsbeirat konnte gleichfalls für die Digitalisierung als Maßnahme der Bestandserhaltung und somit als Grundlage für zukünftige Forschungen sensibilisiert werden. Entsprechende finanzielle und personelle Mittel wurden im Institut eingeplant, um besonders gefährdete Quellen rechtzeitig zu digitalisieren, auch wenn kein unmittelbarer Forschungsbedarf vorliegt. Inzwischen stehen in der digitalen Schulbuchbibliothek GEI-Digital (<https://gei-digital.gei.de/viewer/index/>) 5.000 historische Schulbücher im Umfang von rund 1,6 Millionen Seiten überwiegend im Volltext frei zugänglich zur Verfügung.⁹ Die Digitalisierung von Geographie-, Realien- und Geschichtsschulbüchern aus der Zeit des 17. Jahrhunderts bis 1918 sowie die Digitalisierung von Lesebüchern aus der Zeit des Deutschen Kaiserreichs sind nunmehr abgeschlossen. Neben einer Metadaten- und Volltextsuche im Gesamtbestand bietet GEI-Digital eine differenzierte Recherche in den einzelnen Korpora und verschiedene Ausgabeformate für die Volltexte (u.a. ALTO) und Metadaten (u.a. Dublin Core, METS/MODS, MARC XML).

3. Texterkennung partizipativ und iterativ

In der Zusammenarbeit mit dem Digitalisierungsbeirat wurde zu einem für das Bibliothekswesen recht frühen Zeitpunkt deutlich, dass Volltexte eine unverzichtbare Grundlage und Nukleus für die wissenschaftliche Analyse und Weiterverarbeitung von digitalen Quellen sind. Aus Sicht der Forschungsbibliothek des GEI war es nach zehn Jahren Digitalisierung eine programmatische Notwendigkeit, die OCR-Qualität ihrer Digitalisate zu evaluieren. Eine schon 2014 am GEI durchgeführte Online-Befragung machte die Bedeutung qualitativ hochwertiger Volltexte deutlich. Gegenstand der Umfrage waren die mit Abbyy FineReader Engine 10 und Engine 11 behandelten und auf GEI-Digital bereitgestellten Volltexte. An der Befragung haben 106 GEI-Digital-Nutzer*innen teilgenommen, wobei 70% der Befragten die Qualität der Volltexte generell als „sehr wichtig“ beurteilten.

8 Jacobmeyer, Wolfgang: Das deutsche Schulgeschichtsbuch 1700-1945. Die erste Epoche seiner Gattungsgeschichte im Spiegel der Vorworte, Bd. 1, Berlin 2011, S. 19.

9 Rund 113.000 Seiten wurden nicht mit OCR behandelt, darunter Bestände mit überwiegend graphischen Darstellungen wie zum Beispiel Atlanten.

29% der Befragten waren mit der auf GEI-Digital zugänglichen OCR-Qualität *sehr zufrieden*, jeweils weitere 25% waren *weitestgehend zufrieden* bzw. *zufrieden*. Die Antworten legen nahe, dass ca. $\frac{3}{4}$ der Nutzer*innen von GEI-Digital mit der Texterkennung zufrieden waren und diese als Mehrwert erkannten. Im Rahmen der Befragung wurden darüber hinaus Perspektiven zur Nachnutzung der Texterkennung erfragt. *Exportmöglichkeiten von Volltexten* wünschten sich 76% der Befragten, wobei 54% der Umfrageteilnehmer*innen *ein direktes Exportieren in Analysewerkzeuge wie beispielsweise Text-Grid als sehr wichtig* anerkannten. Auf die Frage, ob GEI-Digital-Nutzer*innen fehlerhafte Volltexte selbst korrigieren würden, gaben 61% der Befragten eine positive Rückmeldung. Aufschlussreich waren auch Angaben zur Relevanz bestimmter Funktionen und Suchmöglichkeiten. Eine Funktion oder Möglichkeit zur Analyse von Volltexten-Fehlerraten wurde von 42% als *sehr wichtig* und von weiteren 42% als *wichtig* angesehen. Bei den Suchmöglichkeiten wurde der fehlertoleranten Suche nach Volltexten, bei der nicht nur eine exakte, sondern zudem eine ähnliche Zeichenfolge als Suchoption fungiert, ein hoher Stellenwert eingeräumt. Die Bedeutung der Volltexterkennung für die Forschung zeigte auch die 2016 von der Bayerischen Staatsbibliothek (BSB) im Rahmen des OCR-D-Kooperationsprojekts durchgeführte Befragung. Knapp über die Hälfte der 139 Teilnehmer*innen würde fehlerhafte OCR-Texte zu Forschungszwecken verwenden, denn auch fehlerhafte Volltexte seien durchaus hilfreich, u.a. für eine Volltextsuche. 40% der Befragten hielten fehlerhafte OCR-Texte indessen für nutzlose Daten.¹⁰

Bekräftigt wurde der Befund, dass die Volltexte von GEI-Digital als verbesserungswürdig anzusehen sind, durch die am Institut zunehmend durchgeführten Digital-Humanities-Projekte, wie zuletzt das Projekt „DiaCollo für GEI-Digital“¹¹, in dem sprachtechnologische Verfahren auf den GEI-Digital Volltextkorpus eingesetzt wurden. Mit dem im Rahmen der CLARIN-D-Initiative entwickelten Werkzeug DiaCollo¹² wurden typische Wortverbindungen auf Grundlage verschiedener digitaler Schulbuchkorpora ermittelt und die Ergebnisse visuell aufbereitet. Im Zuge der Anpassung von DiaCollo für seine Anwendung in der historischen Schulbuchforschung wurden Unschärfen bei den GEI-Digital-Volltexten nochmals sehr sichtbar. Fehlerhafte OCR beeinträchtigen die wissenschaftliche Analyse in erheblicher Weise und verfälschen Ergebnisse u.a. bei der Kollokationsanalyse. So wird zum Beispiel in dem *Berlinischen neu eingerichteten Schulbuch* aus dem Jahr 1761 das Wort „und“ häufig als „uiw“ erkannt, das im Korpus selten, dann aber wiederum sehr oft mit dem Suchbegriff „Schule“ vorkommt und deshalb als relevantes Kollokat interpretiert wird. Die Dominanz von Fraktur sowie die für Schulbücher typischen komplexen Layoutstrukturen einschließlich Anschauungstafeln, Inhaltsverzeichnissen, Fußnoten oder Abbildungen sind hier materialspezifische Ursachen für einen Qualitätsverlust bei der Volltexterkennung. Ein Schulbuch wird darüber hinaus vielfach von mehreren Generationen genutzt. In Schulbüchern vorkommende Randnotizen und Anstreichungen sind ebenfalls entscheidende OCR-Fehlerquellen.

10 Vgl. die Umfrage zur Verwendung von OCR-Texten aus dem Jahr 2016: <<https://ocr-d.de/de/umfrage>>, Stand: 26.04.2022.

11 Leibniz-Institut für Bildungsmedien | Georg-Eckert Institut: Diacollo für GEI-digital.<<https://diacollo.gei.de/>>, Stand: 26.04.2022.

12 CLARIN-D: DiaCollo. Kollokationsanalyse in diachroner Perspektive.<<https://www.clarin-d.net/de/kollokationsanalyse-in-diachroner-perspektive>>, (Stand: 26.04.2022).

Auch wenn die DFG-Standards bei der Volltexterkennung eingehalten wurden, wuchs die Unzufriedenheit der Forschung mit der OCR-Qualität von GEI-Digital. Selbst bei 98% Erkennungsrate ist bei einer Seite mit rund 2.000 Zeichen mit etwa 40 Fehlern zu rechnen und werden letztlich bis zu 40 Wörter nicht gefunden, so dass Datenaufbereitungsverfahren notwendig sind, um zum Beispiel CLARIN-D-Werkzeuge verlässlich anwenden zu können.¹³ Angesichts der hohen Nachfrage nach qualitativ hochwertigen Volltexten hat die Forschungsbibliothek des GEI 2019 die Qualität von verschiedenen OCR-Softwaresystemen evaluiert. Bei der Qualitätsmessung wurde auf das von der DFG in ihren Praxisregeln empfohlene „Bernoulli-Experiment“ zurückgegriffen.¹⁴ In einer Stichprobe mit zehn Lesebüchern mit Fraktur-Schriftbild aus der Zeit des Deutschen Kaiserreichs mit einem Umfang von 4.000 Seiten wurde geprüft, ob Zeichen richtig oder falsch erkannt wurden, was erste Rückschlüsse auf die Genauigkeit der OCR-Erkennung und damit die Validität der Daten ermöglicht. Verglichen wurde die Erkennungsquote zwischen dem bislang für den GEI-Digital-Korpus genutzten kommerziellen Software-Anbieter Abbyy FineReader SDK und der Open-Source-Software Tesseract als ein Texterkennungssystem, das Verfahren der Künstlichen Intelligenz in Form sog. rekurrenter neuronaler Netze nutzt und auf trainierbaren Datenmodellen basiert. Tesseract erlaubt demnach das Training spezifischer Modelle u.a. auf Frakturschriften.¹⁵ Bei seinen Tests auf Tesseract 4.0-Basis hat das GEI bei seinen zehn Lesebüchern zwei Modelle eingesetzt: (1) ein Standard-Modell von einem Dienstleister (Dienstleister-Tesseract) sowie (2) ein durch einen Dienstleister spezifisch auf GEI-Daten trainiertes Tesseract-Modell (GEI-Tesseract), wobei durch eine Zufallsfunktion 10.000 Zeichen aus dem Korpus der Lesebücher ausgewählt und durch deren Transkription das Modell trainiert wurde.

Auch wenn das Bernoulli-Experiment nur einen groben Einblick ermöglicht und die zehn Lesebücher bei einem Korpus von insgesamt über 1.351 digitalisierten Lesebüchern nur einen kleinen Ausschnitt darstellen, konnte festgehalten werden, dass die Tesseract-Modelle gegenüber dem kommerziellen Software-Anbieter große Potenziale aufwiesen, was die UB Heidelberg bei ihren 2019 durchgeführten Tests nochmals bekräftigen konnte.¹⁶ Bei den zehn Lesebüchern wurde in wenigen Fällen der Wert von mehr als 496 korrekt erkannten Zeichen überschritten, was einer Tesseract-Erkennungsrate von rund 98% und damit den Vorgaben der DFG entspricht. Mit Blick auf digitale Methoden wie Topic Modeling oder Data Mining wären für die Belastbarkeit der Daten erneute Qualitätsprüfungen und ggf. OCR-Korrekturverfahren notwendig.

13 Nieländer, Maret; Weiß, Andreas: „Schönere Daten“ – Nachnutzung und Aufbereitung für die Verwendung in Digital Humanities-Projekten“, in: Nieländer, Maret; De Luca, Ernesto William (Hg.): Digital Humanities, 2018, S. 91–116. <<https://repository.gei.de/handle/11428/296>> (DOI 10.14220/9783737009539), Stand: 26.04.2022.

14 DFG: Praxisregeln „Digitalisierung“. DFG Vordruck 12.151 – 12/16, S. 35, <https://www.dfg.de/formulare/12_151/12_151_de.pdf> Stand: 26.04.2022.

15 Weil, Stefan: tesseract-ocr / tesstrain, <<https://github.com/tesseract-ocr/tesstrain/wiki>>, Stand: 26.04.2022.

16 Weil, Stefan: Neue Frakturmodelle für Tesseract. Präsentation auf dem Kitodo Anwendertreffen 18.-19. November 2019, S. 3, <<https://madoc.bib.uni-mannheim.de/53748/1/2019-11-18.pdf>>, Stand: 26.04.2022.

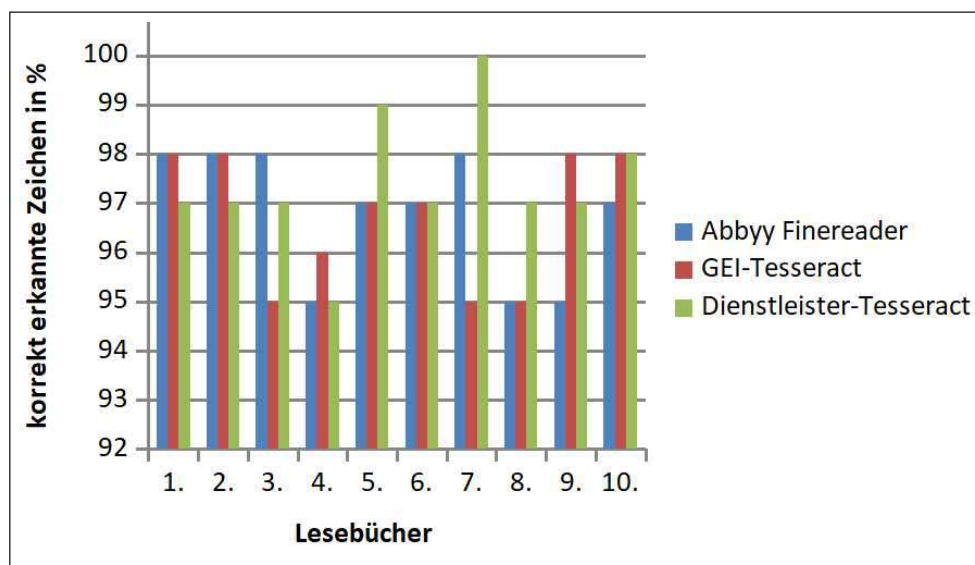


Abbildung: Vergleich OCR-Ergebnisse bei zehn Lesebüchern aus dem Deutschen Kaiserreich durch Bernoulli-Experiment

Neben Tesseract weisen auch andere Open-Source-Engines vielversprechende Ergebnisse auf, die sogar Resultate mit einer Erkennungsrate von über 99% erzielen.¹⁷ Wegweisend ist dabei die Open-Source-Engine Calamari, die die Erkennungsgeschwindigkeit erheblich steigert, die Generierung von Trainingsmaterial erleichtert und auch bei den Erkennungsgenauigkeiten sehr gute Leistungen liefert.¹⁸

Die Erstellung von Volltexten und ihre Qualität sind in einem zunehmenden Maße abhängig vom Stand der jeweiligen Technologien und der Qualität der Trainingsdaten. Vor dem Hintergrund der fortschreitenden OCR-Weiterentwicklungen, wie sie auch im OCR-D Kooperationsprojekt forciert werden, ist Volltexterkennung demnach als iterativ zu verstehen, und das heißt, dass OCR-Prozesse je nach aktuellem Stand der Technologien mehrfach wiederholt durchgeführt werden sollten, zumal sich die Bedarfe aus der Wissenschaft ebenfalls fortwährend verändern. Mit ihren inzwischen vorliegenden großen Datenmengen kommen Bibliotheken allerdings bei umfangreichen OCR-Prüfungen und -Verbesserungen an die Grenzen ihrer Ressourcen. Maßnahmen zur Prüfung und Optimierung auf Korpusebene erscheinen eher realisierbar. So werden am GEI bei der Planung von Digital-Humanities-Projekten, für die Volltexte zur Verfügung gestellt werden sollen, die digitalen Methoden und Werkzeuge für die Forschungsarbeit besprochen und die Voraussetzungen, die die Volltexte erfüllen müssen, eruiert. Liegen digitale Volltexte bereits vor, testet die Bibliothek an Volltext-Stichproben

17 Reul, Christian et al.: State of the Art Optical Character Recognition of 19th Century Fraktur Scripts using Open Source Engines, in: DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts (1.0, p. 366), S. 212-216. Online: <<https://doi.org/10.5281/zenodo.2596095>>, Stand: 26.04.2022.

18 Wick, Christoph; Reul, Christian; Puppe, Frank: Calamari - A High-Performance Tensorflow-based Deep Learning Package for Optical Character Recognition, in: Digital Humanities Quarterly 14 (2), 2020, <<https://doi.org/10.48550/arXiv.1807.02004>>, Stand: 26.04.2022.

gemeinsam mit den Wissenschaftler*innen sowie mit Kolleg*innen, die an der Weiterentwicklung von digitalen Forschungstools arbeiten, die geplanten digitalen Methoden und Werkzeuge. Falls noch keine Digitalisierung und Volltexterkennung erfolgt ist, werden Probedigitalisierungen und entsprechende OCR-Tests durchgeführt. Auf diese Weise können für ein digitales Forschungsprojekt notwendige Zeit- und finanzielle Ressourcen optimiert ermittelt werden. Gegebenenfalls sind Ressourcen zu berücksichtigen, um Volltext-Daten durch Training zu verbessern oder zu korrigieren. Nach Abschluss der Forschungsarbeit stellt die Bibliothek Informationen zur eingesetzten OCR-Technologie sowie OCR-Erkennungsraten für die Forschungsdokumentation zur Verfügung.

Die Sensibilisierung für OCR-Prozesse und insbesondere die Ermittlung und Transparenz von OCR-Fehlerraten waren am GEI Ausgangspunkt für einen über Korpusprioritäten hinausgehenden Verständigungsprozess mit der Forschung. Texterkennung wurde dabei grundlegend für die Evidenz der digitalen wissenschaftlichen Arbeit sowie als partizipative und iterative Daueraufgabe definiert. Diesem Verständnis folgend, erfordert es eine fortwährende enge Zusammenarbeit zwischen Bibliothek und Forschung, wodurch sich die Rolle von Bibliotheken als Dienstleister, die durch Digitalisierung ihre Quellen für die Forschung zugänglich machen, hin zur Rolle, wie sie die Konzepte des „Embedded“ oder des „Liaison Librarian“ vorsehen, erweitert. Im Hinblick auf die Aufgabenbereiche für „Liaison Librarians“¹⁹ liegen die Potentiale, um sich im Bereich der Digitalisierung neu zu positionieren, insbesondere in der Vermittlung von Kenntnissen von OCR-Prozessen und damit verbundenen Resultaten („Teaching and Learning“), im Austausch über Anforderungen aber auch Grenzen der Volltexterkennung („Scholarly Communication“), im Erkennen und Bereitstellen von neuen OCR-Technologien („Digitale Tools“) sowie in der Unterstützung der Forschung bei der Beantragung von Fördermitteln („Fund Raising“). Um die Volltexterkennung entlang digitaler Forschungsfragen und -methodiken zu planen, zu konfigurieren, zu realisieren und zu dokumentieren, sollten Bibliotheken Digital-Humanities-Kompetenzen aufweisen und mit der Forschung auf Augenhöhe kommunizieren und agieren. Die fachwissenschaftliche Forschung wiederum sollte die Volltexterkennung verstärkt als Teil des digitalen Forschungsprozesses anerkennen und dessen kritische Reflexion als signifikanten Bestandteil einer digitalen Quellenkritik verstehen.

4. Texterkennung on Demand

Am GEI bringt die Bibliothek ihre OCR-Expertise nicht nur projektspezifisch ein, sie ist darüber hinaus strukturell in das Forschungsteam „Digital Humanities“ am Institut eingebunden und kann somit Bedarfe unmittelbar mit der Forschung diskutieren und Impulse setzen. Basierend auf den Digital-Humanities-Forschungen, die besonders auf hochwertige Volltexte angewiesen sind, wurde gemeinsam die Perspektive einer korpuspezifischen Volltexterkennung im Sinne einer *Volltexterkennung on Demand* entwickelt. Ideal wäre eine gezielte Auswahl und Nutzung von OCR-Anwendungen, die sich an Materialbesonderheiten von Korpora und korpusbasierten Forschungsfragen orientieren. Die

19 Fühles-Ubach, Simone: Vom „embedded“ zum „liaison librarian“ – Was versprechen die neuen Konzepte?, in: Mittermaier, Bernhard (Hg.): Vernetztes Wissen – Daten, Menschen, Systeme. 6. Konferenz der Zentralbibliothek Forschungszentrum Jülich. 5.–7. November 2012, Proceedingsband (Schriften des Forschungszentrums Jülich Reihe Bibliothek / Library, Band / Volume 21), Jülich 2012, S. 337–350, hier S. 346. Online: <<https://juser.fz-juelich.de/record/126960/files/FZJ-2012-00028.pdf>>, Stand: 26.04.2022.

Organisation und Durchführung der Volltexterkennung würde dabei nicht mehr ausschließlich in der Hand der Bibliothek liegen, Wissenschaftler*innen initiieren vielmehr OCR-Prozesse selbst. Dieser on-Demand-Ansatz hätte den Vorteil, dass die Volltexterkennung gezielt auf Korpus- und Forschungsbedarfe ausgerichtet wäre. Open-Source-Tools speziell für historische Drucke, wie sie im Rahmen des Kooperationsprojekts OCR-D entwickelt werden, stellen für diesen Ansatz eine gute Grundlage dar. Die meisten OCR-Tools sind allerdings nur mit Programmierkenntnissen zu beherrschen, was beispielsweise Forscher*innen am GEI von ihrer Nutzung abhält.²⁰ An einen dezidiert nicht technisch versierten Nutzerkreis richtet sich die vom Zentrum für Philologie und Digitalität „Kallimachos“ (ZPD) an der Universität Würzburg entwickelte Software OCR4all (<http://www.ocr4all.org/>). OCR4all, in dem Calamari vollständig integriert ist, ermöglicht das Preprocessing, also u.a. die Erkennung der Schriftbereiche in binären und graustufigen Bildern und das Umrechnen schief gescannter Textbereiche in gerade Textblöcke, die Region und Line Segmentation, d.h. die Extrahierung der klassifizierten Layoutregionen und Textzeilen, die Textzeichenerkennung (Character Recognition) sowie die Korrektur der erkannten Texte und die Erstellung werkspezifischer OCR-Modelle in einem Trainingsmodul. OCR4all deckt somit den gesamten OCR-Workflow ab.

Nach Tests mit Wissenschaftler*innen aus dem GEI, die die Anwenderfreundlichkeit und Ergebnisqualität von OCR4all an historischen Schulbuchbeständen eruierten, intensivierte die Forschungsbibliothek des GEI die Zusammenarbeit mit den OCR4all-Entwickler*innen. Gemeinsam mit dem ZPD sowie dem Würzburger Lehrstuhl für Mensch-Computer-Interaktion (HCI) wurde das Projekt „OCR4all libraries“ initiiert, das von der DFG im Rahmen der OCR-D Förderlinie gefördert wird. Ziel des Projekts ist es, die OCR4all-Software so zu erweitern und anzupassen, dass die OCR-D-Module niederschwellig und eigenständig sowohl auf Korpus- und Werkebene als auch bei größeren Mengen im bibliothekarischen Kontext eingesetzt werden können. Ein hierfür im Projekt geplantes Graphical User Interface (GUI) soll nicht nur Wissenschaftler*innen befähigen, die um die OCR-D-Module angepasste OCR4all-Software intuitiv zu nutzen. Die GUI und eine visuelle Erklärungskomponente zur Erstellung und Konfiguration optimaler OCR-Workflows soll auch Bibliotheken unterstützen, ihre Volltexterkennung flexibler durchzuführen.

Zur Verbesserung der OCR-Qualität bei Drucken in Fraktur und unnormierten Schriftbildern wird im Projekt ein granularer Ansatz erprobt. Es wird dabei ein Verfahren entwickelt, das eine nach Korpora mit jeweils ähnlicher Materialgrundlage organisierte Volltexterkennung erlaubt. Bei Lesebüchern aus dem Deutschen Kaiserreich sind beispielsweise Inhaltsverzeichnisse eine signifikante OCR-Fehlerquelle. Wurden solche Materialspezifika eruiert, können entsprechende OCR-Modelle für eine Datenoptimierung trainiert werden. Von der Generierung und dem Training entlang von Materialspezifika erhoffen sich die Projektpartner*innen von OCR4all libraries eine vielversprechende OCR-Optimierung auch für Massenverfahren. Die Volltexterkennung auf Forschungs- und Materialbedarfe auszurichten, bedeutet, sie nicht nur wie beschrieben partizipativ und iterativ, sondern verstärkt agil zu gestalten. Die Auszeichnung und Erzeugung fehlerfreier Zeichen und Wörter (Ground-Truth-Daten) als Goldstandard sowie eine breite Bereitstellung trainierter OCR-Modelle ist

20 Nieländer; Weiß: Schönere Daten, 2018, S. 91–116, hier S. 97, <<https://repository.gei.de/handle/11428/296>> (DOI 10.14220/9783737009539), Stand: 26.04.2022.

dabei essenziell, denn dadurch lassen sich Texterkennungsprozesse umfassender bewerten und verbessern. Inwieweit OCR4all-libraries an von Bibliotheken genutzte Digitalisierungsworkflowsysteme angebunden werden kann, wird in einer im Projekt ebenfalls geplanten Machbarkeitsstudie geprüft. In der Entwicklung und Bereitstellung entsprechender Schnittstellen, etwa zu Goobi, Kitodo und DWork, sehen die Projektpartner*innen einen zentralen Meilenstein zur optimalen Unterstützung der Digitalisierung in Bibliotheken.

OCR-Workflows durch eine intuitive Usability eigenständig zu initiieren und spezifisch zu parametrisieren, hätte für Bibliotheken und Wissenschaftler*innen den Vorteil, flexibel auf Forschungsbedarfe und Materialbesonderheiten reagieren zu können. Auch am GEI noch zu klären ist die Nachnutzung bzw. Integration von optimierten Volltext-Daten in Digitalisierungsworkflow- und Präsentationssysteme und die Sicherung und Dokumentation von Volltextversionen. Verschiedene OCR-Textversionen zu dokumentieren, wäre für die Forschung vor allem im Hinblick auf die Zitierbarkeit von Volltexten und damit die Nachvollziehbarkeit der Forschungsergebnisse wichtig.²¹ Die Versionen der Volltexte fungieren in diesem Sinne als Forschungsdaten und sind demnach für ein Forschungsdatenmanagement relevant.

Literaturverzeichnis

- DFG: Praxisregeln „Digitalisierung“. DFG Vordruck 12.151 – 12/16, <https://www.dfg.de/formulare/12_151/12_151_de.pdf>, Stand: 26.04.2022.
- DFG: Merkblatt und ergänzender Leitfaden – Digitalisierung und Erschließung, DFG Vordruck 12.15 – 09/21, <https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/digitalisierung_erschliessung/formulare_merkblaetter/index.jsp>, Stand: 26.04.2022.
- DFG: Implementierung der OCR-D-Software zur Volltextdigitalisierung. Information für die Wissenschaft Nr. 15 | 27. Februar 2020, <https://www.dfg.de/foerderung/info_wissenschaft/2020/info_wissenschaft_20_15/index.html>, Stand: 26.04.2022.
- DHd (Digital Humanities im deutschsprachigen Raum): AG OCR – Punkt 2: Arbeitsschwerpunkte, <<https://dig-hum.de/ag-ocr>>, Stand: 26.04.2022.
- Engl, Elisabeth: OCR-D kompakt. Ergebnisse und Stand der Forschung in der Förderinitiative, in: Bibliothek – Forschung und Praxis 44 (2), 2020, S. 218–230. Online: <<https://doi.org/10.1515/bfp-2020-0024>>.
- Fühles-Ubach, Simone: Vom „embedded“ zum „liaison librarian“ – Was versprechen die neuen Konzepte?, in: Mittermaier, Bernhard (Hg.): Vernetztes Wissen – Daten, Menschen, Systeme. 6. Konferenz der Zentralbibliothek Forschungszentrum Jülich. 5.-7. November

21 Siehe die Umfrage zur Verwendung von OCR-Texten: <<https://ocr-d.de/de/umfrage>>, Stand: 26.04.2022.

2012, Proceedingsband. (Schriften des Forschungszentrums Jülich Reihe Bibliothek / Library Band / Volume 21). Jülich 2012, S. 337-350. Online: <<https://juser.fz-juelich.de/record/126960/files/FZJ-2012-00028.pdf>>, Stand: 26.04.2022.

- Gasser, Sonja: Das Digitalisat als Objekt der Begierde. Anforderungen an digitale Sammlungen für Forschung in der Digitalen Kunstgeschichte, in: Andraschke, Udo; Wagner, Sarah (Hg.): Objekte im Netz. Wissenschaftliche Sammlungen im digitalen Wandel, Bielefeld 2020, S. 261–276. Online: <<https://doi.org/10.14361/9783839455715>>.
- Hertling, Anke; Klaes, Sebastian: Historische Schulbücher als digitales Korpus für die Forschung. Auswahl und Aufbau einer digitalen Schulbuchbibliothek, in: Nieländer, Maret; De Luca, Ernesto William (Hg.): Digital Humanities in der internationalen Schulbuchforschung. (Eckert. Expertise 9). Göttingen 2018, S. 22-44. Online: <<https://repository.gwi.de/handle/11428/296>> (DOI 10.14220/9783737009539), Stand: 26.04.2022.
- Jacobmeyer, Wolfgang: Das deutsche Schulgeschichtsbuch 1700-1945. Die erste Epoche seiner Gattungsgeschichte im Spiegel der Vorworte, Bd. 1, Berlin 2011.
- Jäger, Georg: Der Schulbuchverlag, in: Ders. et al. (Hg.): Geschichte des deutschen Buchhandels im 19. und 20. Jahrhundert, Bd. 1: Das Kaiserreich 1870-1918, Teil 2, Frankfurt am Main 2003.
- Nieländer, Maret; Weiß, Andreas: »Schönere Daten« – Nachnutzung und Aufbereitung für die Verwendung in Digital-Humanities-Projekten, in: Nieländer, Maret; De Luca, Ernesto William (Hg.): Digital Humanities in der internationalen Schulbuchforschung. (Eckert. Expertise 9), Göttingen 2018, S. 91–116. Online: <<https://repository.gwi.de/handle/11428/296>> (DOI 10.14220/9783737009539), Stand: 26.04.2022.
- Reul, Christian; Springmann, Uwe; Wick, Christoph; Puppe, Frank: State of the Art Optical Character Recognition of 19th Century Fraktur Scripts using Open Source Engines, in: DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts (1.0, p. 366), S. 212–216, <<https://doi.org/10.5281/zenodo.2596095>>.
- Weil, Stefan: Neue Frakturmodelle für Tesseract. Präsentation auf dem Kitodo Anwendertreffen 18.–19. November 2019, S. 3. Online: <<https://madoc.bib.uni-mannheim.de/53748/1/2019-11-18.pdf>>, Stand: 26.04.2022.
- Weil, Stefan: tesseract-ocr / tesstrain, <<https://github.com/tesseract-ocr/tesstrain/wiki>>, Stand: 26.04.2022.
- Wick, Christoph; Reul, Christian; Puppe, Frank: Calamari – A High-Performance Tensorflow-based Deep Learning Package for Optical Character Recognition, in: Digital Humanities Quarterly 14 (2), 2020. Online: <<https://doi.org/10.48550/arXiv.1807.02004>>.

Konferenzbeiträge strategisch publizieren

Automatisierte Workflows zur individuellen Veröffentlichung von Konferenzbeiträgen am Beispiel des Verbands Digital Humanities im deutschsprachigen Raum e.V.

Patrick Helling, Data Center for the Humanities (DCH), Universität zu Köln

Anke Debbeler, Data Center for the Humanities (DCH), Universität zu Köln

Rebekka Borges, Rheinische Friedrich-Wilhelms-Universität Bonn

Zusammenfassung

Wissenschaftliche Konferenzen stellen ein wichtiges Forum für den fachlichen Austausch zwischen Forscher*innen dar. Sie dienen häufig als Schaufenster einzelner Fachbereiche. Die digitale Publikation von präsentierten Konferenzbeiträgen dokumentiert dabei zusätzlich aktuelle Entwicklungen eines Forschungsfeldes und erhöht die Sichtbarkeit der wissenschaftlichen Arbeit. Die tatsächliche Publikationspraxis einzelner Konferenzen unterscheidet sich mitunter deutlich. In diesem Beitrag stellen wir automatisierte Prozesse zur Publikation von Konferenzbeiträgen am Beispiel der Jahreskonferenzen des Verbands „Digital Humanities im deutschsprachigen Raum“ e.V. vor. Im Fokus steht dabei die Weiterentwicklung der bisherigen Publikationsprozesse des Verbands vom Sammelband zur individuellen Einzelpublikation der Konferenzbeiträge.

Abstract

Academic conferences are an important platform for scientific exchange between researchers. They often serve as a showcase for individual disciplines. The digital publication of presented conference papers also documents current developments in a research field and increases the visibility of scientific work. The actual publication practice of individual conferences sometimes differs significantly. In this article, we present automated processes for the publication of conference papers based on the example of the annual conferences of the association “Digital Humanities im deutschsprachigen Raum” e.V. The focus lies on the further development of the association’s publication processes from collective volumes to the individual publication of conference papers.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5835>

Schlagwörter: Kongressbericht, Elektronische Publikation, Bibliografie, Workflow, Arbeitsablauf, Digital Humanities

Autorenidentifikation:

Patrick Helling: ORCID: [0000-0003-4043-165X](https://orcid.org/0000-0003-4043-165X);

Anke Debbeler: ORCID: [0000-0001-5864-8465](https://orcid.org/0000-0001-5864-8465);

Rebekka Borges: ORCID: [0000-0002-4651-5638](https://orcid.org/0000-0002-4651-5638)

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Einleitung

Die Einreichung und Präsentation von Beiträgen auf wissenschaftlichen Konferenzen und Fachtagungen stellen einen wichtigen Bestandteil wissenschaftlicher Arbeit dar. Sie ermöglichen einerseits die Kommunikation von Ergebnissen aus Forschungsprozessen, andererseits dienen sie als Basis für einen fachwissenschaftlichen Austausch mit Kolleg*innen. Dementsprechend fungieren sie als ein wichtiges Instrument zur Beförderung wissenschaftlichen Fortschritts. Die Publikation dieser Beiträge durch Konferenzveranstalter bzw. Fachverbände stellt zusätzlich ein Organ zur Dissemination von Ergebnissen und Themen einer wissenschaftlichen Veranstaltung dar.

Die Praxis zur Veröffentlichung von Konferenzbeiträgen unterscheidet sich dabei häufig zwischen verschiedenen Konferenzen und Fachtagungen: Neben klassischen Printpublikationen werden Konferenzbeiträge mittlerweile auch zunehmend in digitaler Form publiziert. Digitale Konferenzbände i.S.v. Proceedings oder sogenannter Books of Abstracts bündeln durch die gesammelte Publikation von überarbeiteten Konferenzbeiträgen und Abstracts den Stand aktueller Forschungen. Aber auch andere Publikationsformen für Konferenzbeiträge wie bspw. Sondersammelbände als Teil (fachspezifischer) Journals sowie die Veröffentlichung von Konferenzbeiträgen als Einzelpublikationen, bspw. im PDF-Format oder als HTML-Website, werden zur Dokumentation und Präsentation einer Veranstaltung bzw. eines Forschungsfeldes verwendet. Jede dieser Publikationsformen konstituiert in ihrem Sinne einen Fachbereich sowohl nach innen als auch nach außen und dokumentiert die Entwicklungen einer Konferenz bzw. eines Forschungsfeldes.¹

Die Beiträge der Jahreskonferenzen des Verbands „Digital Humanities im deutschsprachigen Raum“ e.V. (DHD) werden seit der zweiten Jahreskonferenz gesammelt in Form eines Book of Abstracts im PDF-Format publiziert.² Wenngleich diese Books of Abstracts ein wichtiges Schaufenster des Fachbereichs der Digital Humanities (DH) im deutschsprachigen Raum darstellen, ermöglichen sie weder eine individuelle Zitier- und Referenzierbarkeit einzelner fachlicher Beiträge im Rahmen weiterführender Arbeiten, noch können die enthaltenen Einzelbeiträge in digitalen Katalogen und Findmitteln individuell indiziert werden.³ Um die Auffindbarkeit, Zugänglichkeit und Nutzbarkeit der DHD-Konferenzbeiträge im Sinne der FAIR-Prinzipien zu verbessern, ist eine zusätzliche Veröffentlichung von Tagungs- und Konferenzbeiträgen als einzeln zitier- und referenzierbare Publikationen notwendig.⁴

1 Sahle, Patrick (Hg.): DHD 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts, Vorwort, Frankfurt & Mainz 2019. Online: <<https://doi.org/10.5281/zenodo.2596095>>.

2 Die im Book of Abstracts publizierten Beiträge stellen eine Zwischenform dar. Sie sind weder kurze Zusammenfassungen noch voll über-/ausgearbeitete Konferenzbeiträge i.S.v. Proceedings. Es handelt sich um die eingereichten und begutachteten Abstracts zur Konferenz mit einem durchschnittlichen Umfang zwischen 500 und 2000 Wörtern. In diesem Sinne erscheint die historisch bedingte Benennung der Publikationsform als „Book of Abstracts“ irreführend. Es gibt Bestrebungen innerhalb der DHD-Community, das Book of Abstracts zu einem umfangreicheren Proceedings-Band auszubauen, was in dem vorliegenden Beitrag allerdings nicht problematisiert wird. Trotz dieser begrifflichen Ungenauigkeit werden die in den Books of Abstracts publizierten Texte im Folgenden als Konferenzbeiträge bezeichnet.

3 Schöch, Christof (Hg.): DHD 2020 Spielräume. Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts, Vorwort, Paderborn 2020. Online: <<https://doi.org/10.5281/zenodo.3666690>>.

4 Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific Data 2016 (3). Online: <<https://doi.org/10.1038/sdata.2016.18>>.

Vor diesem Hintergrund hat der DHd-Verband auf seiner Jahreskonferenz 2020 in Paderborn die Stelle eines DHd Data Stewards eingerichtet.⁵ Eine zentrale Aufgabe des DHd Data Stewards ist die Optimierung der Publikationsprozesse für wissenschaftliche Beiträge von DHd-Jahreskonferenzen. Im vorliegenden Beitrag soll die Arbeit des DHd Data Stewards vorgestellt werden.⁶ Im Fokus steht dabei die Vorstellung von Workflows zur nachhaltigen sowie individuell referenzier- und zitierbaren Veröffentlichung von Konferenzbeiträgen als Einzelpublikationen am Beispiel der Jahreskonferenzen des DHd-Verbands.

2. Ausgangssituation im DHd-Verband

Der Verband „Digital Humanities im deutschsprachigen Raum“ e.V. vertritt seit seiner Gründung auf der „DHd Unconference“ 2012 in Hamburg die digitalen Geisteswissenschaften / Digital Humanities in Deutschland, Österreich und der Schweiz.⁷ Als eingetragener Verein verfügt der DHd-Verband über eine eigene Satzung und einen Vorstand, welcher aus insgesamt sieben Hauptmitgliedern unterschiedlicher geisteswissenschaftlicher Fachrichtungen sowie aktuell drei kooptierten Mitgliedern für spezifische Aufgabenbereiche besteht (Stand: Mai 2022).⁸ Der DHd-Verband, der mittlerweile rund 400 Mitglieder führt, agiert als transnationale Interessensvertretung im Kontext unterschiedlicher internationaler Digital-Humanities-Verbände, insbesondere der European Association for Digital Humanities (EADH) und der Alliance of Digital Humanities Organizations (ADHO).⁹

Seit 2014 findet jedes Jahr (mit einer COVID-19 bedingten Ausnahme 2021) eine zentrale DHd-Jahreskonferenz statt, die durch wechselnde DH-Institutionen, -Lehrstühle und -Institute in Deutschland, Österreich und der Schweiz organisiert wird. Mittlerweile nehmen jährlich zwischen 500 und 600 Wissenschaftler*innen an den Tagungen teil. Im Durchschnitt werden ca. 130 Beiträge in den Formaten Vortrag (Umfang: 1500-2000 Wörter), Panel (Umfang: 1200-1500 Wörter), Workshop (Umfang: 1200-1500 Wörter), Poster (Umfang: 500-750 Wörter) und Doctoral Consortium (Umfang: 500-750 Wörter) (seit 2020) präsentiert. Mit Ausnahme der ersten DHd-Jahreskonferenz 2014 in Passau wurde zu jeder Jahreskonferenz ein Book of Abstracts mit allen präsentierten und durch ein Peer Review Verfahren begutachteten Beiträgen digital publiziert.¹⁰ Die Erstellung und Veröffentli-

5 Digital Humanities im deutschsprachigen Raum, <<https://dig-hum.de/>>, Stand: 12.05.2022; Website zur DHd-Jahreskonferenz 2020 in Paderborn, <<https://dhd2020.dig-hum.de/>>, Stand: 12.05.2022.

6 DHd Data Steward, <<https://dig-hum.de/dhd-data-steward>>, Stand: 19.08.2022.

7 Website zur DHd Unconference 2012 in Hamburg, <<https://dig-hum.de/dhd-gr%C3%BCndung-2012>>, Stand: 12.05.2022.

8 Satzung des DHd-Verbands, <<https://dig-hum.de/dhd-satzung>>, Stand: 12.05.2022; Der Vorstand des DHd-Verbands, <<https://dig-hum.de/dhd-vorstand>>, Stand: 12.05.2022.

9 European Association of Digital Humanities, <<https://eadh.org/>>, Stand: 12.05.2022; Alliance of Digital Humanities Organizations, <<https://adho.org/>>, Stand: 12.05.2022.

10 Siehe Stiegler, Johannes (Hg.): DHd 2015 Von Daten zu Erkenntnissen. Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation. Book of Abstracts, Graz 2015. Online: <<https://dhd2015.uni-graz.at/de/nachlese/book-of-abstracts/>>, Stand: 22.07.2022; Burr, Elisabeth (Hg.): DHd 2016 Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma, Leipzig 2017. Online: <<http://doi.org/10.5281/zenodo.3679331>>; Stolz, Michael (Hg.): DHd 2017 Digitale Nachhaltigkeit. Konferenzabstracts, Bern 2017. Online: <<http://doi.org/10.5281/zenodo.3684825>>; Vogeler, Georg (Hg.): DHd 2018 Kritik der digitalen Vernunft, Köln 2018. Online: <<http://doi.org/10.5281/zenodo.3684897>>; Sahle, Patrick (Hg.): DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts, Frankfurt & Mainz 2019. Online: <<https://doi.org/10.5281/zenodo.2596095>>; Schöch, Christof (Hg.): DHd 2020 Spielräume:

chung des Book of Abstracts ist dabei eine zentrale Aufgabe des jeweiligen Organisationskomitees der Jahreskonferenz.

2.1. Der Entstehungs- und Publikationsprozess der DHd Books of Abstracts

Die technischen und redaktionellen Prozesse zur Einreichung von Konferenzbeiträgen sowie zur Erstellung der Books of Abstracts zu den einzelnen DHd-Jahreskonferenzen durch die jeweiligen Organisationskomitees sind historisch gewachsen und haben sich mittlerweile verfestigt: Um einen Konferenzbeitrag zu einer DHd-Jahreskonferenz einzureichen, müssen Autor*innen in einer individuellen Instanz des digitalen Verwaltungssystems ConfTool, welches für jede Jahreskonferenz aufgesetzt wird, zunächst (1) einen Konferenzbeitrag anlegen.¹¹ In diesem Prozess müssen grundsätzliche Metadaten zum geplanten Konferenzbeitrag, bspw. Titel, Zusammenfassung, Autor*innen-Namen, Affiliation, sofern vorhanden die entsprechenden ORCID (seit der Jahreskonferenz 2022) sowie von den Autor*innen selbst zu vergebene Schlagwörter, eingetragen werden.¹² Zusätzlich müssen Autor*innen mit Hilfe der *Taxonomy of Digital Research Activities in the Humanities* (TaDiRAH), deren einzelne Elemente durch Checkboxes auswählbar sind, den Konferenzbeitrag kategorisieren.^{13,14}

Die im ConfTool eingetragenen Metadaten werden über eine Schnittstelle automatisch in das Transformationstool DH-Convalidator eingespeist. In diesem kann schließlich durch die Autor*innen (2) ein individuelles Word-Template zu dem zuvor im ConfTool eingerichteten Konferenzbeitrag erzeugt werden, welches die entsprechenden Metadaten zum Konferenzbeitrag beinhaltet.¹⁵

Wenn das Template durch die Autor*innen mit dem eigentlichen Konferenzbeitrag (einschließlich der Bibliografie) befüllt wurde, müssen sie dieses (3) wieder in den DH-Convalidator hochladen. Dieser prozessiert die Daten des Templates und erzeugt schließlich eine .dhc-Datei. Sie enthält das befüllte Template, eine HTML- und eine TEI-codierte XML-Version des Konferenzbeitrags sowie, falls vorhanden, verwendete Abbildungen.

Zum Abschluss der Einreichung müssen die Autor*innen diese .dhc-Datei wieder (4) im ConfTool zum angelegten Konferenzbeitrag hochladen.

Die TEI-XML-Dateien der nach dem Reviewprozess angenommenen Konferenzbeiträge werden schließlich vom Organisationskomitee der jeweiligen DHd-Jahreskonferenz für die (weitestgehend)

Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts, Paderborn 2020. Online: <<https://doi.org/10.5281/zenodo.3666690>>; Geierhos, Michaela (Hg.): DHd 2022 Kulturen des digitalen Gedächtnisses. Konferenzabstracts, Potsdam 2022. Online: <<https://doi.org/10.5281/zenodo.6304590>>.

11 ConfTool, <<https://www.conftool.net/de/startseite.html>>, Stand: 12.05.2022.

12 ORCID, <<https://orcid.org/>>, Stand: 12.05.2022.

13 Taxonomy of Digital Research Activities in the Humanities (TaDiRAH), <<https://de.dariah.eu/tadirah>>, Stand: 30.03.2022.

14 Reviewer*innen, die Konferenzbeiträge für die DHd-Jahreskonferenzen begutachten, müssen vor Beginn des Review-Prozesses ihre fachliche Expertise auch anhand der TaDiRAH-Taxonomie angeben. Entsprechend wird die Vergabe von TaDiRAH-Kategorien für einen Konferenzbeitrag im Reviewprozess für die Zuordnung von Konferenzbeiträgen zu fachlich kompetenten Reviewer*innen genutzt.

15 DHConvalidator, <<https://github.com/ADHO/dhconvalidator>>, Stand: 13.05.2022.

automatisierte Generierung des Book of Abstracts genutzt.¹⁶ Sie enthalten die eingetragenen Metadaten der Beiträge aus dem ConfTool, Verweise auf mitgelieferte Abbildungen sowie den Beitrag selbst inklusive Bibliografie und Fußnoten.

Diese formalisierte Datenbasis stellt grundsätzlich einen soliden Ausgangspunkt für die automatisierte Generierung eines Book of Abstracts dar. Aufgrund individueller Nutzung unterschiedlicher Textverarbeitungsprogramme der Autor*innen werden jedoch häufig zusätzliche Darstellungs- und Layoutinformationen in das Template und somit in den TEI-XML-Output übernommen. Als Folge weisen die TEI-XML-Strukturen Ungenauigkeiten auf und es bedarf i.d.R. einer grundsätzlichen Datenbereinigung vor der Erstellung des Book of Abstracts.

Für die automatisierte Erstellung des Book of Abstracts werden schließlich die gesammelten und bereinigten TEI-XML-Dateien sowie alle Abbildungen mit Hilfe von XSL-Transformationskripten und einem Shell-Skript zu einem entsprechenden Sammelband im PDF-Format zusammengefasst. Die Transformationskripte basieren auf Skripten zur Erstellung des Book of Abstracts zur internationalen Digital Humanities Konferenz 2013.¹⁷

Alle redaktionellen Bearbeitungen und die Darstellung von Sonderschriftzeichen sowie die Hinzufügung eines Deckblatts, die Erstellung von Inhalts- und Autor*innenverzeichnis sind innerhalb der Transformationskripte definiert.

Mit Hilfe dieses Prozesses wurden durch die jeweiligen Organisationskomitees der verschiedenen DHd-Jahreskonferenzen bereits sechs Books of Abstracts über den generischen Online-Speicherdienst Zenodo veröffentlicht.¹⁸ Jedes Book of Abstracts verfügt über einen Digital Object Identifier (DOI) und wurde einer durch den DHd-Verband kuratierten Zenodo-Community für Konferenzpublikationen zugeordnet.¹⁹ Bei der DHd-Jahreskonferenz 2015 wurde das Book of Abstracts aus eingereichten PDF-Dateien generiert und auf der institutionell betriebenen Website zur Jahreskonferenz veröffentlicht.²⁰ Zur ersten Jahreskonferenz 2014 wurde kein Book of Abstracts publiziert.

Die Erstellung und gesammelte Publikation aller Konferenzbeiträge in Books of Abstracts stellt den Status quo in der Publikationsstrategie des DHd-Verbands dar. Aufbauend auf diesem Publikationsprozess wurden im Rahmen der DHd Data Steward Tätigkeit technische und organisatorische Workflows zur gesonderten Veröffentlichung der Konferenzbeiträge als zusätzliche Einzelpublikationen

16 Beispielstruktur für eine TEI-XML-Datei der DHd Jahreskonferenz 2020, <https://github.com/PatrickHelling/DHd_BoA-separated/blob/main/2020tei.xsd>, Stand: 05.05.2022.

17 Digital Humanities Conference 2013, <<http://dh2013.unl.edu/>>, Stand: 12.05.2022; TEI to PDF Skripte, <<https://github.com/karindalziel/TEI-to-PDF>>, Stand: 12.05.2022.

18 Siehe Burr 2017, Online <<http://doi.org/10.5281/zenodo.3679331>>; Stolz 2017, <<http://doi.org/10.5281/zenodo.3684825>>; Vogeler 2018, Online: <<http://doi.org/10.5281/zenodo.3684897>>; Sahle 2019, Online <<https://doi.org/10.5281/zenodo.2596095>>; Schöch 2020, Online: <<https://doi.org/10.5281/zenodo.3666690>>; Geierhos 2022, Online: <<https://doi.org/10.5281/zenodo.6304590>>.

19 DHd-Community auf Zenodo, <<https://zenodo.org/communities/dhd/?page=1&size=20>>, Stand: 12.05.2022.

20 Siehe Stiegler 2015, Online: <<https://dhd2015.uni-graz.at/de/nachlese/book-of-abstracts/>>, Stand: 22.07.2022.

entwickelt. Zum Zweck der wissenschaftlichen Nachnutzung wurde darüber hinaus ein quasi-Standard zur zentralen Veröffentlichung der DHd-Konferenzbeiträge als TEI-XML-Daten eingeführt und umgesetzt.

3. Individuelle Publikation der einzelnen Beiträge der DHd-Jahreskonferenzen

3.1. Community-getriebene Vorüberlegungen

Mögliche Potentiale, die aus weiterführenden Publikationsbestrebungen entstehen könnten, sowie zusätzliche Überlegungen zur verbesserten, wissenschaftlichen Auswertung von Konferenzbeiträgen im Kontext des DHd-Verbands wurden durch die deutschsprachige Digital Humanities Community adressiert. In seinem Tagungsbericht zur DHd-Jahreskonferenz 2018 reflektiert Cremer bereits den aufwändigen Prozess der Beitragseinreichungen der DHd-Jahreskonferenzen und verweist auf das Problem, dass die Konferenzbeiträge (noch zu) selten als verarbeitbare Daten, bspw. in ihrer TEI-XML-Version, veröffentlicht werden bzw. auffindbar sind.²¹ Gleichzeitig illustrieren Hanneschläger und Andorfer mit ihrer Analyse zu Geschlechterverteilungen auf der TEI Konferenz 2016, die sie an den entsprechenden TEI-XML-Dateien der Beiträge zur Konferenz durchgeführt haben, welche Potentiale Analysen von Konferenzbeiträgen haben können.²² Mit der *DH(d) Konferenzbeiträge Bibliographie* konnte schließlich eine bis dahin noch nicht vorhandene Sammlung bibliografischer Einträge zu den DHd-Konferenzbeiträgen der Jahreskonferenzen 2016, 2018, 2019 und 2020 aufgebaut werden.²³ Zusätzlich wurde mit der *dhd-boas-app* eine analysierbare Datenbank zu den DHd-Konferenzbeiträgen der Jahreskonferenzen 2016, 2018, 2019 und 2020 entwickelt, die unter anderem Topic Modelling Analysen ermöglicht und das grundsätzliche Analysepotential der DHd-Konferenzbeiträge noch einmal unterstreicht.²⁴ Eine durch die Community getragene Diskussion über den Umgang mit DHd-Konferenzbeiträgen sowohl aus einer Publikations- als auch aus einer Datenanalyse-Perspektive wurde schließlich seit der DHd-Jahreskonferenz 2019 kontinuierlich in verschiedenen Konferenzbeitrags- und Veranstaltungsformaten geführt.²⁵

Vor dem Hintergrund dieser Vorüberlegungen ist das zentrale Ziel einer zusätzlichen Einzelpublikation der DHd-Konferenzbeiträge durch den DHd Data Steward die Erhöhung der Sichtbarkeit und Zugänglichkeit eben dieser selbst. Sie bietet die Möglichkeit, die DHd-Konferenzbeiträge mit

21 Cremer, Fabian: Nun sag, wie hältst Du es mit dem Digitalen Publizieren, Digital Humanities?, Digitale Redaktion, 2018. Online: <<https://editorial.hypotheses.org/113>>, Stand: 06.07.2021.

22 Hanneschläger, Vanessa; Andorfer, Peter: Menschen gendern? Einige Gedanken über Datenmodellierung zur Erhebung von Geschlechterverteilung anhand der TEI2016 Abstracts App, Köln 2018. Online: <<https://doi.org/10.5281/zenodo.4622405>>.

23 Lordick, Harald: DH(d) Konferenzbeiträge, 2020, <<http://www.steinheim-institut.de/dhd/>>, Stand: 02.05.2022.

24 Andorfer, Peter: dhd-boas-app, 2019, <<https://dhd-boas-app.acdh-dev.oew.ac.at/>>, Stand: 02.05.2022.

25 Andorfer, Peter; Cremer, Fabian; Steyer, Timo: DHd 2019 Book of Abstracts Hackathon, Frankfurt & Mainz 2019. Online: <<http://doi.org/10.5281/zenodo.4622102>>; Steyer, Timo; Andorfer, Peter; Cremer, Fabian: Abstract Enhancement. Potentiale der DHd-Konferenzabstracts als Daten/Publikation, Paderborn 2020. Online: <<http://doi.org/10.5281/zenodo.4621706>>; Andorfer, Peter; Busch, Anna; Cremer, Fabian u.a.: Bericht zur vDHd2021-Veranstaltung, Zukunftslabor DHd-Abstracts, DHd-Blog, 08.05.2021, Online: <<https://dhd-blog.org/?p=15980>>, Stand: 02.05.2022; Busch, Anna; Cremer, Fabian; Lordick, Harald u.a.: Strukturen und Impulse zur Weiterentwicklung der DHd-Abstracts, Potsdam 2022. Online: <<https://doi.org/10.5281/zenodo.6328089>>.

individuellen Metadaten zu beschreiben und, im Gegensatz zur Veröffentlichung innerhalb eines Book of Abstracts im PDF-Format, besser in Online-Katalogen und Findmitteln indexiert und auffindbar zu machen. Eine eindeutige und direkte Zitation einzelner Konferenzbeiträge über individuelle Persistent Identifier wird ermöglicht. Darüber hinaus konnten zusätzlich Strukturen geschaffen werden, die nicht nur eine Nachnutzung der Konferenzbeiträge als Publikationen verbessern, sondern auch als Daten für weitere Analysen ermöglichen.

3.2. Vorüberlegungen des DHD Data Stewards

Aufgrund der Tatsache, dass die meisten Books of Abstracts der DHD-Jahreskonferenzen (2016-2020) bereits auf dem Online Speicherdienst Zenodo publiziert und der DHD-Zenodo Community zugeordnet wurden, wurde auch für die Publikation der einzelnen DHD-Konferenzbeiträge Zenodo als Publikationsplattform gewählt.

Der Online-Speicherdienst wird langfristig von der European Organization for Nuclear Research (CERN) in der Schweiz betrieben, ist generisch auf die Publikation unterschiedlichster digitaler Objekte ausgelegt und verpflichtet sich gegenüber den FAIR-Prinzipien.²⁶ Er erfüllt wesentliche Grundbedingungen für die auffindbare, zugängliche, interoperable und nachnutzbare Ablage von Publikationen und verfügt über eine dokumentierte REST API-Schnittstelle, die für den automatisierten Upload von Publikationen genutzt werden kann.²⁷

Veröffentlichungen auf Zenodo können mit einem reichhaltigen Metadatenschema, welches sich am DataCite-Metadatenschema orientiert, beschrieben und ausgezeichnet werden.²⁸ Auf diese Weise können beispielsweise auch Konferenzen, auf denen die zu publizierenden Beiträge präsentiert wurden, beschrieben werden. Alle Publikationen auf Zenodo erhalten einen versionierten Digital Object Identifier (DOI), wodurch auch Versionen von Publikationen einzeln oder gesammelt referenziert werden können. Zusätzlich besteht die Möglichkeit, Lizenzen zu vergeben und Zugriffsrechte zu definieren.

3.3. Sammlung, Sichtung und Vorverarbeitung der DHD-Konferenzbeiträge

Zur Einzelpublikation der DHD-Konferenzbeiträge der vergangenen DHD-Jahreskonferenzen mussten zunächst alle entsprechenden TEI-XML-Dateien gesammelt werden. Für die Jahrgänge 2016, 2018, 2019 und 2020 waren zum Zeitpunkt der Vorbereitungen alle TEI-XML-Dateien sowie alle Abbildungen zu den Konferenzbeiträgen über verschiedene GitHub-Repositorien öffentlich verfügbar.²⁹ Die TEI-XML-Dateien zur DHD-Jahreskonferenz 2017 konnten über die lokalen Organisator*innen der Konferenz zur Verfügung gestellt werden.

26 Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a. 2016, Online: <<https://doi.org/10.1038/sdata.2016.18>>.

27 Zenodo Dokumentation, <<https://developers.zenodo.org/>>, Stand: 12.05.2022.

28 DataCite Metadatenschema, <<https://schema.datacite.org/meta/kernel-4.3/>>, Stand: 12.05.2022.

29 TEI-XML-Dateien zur DHD-Jahreskonferenz 2016, <<https://github.com/stefaniegehrke/dhd2016-boa>>, Stand: 31.03.2022; TEI-XML-Dateien zur DHD-Jahreskonferenz 2018, <<https://github.com/GVogeler/DHd2018>>, Stand: 31.03.2022; TEI-XML-Dateien zur DHD-Jahreskonferenz 2020, <<https://github.com/NinaSeemann/DHd2020-BoA>>, Stand: 31.03.2022. Das Repositorium mit den TEI-XML-Dateien zur DHD-Jahreskonferenz 2019 steht mittlerweile nicht mehr zur Verfügung.

Für die DHd-Jahreskonferenzen 2014 und 2015 existieren keine TEI-XML-Dateien. Für die Jahreskonferenz 2014 konnten die meisten Konferenzbeiträge durch die lokalen Organisator*innen sowie durch einen öffentlichen Aufruf auch durch die Community selbst in PDF- und Word-Formaten zur Verfügung gestellt werden. Für die Jahreskonferenz 2015 liegt ein Book of Abstracts im PDF-Format vor, welches manuell in die einzelnen Konferenzbeiträge getrennt wurde (siehe Tab. 1). Auf diese Weise konnten insgesamt 917 Konferenzbeiträge der DHd-Jahreskonferenzen 2014-2020 gesammelt werden (siehe Tab. 2).

Tab. 1: Übersicht der Verfügbarkeit der Daten zu den Konferenzbeiträgen der DHd-Jahreskonferenzen.

Jahr	Verfügbarkeit der Daten
2014	Word-/PDF-Dateien via lokale Organisator*innen der Jahreskonferenz/DHd-Community
2015	Konferenzbeiträge als Book of Abstracts im PDF-Format
2016	TEI-XML-Dateien via GitHub
2017	TEI-XML-Dateien via lokale Organisator*innen der Konferenz
2018	TEI-XML-Dateien via GitHub
2019	TEI-XML-Dateien via GitHub
2020	TEI-XML-Dateien via GitHub

Zur Vorverarbeitung der TEI-XML-Dateien der Jahreskonferenzen 2016-2020 wurden die bereits existierenden Transkriptionsskripte zur Erstellung von Books of Abstracts aus TEI-XML-Dateien (s.o.) angepasst, sodass aus den Ausgangsdaten einzelne PDF-Dateien zu jedem Konferenzbeitrag generiert werden konnten.³⁰ Zwecks Zuordnung entspricht der Dateiname der PDF-Dateien jeweils dem Namen der jeweiligen TEI-XML-Datei.³¹ Bei den einzelnen PDF-Dateien zu den DHd-Jahreskonferenzen 2014 und 2015 wurde jeweils ein individuelles Schema zur Dateibenennung gewählt (siehe auch Tab. 3).

Tab. 2: Anzahl der gesammelten Konferenzbeiträge der DHd-Jahreskonferenzen 2014-2020.

Jahr	Anzahl der zu publizierenden Konferenzbeiträge
2014	100 Konferenzbeiträge im PDF-Format
2015	131 Konferenzbeiträge im PDF-Format
2016	160 Konferenzbeiträge sowohl im TEI-XML- als auch im PDF-Format
2017	92 Konferenzbeiträge sowohl im TEI-XML- als auch im PDF-Format

30 Skripte für Einzel-PDF, <https://github.com/PatrickHelling/DHd_BoA-separated>, Stand: 12.05.2022.

31 Die TEI-XML-Dateien wurden bei ihrer Erstellung innerhalb des Einreichungsprozesses nach dem jeweiligen xml:id-Attribut benannt.

2018	160 Konferenzbeiträge sowohl im TEI-XML- als auch im PDF-Format
2019	137 Konferenzbeiträge sowohl im TEI-XML- als auch im PDF-Format
2020	137 Konferenzbeiträge sowohl im TEI-XML- als auch im PDF-Format

Darüber hinaus wurden mit weiteren XSL-Transformationskripten aus den TEI-XML-Dateien für die Jahrgänge 2016-2020 jeweils eine Konferenz-Metadatendatei im XML-Format generiert, die dem DataCite-Metadatenchema und somit den Anforderungen von Zenodo entsprechen.³² Um dies zu realisieren, wurden die TEI-XML-Dateien der einzelnen Jahreskonferenzen zunächst zu einer Gesamtdatei zusammengeführt, wobei das *xml:id*-Attribut bei einigen DHd-Jahreskonferenzen aus den TEI-Headern entfernt werden musste.³³ Aus den jeweiligen Gesamtdateien konnten dann die benötigten individuellen Metadaten zu den Konferenzbeiträgen extrahiert und in einer Konferenz-Metadatendatei zusammengefasst werden. Schließlich wurde aus den Konferenz-Metadatendateien für jeden Konferenzbeitrag eine einzelne JSON-Datei mit den entsprechenden Metadaten erstellt. Für die DHd-Jahreskonferenzen 2014 und 2015 wurden entsprechende Konferenz-Metadatendateien manuell erstellt.

Die verwendeten Metadaten umfassen inhaltliche und bibliografische Daten, Angaben zur jeweiligen Jahreskonferenz sowie eine generische *<description>* (bspw. „A single abstract from the DHd-2018 Book of Abstracts.“). Zusätzlich wurde eine *<reference>* zum jeweiligen Book of Abstracts, in dem der Konferenzbeitrag bereits veröffentlicht wurde, hinzugefügt. Auch Angaben zu Editor*innen sowie redaktionelle Hinweise zur Verarbeitung als *<note>* wurden angegeben. Bei der Verschlagwortung der Konferenzbeiträge, für die die durch die Autor*innen selbst vergebenen Schlagworte sowie die TaDiRAH-Auswahl verwendet wurden, wurde zusätzlich das Akronym der jeweiligen DHd-Jahreskonferenz (bspw. „DHd2018“) hinzugefügt. Auf diese Weise können nach Abschluss des Publikationsprozesses über das Interface von Zenodo alle einzelnen Publikationen der jeweiligen DHd-Jahreskonferenz ausgewählt und angezeigt werden. Um die Publikationen im Upload-Prozess direkt der durch den DHd-Verband kuratierten Zenodo Community zuordnen zu können, wurde auch der entsprechende Community-Identifizier „dhd“ hinzugefügt. Alle Konferenzbeiträge verfügen über eine CC-BY Lizenz und wurden als *open access* getagged.³⁴

3.4. Der Publikationsprozess

Für die technische Umsetzung des Publikationsworkflows wurde eine Python-Anwendung geschrieben, die zur Nachnutzung auch auf GitHub unter einer MIT-Lizenz veröffentlicht wurde.³⁵ Die Anwendung umfasst zwei Hauptfunktionen: Im ersten Schritt werden (1) einzelne Ordner (Bundles) pro Konferenzbeitrag erstellt, die eine PDF-Version und (sofern vorhanden) eine TEI-XML-Version eines

32 Skripte für Metadaten, <https://github.com/reborg789/zenodup>, Stand: 09.05.2022.

33 Der Attributinhalt wird i.d.R. automatisch aus dem ConfTool generiert. Bei einigen DHd-Jahreskonferenzen beginnt der Wert mit einer numerischen ID (bspw. „145_Helling_Der_DHd_Data_Steward“). Da ein Attribut, das mit numerischen Ziffern beginnt, in diesem Fall nicht valide ist, musste das *xml:id*-Attribut für die Verarbeitung entsprechend entfernt werden.

34 2020er Metadaten-Liste, <https://github.com/reborg789/zenodup/blob/main/2020metadata.xsd>, Stand: 05.05.2022.

35 Zenodup Skripte, <https://github.com/cceh/zenodup> Stand: 09.05.2022.

einzelnen DHd-Konferenzbeitrags sowie eine entsprechende Metadatendatei im JSON-Format, die aus der jeweiligen Konferenz-Metadatendatei erstellt wurde, enthalten (siehe Abb. 1). Anschließend wird über die Anwendung (2) die Anbindung an Zenodo über die Zenodo REST API realisiert und der automatisierte Upload sowie die Publikation der Beiträge durchgeführt.

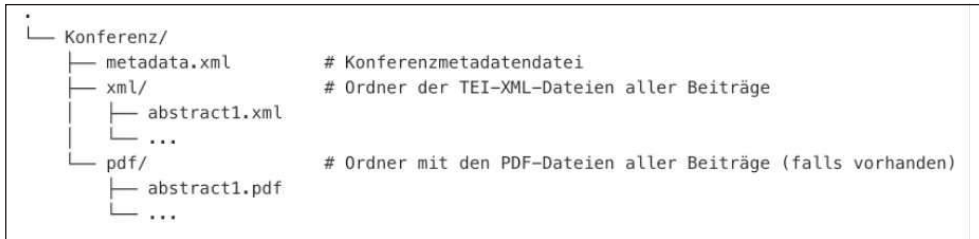


Abb. 1: Zugrundeliegende Ordnerstruktur eines Bundles für den Publikations-Workflow

3.4.1. Erstellung der Bundles pro Konferenzbeitrag

Das Zusammenführen der zu den einzelnen Konferenzbeiträgen zugehörigen Dateien in Bundles wurde zunächst über Konventionen der Dateibenennung realisiert. Die Dateinamen der PDF-Dateien sowie der TEI-XML-Dateien der DHd-Jahreskonferenzen 2016-2020 entsprechen den *xml:id*-Attributen der Konferenzbeiträge. Bei den einzelnen PDF-Dateien der DHd-Jahreskonferenzen 2014 und 2015, zu denen keine Konventionen in der Dateibenennung existieren, wurde ein eigenes Schema zur Dateibenennung gewählt. Die Benennungen unterscheiden sich je DHd-Jahreskonferenz; innerhalb einer Jahreskonferenz sind sie allerdings einheitlich (siehe Tab. 3).

Tab. 3: Unterschiedliche Konventionen bei der Dateibenennung der Beiträge zu den verschiedenen DHd-Jahreskonferenzen.

Jahr	Schema Dateibenennung
2014	[LAST_NAME_OF_CREATOR]_[FIRST_NAME_OF_CREATOR]_[TITLE]*
Beispiel: ARZHANOV_Yuri_The_Glossarium.pdf*	
2015	[ID]_[SLOT]_[LAST_NAME_OF_CREATOR]_[TITLE]_[NUMBER]*
Beispiel: 150225_Philologie_III_3_de_Kok-Weblicht-1151162.pdf*	
2016	[PUBLICATION_TYPE]_[NUMBER]
Beispiel: Panels-006.xml	
2017	[PUBLICATION_TYPE]_[LAST_NAME_OF_CREATOR]
Beispiel: Panel-FEIGE.xml	
2018	[LAST_NAME_CREATOR]_[FIRST_NAME_CREATOR]_[TITLE]

Beispiel: BARZEN_Johanna_SustainLife_Erhalt_Lebender_digitaler_Syst.xml

2019 [ID]_[LAST_NAME_CREATOR]_[FIRST_NAME_CREATOR]_[TITLE]

Beispiel: 119_SCHASSAN_Torsten_Erneuerung_der_Digitalen_Editionen_an_der_H.xml

2020 [ID]_[LAST_NAME_CREATOR]_[FIRST_NAME_CREATOR]_[TITLE]

Beispiel: 107_HADERSBECK_Maximilian_Nachlass_Ludwig_Wittgenstein_Software.xml

**Das Schema zur Dateibenennung wurde im Vorverarbeitungsprozess manuell gewählt.*

Für die Konferenzbeiträge der DHd-Jahreskonferenzen 2018-2020 konnten zunächst mit einem automatisierten Abgleich der Dateinamen entsprechend zusammengehörende PDF- und TEI-XML-Dateien identifiziert und in einem Bundle zusammengeführt werden. Die Informationen aus den Dateinamen wurden schließlich auch dazu verwendet, den passenden Metadaten-Eintrag in der jeweiligen Konferenz-Metadatendatei zu identifizieren und im JSON-Format dem entsprechenden Bundle zuzuordnen. Aufgrund von unzuverlässiger Ersetzung von Sonder- sowie Leerzeichen durch '_' in den Dateinamen, konnte allerdings keine fehlerfreie Zuweisung aller Dateien zu einem Konferenzbeitrag erfolgen. Entsprechend wurde dazu übergegangen, eine Zuordnung von PDF- und TEI-XML-Dateien sowie passendem Metadaten-Eintrag im JSON-Format über eine chronologische bzw. alphabetische Reihenfolge der Dateien zu realisieren. Zu diesem Zweck wurde die Reihenfolge der Metadaten-Einträge in der jeweiligen Konferenz-Metadatendatei der Reihung von PDF- bzw. TEI-XML-Dateien angepasst. Der manuelle Kontrollaufwand beider Vorgehensweisen ist identisch, die Menge an zu korrigierenden Zuweisungen ist beim zweiten Vorgehen allerdings deutlich geringer.

Darüber hinaus existieren für die Jahreskonferenzen 2014 und 2015 keine vorgegebenen Konventionen für die Dateibenennung. Bei den DHd-Jahreskonferenzen 2016 und 2017 wurde ein Schema zur Dateibenennung der TEI-XML-Dateien verwendet, das keine Informationen über Autor*innen oder Beitragstitel enthält. Entsprechend hat sich auch für diese vier Jahrgänge die Bundle-Erstellung nach chronologischer bzw. alphabetischer Ordnung empfohlen.

Nach der Zuordnung der PDF-Dateien, der TEI-XML-Dateien (sofern vorhanden) sowie der Metadateneinträge zu individuellen Bundles pro Konferenzbeitrag wird schließlich automatisch überprüft, ob jedes erstellte Bundle die korrekte Anzahl an Dateien enthält. Für eine zusätzliche manuelle Überprüfung wird eine CSV-Datei für jede DHd-Jahreskonferenz mit einer Übersicht der generierten Bundles erstellt. Die CSV-Dateien enthalten für jedes Bundle eine Übersicht mit dem Titel der Publikation aus der JSON-Metadatendatei, den Dateinamen der PDF- und ggf. der TEI-XML-Dateien sowie die aus den TEI-XML-Dateien extrahierten Titel der Publikation (Workflow zur Bundle-Erstellung siehe Abb. 2).

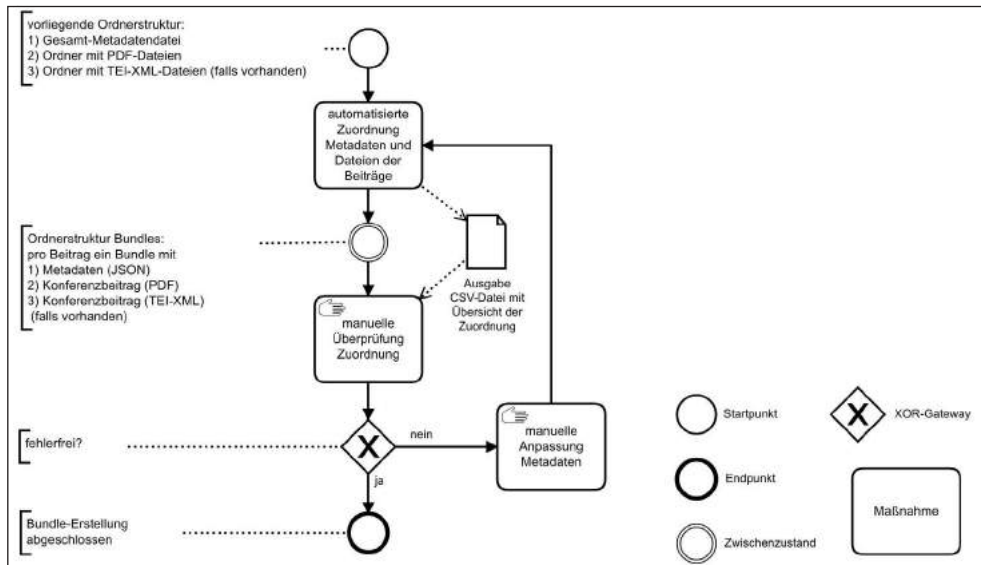


Abb. 2: Workflow zur Bundle-Erstellung.

3.4.2. Automatisierter Upload der Bundles auf Zenodo

Der Upload der einzelnen Bundles über die Zenodo REST API erfolgt mit einem Funktionsaccount des DHD-Verbands, über den auch die Zenodo-Community des Verbands kuratiert wird (Publikationsworkflow siehe Abb. 3). Beim Upload der Konferenzbeiträge auf Zenodo werden für jedes Bundle die Metadaten aus der JSON-Datei gelesen und zusammen mit den jeweiligen PDF- und ggf. TEI-XML-Dateien als Draft-Deposits gespeichert. Die dabei von Zenodo generierte *Deposition-ID* wird für alle Bundles einer Jahreskonferenz in einer Textdatei gespeichert.³⁶ Schließlich werden die hochgeladenen Drafts über das User Interface von Zenodo ein letztes Mal manuell überprüft und bei Bedarf korrigiert: Individuelle Fehler wie bspw. falsch oder nicht übertragene einzelne Metadaten können manuell korrigiert werden. Fallen bei der manuellen Überprüfung systematische Fehler auf, können die Drafts auf Zenodo automatisch gelöscht, die Fehler in der jeweiligen Metadaten-datei oder in den Python-Skripten korrigiert und der Workflow erneut gestartet werden.

36 Über die Deposition IDs können die Uploads über die Zenodo REST API gelöscht, publiziert oder eine Übersicht der Deposits erstellt werden. Siehe Zenodo REST API Dokumentation, <<https://developers.zenodo.org/#introduction>>, Stand: 25.03.2022.

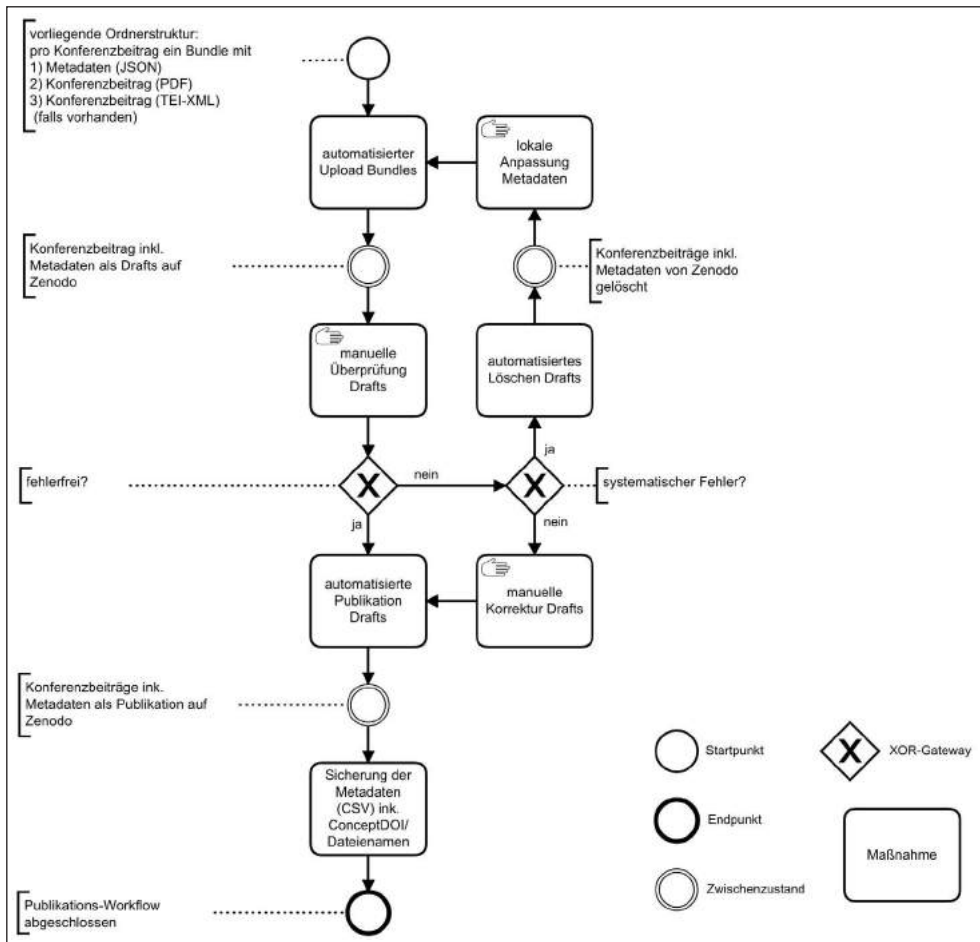


Abb. 3: Der Publikationsworkflow.

Nach einer letzten Überprüfung und ggf. Korrektur der Drafts werden die Uploads zur jeweiligen DHd-Jahreskonferenz schließlich automatisiert veröffentlicht. Zum Zweck der Dokumentation wird zuletzt eine Übersicht der publizierten Konferenzbeiträge, der final verwendeten Metadaten sowie der zugeordneten DOIs, über die Zenodo REST API in einer CSV-Datei extrahiert.

Zu jeder DHd-Jahreskonferenz, zu der die Konferenzbeiträge über die beschriebenen Workflows vorbereitet und publiziert wurden, wurde zusätzlich ein GitHub-Repository über den GitHub-Account des DHd-Verbands eingerichtet. Diese Repositorien enthalten alle PDF-Dateien der einzelnen Konferenzbeiträge, sofern vorhanden die zugehörigen TEI-XML-Dateien, die für den Upload verwendete Konferenz-Metadatendatei sowie die am Ende des Prozesses generierte CSV-Datei zu

den Publikationen.³⁷ Neben der Dokumentation der Vorverarbeitungs- und Publikationsprozesse ermöglichen diese GitHub-Repositorien auch explizit die Nachnutzung der Daten und Prozesse für weitere inhaltliche Analysen und Auswertungen der Konferenzbeiträge zu DHd-Jahreskonferenzen.

4. Erfahrungen im Rahmen der DHd-Jahreskonferenz 2022

Die beschriebenen Workflows und Prozesse basieren auf der Anforderung, die Konferenzbeiträge der bereits abgeschlossenen DHd-Jahreskonferenzen 2014-2020 zu verarbeiten und persistent zu publizieren. Im Jahr 2022 wurden diese Workflows und Prozesse schließlich das erste Mal im Vorfeld einer anstehenden DHd-Jahreskonferenz durchgeführt. Ziel dabei war es, die Konferenzbeiträge bereits vor der eigentlichen Jahreskonferenz persistent sowohl gesammelt in einem Book of Abstracts als auch als individuelle Einzelpublikationen zu veröffentlichen. Die bis hierhin entwickelten Workflows und Prozesse haben sich auch im Szenario einer bevorstehenden Jahreskonferenz grundsätzlich bewährt. Neben dem Book of Abstracts konnten insgesamt 154 Konferenzbeiträge zur DHd-Jahreskonferenz 2022 in der DHd-Community auf Zenodo publiziert werden.³⁸ Zusätzlich wurde der Publikationsprozess genutzt, um das erste Mal in der Geschichte der DHd-Jahreskonferenzen auch die Posterpräsentation zu den Posterbeiträgen persistent auf Zenodo zu veröffentlichen.³⁹ Insgesamt konnten auf diese Weise 61 Posterpräsentationen publiziert und mit den jeweiligen Konferenzbeiträgen verknüpft werden.

Dennoch haben sich, nicht zuletzt resultierend aus der Terminierung der Publikation vor Konferenzbeginn, einige *lessons learned* ergeben: Vor dem Hintergrund, dass die Erstellung eines gesammelten Book of Abstracts im Aufgabenbereich des jeweiligen lokalen Konferenz-Organisationskomitees liegt und die Datenbasis für das Book of Abstracts sowie für die Veröffentlichung der Konferenzbeiträge als Einzelpublikationen identisch ist, bedarf es eines frühzeitigen und regelmäßigen Austauschs zwischen dem DHd Data Steward und dem lokalen Organisationskomitee. Hierbei ist eine Zusammenarbeit insbesondere bei der Bereinigung der TEI-XML-Dateien sowie erweiterter/veränderter Metadaten, welche durch das lokale Konferenz-Organisationskomitee erfasst werden, von zentraler Bedeutung.

So wurde bspw. im Rahmen der DHd-Jahreskonferenz 2022 im Einreichungsprozess für Konferenzbeiträge zum ersten Mal die ORCID der Autor*innen abgefragt und in die Metadaten der TEI-XML-Dateien übernommen. Folglich war es auch erforderlich, die ORCID-Angaben bei der Erstellung der Metadaten für den Publikationsprozess der Konferenzbeiträge als Einzelpublikationen zu integrieren und bei der Veröffentlichung mit zu berücksichtigen. Darüber hinaus müssen grundsätzlich frühzeitig

37 GitHub-Repositorium zur DHd-Jahreskonferenz 2014, <<https://github.com/DHd-Verband/DHd-Abstracts-2014>>, Stand: 05.04.2022; GitHub-Repositorium zur DHd-Jahreskonferenz 2015, <<https://github.com/DHd-Verband/DHd-Abstracts-2015>>, Stand: 05.04.2022; GitHub-Repositorium zur DHd-Jahreskonferenz 2016, <<https://github.com/DHd-Verband/DHd-Abstracts-2016>>, Stand: 05.04.2022; GitHub-Repositorium zur DHd-Jahreskonferenz 2017, <<https://github.com/DHd-Verband/DHd-Abstracts-2017>>, Stand: 05.04.2022; GitHub-Repositorium zur DHd-Jahreskonferenz 2018, <<https://github.com/DHd-Verband/DHd-Abstracts-2018>>, Stand: 05.04.2022; GitHub-Repositorium zur DHd-Jahreskonferenz 2019, <<https://github.com/DHd-Verband/DHd-Abstracts-2019>>, Stand: 05.04.2022; GitHub-Repositorium zur DHd-Jahreskonferenz 2020, <<https://github.com/DHd-Verband/DHd-Abstracts-2020>>, Stand: 05.04.2022.

38 Geierhos 2022, Online: <<https://doi.org/10.5281/zenodo.6304590>>.

39 DHd-Zenodo-Community, <<https://zenodo.org/communities/dhd/?page=1&size=20>>, Stand: 09.05.2022.

redaktionelle Abstimmungen und Anpassungen im Layout der Publikationen gemeinsam geklärt und umgesetzt werden.

5. Ausblick

Mit Hilfe der in diesem Beitrag beschriebenen Workflows und Prozesse konnten die Auffindbarkeit, Zitierbarkeit und individuelle Referenzierbarkeit der Konferenzbeiträge zu DHd-Jahreskonferenzen via Zenodo grundsätzlich verbessert werden. Zusätzlich konnten auf diese Weise die DHd-Konferenzbeiträge mittlerweile durch das *Digital Bibliography & Library Project (DBLP)* und den *Index of Digital Humanities Conferences* indexiert werden.⁴⁰

Die entwickelten Workflows stellen dabei einen pragmatischen Ansatz dar, der explizit auf die Ausgangssituation und insbesondere die Datengrundlage ausgerichtet ist, die sich im Rahmen der DHd-Jahreskonferenzen ergeben. Nichtsdestotrotz konnten der Vorverarbeitungs- und Publikationsprozess in Teilen bereits im Rahmen weiterer Konferenzen nachgenutzt werden: So wurden mittlerweile auch die Konferenzbeiträge der *FORGE 2021 Konferenz - Forschungsdaten in den Geisteswissenschaften* sowie der durch die DHd-Community organisierten virtuellen *vDHd Konferenz 2021* über die beschriebenen Workflows und Prozesse als Einzelpublikationen auf Zenodo publiziert werden.^{41, 42, 43}

Wenngleich die Entwicklung und Umsetzung der hier beschriebenen Vorverarbeitungs- und Publikationsprozesse als grundsätzliche Verbesserung der Publikationsstrategie des DHd-Verbands verstanden werden können, gibt es noch weitere Herausforderungen, die in der zukünftigen Arbeit des DHd Data Stewards angegangen werden sollen: Neben der kontinuierlichen Integration von Normdaten zur persistenten Identifikation und Referenzierbarkeit von bspw. Personen und Einrichtungen in den TEI-XML-Dateien sowie den Metadaten zu den Konferenzbeiträgen bedarf es auch einer Optimierung von Zitationsstil und Bibliografieangaben orientiert an aktuellen Community-Standards.

Vor dem Hintergrund, dass insbesondere der in diesem Beitrag beschriebene Einreichungsprozess für Konferenzbeiträge auf DHd-Jahreskonferenzen einerseits aus Nutzer*innen-Perspektive umständlich und wenig intuitiv gestaltet ist und andererseits das aus diesem Prozess generierte TEI-XML i.d.R. grundsätzliche manuelle Bereinigungen erfordert, gilt es auch diesen Prozess zu überprüfen und potenzielle Tools zur Optimierung zu evaluieren.⁴⁴

40 Digital Bibliography & Library Project (DBLP), <<https://dblp.org/db/conf/dhd/index.html>>, Stand: 12.05.2022; Index of Digital Humanities Conferences, <https://dh-abstracts.library.cmu.edu/conference_series/30>, Stand: 12.05.2022.

41 Sowohl bei der FORGE 2021 Konferenz als auch bei der vDHd Konferenz 2021 standen keine TEI-XML-Dateien als Datenbasis zur Verfügung. Entsprechend mussten die jeweiligen Metadaten zu den Konferenzbeiträgen manuell in eine Gesamt-Metadaten-datei geschrieben werden. Die Generierung einzelner Metadaten-Dateien im JSON-Format für jeden Konferenzbeitrag konnten wiederum automatisch umgesetzt werden. Publiziert wurden schließlich die einzelnen Konferenzbeiträge im PDF-Format.

42 FORGE 2021, <<https://forge2021.uni-koeln.de/>>, Stand: 17.05.2022; vDHd 2021, <<https://vdhd2021.hypotheses.org/>>, Stand: 17.05.2022.

43 FORGE-Zenodo-Community, <<https://zenodo.org/communities/forge/?page=1&size=20>>, Stand: 17.05.2022; DHd-Zenodo-Community, <<https://zenodo.org/communities/dhd/?page=1&size=20>>, Stand: 17.05.2022.

44 Fidus Writer, <<https://www.fiduswriter.org/>>, Stand: 12.05.2022; Gebhard, Henning: Fidus Writer als Alternative zum DH ConValidator? Ein Prototyp, Potsdam 2022, Online: <<http://doi.org/10.5281/zenodo.6327991>>.

Darüber hinaus gibt es einen Bedarf, eine grundsätzliche Dokumentation des gesamten Publikationsprozesses inklusive der Vorarbeiten für die Erstellung eines Book of Abstracts zu forcieren, um die Umsetzung im Rahmen zukünftiger DHd-Jahreskonferenzen zu erleichtern.

Literaturverzeichnis

- Andorfer, Peter; Cremer, Fabian; Steyer, Timo: DHd 2019 Book of Abstracts Hackathon, Frankfurt & Mainz 2019. Online: <<http://doi.org/10.5281/zenodo.4622102>>.
- Andorfer, Peter; Busch, Anna; Cremer, Fabian u.a.: Bericht zur vDHd2021-Veranstaltung. Zukunftslabor DHd-Abstracts, DHd-Blog, 08.05.2021. Online: <<https://dhd-blog.org/?p=15980>>, Stand: 02.05.2022.
- Burr, Elisabeth (Hg.): DHd 2016 Modellierung – Vernetzung – Visualisierung. Die Digital Humanities als Fächerübergreifendes Forschungsparadigma, Leipzig 2017. Online: <<http://doi.org/10.5281/zenodo.3679331>>.
- Busch, Anna; Cremer, Fabian; Lordick, Harald u.a.: Strukturen und Impulse zur Weiterentwicklung der DHd-Abstracts, Potsdam 2022. Online: <<https://doi.org/10.5281/zenodo.6328089>>.
- Cremer, Fabian: Nun sag, wie hältst Du es mit dem Digitalen Publizieren, Digital Humanities?, Digitale Redaktion, 2018. Online: <<https://editorial.hypotheses.org/113>>, Stand: 06.07.2021.
- Gebhard, Henning: Fidus Writer als Alternative zum DH ConValidator? Ein Prototyp, Potsdam 2022. Online: <<http://doi.org/10.5281/zenodo.6327991>>.
- Geierhos, Michaela (Hg.): DHd 2022 Kulturen des digitalen Gedächtnisses. Konferenzabstracts, Potsdam 2022. Online: <<https://doi.org/10.5281/zenodo.6304590>>.
- Hanneschläger, Vanessa; Andorfer, Peter: Menschen gendern? Einige Gedanken über Datenmodellierung zur Erhebung von Geschlechterverteilung anhand der TEI2016 Abstracts App, Köln 2018. Online: <<https://doi.org/10.5281/zenodo.4622405>>.
- Sahle, Patrick (Hg.): DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts, Frankfurt & Mainz 2019. Online: <<https://doi.org/10.5281/zenodo.2596095>>.
- Schöch, Christof (Hg.): DHd 2020 Spielräume. Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts, Paderborn 2020. Online: <<https://doi.org/10.5281/zenodo.3666690>>.

- Steyer, Timo; Andorfer, Peter; Cremer, Fabian: Abstract Enhancement. Potentiale der DHd-Konferenzabstracts als Daten/Publikation, Paderborn 2020. Online: <<http://doi.org/10.5281/zenodo.4621706>>.
- Stiegler, Johannes (Hg.): DHd 2015 Von Daten zu Erkenntnissen. Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation. Book of Abstracts, Graz 2015. Online: <<https://dhd2015.uni-graz.at/de/nachlese/book-of-abstracts/>>, Stand: 22.07.2022.
- Stolz, Michael (Hg.): DHd 2017 Digitale Nachhaltigkeit. Konferenzabstracts, Bern 2017. Online: <<http://doi.org/10.5281/zenodo.3684825>>.
- Vogeler, Georg (Hg.): DHd 2018 Kritik der digitalen Vernunft, Köln 2018. Online: <<http://doi.org/10.5281/zenodo.3684897>>.
- Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u.a.: The FAIR Guiding Principles for scientific data management and stewardship, in: Scientific Data 2016 (3). Online: <<https://doi.org/10.1038/sdata.2016.18>>.

Infrastrukturen und Services für die wissenschaftliche Nutzung von Webarchiven

Ein Überblick

Tobias Beinert, Bayerische Staatsbibliothek, München

Katharina Schmid, Bayerische Staatsbibliothek, München

Konstanze Weimer, web all in One, München

Zusammenfassung

Der Beitrag gibt zunächst einen kurzen Überblick über den aktuellen Stand der Webarchivierung in deutschen Bibliotheken und beleuchtet dabei auch die rechtlichen Rahmenbedingungen. Darauf aufbauend werden die derzeitige Praxis der Erschließung und Nutzung von Webarchiven sowie die Anforderungen an die Dokumentation der Prozesse der Webarchivierung beschrieben. Eine zusammenfassende Analyse von weitergehenden Formen der Datenbereitstellung aus Webarchiven sowie von unterstützenden Services zur wissenschaftlichen Nutzung mit computergestützten Analysemethoden anhand von Beispielen aus der internationalen Webarchivierungs-Community bildet den Schwerpunkt des Artikels.

Summary

The article first gives a brief overview of the current state of web archiving in German libraries and also sheds light on the legal framework. Based on this, the current practice of indexing and using web archives as well as the requirements for the documentation of web archiving processes are described. The focus of the article is a comprehensive analysis of additional forms of data provision from web archives and of supporting services for scientific use with computer-aided analysis methods using examples from the international web archiving community.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5821>

Autorenidentifikation:

Beinert, Tobias: ORCID: <https://orcid.org/0000-0003-0302-0536>;

Schmid, Katharina: ORCID: <https://orcid.org/0000-0001-6057-6640>;

Weimer, Konstanze: ORCID: <https://orcid.org/0000-0001-5080-7353>

Schlagwörter: Webarchivierung

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Aktuelle Entwicklungen und rechtliche Situation der Webarchivierung in Deutschland

Websites sind ein ständig wachsender Teil unseres kulturellen und wissenschaftlichen Erbes. Eine zunehmende Anzahl an Gedächtnisinstitutionen sammelt ausgewählte Websites. Im folgenden Beitrag werden die daraus entstehenden institutionellen Webarchive als die Gesamtheit aller archivierten

Websites, die eine Einrichtung verwaltet und für die Nachwelt aufbewahrt, verstanden. Am internationalen Maßstab gemessen sind die Aktivitäten zur Archivierung von Websites in deutschen Einrichtungen jedoch vergleichsweise überschaubar.¹ Im Bereich der deutschen Bibliotheken ist es eine kleine Anzahl an Akteur*innen, die die Archivierung und Bereitstellung von als wissenschaftlich oder kulturell relevant bewerteten Websites betreibt. Insbesondere aufgrund der geltenden rechtlichen Rahmenbedingungen und angesichts endlicher Ressourcen geschieht dies tendenziell in begrenztem Umfang.² In Deutschland sind Websites wie auch alle anderen Formen von Netzpublikationen urheberrechtlich geschützt, weshalb eine Vervielfältigung im Rahmen der Webarchivierung ohne die Einwilligung der Rechteinhaber*innen grundsätzlich verboten ist.³ Allerdings decken die Pflichtexemplargesetze auf Bundes- und Landesebene inzwischen zunehmend auch Netzpublikationen ab und erlauben Institutionen die Archivierung von Websites in ihrem jeweiligen Zuständigkeitsbereich. Netzpublikationen, die unter den Pflichtexemplarregelungen archiviert wurden, dürfen vor Ort in den Lesesälen zugänglich gemacht werden, nicht jedoch öffentlich im Netz, um die Interessen der Rechteinhaber*innen zu wahren.

Eine neue Entwicklung ist die anlaufende Zusammenarbeit der Deutschen Nationalbibliothek (DNB) mit einigen Landesbibliotheken wie der Thüringer Universitäts- und Landesbibliothek (ThULB) Jena oder der Staats- und Universitätsbibliothek (SUB) Hamburg.⁴ Die Anpassung des Gesetzes über die Deutsche Nationalbibliothek aus dem Jahr 2017 ermöglicht eine Form der Kooperation, bei der die Landesbibliotheken jeweils landeskundlich relevante Ressourcen vorschlagen und diese von der DNB in Zusammenarbeit mit einem*r Dienstleister*in archiviert werden.⁵ Die Bereitstellung erfolgt dann den rechtlichen Vorgaben entsprechend sowohl im Lesesaal der DNB als auch im Lesesaal der jeweiligen Landesbibliothek, wobei dort jeweils auf den Gesamtbestand des Webarchivs der Deutschen Nationalbibliothek zugegriffen werden kann.

Generell stellen die in Deutschland aktiven Bibliotheken ihre Webarchive überwiegend vor Ort in den Lesesälen bereit:

- Deutsche Nationalbibliothek
- Badische Landesbibliothek und Württembergische Landesbibliothek

1 Vgl. Altenhöner, Reinhard: Noch immer am Anfang? Stand und Perspektiven der Webarchivierung in Deutschland 2019, in: Fühles-Ubach, Simone; Georgy, Ursula (Hg.): Bibliotheksentwicklung im Netzwerk von Menschen, Informationstechnologie und Nachhaltigkeit. Festschrift für Achim Oßwald, Bad Honnef 2019, S. 237–250. Online: <<https://nbn-resolving.org/urn:nbn:de:hbz:79pbc-opus-16232>>; Beinert, Tobias; Schoger, Astrid: Vernachlässigte Pflicht oder Sammlung aus Leidenschaft. Zum Stand der Webarchivierung in deutschen Bibliotheken, in: Zeitschrift für Bibliothekswesen und Bibliographie 62 (3/4), 2015, S. 172–183. Online: <<http://dx.doi.org/10.3196/1864295015623459>>.

2 Ebd. Zu den bereits seit langer Zeit im Bereich Webarchivierung aktiven Bibliotheken in Deutschland zählen die Deutsche Nationalbibliothek, die Landesbibliotheken in Baden-Württemberg, das Landesbibliothekszentrum Rheinland-Pfalz, die Saarländische Universitäts- und Landesbibliothek sowie die Bayerische Staatsbibliothek.

3 Vgl. Gesetz über Urheberrecht und verwandte Schutzrechte, § 15. Online: <https://www.gesetze-im-internet.de/urhg/_15.html>, Stand: 01.03.2022.

4 Vgl. Mutschler, Thomas: Zum Stand der kooperativen Webarchivierung in Thüringen. Gemeinsames Sammeln von landeskundlich relevanten Websites der Thüringer Universitäts- und Landesbibliothek und der Deutschen Nationalbibliothek, in: O-Bib. Das Offene Bibliotheksjournal 7 (4), 2020, S. 1–12, <<https://doi.org/10.5282/o-bib/5632>>.

5 Die Anpassungen des Gesetzes ermöglichen es den Landes- und Regionalbibliotheken, die das Recht zum Sammeln und Archivieren von elektronischen Pflichtexemplaren haben, grundsätzlich auch Webarchive in Eigenregie zu betreiben.

- Landesbibliothekszentrum Rheinland-Pfalz
- Thüringer Universitäts- und Landesbibliothek Jena
- Staats- und Universitätsbibliothek Hamburg
- Saarländische Universitäts- und Landesbibliothek

Diese Art der Bereitstellung erschwert Forscher*innen jedoch die Arbeit und schränkt die derzeitigen Nutzungsmöglichkeiten stark ein. Die Institutionen bemühen sich teilweise aber bereits um eine zusätzliche Einholung von Genehmigungen für die öffentliche Bereitstellung der Archivbestände im Web, sodass einige Teile dieser Webarchive dort auch zugänglich sind.

Neben Pflichtexemplargesetzen ermöglicht der 2018 in das deutsche Urheberrechtsgesetz aufgenommene Paragraph 60d die Vervielfältigung von urheberrechtlich geschützten Daten für Data Mining zu Forschungszwecken, wobei die Daten nur einem ausgewählten Personenkreis im Rahmen der wissenschaftlichen Forschung bzw. Dritten zur Überprüfung der erzielten Forschungsergebnisse zugänglich gemacht werden dürfen.⁶ Dies ermöglicht auch das Kopieren und die Archivierung von Websites im Kontext eines Forschungsprojekts. Nach Abschluss der Forschungsarbeiten muss die Zugänglichmachung beendet werden und die Daten können von der Forschungseinrichtung bzw. einer Gedächtniseinrichtung so lange archiviert werden, wie dies für wissenschaftliche Zwecke oder zur Prüfung der Forschung nötig ist. Eine Nutzung der Daten für andere Forschungsvorhaben ist nach derzeitigem Stand rechtlich noch nicht eindeutig geregelt.⁷

Außerhalb von konkreten Forschungsprojekten oder im Rahmen von Pflichtexemplarregelungen können Bibliotheken wie bereits erwähnt mit einem Genehmigungsverfahren arbeiten, bei dem vorab die Zustimmung der Websitebetreiber*innen für die Archivierung und Bereitstellung eingeholt wird. Anders als bei Pflichtexemplarregelungen ist es auf der Basis von Genehmigungen in der Regel möglich, die Daten öffentlich im Netz bereitzustellen. Aktuell ist die Bayerische Staatsbibliothek die einzige bekannte deutsche Einrichtung, die ausschließlich Webarchivierung auf diesem Weg betreibt. Insgesamt betrachtet stehen in Deutschland nach wie vor der Aufbau und Betrieb von Infrastrukturen für die Webarchivierung sowie der Sammlungsaufbau im Vordergrund. Die wissenschaftliche Nutzung von Webarchiven ist dagegen bislang gering und beschränkt sich in der Regel auf einen lesenden Zugriff (Close-Reading) auf die archivierten Ressourcen in der sogenannten Wayback-Machine, dem in den meisten Institutionen eingesetzten Viewer zur Darstellung von archivierten Websites.⁸

6 Vgl. Gesetz über das Urheberrecht und verwandte Schutzrechte, § 60d. Online: <https://www.gesetze-im-internet.de/urhg/_60d.html>, Stand: 01.03.2022.

7 Vgl. Kleinkopf, Felicitas; Jacke, Janina; Gärtner Markus: Text- und Data-Mining: Urheberrechtliche Grenzen der Nachnutzung wissenschaftlicher Korpora und ihre Bedeutung für die Digital Humanities, 2021, S. 6–10, <https://elib.uni-stuttgart.de/bitstream/11682/11462/1/Urheberrechtliche_%20Nachnutzbarkeit_TDM_Korpora_KleinkopfJackeGaertner.pdf>, Stand: 01.03.2022.

8 Eine Ausnahme hiervon ist das gemeinsame Projekt des Lehrstuhls für Digital Humanities der Universität Passau, des Jean-Monnet-Lehrstuhls für Europäische Politik der Universität Passau und des Münchener Digitalisierungszentrums der Bayerischen Staatsbibliothek. Hier wurden explorativ Methoden der digitalen Geisteswissenschaften, wie z.B. das Text- und Data-Mining, auf Webarchivbestände angewendet. Vgl. <<https://dh.uni-passau.de/webarchive-dh-dfg/>>, Stand: 01.03.2022.

Abseits von den derzeit konkret geltenden rechtlichen Rahmenbedingungen in Deutschland möchte dieser Beitrag zeigen, wie Angebote und Services von webarchivierenden Einrichtungen zur wissenschaftlichen Nutzung von Archivdaten aktuell aussehen bzw. zukünftig aussehen könnten. Dies soll im Folgenden anhand einiger aktueller Beispiele aus der Webarchivierungs-Community beleuchtet werden.

2. Praxis der Erschließung für die Nutzung von Webarchiven

Webarchive bieten in der Regel große Datenmengen mit einer Vielzahl von Dateiformaten. Traditionell unterstützen sie vor allem die qualitative Analyse einzelner Webangebote, indem sie archivierte Websites in einem speziellen Viewer (z.B. OpenWayback, pywb) anzeigen.⁹ Auch wenn sich mittlerweile erste Formen einer auf quantitativen Analyseverfahren beruhenden Auswertung von größeren Datenbeständen aus Webarchiven entwickelt haben, ist der primäre Zugang in den meisten Fällen nach wie vor der lesende Zugriff auf einzelne archivierte Webpages. Dabei beginnt die Suche einer*ines Nutzer*in nach einer archivierten Website entweder in einem Bibliothekskatalog bzw. Discovery-System oder in einem eigenen Portal zur Suche und Präsentation von archivierten Websites, wobei in der Regel nach der URL der Original-Website oder mit beschreibenden Metadaten (z.B. Autor*in, Titel, Medientyp) gesucht wird. In den Portalen ist neben der Suche oftmals auch ein Browsing in thematischen Kollektionen möglich, wobei eine Kurzbeschreibung der Kollektionen und teilweise auch der einzelnen archivierten Websites angeboten wird. Beim Aufruf des Links zu einer archivierten Website wird in vielen Fällen zunächst eine kalendarische Übersicht aller archivierten Zeitschnitte der Ressource angezeigt. Nach der Auswahl gelangt man im Viewer auf die archivierte Version der Website vom entsprechenden Datum und kann dort einzelne Webpages oder Inhalte aufrufen.

The screenshot shows a search results page for the URL <https://www.bsb-muenchen.de/>. The page displays a calendar view of search results for the period from January 1, 2010, to December 31, 2021. The results are organized by year and month, with each entry indicating the number of pages found and the specific date of the snapshot.

Search Results for Jan 1, 2010 - Dez 31, 2021											
2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
0 pages	2 pages	2 pages	2 pages	2 pages	5 pages	4 pages	4 pages	4 pages	4 pages	4 pages	4 pages
	Mar 27, 2011 *	Jan 1, 2012	Jan 1, 2013	Jan 7, 2014	Jan 9, 2015	Jan 1, 2016 *	Jan 1, 2017 *	Jan 1, 2018 *	Jan 1, 2019 *	Jan 1, 2020 *	Jan 1, 2021 *
	Jul 11, 2011	Jul 1, 2012	Jul 1, 2013	Jul 1, 2014	Feb 25, 2015	Jan 3, 2016 *	Jan 1, 2017 *	Jan 1, 2018 *	Jan 1, 2019 *	Jan 1, 2020 *	Jan 1, 2021 *
					Feb 28, 2015	Jul 2, 2016 *	Jul 10, 2017 *	Jul 1, 2018 *	Jul 1, 2019 *	Jul 1, 2020 *	Jul 1, 2021 *
					Jul 10, 2015	Jul 2, 2016 *	Jul 10, 2017 *	Jul 1, 2018 *	Jul 1, 2019 *	Jul 1, 2020 *	Jul 2, 2021 *
					Jul 10, 2015						

Abb. 1: Kalendarische Übersicht am Beispiel der von der Bayerischen Staatsbibliothek archivierten Zeitschnitte mit der URL <https://www.bsb-muenchen.de>

9 IIPC OpenWayback, <<https://github.com/iipc/openwayback/wiki>> und Webrecorder pywb documentation!, <<https://pywb.readthedocs.io/en/latest/>>, Stand: 01.03.2022.

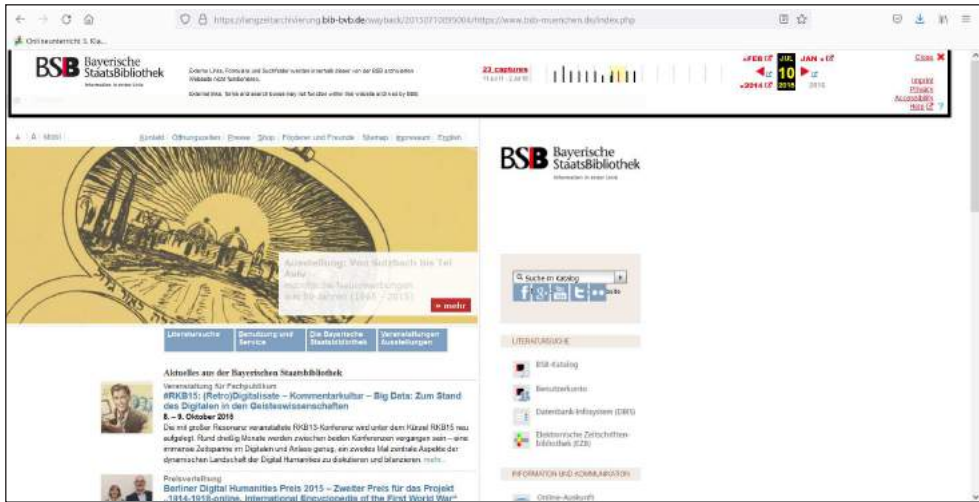


Abb. 2: Beispiel eines Zeitschnitts: die Website der Bayerischen Staatsbibliothek vom 10.07.2015

Viele Webarchive, beispielsweise die institutionellen Kollektionen des Dienstleisters Archive-It, das UK Web Archive, das dänische Webarchiv Netarkivet oder das portugiesische Webarchiv unter arquivo.pt bieten mittlerweile auch eine Volltextsuche an.¹⁰ Es kann dabei wahlweise nach einer spezifischen URL oder nach einem bzw. mehreren Schlüsselwörtern im Volltext des Gesamtbestandes der archivierten Websites oder in ausgewählten Kollektionen gesucht werden. Die gängige Verknüpfung von mehreren Suchbegriffen über Boolesche Operatoren und die Option nach einer Phrase zu suchen sind etabliert. Je nach Datenbestand und Indexierung können die Suchergebnisse bei der Volltextsuche über verschiedene Facetten, z.B. auf eine bestimmte Top-Level-Domain wie .com oder .net, einen Zeitraum oder Herausgeber*in eingeschränkt werden. Gelistet werden die Treffer bei den einzelnen Webarchiven unterschiedlich detailliert, idealerweise zumindest mit Angabe des Titels, der URL und dem Erfassungsdatum des Zeitschnitts, in einigen Fällen kommen beschreibende Metadaten oder kurze Textauszüge der archivierten Website hinzu. Trefferlisten können nach zeitlichen oder alphabetischen Kriterien sortiert werden, in der Regel wird eine Sortierung nach der Relevanz der Treffer angeboten. Wegen der hohen Redundanz von Texten in Webarchiven kann der gleiche Text nicht nur auf einer archivierten Website in unterschiedlichen Dateiformaten vorliegen, sondern auch jeweils in mehreren der archivierten Zeitschnitte vorkommen. Ein Suchbegriff wird in der Regel sehr viele und teilweise redundante Treffer produzieren und damit die Trefferliste aus Sicht der Nutzer*innen unübersichtlich machen. Insgesamt bleiben Ranking und Sortierung der Treffer einer Volltextsuche in Webarchiven für die Nutzer*innen bislang vielfach nicht nur intransparent, sondern auch wenig nutzerfreundlich im Vergleich zu herkömmlichen Internetsuchmaschinen.¹¹

10 Archive-It, <<https://archive-it.org/>>, UK Web Archive, <<https://www.webarchive.org.uk/>>, Netarkivet, <<https://www.kb.dk/en/find-materials/collections/netarkivet>>, Arquivo.pt, <<https://arquivo.pt/>>, Stand: 01.03.2022.

11 Vgl. Costa, Miguel: Full-Text and URL Search Over Web Archives, [2021]. Online: <<https://doi.org/10.48550/arXiv.2108.01603>>.

Die meisten Werkzeuge zur Auswertung von Webarchiven sind derzeit noch auf Textdaten ausgerichtet, andere Datentypen – wie Bilder, Audiodateien oder Videos – rücken erst allmählich in den Fokus.¹² In Webarchiven kann zumindest teilweise in der Facettensuche nach Datentyp gefiltert werden, vereinzelt wird auch eine Bildersuche anhand von Begriffen aus der übergeordneten Website oder anhand von Metadaten zum Ort der Aufnahme angeboten.¹³ Innovativere Suchverfahren, wie beispielsweise eine Bildähnlichkeitssuche, befinden sich noch in der Entwicklung und kommen noch nicht zum Einsatz.¹⁴

3. Webarchive als wissenschaftliche Quelle

Für eine wissenschaftliche Nutzung von Webarchiven ist es eine Herausforderung, dass die inhaltlichen Auswahlkriterien und technischen Parameter der Erstellung von thematischen Sammlungen oder Event-Crawls oftmals kaum oder gar nicht mehr nachzuvollziehen sind. Die Informationswissenschaftlerin Emily Maemura verweist deshalb im Zusammenhang mit der Webarchivierung und Korpusbildung von Sammlungen darauf, wie wichtig neben Hintergrundinformationen zu inhaltlichen Archivierungsentscheidungen und zur Qualitätssicherung die detaillierte Dokumentation der Crawls für eine spätere wissenschaftliche Nutzung ist.¹⁵

Neben der Beschreibung des organisatorischen Kontexts der Sammlung sollte sowohl eine inhaltliche als auch eine methodische Beschreibung mit Einzelheiten zu den ausgewählten Websites, zur technischen Infrastruktur, zu Zeiträumen und Dauer, Frequenz und Tiefe der Crawls zugänglich sein. Vor allem Lücken im Datenbestand sollten dokumentiert werden. Maemura plädiert zudem dafür, dass Wissenschaftler*innen nicht nur auf die WARC-Dateien oder auf die extrahierten Datensets Zugriff haben sollten, sondern auf alle (zusätzlich) erfassten Metadaten, alle Skripte, Reports bzw. relevanten Logs.¹⁶ Für eine Auswertung, die wissenschaftlichen Ansprüchen genügen soll, können darüber hinaus weitere technische Spezifikationen wichtig sein, beispielsweise ob die Websites jedes Mal vollständig oder inkrementell gecrawlt wurden, ob Medientypen beim Crawl von vornherein ausgeschlossen

12 Vgl. Huurdeman, Hugo C.; Ben-David, Anat; Sammar, Thaer: Sprint Methods for Web Archive Research, in: Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13, Paris 2013, S. 182–90. Online: <<https://doi.org/10.1145/2464464.2464513>>; Hockx-Yu, Helen: Access and Scholarly Use of Web Archives, in: Alexandria. The Journal of National and International Library and Information Issues 25 (1-2), 2014, S. 113–127. Online: <<https://doi.org/10.7227/ALX.0023>>; Adewoye, Tobi et al.: Content-Based Exploration of Archival Images Using Neural Networks, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, China 2020, S. 489–490. Online: <<https://doi.org/10.1145/3383583.3398577>>.

13 Vgl. Jackson, Andrew et al.: Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities, in: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries, Newark, New Jersey USA 2016, S. 103–110, hier S. 103–106. Online: <<https://doi.org/10.1145/2910896.2910912>>; Image Advanced Search, 2021, <<https://sobre.arquivo.pt/en/help/advanced-image-search/>>; Lauridsen, Jesper: SolrWayback 4.0 Release! What's It All about?, 2021, <<https://netpreserveblog.wordpress.com/2021/02/25/solrwayback-4-0-release-whats-it-all-about/>>, Stand: 01.03.2022.

14 Vgl. Adewoye et al.: Content-Based Exploration of Archival Images Using Neural Networks, 2020.

15 Vgl. Maemura, Emily; Worby, Nicholas; Milligan, Ian; Becker, Christoph: If these crawls could talk. Studying and documenting web archives provenance, in: Journal of the Association for Information Science and Technology, 69 (10), 2018, S. 1223–1233. Online: <<https://doi.org/10.1002/asi.24048>>.

16 WARC ist ein Containerformat, in dem verschiedene Datentypen, die im Rahmen der Webarchivierung erfasst werden, zusammen mit zugehörigen Metadaten gebündelt abgelegt werden können. Vgl. Schoger, Astrid; Weimer, Konstanze: Das Dateiformat WARC für die Webarchivierung, in: Nestor Thema 15, 2021. Online: <https://files.dnb.de/nestor/kurzartikel/thema_15-WARC.pdf>, Stand: 01.03.2022.

wurden oder ob und mit welchen Verfahren Inhalte dedupliziert wurden. Deutlich wird hier auch, dass die Beurteilung und Analyse einer Sammlung von Webarchiven bei den Wissenschaftler*innen sowohl ein vertieftes Verständnis von Webarchiven als Quelle als auch zusätzliches technisches Knowhow zur Klärung fachwissenschaftlicher Fragen voraussetzt.

4. Datenbereitstellung von Webarchivdaten jenseits der Webpräsentation

Seit Forschende in den Geisteswissenschaften zunehmend auch mit digitalen Analysemethoden arbeiten, wird an archivierende Einrichtungen der Wunsch herangetragen, die erfassten Webdaten in einer Form anzubieten, die sich für computergestützte Analysen eignet.¹⁷ Beispielhaft für die Auswertung großer Mengen von Webarchivdaten mit digitalen Methoden ist das Forschungsprojekt „Probing a Nation's Web Domain“, das die Domain-Crawls des dänischen Webarchivs Netarkivet aus der Zeit von 2005 bis 2015 untersucht. Anhand von Metadaten aus Crawls Logs, extrahierten Hyperlinks und Textinhalten wird versucht, die historische Entwicklung der dänischen Domain nachzuvollziehen. Die groß angelegte Studie einer gesamten nationalen Webdomain dient dazu, den Gesamtkontext zu beschreiben, in dem einzelne Webinhalte entstanden sind, um diese besser verstehen zu können.¹⁸

Untersuchungen wie „Probing a Nation's Web Domain“ setzen eine Datenbereitstellung voraus, die über die Anzeige im Viewer hinausgeht. Einige Institutionen haben auf diese Anforderung reagiert, indem sie kuratierte Datensets für ausgewählte Sammlungen zur Verfügung stellen. Diese Datensets sind vielfältig und folgen keinem einheitlichen Schema. Die Österreichische Nationalbibliothek beispielsweise stellt für ihre selektiven und Eventcrawls die Namen der Sammlungen und Listen der URLs der archivierten Websites zur Verfügung.¹⁹ Die Koninklijke Bibliotheek, die niederländische Nationalbibliothek, hat für ausgewählte thematische Sammlungen ebenfalls eine Liste der URLs erstellt, die zusätzliche Metadaten wie thematische Schlagwörter oder Angaben zu den Verfasser*innen enthält.²⁰ Für die thematische Sammlung JISC UK Web Domain Dataset (1996–2013) bietet das UK Web Archive verschiedene abgeleitete Datensätze an, darunter einen Geoindex mit den Postleitzahlen, die auf den archivierten Websites vorkommen.²¹ Die Library of Congress wiederum veröffentlicht Datensets der .gov-Domain, die jeweils 1.000 zufällig ausgewählte Dateien eines bestimmten Dateityps (Tabellen, PDF-Dateien, Audio-Dateien) enthalten. Des Weiteren werden dort Datensets mit Metadaten zu den US-amerikanischen Wahlen seit dem Jahr 2000, Bilder zu Memes aus dem Web Cultures Web

17 Vgl. Brügger, Niels: Digital Humanities and Web Archives. Possible New Paths for Combining Datasets, in: International Journal of Digital Humanities 2, 2021. Online: <<https://doi.org/10.1007/s42803-021-00038-z>>.

18 Brügger, Niels; Nielsen, Janne; Laursen, Ditte: Big Data Experiments with the Archived Web. Methodological Reflections on Studying the Development of a Nation's Web, in: First Monday 25 (3), 2020. Online: <<https://doi.org/10.5210/fm.v25i3.10384>>.

19 Webarchive Austria, 2022, <<https://labs.onb.ac.at/en/dataset/webarchive/>>, Stand: 01.03.2022.

20 Bode, Peter de; Geldermans, Iris; Teszelszky, Kees: Web collection NL-blogsfeer, 2021. Online: <<https://doi.org/10.5281/zenodo.4593479>>.

21 JISC UK Web Domain Dataset (1996–2010), 2013, <<https://doi.org/10.5259/UKWA.DS.2/1>>.

Archive sowie zum Irakkrieg bereitgestellt.²² Auch das Internet Archive hat mittlerweile Datensets für die wissenschaftliche Nutzung veröffentlicht.²³

In diesen abgeleiteten Datensets wird auf das außerhalb der Webarchivierung weitgehend unbekannte Containerformat WARC verzichtet und auf Standardformate wie JSON oder CSV zurückgegriffen. Diese Datensätze sind weniger umfangreich und damit leichter herunterzuladen und lokal zu verarbeiten als die ursprünglichen Containerdateien. Neben kuratierten Datensets bieten einzelne Institutionen auch Zugang zu den archivierten Daten oder Metadaten über Programmierschnittstellen. Ein einheitlicher Standard hat sich hier noch nicht etabliert, allenfalls die sogenannte CDX-API wird von mehreren Institutionen angeboten.²⁴ Diese API wird von Viewern wie der OpenWayback-Machine als Index genutzt, erlaubt es Nutzer*innen aber auch, Metadaten wie MIME-Typ, http-Status oder Umfang zu einer Ressource abzufragen und Ressourcen anhand dieser Metadaten zu filtern.

Einzelne Institutionen wie das portugiesische Webarchiv bieten zusätzlich eine eigene Programmierschnittstelle für ihre Volltextsuche. Über die Schnittstelle des Arquivo.pt können nicht nur die Metadaten der Suchergebnisse abgefragt werden, sondern auch die extrahierten Volltexte für weiterführende Textanalysen.²⁵ Die Funktion sich anhand der Suche im Volltext oder in den Metadaten ein eigenes Korpus zusammenzustellen und ihn für weitere Auswertungen zu exportieren ist auch Teil der SolrWayback, einer Webanwendung für die Suche und Anzeige von Webarchivdaten, die am dänischen Netarkivet entwickelt wurde. Anders als bei der Volltextextraktion des Arquivo.pt können Nutzer*innen wählen, ob sie die Daten in ihrer originalen Form im WARC-Format oder die daraus abgeleiteten Felder aus dem Index im CSV-Format exportieren wollen.²⁶ Am Netarkivet ist diese Exportfunktion aus rechtlichen Gründen für die Nutzer*innen jedoch nicht freigeschaltet: Die Suche über die SolrWayback soll Forschenden vielmehr dazu dienen, sich einen Überblick über die Daten zu verschaffen und erste Machbarkeitsstudien durchzuführen. Falls diese Erkundungen vielversprechend sind, können Forschende dann im direkten Austausch mit dem Bibliothekspersonal definieren, welche Daten sie für ihre Fragestellung benötigen, und in einer sogenannten Extraktionsvereinbarung den Umfang der Datenlieferung und die Nutzungsbedingungen individuell festlegen.²⁷

22 Web Archive Datasets, 2022, <<https://labs.loc.gov/work/experiments/webarchive-datasets/>>, Stand: 03.03.2022.

23 Vgl. Bailey, Jefferson: Early Web Datasets & Researcher Opportunities, 12.03.2021, <<http://blog.archive.org/2021/03/12/early-web-datasets-researcher-opportunities/>>, Stand: 01.03.2022.

24 Vgl. Blumenthal, Karl-Reiner: Access Archive-It's Wayback Index with the CDX/C API, 2022, <<https://support.archive-it.org/hc/en-us/articles/115001790023-Access-Archive-It-s-Wayback-index-with-the-CDX-C-API>>, Stand: 01.03.2022; CDXJ Server API, 2022, <https://pywb.readthedocs.io/en/latest/manual/cdxserver_api.html>. Stand: 01.03.2022.

25 Arquivo.Pt API, 2021, <<https://github.com/arquivo/pwa-technologies/wiki/Arquivo.pt-API>>, Stand: 01.03.2022.

26 Vgl. Egense, Thomas: SolrWayback 4.0 Release! What's It All about? Part 2, 2021, <<https://netpreserveblog.wordpress.com/2021/03/04/solrwayback-4-0-release-whats-it-all-about-part-2/>>, Stand: 01.03.22.

27 Research access to Netarkivet, 2022, <<https://www.kb.dk/en/find-materials/collections/netarkivet/research-access>>, Stand: 01.03.2022.

5. Computergestützte Analysewerkzeuge als Serviceangebote

Computergestützte Analysen setzen häufig ein hohes Maß an technischen Kenntnissen und verfügbaren Infrastrukturen voraus: „Yet when it comes to analysis, options are rather limited. Users are required to open up command line terminals, install software with complicated dependencies, have access to either powerful standing infrastructure or the ability to use cloud services such as Amazon Web Services or Microsoft Azure, if they want to work with web archives at scale beyond replay.“²⁸ Um hier Hürden für Nutzer*innen abzubauen, experimentieren archivierende Einrichtungen mit Benutzerschnittstellen, die über die Suche und Darstellung der Websites im Viewer hinausgehen und zusätzliche Analyseoptionen für Webarchivdaten bieten.

Ein verbreitetes Verfahren sind Frequenzanalysen für Textdaten. In der prototypischen Benutzeroberfläche SHINE des UK Web Archive können Nutzer*innen beispielsweise untersuchen, wie sich die Häufigkeit bestimmter Wortfolgen (N-Gramme) in einer Sammlung im Laufe der Zeit verändert.²⁹ Für eine qualitative Auswertung einzelner Vorkommnisse können sie sich das N-Gram zusätzlich in seinem ursprünglichen Kontext anzeigen lassen. In der SolrWayback wiederum können Nutzer*innen die häufigsten Begriffe einer Domain in einer Wortwolke visualisieren, um sich einen Überblick über die Inhalte zu verschaffen oder verschiedene Domains in dieser Hinsicht zu vergleichen.³⁰ Nicht immer ist jedoch klar erkennbar, wann solche Häufigkeitsanalysen tatsächlich sprachliche Besonderheiten aufzeigen und wann die beobachteten Phänomene durch die Art der Datenauswahl und -archivierung bedingt sind. Die Entwickler*innen von SHINE plädieren deshalb dafür, weitere Metadaten als Facetten in die Trendanalyse einzubeziehen, um die Daten besser eingrenzen und einordnen zu können.³¹

Neben Textdaten werden auch Daten zur Netzwerkstruktur bereitgestellt. Die SolrWayback beispielsweise bietet einen interaktiven Linkgraphen. Zur Darstellung besonders großer Netzwerke, die die Kapazitäten von Analysewerkzeugen wie Gephi sprengen, arbeiten die ägyptische Bibliotheca Alexandrina und das neuseeländische Webarchiv an der Webanwendung LinkGate.³² Durch ihre modulare und verteilte Struktur kann LinkGate auch umfangreiche Linkgraphen, wie sie im Rahmen der Webarchivierung erfasst werden, verarbeiten und interaktiv visualisieren. Netzwerkvisualisierungen laufen bei großen Datenmengen rasch Gefahr, unübersichtlich und uninterpretierbar zu werden, weshalb die Anwendung entsprechende Filteroptionen und verschiedene Layouteinstellungen bietet.

28 Ruest, Nick; Fritz, Samantha; Deschamps, Ryan et al.: From archive to analysis: accessing web archives at scale through a cloud-based interface, in: International Journal of Digital Humanities 2, 2021. Online: <<https://doi.org/10.1007/s42803-020-00029-6>>.

29 SHINE, 2022, <<https://www.webarchive.org.uk/shine>>, Stand: 01.03.2022.

30 Vgl. Lauridsen, Jesper: SolrWayback 4.0 Release! What's It All about?, 2021; Egense, Thomas: SolrWayback 4.0 Release! What's It All about? Part 2, 2021.

31 Vgl. Jackson, Andrew et al.: Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities, 2016.

32 Gephi ist eine quelloffene und kostenfreie Software für die Analyse und Visualisierung von Netzwerkdaten. Vgl. Gephi, 2022, <<https://gephi.org/>>, Stand: 01.03.2022. Zu LinkGate vgl. Eldakar, Youssef; Alsabbagh, Lana: LinkGate: Let's Build a Scalable Visualization Tool for Web Archive, 2020, <<https://netpreserveblog.wordpress.com/2020/04/23/linkgate-update/>>, Stand: 01.03.2022.

Statt ausgewählte Analysedienste anzubieten, entscheiden sich manche Webarchive dafür, ihre Datenschnittstellen und mögliche Analysen in Form von sogenannten Notebooks zu dokumentieren. Notebooks werden häufig als interaktive Dokumente beschrieben. Sie enthalten Programmierbefehle, die von erklärenden Textfeldern begleitet werden und von Nutzer*innen selbst ausgeführt und verändert werden können. Das Ausführungsergebnis wird ebenfalls unmittelbar in das Dokument eingebettet und erscheint als Text- oder grafische Ausgabe unterhalb des jeweiligen Codeblocks. Durch diesen besonderen Aufbau erfüllen Notebooks mehrere Funktionen: Sie dokumentieren Analysen und stellen die Ergebnisse anschaulich dar. Außerdem erlauben sie es, Auswertungen zu reproduzieren und anzupassen, um sie auf andere Daten zu übertragen. Dadurch fungieren sie als Tutorials für Einsteiger*innen, die sich anhand der interaktiven Beispiele eine Analyseverfahren oder einen Datensatz erschließen können. Über Cloud-Dienste wie Binder lassen sich Ausführungsumgebungen für Notebooks auch dynamisch bereitstellen, sodass Nutzer*innen über den Browser mit dem Notebook interagieren können. Dafür müssen jedoch auch die zu analysierenden Daten für den Cloud-Dienst zugänglich gemacht werden, was bei urheberrechtlich geschützten Daten problematisch ist.

Ein Beispiel für Notebooks in der Webarchivierung ist die GLAM-Workbench, die in einem Projekt des International Internet Preservation Consortium (IIPC) entstanden ist und Daten von verschiedenen Mitgliederinstitutionen über öffentliche Schnittstellen wie Memento oder CDX API abfragt.³³ Die Archives Unleashed Notebooks wiederum arbeiten mit Derivaten, wie sie das Archives Unleashed Toolkit aus WARC-Dateien erzeugt.³⁴ Mit dem Toolkit können beispielsweise Links oder Volltexte aus archivierten HTML-Seiten, aber auch Metadaten zu Binärdateien wie Bild- oder Audiodateien extrahiert und als abgeleitete Datensets gespeichert werden. Die Notebooks des Archives-Unleashed-Projekts und der GLAM-Workbench führen in die Nutzung der verschiedenen Datensets und APIs ein und untersuchen den Textinhalt sowie Metadaten wie Crawldatum, MIME Type und die Verteilung der Daten auf verschiedene Domains.

Im selben Forschungsprojekt wie die Archives Unleashed Notebooks ist auch die Archives Unleashed Cloud entstanden, die den Import und die anschließende Auswertung von Webarchivsammlungen aus dem Webarchivierungsdienst Archive-It ermöglichte. Das mittlerweile nicht mehr verfügbare kostenlose Angebot war besonders darauf ausgerichtet, technische Hürden abzubauen: So mussten Nutzer*innen keine Software lokal installieren und benötigten keine besonderen Rechnerkapazitäten, da die Auswertungen ausschließlich auf den Servern der Cloud stattfanden. Nutzer*innen konnten über ihre Browser Daten wie Textinhalte oder Links aus eigenen Webarchivsammlungen extrahieren, visualisieren und in begrenztem Umfang auch statistisch auswerten. Da die Cloud als quelloffene Software mit der entsprechenden Lizenz bereitgestellt wurde, konnte sie prinzipiell von anderen Institutionen angepasst und auf die eigenen Datenbestände angewendet werden. Mittlerweile ist die prototypische Instanz der Archives Unleashed Cloud nicht mehr zugänglich, die entwickelten Extraktions- und Auswertungsmöglichkeiten sollen aber in das Angebot des Archivierungsdienstleisters

33 Sherratt, Tim; Jackson, Andrew: GLAM-Workbench/web-archives, 2021. Online: <<https://doi.org/10.5281/zenodo.5584126>>.

34 Vgl. Ruest, Nick; Lin, Jimmy; Milligan, Ian; Fritz, Samantha: The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, China 2020, S. 157–166. Online: <<https://doi.org/10.1145/3383583.3398513>>.

Archive-It integriert und erweitert werden.³⁵ Die Archive-It Research Services bieten ihren Kund*innen bereits erste Analysedienste und extrahieren Links oder Eigennamen von Personen, Orten oder Organisationen (Named Entity Recognition, NER) aus den archivierten Daten.³⁶ Durch die Zusammenarbeit von Archive-It und dem Archives Unleashed Project sollen die verschiedenen Ansätze zusammengeführt und eine Plattform geschaffen werden, die von der Datenerfassung und -speicherung bis zur Analyse alle Dienste bündelt.

Insgesamt bleibt festzuhalten, dass die Möglichkeiten zur Auswertung von Webarchivdaten mit computergestützten Verfahren auch in den weltweit führenden Institutionen noch im Aufbau sind. Dabei hat sich ein experimenteller, iterativer und kollaborativer Ansatz etabliert, der sich mit dem Begriff des „Library Lab“ beschreiben lässt. Auf der IIPC Web Archiving Conference 2021 widmete sich ein ganzes Panel unter dem Titel „Supporting Research Use for Web Archives: A ‚Labs‘ Approach“ den Aktivitäten verschiedener Institutionen in diesem Bereich. Kennzeichnend für diesen Ansatz ist der enge Austausch zwischen Forschenden unterschiedlicher Fachrichtungen und Bibliotheksmitarbeiter*innen, seien es Kurator*innen oder IT-Fachkräfte. Diese interdisziplinären Kooperationen werden immer mehr als unerlässlich für die Arbeit mit digitalen Sammlungen im Allgemeinen und Webarchivdaten im Besonderen erkannt. Wichtig ist dabei der Austausch, von dem alle Beteiligten profitieren können, weil Labs fachliche, technische und sammlungsspezifische Expertise bündeln. Bibliotheksmitarbeiter*innen erwerben dort neue Kenntnisse im Bereich Big Data oder Data Science, Forschende greifen auf bibliothekarische Expertise zur Entstehung von Webarchivsammungen zurück oder stützen sich auf Empfehlungen für Werkzeuge und Methoden zur Sammlung und Auswertung von Webarchivdaten.

6. Fazit

Die aktuellen Aktivitäten in der Webarchivierung fokussieren in Deutschland vor allem auf die Entwicklung von Betriebsmodellen sowie den Bestandsaufbau. Die von den Institutionen angebotenen Nutzungsmöglichkeiten beschränken sich vielfach auf den lesenden Zugriff auf einzelne Zeitschnitte der archivierten Websites, dieser ist zum Teil durch rechtliche Vorgaben auf die Lesesäle der Einrichtungen beschränkt. Ein Blick auf die internationalen Akteur*innen zeigt, dass in den letzten Jahren verbesserte Nutzungsmöglichkeiten und spezielle Services für Webarchive aufgebaut, getestet und etabliert wurden. Dabei etabliert sich die Volltextsuche zunehmend als Standard und verbessert die Auffindbarkeit der Inhalte in Webarchiven deutlich. Weitergehende Ansätze für eine Nutzung im Kontext der Digital Humanities umfassen die Bereitstellung von kuratierten Datensets in gut zu verarbeitenden Größen und Formaten, spezielle Programmierschnittstellen (APIs), Analysetools z.B. für Frequenzanalysen sowie spezielle Notebooks, die Analyseverfahren beschreiben und für eigene Forschungszwecke nachgenutzt und angepasst werden können.

35 Vgl. Bailey, Jefferson: Archive-It and Archives Unleashed Join Forces to Scale Research Use of Web Archives, 2020, <<http://blog.archive.org/2020/07/28/archive-it-and-archives-unleashed-join-forces-to-scale-research-use-of-web-archives/>>, Stand: 01.03.2022.

36 Archive-It: Research Services, 2022, <<https://webarchive.jira.com/wiki/spaces/ARS/overview>> Stand: 01.03.2022.

Literaturverzeichnis

- Adewoye, Tobi et al.: Content-Based Exploration of Archival Images Using Neural Networks, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, China 2020, S. 489–490. Online: <<https://doi.org/10.1145/3383583.3398577>>.
- Altenhöner, Reinhard: Noch immer am Anfang? Stand und Perspektiven der Webarchivierung in Deutschland 2019, in: Fühles-Ubach, Simone; Georgy, Ursula (Hg.): Bibliotheksentwicklung im Netzwerk von Menschen, Informationstechnologie und Nachhaltigkeit. Festschrift für Achim Oßwald, Bad Honnef 2019, S. 237-250. Online <<https://nbn-resolving.org/urn:nbn:de:hbz:79pbc-opus-16232>>.
- Bailey, Jefferson: Archive-It and Archives Unleashed Join Forces to Scale Research Use of Web Archives, 28.07.2020, <<http://blog.archive.org/2020/07/28/archive-it-and-archives-unleashed-join-forces-to-scale-research-use-of-web-archives/>>, Stand: 01.03.2022.
- Bailey, Jefferson: Early Web Datasets & Researcher Opportunities, 12.03.2021, <<http://blog.archive.org/2021/03/12/early-web-datasets-researcher-opportunities/>>, Stand: 01.03.2022.
- Beinert, Tobias; Schoger, Astrid: Vernachlässigte Pflicht oder Sammlung aus Leidenschaft – Zum Stand der Webarchivierung in deutschen Bibliotheken, in: Zeitschrift für Bibliothekswesen und Bibliographie 62 (3/4), 2015, S. 172–183. Online: <<http://dx.doi.org/10.3196/1864295015623459>>.
- Blumenthal, Karl-Reiner: Access Archive-It's Wayback Index with the CDX/C API, 2022, <<https://support.archive-it.org/hc/en-us/articles/115001790023-Access-Archive-It-s-Wayback-index-with-the-CDX-C-API>>, Stand: 01.03.2022.
- Bode, Peter de; Geldermans, Iris; Teszelszky, Kees: Web collection NL-blogsfeer, 2021. Online: <<https://doi.org/10.5281/zenodo.4593479>>.
- Brügger, Niels: Digital Humanities and Web Archives. Possible New Paths for Combining Datasets, in: International Journal of Digital Humanities 2, 2021. Online: <<https://doi.org/10.1007/s42803-021-00038-z>>.
- Brügger, Niels; Nielsen, Janne; Laursen, Ditte: Big Data Experiments with the Archived Web. Methodological Reflections on Studying the Development of a Nation's Web, in: First Monday 25 (3), 2020. Online: <<https://doi.org/10.5210/fm.v25i3.10384>>.
- Costa, Miguel: Full-Text and URL Search Over Web Archives. Online: <<https://doi.org/10.48550/arXiv.2108.01603>>.

- Egense, Thomas: SolrWayback 4.0 Release! What's It All about? Part 2, 2021, <<https://netpreserveblog.wordpress.com/2021/03/04/solrwayback-4-0-release-whats-it-all-about-part-2/>>, Stand: 01.03.22.
- Eldakar, Youssef; Alsabbagh, Lana: LinkGate: Let's Build a Scalable Visualization Tool for Web Archive, 2020, <<https://netpreserveblog.wordpress.com/2020/04/23/linkgate-update/>>, Stand: 01.03.2022.
- Hockx-Yu, Helen: Access and Scholarly Use of Web Archives, in: Alexandria: The Journal of National and International Library and Information Issues 25 (1-2), 2014, S. 113-127. Online: <<https://doi.org/10.7227/ALX.0023>>.
- Huurdeman, Hugo C.; Ben-David, Anat; Sammar, Thaer: Sprint Methods for Web Archive Research, in: Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13, Paris 2013, S. 182-90. Online: <<https://doi.org/10.1145/2464464.2464513>>.
- Jackson, Andrew et al.: Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities, in: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries, Newark, New Jersey USA 2016, S. 103-106. Online: <<https://doi.org/10.1145/2910896.2910912>>.
- Kleinkopf, Felicitas; Jacke, Janina; Gärtner Markus: Text-und Data-Mining. Urheberrechtliche Grenzen der Nachnutzung wissenschaftlicher Korpora und ihre Bedeutung für die Digital Humanities, 2021, <https://elib.uni-stuttgart.de/bitstream/11682/11462/1/Urheberrechtliche_%20Nachnutzbarkeit_TDM_Korpora_KleinkopfJackeGaertner.pdf>, Stand: 01.03.22.
- Lauridsen, Jesper: SolrWayback 4.0 Release! What's It All about?, 2021, <<https://netpreserveblog.wordpress.com/2021/02/25/solrwayback-4-0-release-whats-it-all-about/>>, Stand: 01.03.2022.
- Maemura, Emily; Worby, Nicholas; Milligan, Ian; Becker, Christoph: If these crawls could talk: Studying and documenting web archives provenance. Journal of the Association for Information Science and Technology, 69 (10), 2018, S. 1223-1233. Online: <<https://doi.org/10.1002/asi.24048>>.
- Mutschler, Thomas: Zum Stand der kooperativen Webarchivierung in Thüringen. Gemeinsames Sammeln von landeskundlich relevanten Websites der Thüringer Universitäts- und Landesbibliothek und der Deutschen Nationalbibliothek, in: O-Bib. Das Offene Bibliotheksjournal 7 (4), 2020, S. 1-12. Online: <<https://doi.org/10.5282/o-bib/5632>>.

- Ruest, Nick; Fritz, Samantha; Deschamps, Ryan et al.: From archive to analysis: accessing web archives at scale through a cloud-based interface, in: International Journal of Digital Humanities 2, 2021. Online: <<https://doi.org/10.1007/s42803-020-00029-6>>.
- Ruest, Nick; Lin, Jimmy; Milligan, Ian; Fritz, Samantha: The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, China 2020, S. 157–166. Online: <<https://doi.org/10.1145/3383583.3398513>>.
- Schoger, Astrid; Weimer, Konstanze: Das Dateiformat WARC für die Webarchivierung, in: Nestor Thema 15, 2021, <https://files.dnb.de/nestor/kurzartikel/thema_15-WARC.pdf>, Stand: 01.03.2022.
- Sherratt, Tim; Jackson, Andrew: GLAM-Workbench/web-archives, 2021, <<https://doi.org/10.5281/zenodo.5584126>>.

Open Access zwischen Community-Arbeit und Sammlungsmanagement

Ein Praxisbericht

Tabea Lurk, Hochschule für Gestaltung und Kunst FHNW Basel

Zusammenfassung

Der Beitrag geht auf Effekte von Open Access an Kunsthochschulen ein. Dabei rückt er das Wechselspiel von Community-Arbeit und Sammlungsaufbau ins Zentrum der Betrachtung. Mithilfe von Fallbeispielen lassen sich Chancen und Herausforderungen spezifizieren. Der etablierten Konnotation von „Publish and Read“ wird so eine neue, hochschulinterne Sicht hinzugefügt: Indem Bibliotheken in die Community und die Zugänglichmachung von – in diesem Fall künstlerisch-gestalterischen – Inhalten und Werken investieren (publish), können sie einerseits die eigenen Bestände auf- und ausbauen sowie mit anderen teilen (read). Andererseits schaffen sie Mehrwerte, beispielsweise Referenzierbarkeit, die ihrerseits sowohl die Kunstschaffenden unterstützen als auch der guten wissenschaftlichen Praxis zuträglich sind. Die Relevanz stabil identifizierbarer, zitierbarer und nachnutzbarer Werke sowie deren Nutzungshistorie steigt nicht nur durch die Anforderungen der künstlerischen Forschung, von Promotionsprogrammen und Berufungsverfahren, sondern auch aufgrund bibliometrischer Ansätze, die eine quantitativen Vermessung von „Output“ intendieren.

Summary

This article examines the effects of open access at academies of fine art. It focuses on the interplay between community work and collection development. Based on case studies, chances and challenges are explained. This adds a new, university-centered view to the established connotations of “publish and read”: By investing in dedicated communities and making content (in this case artistic and creative) and works accessible (publish), libraries can not only develop and expand their own collections and share these with others (read). They also create added value, such as citability, which in turn supports the artists and designers and also serves good scientific practice. The relevance of stable persistent identifiers, citability, and re-usability of works, including the history of how they were used and referenced, is increasingly important. This is not only due to the requirements of artistic research, PhD programs, and appointment procedures, but also due to bibliometric methods for measuring performance and scholarly outcome quantitatively.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5830>

Autorenidentifikation:

Lurk, Tabea: GND: 1078962359; ORCID: <https://orcid.org/0000-0001-9848-8136>

Schlagwörter: Open Access; Künste; FAIR; Community Work

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Einleitung

Traditionellerweise bauen Kunsthochschulen (Spezial- und/oder Sonder-)Sammlungen mit künstlerisch-gestalterischen Inhalten, Modellen und/oder vielfältigen Materialien auf.¹ Sie tun dies, weil zum einen, wie auch in den übrigen Disziplinen, Anschauungsmaterialien für die Lehre und/oder Forschung benötigt werden.² Zum anderen fallen hausintern Inhalte an, die zu wertvoll oder relevant erscheinen, um einfach entsorgt zu werden.³ Darüber hinaus ist die Aufbereitung und Bewahrung künstlerischer (Daten-)Inhalte durch die Bibliotheken strategisch wichtig, weil es, im Unterschied zu klassischen Publikationsformen, kaum Anbieter für entsprechende (Werk-)Quellen gibt.⁴

Zwar werden in Zeiten der künstlerischen Forschung und sich öffnender Archive die Grenzen zwischen Werk und Publikation immer durchlässiger: Künstlerische Werke erheben zunehmend den Anspruch, (auch) als publizierter Output wissenschaftlich anerkannt zu werden.⁵ Dennoch landen diese häufig in Galerien, Museen oder im Privatbesitz, wo sie nur beschränkt zugänglich und selten systematisch online einseh- oder studierbar sind. Während die Finanzierungsmodelle der Onlineservices für Videokunst⁶ und die Streaming-Angebote für filmische Werke⁷ kaum zu den Bedürfnissen (Programmvielfalt, Stabilität des Angebots, Deckungsbreite) und Ressourcen der Kunsthochschulen passen, schränkt bei digitalen Werken und Daten das Urheberrecht die technische Option zur Kopie massiv ein.⁸

- 1 Vgl. Mühlenberend, Sandra (Hg.): Sammlungen an Kunsthochschulen. Speichern und vermitteln. Dresden 2020. Online: <<https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-710310>>. International bekannt sind z.B. die Sammlungen der Akademie der Künste Berlin (Online: <<https://www.adk.de/de/archiv/kunstsammlung/index.htm>>, Stand: 20.09.2022), der Hochschule für Bildende Künste Dresden (Online: <<https://www.hfbk-dresden.de/hochschule/einrichtungen/archiv-kustodie>>, Stand: 20.09.2022) oder der Akademie der bildenden Künste Wien (Online: <<https://www.akbild.ac.at/de/museum-und-ausstellungen>>, Stand: 20.09.2022). Aus der Schweiz sei der Verbund der Materialarchive (Online: <<https://materialarchiv.ch/>>, Stand: 20.09.2022) sowie z.B. der künstlerisch-forschende Ansatz des Textilarchivs (Online: <<https://silkmemory.ch/>>, Stand: 20.09.2022) erwähnt.
- 2 Zu den wissenschaftlichen Sammlungen in Deutschland vgl. Portal Wissenschaftliche Sammlungen. Online: <<https://portal.wissenschaftliche-sammlungen.de/>>, Stand: 20.09.2022. Zur FAIRness derselben vgl.: Beer, Anna: Anwendbarkeit der FAIR Data Principles auf digitale Sammlungen in deutschen Forschungsmuseen, Masterarbeit, Fachhochschule Potsdam, Potsdam 2021. Online: <<https://nbn-resolving.org/urn:nbn:de:kobv:525-25310>>.
- 3 Vgl. hierzu die Diskussion um die Entsorgung/Zerstörung der berühmten Fettecke („Raum 3“, 1982-1986) von Joseph Beuys in der Düsseldorfer Kunstakademie.
- 4 Mit Blick auf die Quellenlage kommt dem Alterswert häufig ein deutlich größerer Stellenwert zu, als z.B. der Neuerungsquote.
- 5 Der Recherche- und Sichtungsbedarf entsprechender Quellen steigt mithin.
- 6 Jenseits des freien Zugangs via UBU-Web (<<https://www.ubu.com/>>, Stand: 20.09.2022) wird Videokunst i.d.R. lizenziert und stark reguliert zugänglich gemacht – z.B. durch Electronic Arts Intermix (EAI New York, <<https://www.eai.org/>>, Stand: 20.09.2022), London Video Arts (LUX, <<https://lux.org.uk/tag/london-video-arts-lva>>, Stand: 20.09.2022), LI-MA (ehem. MonteVideo/Time Based, Amsterdam <<https://www.lima.nl/lima/>>, Stand: 20.09.2022), Vtape (Toronto, <<https://vtape.org/>>, Stand: 20.09.2022), Video Gallery Scan (Tokio, <<http://www.processart.jp/?mode=about>>, Stand: 20.09.2022), Stiftung IMAI (Düsseldorf, <<https://stiftung-imai.de/>>, Stand: 20.09.2022), Medien Kunst Netz (Zentrum für Kunst und Medientechnologie Karlsruhe, <<http://www.medienkunstnetz.de/konzept/>>, Stand: 20.09.2022), ARGOS (Brüssel, <<https://www.argosarts.org/>>, Stand: 20.09.2022) sowie teilweise die Memobase (Memoriav Bern, <<https://memobase.ch/de/start>>, Stand: 20.09.2022).
- 7 Vgl. Bohn, Anna: Themenschwerpunkt Video-Streaming. Editorial, in: Bibliothek Forschung und Praxis 44 (3), 2020, S. 309–312. Online: <<https://doi.org/10.1515/bfp-2020-2053>>.
- 8 Hinzu kommen zumeist komplexe konservatorische An- und Herausforderungen – insbesondere, wenn die Übernahme erst zeitverzögert und nicht, wie etwa bei Forschungsdaten, von Anfang an und kontinuierlich erfolgt.

Global betrachtet machen immer mehr Galerien, Museen, Archive und Bibliotheken (GLAM) Abbildungen oder Repräsentationen insbesondere ihrer gemeinfreien (historischen) Bestände im Zuge von sog. Open-GLAM-Engagements frei zugänglich.⁹ Dennoch gestaltet sich der Zugang zu Werken insbesondere der Gegenwartskünste aufgrund der weiterhin urheberrechtlich geschützten oder nicht klar als offen deklarierten, geschlossenen Bestände und restriktiven Nutzungsformen weiterhin komplex.¹⁰ Gebremst wird die Entwicklung von Open Access (OA) in den Künsten aber nicht nur durch die Nähe zum Kunstmarkt, zu Patent- bzw. Urheberrechtskonventionen des Designumfeldes (Produkte, Marken/Labels, Shows/Messen) und die veraltete Vorstellung: „Was nichts kostet, ist nichts (wert)“,¹¹ sondern auch wichtige OA-Anreizsysteme wie etwa steigende (Rück-)Zitationen greifen in den Künsten aufgrund fehlender Praktiken bisher kaum.¹² Es braucht also Anwendungsfälle, die Mehrwerte einer auf Sharing, Open- und FAIRness-basierten Praxis aufzeigen und jenem ‚Verlangen nach Sichtbarkeit‘ entgegenkommen, das in den Künsten besonders zu spüren ist.¹³

Die folgenden Überlegungen verdeutlichen daher, wie a) Wertschätzung und Sichtbarkeit ein Schlüssel zum OA-Erfolg sein können und b) das Instrumentarium der Bibliotheken gerade in den Künsten bisher ungenutzte Potenziale ausschöpfen könnte. Der Weg zur Quelle führt hier häufig weiterhin über Menschen, welche einer Nutzung meist persönlich zustimmen müssen, und nur selten Stakeholder wie etwa juristische Personen.

2. Kollaboration und Kooperation

Neben dem Aufbau umfangreicher Material- und Werksammlungen gehört die enge Vernetzung mit den fach- und themenspezifischen Communities zu den tragenden Säulen der Arbeit an Kunsthochschulen. Dies ist auch in den Bibliotheken spürbar: Als meist öffentliche Organisationseinheiten werden sie gern von Alumni, kulturellen Akteurinnen und Akteuren, Forschenden und Interessierten genutzt. Das hat auch deshalb Tradition, weil die Grenzen zwischen Hochschule, Kunst-/Design-Kosmos und Gesellschaft permeabel wirken, die Anstellungsverhältnisse oft prekär sind („unaffiliated knowledge workers“)¹⁴ und die Kunstschaffenden häufig unterschiedliche Aufgaben und Tätigkeiten

9 Vgl. McCarthy, Douglas; Wallace, Andrea: Survey of GLAM open access policy and practice, 2018 to present. Online: <<http://bit.ly/OpenGLAMsurvey>>, Stand: 20.09.2022, sowie der jüngst erschienene Lotse: Böhme, Moritz; Brand, Joachim; Gräff, Andreas u.a.: Der Lotse. Digitale Ressourcen der Kunstbibliothek, Staatliche Museen zu Berlin – Preußischer Kulturbesitz, Berlin 2022. Online: <<https://doi.org/10.5281/zenodo.6509870>>.

10 Die Herausforderungen z.B. von Prometheus (Online: <<https://www.prometheus-bildarchiv.de/>>, Stand: 20.09.2022) oder dem Bildarchiv Foto Marburg (Online: <<https://www.uni-marburg.de/de/fotomarburg>>, Stand: 20.09.2022) verdeutlichen, dass auch andere Disziplinen mit den bestehenden Zugangsbeschränkungen kämpfen.

11 Dieser Haltung steht die Praxis des Relevanz-Sammelns (z.B. mittels Counts) entgegen, die von sozialen Netzwerken (z.B. Freunde) bis zur wissenschaftlichen Publikation (Journal Impact Factor (JIF), CiteScore (Scopus), SCImago Journal & Country Rank, Impact per publication (IPP) etc.) die digitale Gesellschaft prägen (vgl. hierzu: Stalder, Felix: Kultur der Digitalität. Berlin 2017, S 151ff.).

12 Künstlerische Werke werden zumeist beschrieben, aber nicht zitiert. Sie sind weder in globale Verzeichnisse eingeschrieben noch referenzierbar. Es kann also auch nichts quantitativ gemessen oder ausgewertet werden.

13 Vgl. hierzu z.B. die sog. Vienna Declaration on Artistic Research (Online: <<https://cultureactioneurope.org/news/vienna-declaration-on-artistic-research/>>, Stand: 20.09.2022) sowie Bettel, Florian; Frank, Alexandra; Miljes, Wiebke: Sichtbarkeit, Sicherheit, Usability und Weiterverwendung Benutzer/innenorientierte FIS/CRIS-Entwicklung am Beispiel von „Portfolio/Showroom“, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 71 (1), 2018, S. 136–148. Online: <<https://doi.org/10.31263/voebm.v71i1.1989>>.

14 Vgl. hierzu: Brown, David J.: Information Needs and Habits of Unaffiliated Knowledge Workers, Masterarbeit, University College London, London 2016. Online: <<https://discovery.ucl.ac.uk/id/eprint/1522608>>, Stand: 20.09.2022.

in Personalunion vereinen. Durch den Kontakt innerhalb und außerhalb der Bibliothek etablieren sich kommunikative Netzwerke, die sowohl für die generationenübergreifende Weitergabe von Wissen als auch für die Bestandsentwicklung, so die These, eine wichtige Rolle spielen.¹⁵

Wendet man daher das Konzept der ‚Community of Practice‘ (CoP) auf die Bestandsarbeit an, werden Bibliotheken zu Informationsbrokern¹⁶ die durch ihre Serviceleistungen (Benutzung, Beratungen im Feld der Informations-, Medien- und Digitalkompetenz, Forschungsunterstützung etc.) die Kommunikation aufrechterhalten und/oder aktivieren.¹⁷ Darüber agieren sie ermöglichend und wirken mittelfristig steuernd: Indem sie flüchtige Informationen so aufbereiten, dass diese dauerhaft auffindbar, zugänglich, interoperabel und nachnutzbar (FAIR) werden, bringen sie Stabilität und Referenzierbarkeit (PID, Handle, DOI, ORCID, GND, FIAV, Wikidata-ID etc.) in einen Bereich, der zwar seit Jahrhunderten Gegenstand der Forschung und Vermittlung ist, jedoch ohne selbst systematisch über wissenschaftliche Kanäle (bibliografische Datenbanken, Wissenschafts-, Fachportale) recherchierbar und referenziert zu sein.

CoP, Bestandsentwicklung und FAIRness gemeinsam zu denken, bedeutet letztlich, dass das semantische Geflecht der Daten und Metadaten sowohl die Sammlungen kontextualisiert als auch die Community stabilisiert: Als Träger der Metadaten werden die Werke für die Künste zu idealerweise normierten Knotenpunkten, welche Personennamen und Werketitel über Publikations- bzw. Erscheinungsorte, Anlässe (z.B. Festivals) und Datierungen miteinander in Beziehung setzen.¹⁸ Diese Kontextinformationen erscheinen in den Künsten umso wichtiger, als Kunst häufig bildbasiert arbeitet und die textuelle Erschließung komplex sein kann.¹⁹ Mittels semantischer Interoperabilität tragen kontextuelle Metadaten dazu bei, textuelle Erschließungslücken zu überbrücken.²⁰

15 Eine künstlerische Reflexion auch negativ konnotierter ‚Seilschaften‘ findet sich z.B. in Marina Belobrovajas Projekt „kunstwette.ch“ (2010, Online: <<https://marinabelobrovaja.kleio.com/record/event/7>>, Stand: 20.09.2022).

16 Zur Funktion des Brokers in der CoP vgl. Zboralski, Katja: Wissensmanagement durch Communities of Practice. Eine empirische Untersuchung von Wissensnetzwerken, Wiesbaden 2007, insbesondere S. 258. Online: <<https://doi.org/10.1007/978-3-8350-9530-3>>.

17 Zu CoP-Handlungen für Spezialbibliotheken vgl. Kim, Jong-Ae: Integrating Communities of Practice into Library Services, in: Collaborative Librarianship 7 (2), 2015, S. 47–55, dort insbesondere der Massnahmenkatalog S. 52. Online: <<https://digitalcommons.du.edu/collaborativelibrarianship/vol7/iss2/2/>>, Stand: 20.09.2022.

18 Besonders reichhaltig wird das Gefüge durch jene Metadaten, die durch inhaltliche Aspekte wie verwendete Materialien, angewandte Techniken (Methoden), behandelte Themen, (historische) Werkbezüge etc. formalisierend erhoben werden. Mit Blick auf Datierungen etwa können Muda Mathis (1. Generation), Marion Ritzmann und Iris Ganz (je 2. Generation) erwähnt werden, die mit unterschiedlichen Funktionen (Künstlerin, Sammlerin, Dokumentaristin, Organisatorin) in den Beständen auftauchen. Sie waren vor ihrer Tätigkeit als Dozentinnen Studierende der HGK (bzw. der Vorgängereinrichtung). Derzeit studiert Lysann König (3. Generation) parallel zu ihrer Tätigkeit als kuratorische Leitung des Kaskadenkondensators Basel (kurz Kaskos) an der HGK, dessen videografischer Dokumentationsbestand sich derzeit im Übernahmeprozess durch die Mediathek befindet.

19 Längerfristig sind aufgrund des technischen Fortschritts non-textuelle Metadaten denkbar, welche z.B. Thumbnails, Sonogramme etc. nicht als Referenz anhängen, sondern direkt als Metadatum (Binärobjekt) nutzen.

20 Es mangelt nicht an großartigen Metadaten-Modellen (CIDOC, FRBR, LOM, CDWA) oder umfangreichen schemata (CERIF, Skosmos etc.). Einzig die Implementierung in der alltäglichen Praxis bzw. die Verständigung auf ein Modell (inkl. Crosswalks) fehlt.

3. Teilen von Wissen und Ressourcen

Ein einfaches Beispiel für die Semantik, die durch simple Korrelationen entsteht, verdeutlicht Abbildung 1: Im Hintergrundbild verzeichnet sind die Sammlungstitel (repräsentiert durch Ovale) der digitalen Performancebestände der Mediathek der Hochschule für Gestaltung und Kunst FHNW und alle in diesen Sammlungen nachgewiesenen Kunstschaffenden (repräsentiert durch Rechtecke). Linien weisen die Zugehörigkeiten aus, wobei Werke/Dokumentationen einer Person in unterschiedlichen Sammlungen vorkommen.²¹ Da die Datenmenge in Summe so groß ist, dass sich Rechtecke und Ovale bis zur Unkenntlichkeit verdichten, wurden von Hand einzelne Werkbeiträge (Vordergrundbild unten rechts) herausgegriffen.²²

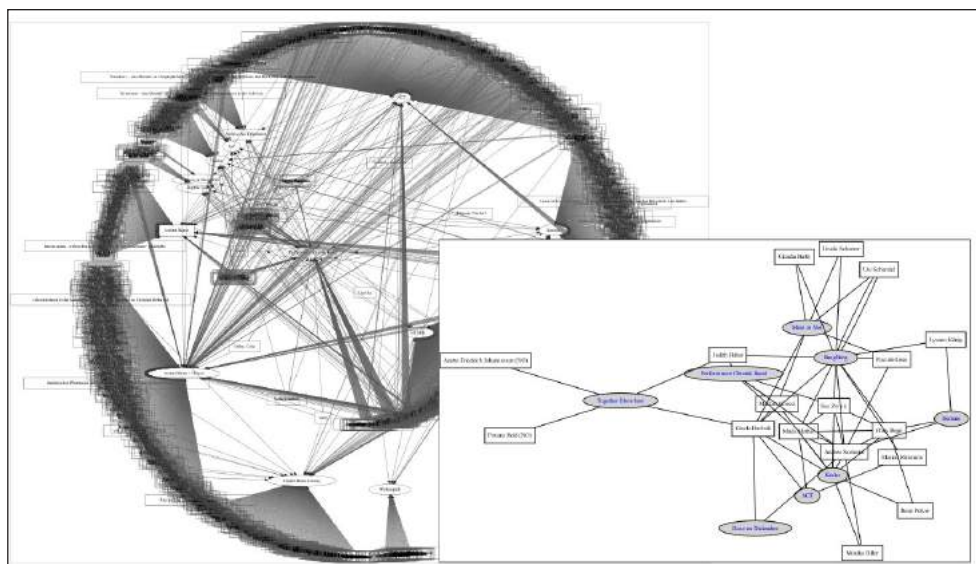


Abb. 1: Auswertung der digitalen Sammlungen zur Performancekunst der Mediathek der Hochschule für Gestaltung und Kunst FHNW Basel (Stand: Sommer 2022)

Für den vorliegenden Kontext ist von Interesse, dass hinter den Vernetzungslinien meist Schlüsselfiguren (CoP: Insider) stehen, die im Unterschied zu den gesammelten Kunstschaffenden (CoP:

21 Analog zum Nachweis der Publikationen von Forschenden in unterschiedlichen Fachzeitschriften steigt die Relevanz in den Künsten durch Mehrfachnennungen (Preise, Ausstellungs- und Festivalteilnahmen, Auftritte etc.). Dass mitunter auch ein und dasselbe Werk in unterschiedlichen Konvoluten nachgewiesen ist, wird hier insofern ausgeklammert, als dies zwar auch der kunstinternen Beurteilung zuträgt, aber ggf. nicht mehr notwendig mit der in dieser Fußnote benannten Zitation korreliert. Erneute oder Wiederaufführungen von Performances wären hingegen eigentlich zu Versionieren.

22 Tatsächlich sind die Künstlerinnen häufig mit mehreren Werken oder Beiträgen in einer Sammlung vertreten, was hier jedoch vernachlässigt wird.

periphere Mitglieder) die Geschehnisse z.B. kuratorisch strukturieren und lenken.²³ Mediatheksseitig schlägt sich dieses Insider-Engagement in (Material- und Werk-)Sammlungen nieder, die zur Übernahme angeboten oder kooperativ durch die Mediathek produziert werden. Mit Blick auf die Sonderbestände der Mediathek gehören René Pulfer,²⁴ Muda Mathis²⁵ und Reinhard Manz²⁶ zu den Schlüsselfiguren der ersten Generation.

Tabelle 1: Selektiver Überblick zu den Sondersammlungen der Mediathek der Hochschule für Gestaltung und Kunst FHNW Basel, den dahinterstehenden Schlüsselfiguren und ihrer Zugänglichkeit (Stand: Sommer 2022)

Titel	Schlüsselperson**	Zugänge
Datum	Beteiligte*	
Videosammlung IKUVID	René Pulfer**	Hochschulinterne Nutzung
1980er Jahre - Gegenwart	Institutsangehörige*	
Videowochen im Wenkenpark	René Pulfer**	https://mediathek.hgk.fhnw.ch/wenkenpark.php
1984, 1986, 1988	Reinhard Manz*	https://www.wikidata.org/wiki/Q111754595 SLSP-Sammlungs-ID: 6140810280005518
Performance Chronik Basel	Muda Mathis**	https://mediathek.hgk.fhnw.ch/pcb.php
1987-2006		https://www.wikidata.org/wiki/Q111803190 SLSP-Sammlungs-ID: 6140850260005518

23 Im vorliegenden Fall wäre weniger die Regelmäßigkeit der Kommunikation als vielmehr die Qualität des Austauschs (z.B. Bereitstellung von Daten) und die Stellung im jeweiligen Subsystem der Community (Expertentum, Kenner-schaft, Vernetzung) ausschlaggebend.

24 Der Kunsthistoriker und Künstler Réne Pulfer hat in den 1980er Jahren die Videoklasse der Hochschule für Gestaltung und Kunst FHNW (HGK) aufgebaut, ein Dreh- und Angelpunkt der Schweizer Medienkunst (inkl. Professionalisierung). Neben der nur hausintern zugänglichen Videokunstsammlung (IKU-Vid) war er 1984, 1986 und 1988 maßgeblich für die Realisierung der internationalen „Videowochen im Wenkenpark“ verantwortlich, deren Bänder 2016/2017, gefördert von Memoriav, unter Mitarbeit von Reinhard Manz an die Mediathek gelangten. Vgl. hierzu z.B. auch online: <<https://www.sikart.ch/kuenstlerinnen.aspx?id=9586812>>, Stand: 20.09.2022.

25 Die Künstlerin, Musikerin und langjährige Dozentin Muda Mathis ist eine der prägenden Figuren u.a. der schweizerischen Performancekunst. Mit dem Bestand „Performance Chronik Basel 1987-2006“ (PCB) hat sie eines der ersten öffentlich zugänglichen Konvolute der Mediathek bereitgestellt. Als Person, die großzügig ihr Wissen teilt und Menschen vernetzt, steht sie quasi hinter allen Performanceprojekten (vgl. Fußnote 29).

26 Der Filmemacher, Dokumentarist und Dozent Reinhard Manz begann seine filmische Laufbahn an der Schule für Gestaltung (Vorgängerinstitution der HGK in diesem Punkt). Neben der Digitalisierung von „Wenkenpark“ war er für die Bereitstellung der Lehroutputs der ehemaligen Grafikklassse (Sammlung „Film + Design 1968-1998“) verantwortlich. Hinzu kommt das 2022 begonnene Förderprojekt „Videonetzwerke der 1980er und 1990er Jahre“. Vgl. hierzu z.B. auch online: <<https://www.sikart.ch/kuenstlerinnen.aspx?id=4005855>>, Stand: 20.09.2022.

ACT Basel ab 2003	Muda Mathis** Marion Ritzmann*	Hochschulinterne Nutzung
Summe 2017 2017	Chris Regn**	https://mediathek.hgk.fhnw.ch/event/Summe2017 https://www.wikidata.org/wiki/Q111754212 SLSP-Sammlungs-ID: 6140850570005518
Archive des Ephemereren 2017-2019	Pascale Grau**	https://mediathek.hgk.fhnw.ch/event/archivedesephemereren https://www.wikidata.org/wiki/Q111754212 SLSP-Sammlungs-ID: 6140850570005518
Film + Design 1968-1998	Reinhard Manz** Peter von Arx*	https://mediathek.hgk.fhnw.ch/I/Film+Design SLSP-Sammlungs-ID: 6140850120005518
Kasko /I + II ab 2001	Chris Regn** Pascale Grau*	Im Übernahmeprozess inkl. Digitalisierung
6x2 Performance Duos 2020	Pascale Grau**	https://mediathek.hgk.fhnw.ch/I/6x2PerformanceDuos https://www.wikidata.org/wiki/Q112063619
Doce en Diciembre - Ozeanische Sache 2020	Andrea Saemann**	https://mediathek.hgk.fhnw.ch/event/archivedesephemereren https://www.wikidata.org/wiki/Q111754212 SLSP-Sammlungs-ID: 6140850570005518
Must or Not 2017-2019	Andrea Saemann** Chris Regn*	https://mediathek.hgk.fhnw.ch/I/MustOrNot SLSP-Sammlungs-ID: 6140850470005518

Together Elsewhere	Gisela Hochuli**	https://mediathek.hgk.fhnw.ch/event/togetherelsewhere
ab 2021	Pavana Reid*	https://www.wikidata.org/wiki/Q111754368 SLSP-Sammlungs-ID: 6140850550005518
Bang Bang. Performance	Chris Regn**	https://mediathek.hgk.fhnw.ch/bangbang/digitale-see/
Hi:stories revolved (2017-2019)	Muda Mathis* Andrea Saemann*	https://www.wikidata.org/wiki/Q111803182

In den letzten Jahren verändert sich die Motivation und die Art des Sammelns.²⁷ Zu den historischen Sammlungskonzepten einer ‚Quellensicherung für Recherchen‘ (z.B. bezogen auf die Werke von Pulfer, Manz oder Mathis) kommt die Relevanz des freien Zugangs via Internet hinzu. Exemplarisch für diese auf unbeschränkte Sichtbarkeit zielende Repräsentationsfunktion sei die Zusammenarbeit mit Pascale Grau,²⁸ Albena Mihaylova,²⁹ Chris Regn,³⁰ Andrea Saemann³¹ sowie Gisela Hochuli (s.u.) erwähnt. Als Künstler*innen und Kurator*innen haben sie die Mediathek als institutionelle Partnerin konsultiert, um ihre Werke, Werkreihen und Veranstaltungen dauerhaft zugänglich und künstlerisch nachnutzbar zu machen. Die Mediathek übernimmt in diesen Fällen z.T. nicht erst nachträglich die Daten, sondern wird bereits in die Erstellung (Erschließung, Live-Streaming mit Regie und Support etc.) einbezogen.

4. Community-basierte Zusammenarbeit und Teilhabe

Die Freigabe von Quellen folgt in diesem Fall auch einer intrinsischen Motivation: Orchesterwerke, Spielfilme, Theateraufführungen oder eben Performancekunst leben von Zusammenarbeit: Sie können als Werk häufig erst in Erscheinung treten und ex post dauerhaft zugänglich bzw. wahrnehmbar werden, wenn unterschiedliche Personen zuvor miteinander kooperiert haben. Abbildung 2 präsentiert eine Entscheidungshilfe, um trotz wachsender Anzahl an Beteiligten die steigende Komplexität der Rechtelage (z.B. im anschließenden Publikationsprozess) im Blick zu behalten.³² Die

27 Die Corona-Pandemie hat diese Entwicklung forciert, weil digitale Quellen und Kommunikationsformate sowohl neu bewertet als auch gefühlt zugänglicher wurden. Die Entwicklung zum Digitalen in den Künsten schließt hier teilweise ältere Strömungen innerhalb der medialen Künste an.

28 Auf die Kooperation mit Pascale Grau gehen die Teilbestände „Kasko I“ (derzeit geschlossen), „Archive des Ephemereren“ (2017-2019), „6x2 Performance-Duos“ (2020) sowie Einzelbeiträge in PCB und der Künstler*innen-Sammlung zurück.

29 Albena Mihaylova hat einen Großteil ihres Videoschaffens mit umfangreichen Kontextinformationen zur freien Publikation via Internet bereitgestellt (Online: <<https://mediathek.hgk.fhnw.ch/l/AlbenaMihaylova>>, Stand: 20.09.2022).

30 Neben dem aktuellen Kasko-Bestand war Chris Regn Drahtzieherin hinter allen Community-Projekten: „Summe 2017“, „Summe 2021“ (in Übernahme), „Feministisches Improvisatorium“ (2018), aktuell, gemeinsam mit Andrea Saemann, „Bang Bang“ (2022) sowie diversen Online-Events und Video-/Kunst-Streamings.

31 Andrea Saemann war federführend für die Zusammenarbeit zu „Must or Not“ (2020), „Doce En Diciembre“ (2021) die Rechteabklärung bei „Bang Bang“, weitere Teilbestände in Diskussion.

32 Die Abbildung sowie eine weitere Grafik, die erläutert, wie der Urheberrechtsstatus eines Werkes geklärt werden kann, sind publiziert in: Sykora, Sandra: Entscheidungsmatrix zur Nutzung von Bildmaterialien. Eine juristische

Mediathek empfiehlt daher, dem Komplexitätsargument folgend, durchgehend die Nutzung freier Creative-Commons-Lizenzen.³³

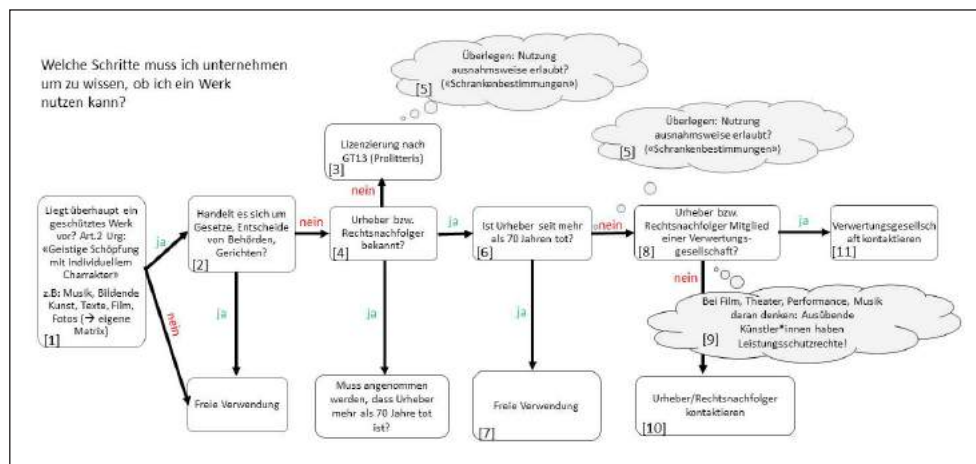


Abb. 2: Klärungsschritte vor der Nachnutzung von Bildmaterialien nach Sandra Sykora (CC BY 4.0)

Im Umfeld der Performancekunst wird dies gern aufgegriffen, weil nach der Aufführung meist nur noch Planungsunterlagen, Performancereликte und Dokumentationen in Bild und Ton als Quelle übrigbleiben.³⁴ Dokumentationen stammen i.d.R. von Zweiten, weil die Performenden während der Aufführung mit der jeweiligen Aktion beschäftigt sind und sich häufig nicht selbst aufzeichnen können. Auch wenn die Urheberrechte am aufgezeichneten Dokument technisch bei den Foto- oder Videografierenden liegen, werden die Performenden (darin quasi Darstellungsgegenstand) weiterhin als Urhebende der Performance genannt.³⁵ Die überlieferten Quellen werden zum Werk. Freie Lizenzen entspannen die Situation für alle Parteien und machen die Resultate für Forschung, Lehre und die Kunstschaffenden selbst dauerhaft nachnutzbar.³⁶

Momentaufnahme, Basel 2022. Online: <<https://www.doi.org/10.26254/med/6341>>.

33 Aufgrund von Kompetenzen und Ressourcen sowie fehlender Anwendungsfälle beschränkt sich die Empfehlung bisher auf Creative-Commons-Lizenzen.

34 Hinzu kämen Erinnerungen, Eindrücke, Beschreibungen etc. In den Beständen der Mediathek finden sich lediglich digital(isiert)e Dokumente, vor allem videografische Dokumentationen.

35 Die Dokumentierenden bleiben zu deren Leidwesen oft ungenannt. Da Festivals eine besondere Rolle in der Performancekunst spielen, kommen Interessen der Veranstaltenden meist hinzu.

36 Freie Lizenzen sind auch deshalb so wichtig, weil jüngere Ankaufmodelle vermehrt den Bereich der Performancekunst kommerzialisieren und so neue Zugangsmonopole (insbesondere für die Museen) entstehen (vgl. Performing Documentation in the Conservation of Contemporary Art, Revista de História da Arte Nr. 04, 2015. Online: <<https://institutodehistoriadaarte.com/revista-de-historia-da-arte-n-o-4-2007/>>, Stand: 20.09.2022).

5. Offenheit als Basis für Teilnahme und Teilhabe

Aspekte der Autorenschaft und Selektion sind im Zuge des Schaffens von Kunstwerken nicht nur Gegenstand der späteren Verhandlung und Publikation, sondern oft Inhalt der künstlerischen Aushandlungsprozesse. Anschaulich wird dies bspw. mit „Must or Not“. Als öffentliche Veranstaltung mit 25 Performances bestand das Ziel dieses Events in der Produktion hochwertiger Quellen, die von Anfang an für freie, bildbasierte Verhandlung und Zirkulation der Werke im Internet, der Lehre und Forschung gedacht waren.³⁷

Das Anlassformat entsprach den üblichen Events des Kaskadenkondensators (Kasko Basel), die Teilnehmenden waren alle seit etlichen Jahren in der Szene sowie untereinander bekannt. Speziell war hingegen die Explizitheit der Rahmenbedingungen, also die gemeinschaftliche, freie Publikation (doppelte Autorenschaft [Performende, Fotografierende] unter CC BY 4.0) via Mediathek,³⁸ die letztlich dazu führte, dass gleich mehrere Fotografinnen und Fotografen die Performances dokumentierten, statt sich abzuwechseln. Während künstlerisch und kunsthistorisch die jeweiligen Sichten und die Relativität vermeintlich objektiver Dokumentationsformen interessant wirken, war und ist die metadatentechnische Modellierung des Projekts mit Bordmitteln komplex und führt letztlich zu einer semantischen Verschiebung bei der Beurteilung der Quellen.

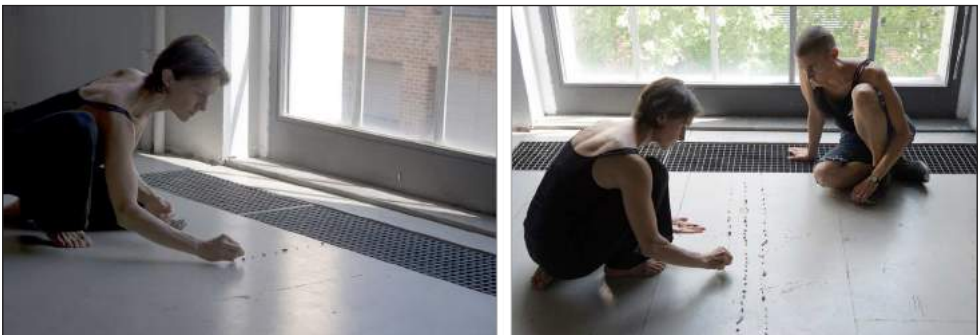


Abb. 3: Bildvergleich zu „insekten sterben“. Links: Ursula Scherrer, Markus Goessi (2020, CC BY 4.0); rechts: Ursula Scherrer, Ute Schendel (2020, CC BY 4.0)

Bestandsarbeit und Publikationsunterstützung nähern sich auch bei der monatlich stattfindenden Serie „Together Elsewhere“ (seit 2021) einander an, die von Gisela Hochuli (Performance Art Network Schweiz – PANCH) und Pavana Reid (Performance Art Bergen – PAB) als internationaler Austausch organisiert wird. Die Mediathek produziert die 30-minütigen Liveperformances, die gleichzeitig in

37 Neben wiederaufgeführten Performances (sog. Reenactments), zu denen die Künstler*innen bspw. kein dokumentarisches Material hatten, finden sich häufig verwendete performative Gesten, die für die Zukunft überliefert werden sollen. Die Quellen verdichten sowohl die Geschichte der Performancekunst als auch die eigenen Bestände. Bis zur Satzlegung wurden nicht zu allen Performances entsprechende Bilder bereitgestellt.

38 Im Rahmen von „Must or Not“ fand zudem ein Hackathon statt, bei dem Wikipedia-Einträge vorbereitet bzw. deren Eingabe erprobt wurden.

Norwegen, der Schweiz und dem digitalen Raum des Internets stattfinden.³⁹ An Doppelprojektionen erinnernd, werden in den CC-BY-4.0-Online-Ressourcen unterschiedliche Haltungen und Methoden der aktuellen Performancekunst wahrnehmbar, was insbesondere in Lehr- und Vermittlungskontexten dankend aufgenommen wird (Reusability im Sinne der FAIR-Prinzipien).

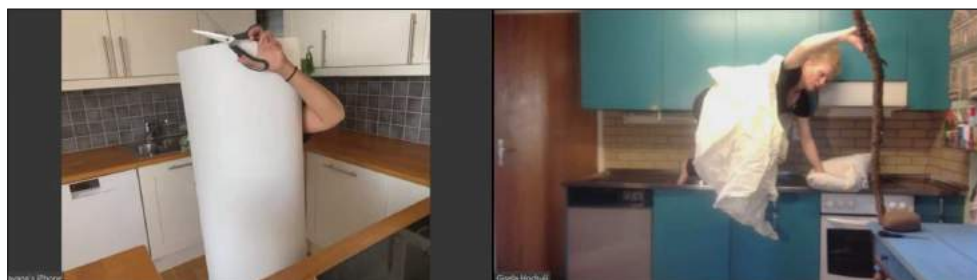


Abb. 4: Screenshot zu „Together Elsewhere I“. Pavana Reid, Gisela Hochuli (2021, CC BY 4.0)

Auch wenn es Aufgabe der Kunstgeschichte sein wird, die Quellen im Detail zu beurteilen, unterstreicht ein Blick auf die Metadaten die Relevanz der Inhalte.⁴⁰ Dieser Nachweis wird künftig noch deutlicher ausfallen, denn mit der Datenbasis zum Ausstellungsprojekt „Bang Bang. Translokale Performance Geschichte:n“ hat die Mediathek Zugang zu einem (Meta-)Datenpool geschaffen, der mit ca. 1.500 Werkdatensätzen die Geschichte der Performancekunst der Schweiz in einer bisher unzugänglichen Breite abbildet und erstmals recherchierbar macht.⁴¹ Als großangelegtes Community-Projekt haben die Kuratorinnen der Performance Chronik Basel⁴² die Performance-Szene dazu aufgerufen, ihre Werke und Projekte über ein auf diesen Anwendungsfall angepasstes Formular der Mediathek einzureichen. Die Einträge sollten während der Art Basel 2022 im Rahmen der „Bang Bang“-Ausstellung im Tinguely Museum gezeigt sowie auf Wunsch als historische Quelle dauerhaft via Mediathek zugänglich bleiben. Zudem wurde vorab festgelegt, dass die Metadaten via Internet frei einsehbar bleiben. Die digital eingereichten Referenzmaterialien dokumentieren neben einzelnen Werken, Festivals (z.B. Jahresausgaben), Dokumentensammlungen und historische Materialien (z.B. digitalisierte Kontaktabzüge) von über 2.000 internationalen Kunstschaffenden auf über 15.000 Dateien (Datenvolumen ca. 2 TB).⁴³

39 Der Service umfasst die Bereitstellung des virtuellen Raums via Zoom, die Betreuung (Regie, Support) der Künstlerinnen und Künstler, das Streaming via Vimeo, die Dokumentation und Publikation via Mediathekskatalog (Online: <<https://mediathek.hgk.fhnw.ch/amp/search>>, Stand: 20.09.2022) sowie dem Bibliothekskatalog (Swisscovery, in Arbeit). Die Mediathek erhält die Quellen kostenneutral.

40 Relevanz entsteht hier, wie auch andernorts im Digitalen, durch Verknüpfung.

41 Während der Ausstellung sind die Rechercestationen öffentlich einsehbar. Anschaulich wird die Vielfalt im Interface der sog. „Digitalen See“, von dem aus auch alle übrigen Zugänge aufgesucht werden können (Online: <<https://mediathek.hgk.fhnw.ch/bangbang/digitale-see/>>, Stand: 20.09.2022). Zur Erläuterung der unterschiedlichen Zugangsformen vgl. das Saalposter (Online: <<http://dx.doi.org/10.26041/fhnw-4198>>).

42 Die Vereinigung setzt die mit PCB (vgl. Fußnote 25) begonnenen Forschungstätigkeiten fort, wobei der Kreis der Aktiven erweitert wurde.

43 Auch wenn die (Meta-)Datenqualität variiert und mit den Maßstäben eines Community-Projektes zu bemessen ist, da mehrere historische wie digitale Generationen umspannt sind, ist ein unerwartet reichhaltiger Querschnitt entstanden.

6. Diskussion und Ausblick

Die Beispiele verdeutlichen, welche Potentiale die offene und offensive Datenpolitik einer Bibliothek entfalten kann. So wird die Mediathek der Hochschule für Gestaltung und Kunst FHNW immer häufiger bei jenen Projekten der Community von Beginn an mitgedacht, die versprechen datenintensiv, technisch komplex und/oder historisch relevant zu werden. Während die Mediathek dauerhaften Zugriff auf sonst unzugängliche Daten erhält, profitieren die verschiedenen Communities von einem reibungslosen, hochwertigen Datenmanagement, an dessen Ende nicht selten dauerhafte Sichtbarkeit (idealerweise durch eine CC-Lizenz) und Referenzierbarkeit (durch persistente Identifikatoren) steht.

Dennoch sollte die Freude über derartige Entwicklungen nicht über beachtliche Desiderate hinwegtäuschen. So gibt es gravierende organisatorische und infrastrukturelle Unterschiede an den deutschsprachigen Kunsthochschul(bibliothek)en: Während die meisten Einrichtungen in Österreich und der Schweiz über Publikationsserver oder entsprechende Zugänge verfügen⁴⁴ und ihre Outputs darüber teils regelmäßig rapportieren (müssen⁴⁵), sind derartige Infrastrukturen bei den deutschen Einrichtungen eher die Ausnahme (z.T. trotz Promotionsrecht). Auch variieren der Umgang mit Drittdaten stark, die häufig eher in Ausnahmefällen und projektbasiert Einzug in die entsprechenden Aufbewahrungssysteme finden, sowie die Bemühungen bspw. zum Kuratieren von Open Access.⁴⁶

Das hat nicht nur Effekte auf die Quellenlage, sondern auch auf Absprachen und die Implementierung von Nachhaltigkeit.⁴⁷ Die Ungleichzeitigkeit führt dazu, dass die einen noch nach Zugängen zu nachhaltigen Publikations- und Dokumentationsinfrastrukturen suchen, während sich andere z.B. fragen:

- Wie umgehen mit unpassenden Werktypologien und/oder Klassifikationen in Erschließungssystemen (DSpace, Fedora, Zenodo etc.) und den daran anschließenden, publikationsverwertenden Systemen (bibliothekarische Verbundsysteme,⁴⁸ Scopus, Web of Science)?
- Wie ungeklärte Rechtesituationen insbesondere bei Altbeständen angehen?
- Wie Clusterobjekte verwalten und erhalten, wenn pro Datensatz viele, große (> 2GB) und/oder Dateien mit heterogenen Formaten angehängt wurden
- Wie dynamische Kontexte stabil abbilden?

44 In der Schweiz sind die Kunsthochschulen, außer der Zürcher Hochschule der Künste (ZHdK), den jeweiligen lokalen Fachhochschulen als Departements eingegliedert. I.d.R. besteht so ein Zugang zu deren Publikationsservern und/oder Repositorien, auch wenn die Policies mitunter sehr restriktiv und wenig kunstkonform sind. ZHdK und HGK verfügen aufgrund der Wesenhaftigkeit der Publikationsserver über weitere Kunst- und Archivinfrastrukturen (Medienarchiv der Künste, Integrierter Katalog).

45 Vgl. hierzu die Wissensbilanzen der österreichischen Universitäten.

46 Nicole Döll hat z.B. damit begonnen, Open Access Publikationen mittels Zotero zu picken und über den Katalog der Galerie für Zeitgenössische Kunst Leipzig zugänglich zu machen (Online: <<https://katalog.gfzk.de/Search/Results?lookfor=volltext&type=Signatur>>, Stand: 20.09.2022).

47 Synergie-Effekte sowie sich wechselseitig stabilisierende Kräfte der jeweiligen Community entfallen. Der Erklärungsaufwand bzw. der Rechtfertigungsdruck (in house) steigen.

48 Auch wenn Resource Description and Access (RDA) und die in Bibliothekssystemen eingebundenen Normdaten eine hochwertige Erschließung künstlerischer Werkformen ermöglichen könnte, zeigt z.B. ein Blick auf die Kataloge der Österreichischen Bibliothekenverbund und Service GmbH (OBVSG), der Swiss Library Service Platform (SLSP) sowie des FIDs Darstellende Künste, dass insbesondere (künstlerische) Werke und Sammlungsobjekte in Dublin Core (statt MARC21) erfasst werden. Bei SLSP hat dies u.a. den Verlust bzw. Verzicht auf Normdaten zur Folge.

- Wie Sichtbarkeit und Referenzierbarkeit schaffen, um den (Gegen-)Wert der PIDs auch auszuschöpfen?⁴⁹

Gleichzeitig produzieren insbesondere kollaborativ ausgelegte Bereiche der Künste immer häufiger frei nachnutzbare Quellen, deren hausinterne Zweitablage (i.d.R. 10 Jahre) allein schon aufgrund der hohen Dynamiken an Kunsthochschulen, dem Kunstmarkt sowie den weiteren digitalen Ökonomien sinnvoll erscheint.⁵⁰ Dennoch kann mit Blick auf die digitalen Bestände der Kunsthochschulbibliotheken der Eindruck entstehen, dass einerseits Vernetzung und Zusammenarbeit noch ausstehen,⁵¹ andererseits könnte der zeitliche Verzug auch Vorteile bergen. So könn(t)en kooperativ aufgebaute DM-, Speicher- und Publikationsinfrastrukturen dort von Anfang an kunstkonform(er) ausgerichtet werden, wo aktuelle Lösungen aus Sicht der Künste zu selten die Bedürfnisse der künstlerischen Fächer abbilden.⁵²

Ein Blick auf die sog. FAIR Data Principles zeigt weitere, wenig genutzte Potentiale. Die Prinzipien spezifizieren auf einer konzeptionellen Ebene, wie die Auffindbarkeit (Findability), Zugänglichkeit (Accessibility), der Datenaustausch (Interoperability) und die Nachnutzbarkeit (Reusability) digitaler Daten und Publikation verbessert sowie dauerhaft gesichert werden kann.⁵³ Andersherum ausgedrückt, lassen sich aus den FAIR Data Principles Hilfestellungen für den systematischen Aufbau der Bestände, ihrer Pflege und Vernetzung ableiten, die zugleich naheliegend und zumindest weitgehend realisierbar wirken: Vom Prinzips der Auffindbarkeit (FAIR:F) sind im vorliegenden Kontext die Absätze eins (FAIR:F1) bezüglich PIDs und vier (FAIR:F4) zwecks Nachweissystemen und Katalogen besonders relevant. Die dauerhafte öffentliche Zugänglichkeit (FAIR:A) dieser Verzeichnisse wird in Absatz 1 (FAIR:A1) des zweiten Prinzips gefordert. Nicht zuletzt rät Absatz 1.3 zur Nachnutzbarkeit (FAIR:R) eine breite Verankerung der verwendeten Begrifflichkeiten, Vokabulare und Systematiken innerhalb der jeweiligen den Fachgesellschaften und Zielgruppen.

Wo Interoperabilität aufgrund der nicht-textuellen und wenig harmonisierten Metadatenlage derzeit komplex scheint,⁵⁴ wirken solche Ansätze vielversprechend, die sie aus dem Umfeld der

49 Auch wenn die Vernetzung und der Aufbau für (Rück)Zitationen noch nicht implementiert sind, wirkt es realistisch, dass der Nachweis von Werken in wissenschaftlichen (vertrauenswürdigen) Repositorien längerfristig äquivalent zu den bestehenden Kompetenzausweisen (Preise/Stipendien, Ausstellungsnachweise, Teilnahme an Messen/Festivals etc.) bewertet werden wird.

50 Vgl. hierzu die Zusammenfassung (insbesondere Argumentation) von: Stäcker, Thomas. Die Sammlung ist tot, es lebe die Sammlung! Die digitale Sammlung als Paradigma moderner Bibliotheksarbeit, in: Bibliothek Forschung und Praxis 43 (2), 2019, S. 304–310. Online: <<https://doi.org/10.1515/bfp-2019-2066>>.

51 Die fachliche Kommunikation findet vor allem über die Arbeitsgemeinschaft für Kunst- und Museumsbibliotheken (AKMB) statt. Bei Kooperations- und Infrastrukturprojekten werden Kunsthochschulen jedoch meist auf die geisteswissenschaftlich organisierten FIDs (z.B. Darstellende Künste) oder NFDIs (4culture, 4objects etc.) verwiesen. In der Fachsystematik der DFG von 2021 fehlen die Künste (vgl. Online: <https://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher/index.jsp>).

52 Lurk, Tabea. Datenmanagement an Kunsthochschulen. Anforderungen zwischen Forschung, Vermittlung und Archiv, Masterarbeit, Technische Hochschule Köln, Köln 2022, Kapitel 4. Online: <<https://nbn-resolving.org/urn:nbn:de:hbz:79pbc-opus-18709>>.

53 Die GoFAIR-Website präsentiert die Rahmenbedingungen zur FAIRness online <<https://www.go-fair.org/fair-principles/>>, Stand: 20.09.2022.

54 Bisher wirkt die Anzahl der auffindbaren OAI-PMH-Schnittstellen sowie die Nachweise von Kunsthochschulrepositorien etc. in re3data und OpenDOAR überschaubar, vgl. Lurk, Tabea: Künstlerische Forschung und Open Access?

Datenpublikationen stammen und Kontextinformationen wertschätzen. Diese fordern, wie übrigens auch wissenschaftliche Datenmanagementpläne, Informationen zur Datenlage, zu Erhebungs- und Auswertungsmethoden sowie zu Forschungsprozessen. Das lässt sich insofern trefflich auf die Künste übertragen, als die verwendeten Materialien, (Herstellungs-)Techniken und Methoden sowie Gestaltungs- bzw. Schaffensprozesse immer schon rege diskutiert werden und den Beteiligten oft sehr viel greifbarer wirken als die üblichen Beschreibungsformen.⁵⁵ Da nicht zuletzt Relevanz durch Verknüpfung entsteht, dürften in Zukunft Kontextinformationen, wie insbesondere Angaben zu Festivals, Ausstellungsteilnahmen (Ortsangaben, Institutionen/Veranstalter) oder Preise, Raum-Zeit-Korrelationen einzelner Kunstschaffender, Kollektivteilnahmen, Aufgabenfelder (innerhalb z.B. der Kunstereignisse) an Bedeutung gewinnen. Sie können dort semantische Lücken schließen, wo die Angaben formalisierbar und harmonisierbar sind. Trotz beachtlicher Baustellen lässt sich aus dem Feld der Künste mithin optimistisch in Richtung OA blicken.

Literaturverzeichnis

- Beer, Anna: Anwendbarkeit der FAIR Data Principles auf digitale Sammlungen in deutschen Forschungsmuseen, Masterarbeit, Fachhochschule Potsdam, Potsdam 2021. Online: <<https://nbn-resolving.org/urn:nbn:de:kobv:525-25310>>.
- Bettel, Florian; Frank, Alexandra; Miljes, Wiebke: Sichtbarkeit, Sicherheit, Usability und Weiterverwendung Benutzer/innenorientierte FIS/CRIS-Entwicklung am Beispiel von „Portfolio/Showroom“, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 71 (1), 2018, S. 136–148. Online: <<https://doi.org/10.31263/voebm.v71i1.1989>>.
- Bohn, Anna: Themenschwerpunkt Video-Streaming – Editorial, in: Bibliothek Forschung und Praxis 44 (3), 2020, S. 309–312. Online: <<https://doi.org/10.1515/bfp-2020-2053>>.
- Brown, David J.: Information Needs and Habits of Unaffiliated Knowledge Workers, Masterarbeit, University College London, London 2016. Online: <<https://discovery.ucl.ac.uk/id/eprint/1522608>>, Stand: 20.09.2022.
- Böhme, Moritz; Brand, Joachim; Gräff, Andreas u.a.: Der Lotse. Digitale Ressourcen der Kunstbibliothek, Staatliche Museen zu Berlin – Preußischer Kulturbesitz, Berlin 2022. Online: <<https://doi.org/10.5281/zenodo.6509870>>.

Übersicht zu Publikationsoptionen und praktischen Herausforderungen, in: Bibliothek Forschung und Praxis 45 (3), 2021, S. 517–526. Online: <<https://doi.org/10.1515/bfp-2021-0038>>.

55 Vgl. hierzu aus dem Bereich der Kunstdokumentation die Verzeichnung von Ausstellungsdaten und die Vernetzung mit Kunstinstitutionen, wie sie z.B. von European-art.net (EAN, <<https://european-art.net/>>, Stand: 20.09.2022), oder mit Blick auf die Medienkunst und ihre Festivals bspw. das International Symposium on Electronic Art Barcelona (ISEA, <<https://isea-archives.siggraph.org/>>, Stand: 20.09.2022) seit vielen Jahren praktiziert werden. Im musealen Umfeld hat sich ferner der besondere Stellenwert von Material und Technik bspw. in die Metadatensets der Sammlungsmanagementsysteme eingeschrieben.

- Kim, Jong-Ae: Integrating Communities of Practice into Library Services, in: Collaborative Librarianship 7 (2), 2015, S. 47–55. Online: <<https://digitalcommons.du.edu/collaborativelibrarianship/vol7/iss2/2/>>, Stand: 20.09.2022.
- Lurk, Tabea: Datenmanagement an Kunsthochschulen. Anforderungen zwischen Forschung, Vermittlung und Archiv, Masterarbeit, Technische Hochschule Köln, Köln 2022. Online: <<https://nbn-resolving.org/urn:nbn:de:hbz:79pbc-opus-18709>>.
- Lurk, Tabea: Künstlerische Forschung und Open Access? Übersicht zu Publikationsoptionen und praktischen Herausforderungen, in: Bibliothek Forschung und Praxis 45 (3), 2021, S. 517–526. Online: <<https://doi.org/10.1515/bfp-2021-0038>>.
- McCarthy, Douglas; Wallace, Andrea: Survey of GLAM open access policy and practice, 2018 to present. Online: <<http://bit.ly/OpenGLAMsurvey>>, Stand: 20.09.2022.
- Mühlenberend, Sandra (Hg.): Sammlungen an Kunsthochschulen. Speichern und vermitteln. Dresden 2020. Online: <<https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-710310>>.
- Performing Documentation in the Conservation of Contemporary Art, Revista de História da Arte Nr. 04, 2015. Online: <<https://institutodehistoriadaarte.com/revista-de-historia-da-arte-n-o-4-2007/>>, Stand: 20.09.2022.
- Stäcker, Thomas. Die Sammlung ist tot, es lebe die Sammlung! Die digitale Sammlung als Paradigma moderner Bibliotheksarbeit, in: Bibliothek Forschung und Praxis 43 (2), 2019, S. 304–310. Online: <<https://doi.org/10.1515/bfp-2019-2066>>.
- Stalder, Felix: Kultur der Digitalität. Berlin 2017.
- Sykora, Sandra: Entscheidungsmatrix zur Nutzung von Bildmaterialien. Eine juristische Momentaufnahme, Basel 2022. Online: <<https://www.doi.org/10.26254/med/6341>>.
- Zboralski, Katja: Wissensmanagement durch Communities of Practice. Eine empirische Untersuchung von Wissensnetzwerken, Wiesbaden 2007. Online: <<https://doi.org/10.1007/978-3-8350-9530>>.

LIBER Architecture Group Seminar “Designing for Learning and Scholarship: a challenge for librarians, architects and all” in Luxemburg

Gefühlt die erste Tagung in Präsenz und die erste Dienstreise nach den starken Einschränkungen durch den Corona-Schutz-Betrieb! Nachdem das traditionsreiche Seminar der LIBER Architecture Group pandemiebedingt 2020 und 2021 ausfallen musste, konnte es endlich im April dieses Jahres nachgeholt werden, noch dazu als 20. Jubiläumsausgabe der Tagung.¹

Durch die Erfahrungen der Bibliotheken im Verlauf der Corona-Pandemie ist in vielerlei Hinsicht eine neue Situation entstanden. Einerseits wurden die Bibliotheken gerade von den Studierenden schmerzlich vermisst, andererseits stellen Bibliotheken fest, dass sie sich im dauerhaft hybriden Alltag neu aufstellen müssen. Das Tagungsprogramm entsprach weitgehend dem Planungsstand von 2020, doch die Diskussion der Konsequenzen der Pandemieerfahrung lief als Subtext gerade beim informellen Austausch immer mit. Es ging um die Grundfragen: Wie entwickelt sich Bildung? Wie entwickelt sich soziale Interaktion? Welche Technologien und welche Raumangebote sind dafür maßgeblich?²

Das 20. Seminar der LIBER Architecture Group (27.-29.04.2022) bot Anlass, neue Bibliotheken und Lernzentren kennenzulernen, die seit der letzten Tagung 2018 in Wien eröffnet wurden. Leider konnten dabei nicht alle interessanten Neubauten berücksichtigt werden, so fehlten zum Beispiel Präsentationen zum markanten Neubau der Universitätsbibliothek Limerick oder zur umfangreichen Sanierung der Staatsbibliothek zu Berlin im Haus Unter den Linden. Konkrete Bau- und Ausstattungsbeispiele verbanden sich mit der generellen Reflexion, wie sich Bibliotheken verändern (müssen), damit sie als Ort und Angebot relevant bleiben. Wie können sie ihre jeweiligen Communities unterstützen bei zukunftsorientierten Veränderungen und angesichts der großen Herausforderungen unserer Zeit, die nach innovativen, nachhaltigen und sozialen Lösungen verlangen?

Besichtigt wurden zwei auf ihre jeweilige Weise sehr eindrucksvolle Luxemburger Bibliotheken: Zum einen das Luxembourg Learning Center als Zentralbibliothek der jungen Luxemburger Universität. Sie wurde 2003 durch Fusion von drei akademischen Bildungsstätten gegründet und befindet sich heute zu großen Teilen in Belval / Esch-sur-Alzette auf dem Gelände einer ehemaligen Stahlfabrik, deren Betrieb 1997 eingestellt wurde und deren Überreste großartig in den Campus integriert sind. Das Gelände wurde komplett saniert, es entstanden hier neben dem Universitätscampus Verwaltungs- und Kultureinrichtungen, eine Shopping Mall sowie Gastronomiebetriebe. Der Kontrast zwischen den Zeugnissen des industriellen Zeitalters und der neuen Architektur für post-industrielle Wissensarbeit ist überaus reizvoll.

1 Vgl. <<https://libereurope.eu/article/20th-liber-lag-architecture-working-group-seminar-2022-designing-for-learning-and-scholarship-a-challenge-for-librarians-architects-and-all/>>, Stand: 01.08.2022.

2 Vgl. eine Nachlese der Tagung von Oliver Blake: <<https://libereurope.eu/article/20th-liber-lag-architecture-working-group-seminar-2022-designing-for-learning-and-scholarship-a-challenge-for-librarians-architects-and-all/>>, Stand: 01.08.2022.



Abb. 1: Impression vom Campus Belval. Foto: Alice Rabeler



Abb. 2: Fassade des Belval Learning Center. Foto: Alice Rabeler

Hier fand auch die Tagung statt, die vom Team des Luxembourg Learning Center hervorragend organisiert wurde. Vor Ort waren ca. 120 Teilnehmende aus 14 europäischen Ländern. Die Vorträge wurden zudem für die mehr als 90 Online-Teilnehmenden live gestreamt und umgekehrt auch nicht anwesende Vortragende in das Auditorium zugeschaltet. So konnte der Eröffnungsvortrag von Alexi Marmot aus London und der Vortrag von Michaela Sheehan aus Australien übertragen werden. Die Vorträge sind über die Webseite der LIBER Architecture Group zugänglich.³

Vor Ort konnten die Teilnehmenden auch die neue Nationalbibliothek im Luxemburger Stadtteil Kirchberg besichtigen. Der Bau der Architektinnen und Architekten Bolles Wilson wurde im Oktober 2019 fertiggestellt.⁴ Der Besichtigung vorangestellt war eine Podiumsdiskussion unter Anwesenheit der früheren Direktorin Monique Kieffer und der Architektin Julia Bolles-Wilson.

3 Vgl. <<https://libereurope.eu/working-group/liber-architecture-working-group-lag/>>, Stand: 01.08.2022.

4 Details zum neuen Gebäude siehe <<https://www.degruyter.com/document/doi/10.1515/abitech-2022-0002/html?lang=de>>, Stand: 01.08.2022.

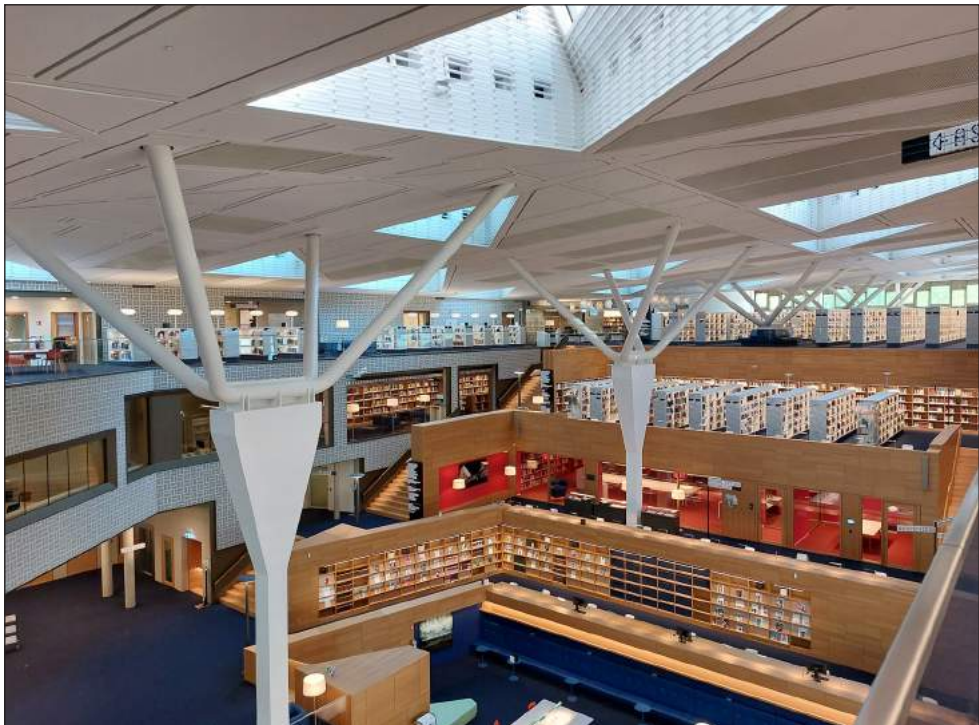


Abb. 3: Nationalbibliothek Luxemburg. Foto: Alice Rabeler

Die Vorträge boten eine große Spannweite von visionären Zukunftsszenarien über begrenzte, konkrete Aufwertungsmaßnahmen existierender Bibliotheksräume bis hin zu Re-Evaluierungen gut genutzter Gebäude, auch hinsichtlich nicht intendierter Effekte, aus denen Andere lernen können.

Den Auftakt machte die Direktorin des Luxembourg Learning Center, Marie-Pierre Pausch, mit ihrer Begrüßungsrede, in der sie zu mutiger, fantasievoller Planung für zukunftsorientierte Bibliotheksgebäude aufrief: „Do not be afraid to dare, do not be afraid to be different!“ Wichtig sei ein Gebäude mit Persönlichkeit und Lebendigkeit, in dem sich Menschen gerne aufhalten, weil sie es als inspirierend erleben. Die Raumstruktur des 2018 eröffneten Belval Learning Center, das in die alte Stahlstruktur der Möllerei, vormals ein Lager für Koks und Eisenerz, eingebaut wurde, bietet abwechslungsreiche Perspektiven auf vielfältige Leseetagen und elliptische, freischwebend wirkende „Pods“ für Gruppenarbeit. Das Gebäude ist 24/7 zugänglich, die Nutzerinnen und Nutzer können im Bereich des flexiblen Mobiliars ihr Arbeits- bzw. Lernsetting beliebig arrangieren.

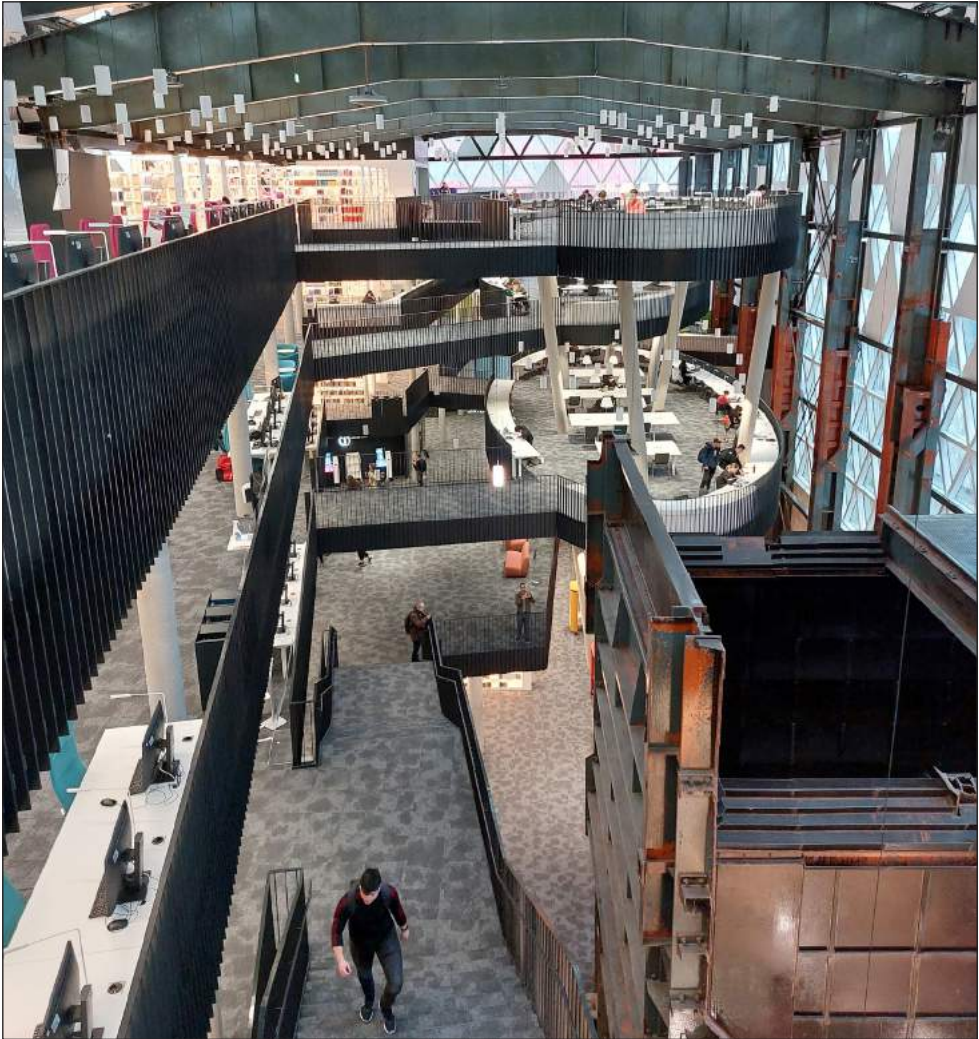


Abb. 4: Luxemburg Learning Center. Foto: Alice Rabeler



Abb. 5: Buchbare Kabinen als Raum-in-Raum-Lösung für Einzelarbeit oder Teilnahme an digitaler Lehre. Foto: Alice Rabeler

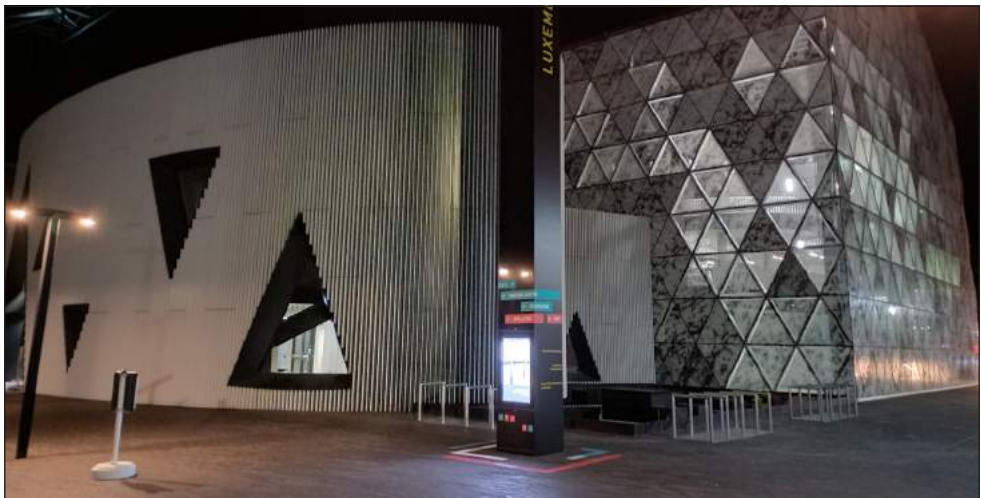


Abb. 6: Luxemburg Learning Center bei Nacht. Foto: Sebastian Lange

In den folgenden Vorträgen wurden neue Bibliotheksbauten vorgestellt: Hier sind an erster Stelle die bereits ikonisch gewordene Zentralbibliothek Oodi in Helsinki und die Bibliothek des Royal College of Surgeons Ireland (RCSI Library), die 2019 den „SCONUL Design Award“ gewann, zu nennen.⁵

Interessant ist auch das Bauprojekt „Lumen Learning Center“ der Université Paris-Saclay. Drei forschungsstarke Hochschulen mit insgesamt 48.000 Studierenden wurden zusammengelegt. Für sie entsteht gerade ein Gebäude, das als Zielvorstellung experimentelles Lernen unterstützt, z. B. durch Labore für Virtual Reality, Videoproduktion oder Materialdatenbanken für Designer.⁶ Raumprogramm und Ausstattung wurden 2019 mit User Experience Design in Workshops mit Studierenden und Institutsvertreterinnen und -vertretern der verschiedenen Fächerkulturen entwickelt. Die Eröffnung ist für November 2022 geplant.

Ein weiteres spannendes Gebäude wurde mit der Biblioteca Gabriel Garcia Marquez in Barcelona vorgestellt. Die Architektinnen und Architekten folgen ökologischen Prinzipien zum einen durch die Baukonstruktion, indem hier die bis dato größte Bibliothek in Holzbauweise entstanden ist. Zum anderen verstehen sie die Bibliothek selbst als Teil eines urbanen Ökosystems („libraries as intensified extensions of their neighbourhoods“). So lassen sich im Erdgeschoss alle Glaswände aufschieben und es entsteht eine Terrasse, wo der Stadtraum in die Bibliothek übergeht, man sich Sitzmöbel nach draußen holen und unter Bäumen lesen, musizieren oder plaudern kann. Die Eröffnung fand Ende Mai statt.⁷

Neben den Neubauten wurden unter der Rubrik „Lessons Learned“ Erfahrungen im täglichen Bibliotheksbetrieb ausgewertet, also der Frage nachgegangen, wie sich ein Gebäude im Alltag bewährt. Das bekannte Gebäude der Bibliothek der TU Delft - mit der markanten mehrere Geschosse hohen Bücherwand - ist bereits 25 Jahre alt, die Inneneinrichtung wurde zuletzt 2012 neu ausgerichtet auf kollaboratives Arbeiten. Inzwischen empfinden die Studierenden jedoch die Akustik als Problem und wünschen sich mehr ruhige Bereiche. In Prüfungsphasen wird daher die ganze Bibliothek auf „silent study“ umgestellt. Außerdem werden zur Entspannung Massage, Achtsamkeitskurse und ein Therapiehund angeboten.

Interessant war auch der Vortrag über das Emily Wilding Davison Building der Royal Holloway University London. Das 2017 eröffnete Gebäude auf dem viktorianischen Campus beherbergt nicht nur die Bibliothek, sondern weitere Serviceeinrichtungen für Studierende. Dies führt zu einer sehr lebendigen Raumnutzung in der Weise, dass auf den gemeinsamen, offenen Flächen die verschiedensten Aktivitäten von studentischen Organisationen stattfinden. Es entsteht aber auch ein neues Governance-Problem: Wer ist für was verantwortlich? Wer hat die Kontrolle über welche Flächen

5 Zu Oodi siehe <https://www.bauwelt.de/dl/1408916/16_bis_23_2_Helsinki.p1_LowRes.pdf> und <<https://b-u-b.de/detail/oodi>>, beide Stand: 01.08.2022.

Die RCSI Library ist Teil des neuen Gebäudes der privaten Medizinhochschule in der Dubliner Innenstadt. <<https://www.rcsi.com/dublin/news-and-events/news/news-article/2019/12/rcsi-library-wins-sconul-design-award>>, Stand: 01.08.2022.

6 Vgl. <<https://www.universite-paris-saclay.fr/lumen>>, Stand: 01.08.2022.

7 Vgl. <<https://www.barcelonaaldia.com/2022/05/28/inaugurada-la-biblioteca-garcia-marquez-la-tercera-mas-grande-de-barcelona/>>, Stand: 01.08.2022.

und Ressourcen? Wie kontrolliert man den Zugang, um Sicherheit für alle zu gewährleisten? Die Partizipation der Studierenden ist erwünscht, erfordert aber zeitintensive Abstimmungs- und Organisationsprozesse.

Zum Schluss stellte der niederländische Architekt Aat Vos, der 2021 mit der Karl-Preusker-Medaille ausgezeichnet wurde, noch einmal ganz prinzipielle Überlegungen an: zu Bibliotheken als Orte zur Stärkung einer freien, demokratischen, inklusiven Gesellschaft. Er gab zu bedenken, dass die digitalen und hier vor allem die sozialen Medien zu einem Suchtphänomen („addiction by design“) führen, das nicht zuletzt durch mangelnde Ausgleichsmöglichkeiten in der Pandemie einen Anstieg bei Depressionen ausgelöst hat. Sein Credo lautet, dass Bibliotheken nicht nur Wissen und Bildung aktivieren sollen, sondern auch Sozialkompetenz und demokratische Werte. Bibliotheken sind also nicht nur Orte der persönlichen Entwicklung des Einzelnen, sondern auch Räume, in denen sich Menschen aus allen Milieus begegnen können und damit insgesamt die sozial immer weiter auseinanderdriftende und durchkommerzialisierte Gesellschaft revitalisieren können. Diese Räume müssen mit partizipativem Design so attraktiv gestaltet werden, dass Menschen dafür gerne das eigene Wohnzimmer verlassen.

Fazit: Die Tagung war rundherum anregend. Nach zwei Jahren Pandemie zeigte sich einmal mehr, wie wichtig zum einen die physische Erfahrbarkeit von neuen Bauten und Raumgestaltungen, zum anderen aber auch der informelle Austausch mit Kolleginnen und Kollegen der europäischen Fachcommunity ist.

Alice Rabeler, Universitäts- und Landesbibliothek Bonn

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5884>

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).

Der Zertifikatskurs Forschungsdatenmanagement in NRW

Eine modular aufgebaute Weiterqualifikation für das professionelle Datenmanagement

1. Ausgangslage

Im Zuge der Digitalisierung der Wissenschaft verändern sich die Methoden zur Erkenntnisgewinnung und Informationserschließung. Während ursprünglich ausschließlich analoge bzw. mechanische Verfahren genutzt wurden, nimmt der Anteil von digitalen Forschungsdaten und damit verbundenen Prozessen rasant zu.¹ In diesem Kontext gewinnt das Handlungsfeld des Forschungsdatenmanagements (FDM) an zentraler Bedeutung, um wertvolle Forschungsdaten nachhaltig zu sichern und sie für eine weitere Verwendung im Forschungszyklus zu halten. Hierdurch werden aber auch neue Herausforderungen für wissenschaftliche und infrastrukturelle Einrichtungen geschaffen: Um digitale Forschungsdaten reproduzierbar zu verarbeiten und für eine Nachnutzung aufzubereiten, bedarf es nicht nur entsprechender technischer Informations- und Analyseinfrastrukturen, sondern vor allem Personen, die über fundierte Kompetenzen verfügen, verschiedene Zielgruppen zu einem nachhaltigen Umgang mit Forschungsdaten und zur Nutzung einschlägiger Infrastrukturen zu beraten und zu schulen. Ein systematischer Auf- und Ausbau solcher Kompetenzen wurde bereits 2016 vom Rat für Informationsstrukturen (RfII) mit der Empfehlung formuliert, neue Berufsbilder, Studiengänge und Qualifizierungswege zu schaffen.² Die stetig wachsende Nachfrage nach Personal mit einschlägigen Expertisen zeigt, dass dieses Desiderat nach wie vor besteht.

Dieser Bedarf ergibt sich aus der Bildung standortbezogener Infrastrukturen und überregionaler Strukturen, Netzwerke und Organisationen, die von einzelnen Hochschulen oder Forschungszentren über FDM-Landesinitiativen und der Einrichtung der Nationalen Forschungsdateninfrastruktur (NFDI) bis zur Bildung internationaler Zusammenschlüsse wie der European Open Science Cloud (EOSC) reichen. Die NFDI hat den Auftrag, für die gesamte deutsche Forschungsgemeinschaft Lösungen für die effiziente Speicherung, Analyse und Nachnutzung von Forschungsdaten zu finden. Der Start der NFDI mit inzwischen 19 der 30 fachlich oder methodisch ausgerichteten Konsortien³ befeuert die Nachfrage nach FDM-versiertem Personal maßgeblich und adressiert das Thema einer entsprechenden Qualifizierung im Rahmen der Querschnittsthemen der NFDI (u. a. die Sektion „Training & Education“).⁴ Auf europäischer Ebene findet die NFDI ihr Pendant in der European Open Science Cloud (EOSC), deren Ziel es ist, Forschenden in Europa eine verteilte, offene und multidisziplinäre

1 Kindling, Maxi; Schirmbacher Peter: „Die digitale Forschungswelt“ als Gegenstand der Forschung, in: *Information - Wissenschaft & Praxis* 64 (2-3), 2013, S. 127-136. Online: <<https://doi.org/10.1515/iwp-2013-0017>>.

2 Rat für Informationsinfrastrukturen: *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*, Göttingen 2016, S. 5-160. Online: <[urn:nbn:de:101:1-2019080711032249706218](https://nbn-resolving.org/urn:nbn:de:101:1-2019080711032249706218)>, Stand: 13.06.2022.

3 NFDI, Nationale Forschungsdaten Infrastruktur. Konsortien, <<https://www.nfdi.de/konsortien/>>, Stand 13.06.2022.

4 Herres-Pawlis, Sonja; Pelz, Peter; Kockmann, Norbert u. a.: *Sektionskonzept zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. (2.0)*. Zenodo, 2022, S. 5-13. Online: <<https://doi.org/10.5281/zenodo.5599769>>.

Forschungsumgebung zu bieten, in der sie Daten, Werkzeuge und Dienste für Forschungs-, Innovations- und Bildungszwecke veröffentlichen, finden und nachnutzen können.⁵

Im Hinblick auf erforderliche Kompetenzen der im FDM-Kontext agierenden Personen gilt es, auf sehr unterschiedliche Bedarfe – abgestimmt auf den jeweiligen beruflichen Kontext – zu reagieren. In Ergänzung zur zuvor erwähnten Empfehlung von 2016 betonte der RfII auch 2019 die Bedeutung vom „Erwerb von Zusatzkompetenzen auf der Basis von Fort- und Weiterbildungen zur Übernahme neuer oder veränderter Aufgaben im Berufsalltag. [...] Denn es ist zu erwarten, dass sich die Anforderungen auch im Rahmen neuer Digitalberufe künftig sehr schnell wandeln.“⁶

Die Komplexität des Forschungsdatenmanagements – vor dem Hintergrund individueller und forschungskontextbezogener Anforderungen gepaart mit der dynamischen Entwicklung der Infrastrukturen im Zuge der NFDI oder EOSC – führt zu einem steigenden Bedarf an Personen, die Forschende umfassend beim Forschungsdatenmanagement unterstützen, Policies und Informationsinfrastrukturen (mit)entwickeln und die Kommunikation zwischen Vertreter*innen der forschungsnahen Infrastrukturen (u. a. Bibliotheken, Rechenzentren, Forschungsförderung) und den Forschenden gestalten. Die damit einhergehenden Herausforderungen lassen sich exemplarisch an der in FDM-Prozessen angestrebten Umsetzung der FAIR-Data-Prinzipien verdeutlichen, nach denen Forschungsdaten auffindbar (Findable), zugänglich (Accessible), interoperabel (Interoperable) und nachnutzbar (Reusable) sein sollen.⁷ Die Einhaltung dieser Kriterien bewegt sich oftmals im Spannungsfeld technischer Möglichkeiten und gegenläufiger Vorgaben. So gibt es Rahmenbedingungen, die eine Publikation von Forschungsdaten verhindern, beispielsweise patentrechtliche Einschränkungen oder die Datenschutzbestimmungen bei personenbezogenen Daten.

Es ist daher ein zentrales Anliegen für die Etablierung eines nachhaltigen Forschungsdatenmanagements, berufsbegleitende und strukturierte Möglichkeiten zum Kompetenzaufbau und -ausbau einzurichten. In den vergangenen Jahren wurden erste Ansätze geschaffen, indem beispielsweise Zertifikatskurse zum Berufsfeld Data Librarian eingerichtet wurden.⁸ Obwohl diese Angebote zum Teil Aspekte aus dem FDM-Kontext behandeln, widmen sie sich nicht exklusiv dem Berufsfeld des Forschungsdatenmanagements. In diese Lücke stößt der Zertifikatskurs „Forschungsdatenmanagement“ in Nordrhein-Westfalen (NRW), der als Säule eines flächendeckenden Aufbaus von FDM in NRW als Arbeitspaket in die Handlungsfelder der Landesinitiative *fdm.nrw* integriert wurde. Neben dem Kursangebot für die Beschäftigten aus NRW werden mit dem Zertifikatskurs

5 European Open Science Cloud (EOSC). What the European Open Science Cloud is, <https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en>, Stand: 13.06.2022.

6 Rat für Informationsinfrastrukturen: Digitale Kompetenzen – dringend gesucht! Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft, Göttingen 2019, S. 20. Online: <<https://rfii.de/?p=3883>>, Stand: 13.06.2022.

7 Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan u. a.: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 2016. Online: <<https://www.nature.com/articles/sdata201618>>, Stand: 13.06.2022.

8 Universität Innsbruck, <<https://www.uibk.ac.at/weiterbildung/universitaetskurse/data-librarian/index.html.de>>, Stand: 13.06.2022 und Technische Hochschule Köln, <https://www.th-koeln.de/weiterbildung/zertifikatskurs-data-librarian_63393.php>, Stand: 13.06.2022.

„Forschungsdatenmanagement“ adaptierbare Strukturen und Modulinhalte vorgelegt.⁹ Diese können und sollen von verschiedenen Einrichtungen und (Bundes)Ländern für eigene Weiterbildungsprogramme genutzt werden, um so kurz- und mittelfristig eine flächendeckende Weiterbildung von aktuellem und künftigen Personal im FDM über NRW hinaus zu erzielen.

2. Ziele und Zielgruppen

Der Zertifikatskurs hilft dabei, das dafür benötigte qualifizierte Personal in ausreichendem Umfang zu gewährleisten und die Hochschulen in NRW „FDM-fit“ zu machen. Zudem wird für das neue Berufsfeld eine kompakte, strukturierte und qualitätsgesicherte Weiterbildung geschaffen, mittels derer die derzeitigen Quereinsteiger*innen und thematischen Pionier*innen sich professionalisieren und ihre gesammelte Erfahrungspraxis in „common knowledge“ verwandeln können.

Der Zertifikatskurs spricht also eine einrichtungsunabhängige, forschungsnahe Zielgruppe an, die – verortet an der Schnittstelle zwischen Wissenschaft, Bibliotheken und Rechenzentren – heterogene Bedarfe in Bezug auf Weiterbildungen formuliert. Um diese näher zu spezifizieren, führte die Landesinitiative *fdm.nrw* zwischen 2019 und 2021 explorative Gespräche mit Beschäftigten aus forschungsnahen Infrastruktureinrichtungen in NRW durch. So konnte ein breites Spektrum an Anforderungen gesichtet werden, welches inhaltlich von FDM-Grundkenntnissen, über Einführungen zu Tools wie Elektronischen Labornotizbüchern und Versionierungssystemen, rechtlichen Aspekten rund um das FDM oder zielgerichteten Förderprogrammen bis hin zu erprobten Strategien reicht. Zudem gilt es, dieses Wissen verschiedenen Zielgruppen über Schulungen und andere Informationsformate zu vermitteln. Insbesondere aus dem NFDI-Kontext ergibt sich darüber hinaus die Anforderung, eine Kombination aus disziplinspezifischen Qualifikationen mit Erfahrungen und Kenntnissen des Datenmanagements zu einem optimalen Kompetenzprofil zu vereinen, da die disziplinspezifischen Anforderungen sowohl in den NFDI-Konsortien als auch im Forschungsdatenmanagement insgesamt so verschieden wie für die Umsetzung ausschlaggebend sind.

Auf der Basis dieser Vorüberlegungen und Abfragen lässt sich die Zielgruppe des Zertifikatskurses in zwei Personenkreise unterteilen:

Der erste Kreis umfasst Beschäftigte aus Infrastruktureinrichtungen, also aus bereits vorhandenen oder aufzubauenden FDM-Kontaktstellen sowie aus weiteren forschungsnahen Arbeitsbereichen wie Bibliotheken, Rechenzentren oder der Forschungsförderung. Innerhalb dieser Personengruppe lassen sich heterogene Bildungshintergründe identifizieren, die von Berufsausbildungen im Bereich Informationsmedien über unterschiedliche Studienabschlüsse bis hin zu abgeschlossenen Promotionen reichen. Je nach Aufgabenbereich, Bereichszugehörigkeit und Position kann der berufserfahrungsbedingte Anknüpfungspunkt zum FDM mehr oder weniger stark ausgeprägt sein. Personen aus dieser Gruppe werden oftmals mit zentralen oder hochschulpolitischen Aufgaben betraut, worunter

⁹ Blümm, Mirjam; Förstner, Konrad U.; Lanczek, Marvin u. a.: *Der Zertifikatskurs Forschungsdatenmanagement als adaptierbares Aus- und Weiterbildungsangebot*, in: Heuveline, Vincent und Bisheh, Nina (Hrsg.): *E-Science-Tage 2021: Share Your Research Data*, Heidelberg: heiBOOKS, 2022. S. 414–420. Online: <<https://doi.org/10.11588/heibooks.979.c13758>>.

generische Beratungs- und Schulungsangebote wie auch die Etablierung einer Forschungsdaten-Policy fallen. Oftmals werden diese Personen als zentrale bzw. generic Data Stewards bezeichnet.¹⁰

Der zweite Personenkreis setzt sich aus Forschenden bzw. Multiplikator*innen aus den Fach- und Forschungsbereichen zusammen. Diese Personen verfügen in der Regel über einen Master (oder vergleichbaren) Abschluss, sind oftmals in ihrer Promotionsphase oder arbeiten als Postdocs. Teilweise verfügen sie über ein Spezialwissen, welches sich auf fach- bzw. datenspezifische Aufgaben und Prozesse bezieht; dabei ist ihre Kenntnis über zentrale bzw. übergeordnete Infrastrukturen unterschiedlich stark ausgeprägt. Je nach Stellenprofil werden diese Personen als dezentrale bzw. embedded Data Stewards bezeichnet, wobei der Übergang zu zahlreichen weiteren Tätigkeitsbeschreibungen fließend ist. (u. a. Data Scientists, Data Analysts).¹¹

Zusammenfassend wird mit dem Zertifikatskurs „Forschungsdatenmanagement“ eine strukturierte Vermittlung von FDM-Grundkenntnissen, FDM-bezogenen Tools und Erfahrungswerten sowohl für Beschäftigte in Infrastruktureinrichtungen als auch von Forschenden aus NRW angestrebt. Der Fokus liegt auf einem generischen Ansatz, der durch Beispiele aus einzelnen Fachdisziplinen ergänzt wird. Zudem wird den Teilnehmenden die Möglichkeit der individuellen Differenzierung gegeben, die an das thematische Interesse, das jeweilige Aufgaben- bzw. Einsatzgebiet oder auch Perspektiven für eine berufliche Weiterentwicklung angepasst werden können. Ein übergeordnetes Ziel des Kurses besteht darin, die der Zielgruppe innewohnende Diversität der beruflichen und institutionellen Kontexte abzubilden und den Austausch unter den Teilnehmenden facettenreich und informativ zu gestalten.

3. Organisation und Konzeption

Die an der Konzeption beteiligten Institutionen, Technische Hochschule Köln, ZBIW – Zentrum für Bibliotheks- und Informationswissenschaftliche Weiterbildung, ZB MED – Informationszentrum Lebenswissenschaften und die Landesinitiative für Forschungsdatenmanagement – fdm.nrw, ergänzen sich in ihren Expertisen, Erfahrungswerten und Zugängen zum FDM, die sie in die gemeinsame Entwicklung und Durchführung des Zertifikatskurses einfließen lassen. Das ZBIW verfolgte als ein bundesweit agierendes Weiterbildungszentrum für Beschäftigte öffentlicher und wissenschaftlicher Bibliotheken und anderer Informationseinrichtungen – eingebunden in das Institut für Informationswissenschaft der TH Köln – bereits seit längerem das Ziel, das Portfolio der Angebote in Richtung Data Science und verwandter Themen zu erweitern. Zusammen mit ZB MED wird seit 2019 der

10 Cremer, Fabian; Engelhardt, Claudia; Neuroth, Heike: Embedded Data Manager. Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale, 2015. Online: <<http://doi.org/10.1515/bfp-2015-0006>>; Verheul, Ingeborg; Imming, Melanie; Ringerma, Jacquelin u. a.: Data Stewardship on the map. A study of tasks and roles in Dutch research institutes, 2019. Online: <<https://doi.org/10.5281/zenodo.2669150>>; Curdt, Constanze; Dierkes, Jens; Helbig, Kerstin u. a.: Data Stewardship im Forschungsdatenmanagement. Rollen, Aufgabenprofile, Einsatzgebiete. Überblick: 11. DINI/nestor Workshop, 16. und 17.11.2020, 2021. Online: <<http://doi.org/10.17192/bfdm.2021.3.8347>>.

11 Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale, 2015; Data Stewardship on the map: A study of tasks and roles in Dutch research institutes, 2019; Data Stewardship im Forschungsdatenmanagement. Rollen, Aufgabenprofile, Einsatzgebiete. Überblick: 11. DINI/nestor Workshop, 16. und 17.11.2020, 2021.

Zertifikatskurs „Data Librarian“ angeboten¹², der sich explizit an Beschäftigte aus wissenschaftlichen Bibliotheken richtet, um Computational Thinking sowie die Fähigkeiten für die praktische Anwendung zu vermitteln. Dabei wird das Forschungsdatenmanagement jedoch nur als Teilaspekt im Rahmen eines Moduls behandelt.

Die dringlichen Rückmeldungen aus den Hochschulen sowie die voranschreitende Entwicklung der NFDI veranlassten die Landesinitiative fdm.nrw dazu, strukturierte Weiterbildungen zu schaffen, mit denen die Beschäftigten der forschungsnahen Dienste der Hochschulen für neue Aufgaben im FDM qualifiziert werden sollen. Im Frühjahr 2019 verständigten sich die TH Köln, ZBIW, ZB MED und fdm.nrw durch einen Letter of Intent auf eine gemeinsame Entwicklung eines Zertifikatskurses „Forschungsdatenmanagement“ und implementierten zur Umsetzung dieser einrichtungsübergreifenden Kooperation einen Lenkungskreis aus Vertreter*innen der Technischen Hochschule Köln (TH Köln), dem dort ansässigen ZBIW, ZB MED sowie fdm.nrw. Von der ersten Absichtserklärung bis hin zum Beginn der ersten Durchführung des Zertifikatskurses „Forschungsdatenmanagement“ dauerte es ungefähr zwei Jahre.

Die Module sollten von ausgewiesenen Expert*innen aus ganz Deutschland für die verschiedenen Themenbereiche im FDM inhaltlich und organisatorisch gestaltet werden. Neben der Konzeption des Kursaufbaus und der Entscheidung, welche Inhalte abzudecken sind, galt es deshalb frühzeitig, geeignete Referent*innen zu finden. Fixiert wurde die Beteiligung auf Basis von Honorarverträgen, die über das ZBIW koordiniert und erstellt wurden. Darüber hinaus bietet das ZBIW zur Qualitätssicherung Schulungen für die Referent*innen an, in denen u. a. didaktische Empfehlungen und Tipps zur Entwicklung von Selbstlerninhalten thematisiert werden.

Das Auswahlverfahren für die 15 Kursteilnehmenden wurde ab Herbst 2020 erarbeitet. Als maßgebliche Kriterien wurde festgehalten, dass jede Kohorte die Diversität der Zielgruppe widerspiegeln soll, dass die Inhalte in der Tätigkeit der Teilnehmenden direkt anwendbar sein sollen und dass die Teilnehmenden an einer Einrichtung in NRW beschäftigt sind. Letzteres bildet die Voraussetzung für den Erhalt eines Stipendiums zur weitgehenden Kompensierung der Teilnahmegebühren (bereitgestellt als Fördermittel der Digitalisierungsinitiative NRW). Weitere Informationen zum Bewerbungsprozess sowie das von allen Organisator*innen und Referent*innen erarbeitete Modulhandbuch sind auf der Website des ZBIW hinterlegt.¹³

Der Zertifikatskurs besteht aus Selbstlernphasen und 12 Live-Terminen, die größtenteils virtuell über die Videoplattform Zoom durchgeführt werden. Das ZBIW hat für den Zertifikatskurs mehrere digitale Kursräume auf Moodle eingerichtet, in denen zum einen die Kommunikation zwischen Teilnehmer*innen, Referent*innen und Organisator*innen stattfindet, zum anderen werden hierüber die Lernmaterialien und Aufgaben der Selbstlernphasen bereitgestellt. Zusätzlich gibt es einen

12 Technische Hochschule Köln. Zertifikatskurs Data Librarian, <https://www.th-koeln.de/weiterbildung/zertifikatskurs-data-librarian_63393.php>, Stand: 13.06.2022.

13 Technische Hochschule Köln. Zertifikatskurs Forschungsdatenmanagement, <https://www.th-koeln.de/weiterbildung/zertifikatskurs-forschungsdatenmanagement_82048.php>, Stand:18.2.2022

Kursraum, in dem sich die Organisator*innen des Kurses und die Referent*innen untereinander austauschen können.

Die umfangreichen Selbstlernphasen werden von den Referent*innen der jeweiligen Module gestaltet. Dahinter steht das Konzept des *flipped classroom*. So erarbeiten sich die Teilnehmenden Lehrinhalte in Vorbereitung auf die gemeinsamen Live-Termine. Hierbei kommen verschiedene Materialien und Medien zum Einsatz, zu denen Lehrvideos, Präsentationen, Skripte, Linksammlungen, Lektüren oder interaktive Whiteboards zum Sammeln von Fragen und Hinweisen zählen. Flankierend gibt es Übungs- und Praxisaufgaben, mit denen die Teilnehmenden Kenntnisse einüben und vertiefen können und die Möglichkeit erhalten, ihren individuellen Wissensstand zu überprüfen.

Die Live-Termine werden von den Referent*innen primär genutzt, um die jeweiligen Inhalte und die Selbstlernphasen aufzugreifen und den Teilnehmenden Raum für Austausch und Diskussionen zu bieten. Präsenzphasen sind zum einen zu Beginn des Kurses vorgesehen, um das Kennenlernen und die Gruppenfindung der Teilnehmenden zu fördern. Zum anderen findet ein gemeinsamer Abschluss statt, in dessen Rahmen die Teilnehmenden ihre Projektarbeiten präsentieren und einen reflektierenden Blick auf den gesamten Zertifikatskurs werfen.

4. Module und Kursinhalte

Die Kursstruktur folgt der Einteilung in einführende Module, themenspezifische Erweiterungsmodule und ein abschließendes Projektmodul. Für die Themenbereiche wurde eine Recherche der bereits vorhandenen strukturierten Weiterbildungen und Studiengänge durchgeführt. Als Orientierung dienten insbesondere der Zertifikatskurs „Data Librarian“ an der TH Köln, der (2020 noch in Planung befindliche) Studiengang „Digitales Datenmanagement“ der HU Berlin und FH Potsdam sowie die Online-Module des „MANTRA - Research Data Management Training“ der University of Edinburgh. Die Kooperationspartner erarbeiteten auf Basis der Recherche und angereichert durch Feedback aus der UAG Schulungen/Fortbildungen der AG Forschungsdaten von DINI/nestor die Modularisierung des Zertifikatskurses. Insgesamt umfasst der Zertifikatskurs einen arbeitszeitlichen Umfang von 8 ECTS, also umgerechnet 240 Zeitstunden (1 ECTS = 30 Zeitstunden), der sich auf drei Basismodule, vier von fünf zu wählenden Aufbaumodulen und ein abschließendes Projektmodul verteilt.

Die drei obligatorischen Basismodule bilden die Grundlage für alle Teilnehmenden: *Modul 1: Grundlagen des FDM* setzt sich zusammen aus dem Train-the-Trainer-Workshop zum Thema Forschungsdatenmanagement¹⁴ und einem Auftakttag. Der Train-the-Trainer-Workshop kann auch unabhängig vom Zertifikatskurs – beispielsweise angeboten von fdm.nrw – besucht und voll angerechnet werden. Der Auftakttag dient dem gegenseitigen Kennenlernen der Teilnehmenden und gibt erste thematische Impulse.

¹⁴ Biernacka, Katarzyna; Buchholz, Petra; Danker, Sarah Ann u. a.: Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement. Version 3.1, Berlin 2020. Online: <<https://doi.org/10.5281/zenodo.1215376>>.

Modul 2: Open Science & rechtliche Aspekte behandelt FDM im Spannungsfeld zwischen Transparenz im Sinne der guten wissenschaftlichen Praxis¹⁵ und des Datenschutzes. Das Modul beleuchtet (insbesondere mit Blick auf Urheberrecht und Lizenzierung), welche Anforderungen, Rechte und Pflichten beim Umgang mit Forschungsdaten existieren, wie diese (institutionell) umgesetzt werden können und welche Akteure und Initiativen sich in diesem Bereich engagieren.

Modul 3: Forschung, FD & FDM in den Fachgebieten vermittelt einen Einblick in die speziellen Anforderungen an das FDM aus der Perspektive verschiedener Disziplinen. Ausgewählt wurden die Fachbereiche (Bio-)Medizin, Biodiversität, Chemie, Digital Humanities, Ingenieurwissenschaften und die Sozial- und Wirtschaftswissenschaften. Für diese befinden sich bereits NFDI-Konsortien im Aufbau, so dass hier auch auf erste bzw. geplante Ergebnisse der Infrastrukturentwicklungen in den fachlichen Konsortien eingegangen werden kann. Zudem erhalten die Teilnehmenden einen Einblick in mehrere Fachbereiche und deren speziellen Aufgaben, Herausforderungen und Lösungen (z. B. bezogen auf Ansätze oder Tools).

An die drei Basismodule schließen sich fünf thematisch-fokussierte Aufbaumodule an, aus denen die Teilnehmenden je nach persönlichem Interesse und Kenntnisstand vier auswählen – oder auf Wunsch auch alle Module besuchen können:

Modul 4: Hacken & experimentieren mit Daten vermittelt Konzepte zum effizienten und reproduzierbaren Bearbeiten von Daten. Ziel ist es hier, die Teilnehmenden mit verschiedenen Werkzeugen vertraut zu machen, aber auch das Bewusstsein für Computational Thinking, Automatisierung und Optimierung der eigenen Prozesse zu schaffen. Das Modul bietet eine Einführung in die Unix-Shell, die Programmiersprache Python und die Versionskontrolle git. Dabei wird auf den Lehrkonzepten von „The Carpentries“¹⁶ aufgebaut. Teile dieses Moduls können ebenfalls durch eine Teilnahme an einer vergleichbaren Weiterbildung abgedeckt werden.

Modul 5: (Meta-)Daten verwalten & teilen gliedert sich in die Themen „Daten strukturieren und organisieren“, „Daten finden und nachnutzen“ sowie „Daten teilen und publizieren“. Ausgehend von (disziplinspezifischen) Metadatenstandards und -vokabularien werden Instrumente und Mechanismen zur Recherche von Forschungsdaten analysiert sowie Initiativen vorgestellt, die sich für eine leichtere Auffindbarkeit und Verfügbarmachung von Forschungsdaten einsetzen.¹⁷

Modul 6: Technische Infrastruktur beleuchtet die technischen Anforderungen an Repositorien und Speicher-Systeme für Forschungsdaten sowie deren Umsetzbarkeit auf institutioneller Ebene. Hierbei spielen Schnittstellen und technische Formate ebenso eine Rolle wie Zugriffsmechanismen und

15 Deutsche Forschungsgemeinschaft e.V.: Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. 2019. Online: <https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf>, Stand: 13.06.2022.

16 Becker, Erin; Michonneau, François: The Carpentries Curriculum Development Handbook, <<https://github.com/carpentries/curriculum-development>>, Stand: 13.06.2022.

17 The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 2016.

Betriebsmodelle. Einen weiteren Schwerpunkt bildet das Thema digitale Langzeitarchivierung bzw. -verfügbarkeit, besonders im Hinblick auf entsprechende Systeme und das OAIS-Referenzmodell als einem der wichtigsten Standards in der digitalen Langzeitarchivierung.¹⁸

Modul 7: Daten- & Projektmanagement in der Forschung adressiert die Planung des Forschungsdatenmanagements in Projekten von der Antragsphase bis zur Umsetzung. Dies umfasst zum einen den Aufbau und die Planung eines Förderantrags sowie die Erstellung und Integration eines Datenmanagementplans in diesen Prozess. Zum anderen werden Methoden des Projektmanagements und Werkzeuge für das FDM im Projekt vorgestellt. Dabei werden auch Aspekte der institutionellen Implementierung von FDM berücksichtigt.

Modul 8: FDM-Beratung & Schulung vermittelt zum einen Ansätze, Settings und Strategien für FDM-bezogene Beratungen, zum anderen erhalten die Teilnehmenden einen Überblick zu didaktischen Konzepten und Methoden und wenden diese anhand praktischer Übungen und Fallbeispiele an. Ziel ist, die Teilnehmenden zu befähigen, an ihren jeweiligen Einrichtungen eigenständig zielgruppenorientierte Beratungs- und Schulungsangebote zum Forschungsdatenmanagement zu konzipieren, anzubieten und zu evaluieren.

Um den Kurs erfolgreich mit einem Zertifikat abzuschließen, bearbeiten die Teilnehmenden im *Projektmodul (Modul 9)* ein Thema mit Bezug zum FDM. Auf diese Weise sollen die Inhalte des Zertifikatskurses vertieft und gleichzeitig das erworbene Wissen angewendet werden. Idealerweise stammt das Projekt aus dem konkreten Arbeitsalltag der Teilnehmenden. Dies kann beispielsweise ein Schulungskonzept, ein Video oder eine Softwarelösung sein. Die Projekte werden in einer Abschlussveranstaltung in Kurzvorträgen den anderen Teilnehmenden vorgestellt und gemeinsam diskutiert. Dies bildet gleichzeitig den Abschluss des Zertifikatskurses.

5. Weiterentwicklung und Vernetzung

Eine ständige Aktualisierung und Weiterentwicklung des Zertifikatskurses „Forschungsdatenmanagement“ ist erforderlich, um auch langfristig den Anforderungen des FDM gerecht zu werden. Neben dem generellen Bedarf, Kursmaterialien auf Basis aktueller Entwicklungen und Standards fortlaufend zu aktualisieren, spielen hier auch zwei Faktoren hinein, die für diesen Arbeitsbereich charakteristisch sind:

Durch die verschiedenen Initiativen und Projekte entwickelt sich das FDM in verschiedene Richtungen, schnell und dynamisch. Das Forschungsdatenmanagement ist und bleibt ein komplexes Arbeitsfeld mit wachsender Spezifikation. Die inhaltliche Gestaltung der Module durch einen Pool ausgewiesener Expert*innen trägt entscheidend dazu bei, den aktuellen Wissensstand im FDM in den Zertifikatskurs einzubinden.

¹⁸ ISO 14721:2012. Space data and information transfer systems. Open archival information system (OAIS). Reference model. 2012. Online: <<https://www.iso.org/standard/57284.html>>, Stand: 13.06.2022.

Basierend auf den Initiativen, Projekten und Entwicklungen ergibt sich als zweites Charakteristikum ein hoher, sich fachlich spezialisierender Bedarf an Personal. Übergreifende FDM-Initiativen (wie die NFDI) befinden sich weitgehend noch im Aufbau. Zudem ist das FDM noch nicht flächendeckend an den Hochschulen und Forschungseinrichtungen Nordrhein-Westfalens etabliert. Gleichzeitig wachsen die Anforderungen an die Forschung, was sich u. a. in neuen DFG-Anforderungen zur Einreichung von Antragsvorhaben und zum Umgang mit Forschungsdaten zeigt.¹⁹ Es ist zu erwarten, dass künftige Sonderforschungsbereiche häufiger eigene Expert*innen für FDM einstellen werden. Im Falle der NFDI gibt es nicht nur den Bedarf nach geschultem Personal, sondern auch Ambitionen, diesen durch eigene Weiterbildungsangebote (zumindest teilweise) zu decken. Neben der NFDI-Sektion „EduTrain“²⁰ arbeitet beispielsweise die KonsortSWD-Arbeitsgruppe „Developing and exchanging RDM skills“²¹ an einer Online-Plattform²², die neben der Bereitstellung von zentralen FDM-Informationen und FDM-Weiterbildungsmodulen auch einen FDM-Zertifikatskurs für (angehende) Beschäftigte in vom RatSWD akkreditierten Forschungsdatenzentren anbieten wird. Hier – wie auch mit anderen Initiativen – bieten sich Möglichkeiten zum Austausch und zu Kooperationen, um die Personalentwicklung und -gewinnung im FDM proaktiv zu gestalten.

Ein Beispiel für den Mehrwert solcher Kooperationen zeigt sich bereits in Österreich: Die Universität Wien und das Projekt FAIR Data Austria²³ adaptieren das Konzept des Zertifikatskurses „Forschungsdatenmanagement“ aus NRW derzeit für Österreich. So werden den Teilnehmenden des ab dem Wintersemester 2022 startenden „Zertifikatskurses Data Steward“ in zwei Semestern Kenntnisse und Kompetenzen vermittelt, die für die Rolle von zentral sowie dezentral arbeitenden Data Stewards benötigt werden.²⁴

Die sich ergänzende Expertise der Kooperationspartner kommt auch der Überarbeitung des Zertifikatskurses „Forschungsdatenmanagement“ zugute. Durch ihre Vernetzungsarbeit ist die Landesinitiative fdm.nrw über die Entwicklungen im Land und darüber hinaus informiert; die TH Köln, ZBIW und ZB MED wiederum sind stark in verschiedene fach- und methodenspezifische Communities eingebunden. Wahrnehmbare Trends mit Auswirkung auf die Personalentwicklung im FDM in NRW fließen so in den Zertifikatskurs ein. Als Teil der didaktischen Methodik wird das Format systematisch durch das ZBIW evaluiert. Die Ergebnisse werden zum einen kontinuierlich zur inhaltlichen und organisatorischen Weiterentwicklung des laufenden Kurses eingesetzt. Zum anderen fließt

19 Deutsche Forschungsgemeinschaft e.V.: Umgang mit Forschungsdaten, <https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/index.html>, Stand: 06.05.2022

20 Sektionskonzept zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. (2.0). Zenodo, 2022. Online: <<https://doi.org/10.5281/zenodo.5599770>>.

21 KonsortSWD. Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften: Developing and exchanging RDM Skills. Online-Plattform für Lehr- und Informationsmaterialien, <<https://www.konsortswd.de/konsortswd/das-konsortium/services/rdm-skills/>>, Stand: 13.06.2022.

22 RDMCompas. Research Data Management Competence Base. Forschungsdatenmanagement Kompetenzen für Datenkuratierung, <<https://rdm-compas.org/>>, Stand: 13.06.2022.

23 FAIR Data Austria. Workshop zum FAIR National Office, <<https://forschungsdaten.at/fda/>>, Stand: 13.06.2022

24 Kalová, Tereza: 5 Data Stewards per 100 Researchers?! The Development of a Certificate Course “Data Steward” at the University of Vienna, 2021. Online: <<https://doi.org/10.5281/zenodo.5544024>>; Kalová, Tereza: Teaching Data Stewardship in Austria: Setting up a Certificate Course for Data Stewards at the University of Vienna, 2022. Online: <<https://doi.org/10.5281/zenodo.6413512>>; Universität Wien. Certificate Course Data Steward, <<https://www.postgraduatecenter.at/en/programs/communication-media/data-steward/>>, Stand: 13.06.2022.

das Gesamtfazit, welches im Sommer 2022 vorliegen wird, in den zweiten Durchgang des Kurses ein. Insgesamt trägt die Evaluation dazu bei, den Zertifikatskurs FDM aktuell zu halten und auf die Bedarfe der Zielgruppe abzustimmen.

Basierend auf den Kursevaluationen, entlang der Bedarfe an FDM-Personal (in NRW) und im Austausch mit Initiativen, die vergleichbare Weiterbildungsangebote im FDM erarbeiten, soll der Zertifikatskurs „Forschungsdatenmanagement“ mit der Perspektive, die Zukunft des FDM-Handlungsfeldes zu gestalten, stetig weiterentwickelt werden.

Benjamin Slowig, Landesinitiative fdm.nrw, Duisburg-Essen

Mirjam Blümm, Technische Hochschule Köln

Konrad U. Förstner, ZB MED – Informationszentrum Lebenswissenschaften und Technische Hochschule Köln

Marvin Lanczek, ZBIW, Technische Hochschule Köln

Birte Lindstädt, ZB MED – Informationszentrum Lebenswissenschaften, Köln

Rabea Müller, ZB MED – Informationszentrum Lebenswissenschaften, Köln

Ulrike Nickenig, Landesinitiative fdm.nrw, Duisburg-Essen

Stephanie Rehwald, Landesinitiative fdm.nrw, Duisburg-Essen

Lioba Schreyer, Landesinitiative fdm.nrw, Duisburg-Essen

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5833>

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).

Multiplikator*innen für eine offene Wissenschaft

Das Konzept der Train-the-Trainer-Workshops von open-access.network

Immer noch existiert eine Diskrepanz zwischen dem Wunsch der Wissenschaft nach freiem Zugang zu wissenschaftlicher Information auf der einen und der eigenen Veröffentlichungspraxis auf der anderen Seite. Eine wesentliche Ursache dafür ist die Unsicherheit der Wissenschaftler*innen in ihrer Rolle als Autor*innen über die Rahmenbedingungen, Voraussetzungen und Möglichkeiten von Open-Access-Publikationen in ihren Einrichtungen. Hier setzt das vom Bundesministerium für Bildung und Forschung von Dezember 2019 bis November 2022 geförderte Projekt open-access.network an. Es unterstützt die Vernetzung und den Kompetenzaufbau von Open Access (OA) Professionals und Wissenschaftler*innen durch zielgruppenorientierte Veranstaltungsformate, Austauschangebote sowie Hilfestellungen in Form von Informationsseiten und Services wie einem Forum¹, einem Open Access Helpdesk² und einem Recherchetool für die Suche nach Publikationsorten³. Zu den begleitenden Fort- und Weiterbildungsangeboten gehört neben Barcamps, Staff Weeks und thematischen Workshops auch eine Veranstaltungsreihe von zwölf Train-the-Trainer-Workshops.

1. Was bedeutet Train-the-Trainer?

Die Train-the-Trainer-Workshops sind ein Angebot zur intensiven Weiterbildung von Open-Access-Multiplikator*innen auf der Basis eines „Train-the-Trainer“-Ansatzes. Der Schwerpunkt der zweitägigen Intensiv-Veranstaltungen liegt darauf, die didaktischen Kompetenzen von Personen, die beratend bei einer Publikation unterstützen, zu stärken und ihnen Ideen zur strukturierten Planung und Konzeption eigener Veranstaltungen zu vermitteln. Darüber hinaus fördert das Format den Austausch und die Bildung einer Open-Access-Community of Practice.

Aufgrund der pandemischen Situation werden die Train-the-Trainer-Workshops im Online-Format mit maximal 20 Teilnehmenden durchgeführt. Die Durchführung der Veranstaltung an zwei Vormittagen hat sich sowohl für die Aufnahmefähigkeit als auch generell für die Möglichkeit zur Teilnahme als sinnvoll erwiesen. Hauptzielgruppe sind Open-Access-Multiplikator*innen aus Bibliotheken (Open-Access-Beauftragte, Erwerbungsleiter*innen, Mitarbeiter*innen im Bereich elektronisches Publizieren etc.) sowie aus weiteren forschungsunterstützenden Bereichen (Forschungsförderung, Drittmittelverwaltung). Die Workshops dienen dem Empowerment der Multiplikator*innen. Über die Vermittlung didaktischer Kenntnisse und Trainingsmethoden hinausgehend, wird auch der Aufbau einer Community of Practice von Open-Access-Trainer*innen und -Multiplikator*innen ermöglicht, in der Erfahrungen ausgetauscht, voneinander gelernt und Materialien miteinander geteilt werden können.

Die Train-the-Trainer-Workshops folgen einem ähnlichen Ablauf. Vor Veranstaltungsbeginn wird eine sogenannte Vorabfrage über ein Pad (kollaboratives Schreiben im Online-Format) an die Teilnehmenden gestellt, in der diese ihre Wünsche und Erwartungen bzw. Fragen zu der Veranstaltung benennen

1 Forum open-access.network unter <<https://forum.open-access.network>>, Stand: 16.05.2022.

2 Helpdesk open-access.network unter <<https://open-access.network/services/helpdesk>>, Stand: 16.05.2022.

3 Recherchetool oa.finder unter <<https://finder.open-access.network>>, Stand: 16.05.2022.

können. Aufgrund der möglichst offenen Frageformulierung variieren die Antworten häufig stark von Wünschen nach einer Didaktik- und Methodikeinführung oder Good-Practice-Beispielen für ganz konkrete Probleme (wie z.B. dem Aufbau eines Publikationsfonds) bis hin zu grundsätzlichen Fragen nach zielgruppenspezifischen Schulungsangeboten in der eigenen Einrichtung. Diese Vorabfrage dient zum einen einer Orientierung zur Vorbereitung für den*die Trainer*in und zum anderen ermöglicht sie eine (Teil-)Ausrichtung der Veranstaltung an Bedürfnissen der Teilnehmenden. Als didaktische Grundlagen werden die Teilnehmenden für ihre eigene Rolle als Trainer*in und ihre Zielgruppe(n) sensibilisiert. Des Weiteren werden Lern- und Motivationsstrategien behandelt sowie verschiedene Veranstaltungsformate und Wege der Vermittlung thematisiert, um diese, je nach Zielgruppe und Lernziel, selbst einzusetzen. Didaktische Elemente werden in kurzen Intervallen im Wechsel mit aus-gesuchten Open-Access-Inhalten präsentiert. Ziel ist es, didaktische Methoden und Tools „erlebbar“ zu machen, damit die Teilnehmenden die didaktischen Grundlagen in Kombination von Online-Tools für Umfragen, virtuelle Pinnwände, Whiteboards oder kollaborative Schreibumgebungen (z.B. mit-hilfe von Mentimeter, Flinga, Tricider, Padlet) direkt anwenden können. Wichtige Bestandteile eines Train-the-Trainer-Workshops von open-access.network sind: ein Praxisbeispiel oder ein Inputvortrag zu Veranstaltungsbeginn, mehrere interaktive Einzel- und Gruppenarbeitsphasen im Wechsel und ein abschließendes Mini-Training, um Gelerntes direkt anzuwenden.⁴ Im Nachgang erhalten die Teilnehmenden eine aufbereitete Workshop-Dokumentation.

Als Inspiration für die Ausgestaltung dieses Formats dienten die Trainingshandbücher der Projekte FOSTER Open Science⁵ und FDMentor⁶, aus denen für die Community nachnutzbare Trainingshandbücher hervorgingen. Als weitere nachnutzbare Open Educational Resource reiht sich das kollaborativ erarbeitete und jüngst veröffentlichte “Open Access und wissenschaftliches Publizieren: Train-the-Trainer-Konzept”⁷ ein.

2. Einbettung im Projekt open.access.network – Warum sind Multiplikator*innen so wichtig?

20 Jahre nach der Berliner Erklärung und dem Bekenntnis zu Open Access fordert der Wissenschaftsrat in 2022 immer noch, Open Access als Standard des wissenschaftlichen Publizierens zu etablieren.⁸ Die Unterstützung der Transformation hin zu mehr Offenheit in der Wissenschaft an den Wissenschaftseinrichtungen zeichnet sich in Folge dessen auch aktuell in der Zunahme der Zahl von für die Themenfelder Open Access und Open Science zuständigen Personen ab.

4 Martin, Linda: Nachhaltig für Open Access begeistern – Oder: Wie erreiche ich die Wissenschaftler*innen meiner Einrichtung?, Zenodo, 2021, <<https://doi.org/10.5281/zenodo.5535297>>.

5 Bežjak, April; Clyburne-Sherin u.a.: Open Science Training Handbook (1.0), Zenodo, 2018, <<https://doi.org/10.5281/zenodo.1212496>>.

6 Biernacka, Katarzyna u.a.: Train-the-Trainer-Konzept zum Thema Forschungsdatenmanagement (4.0), Zenodo, 2021, <<https://doi.org/10.5281/zenodo.5773203>>.

7 Biernacka, Katarzyna u.a.: Open Access und wissenschaftliches Publizieren: Train-the-Trainer-Konzept, Zenodo, 2022, <<https://doi.org/10.5281/zenodo.6034407>>.

8 Wissenschaftsrat: Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access, 2022, <<https://doi.org/10.57674/fyrc-vb61>>.

Das Train-the-Trainer-Konzept bietet die Möglichkeit, die teilnehmenden Multiplikator*innen für die Bedürfnisse der Wissenschaftler*innen und die Angebote zu Open Access an der eigenen Einrichtung zu sensibilisieren. Gleichmaßen werden universell abrufbare Informationsangebote und Tools vorgestellt, die direkt in eigenen Fortbildungsangeboten vor Ort eingesetzt werden können. In den Einrichtungen, die aufgrund von bspw. Ressourcenknappheit wenige Anlaufstellen zur Verfügung stellen können, böte sich wiederum eine gut präsentierte, ausführliche Homepage zur Beantwortung erster Fragen der Publizierenden an. Ein wichtiger, stetig im Workshop vermittelter Leitgedanke betont die Bedeutung der Sichtbarkeit und Ansprechbarkeit zuständiger Personen für Fragen der Wissenschaftler*innen an der eigenen Einrichtung. Um Wissen zu vermitteln und sich als Ansprechperson für Fragen zu Open Access vorzustellen, empfehlen wir den Teilnehmenden im Besonderen kurze Informationseinheiten, die die Zielgruppen bedürfnisorientiert ansprechen. Auch eignen sich fach-, themen- oder karrierespezifische Veranstaltungen, um über die Relevanz von bzw. die Angebote zu Open Access an der eigenen Einrichtung zu informieren. Innerhalb der Veranstaltungen lernen die Teilnehmenden Wege kennen, Wünsche und Fragen oder Herausforderungen der eigenen Zielgruppe zu erkennen. Dazu gehören unter anderem die Sondierung der Fragen aus der alltäglichen Beratungspraxis, das Versenden von Umfragen über Mailinglisten oder offen gestaltete, kurze Veranstaltungen à la “bring your own problem”.

Open Access ist ein Querschnittsthema, dessen mannigfaltige Bereiche häufig nicht durch einzelne Open-Access-Beauftragte einer wissenschaftlichen Institution abgedeckt werden können. So ist es lohnenswert, Multiplikator*innen aus den Fachcommunitys oder aus administrativen Bereichen zu akquirieren und einzubinden, um fachspezifischen, detaillierten Anfragen begegnen zu können. Ein regelmäßiger Austausch dieser (fach- und themenübergreifenden) Multiplikator*innen mit den Open-Access-Beauftragten vor Ort bleibt unerlässlich. Die Train-the-Trainer-Workshops von open-access.network möchten darüber hinaus den institutionenübergreifenden Austausch von Erfahrungswerten in der Beratungspraxis einerseits und den Austausch von Good Practices andererseits befördern. Eine überregionale Vernetzung unterstützt Synergien, schafft Raum für neue Ideen der Vermittlung und verweist auf nachnutzbare Angebote.

3. Stimmen der Teilnehmenden

In den bisher zehn durchgeführten Workshops konnten über 150 Teilnehmende zu Multiplikator*innen weitergebildet werden. Zehn Gastredner*innen haben die Workshops mit ihren Vorträgen – direkt aus der Praxis oder aus laufenden Projekten – bereichert und sich den interessierten Nachfragen aus dem Plenum gestellt. Zielgruppen waren Open-Access-Beauftragte und Personen aus publikationsunterstützenden Bereichen von Universitäten, Fachhochschulen und Hochschulen für angewandte Wissenschaften, Fachinformationsdiensten, außeruniversitären Forschungseinrichtungen sowie jüngst Fachreferent*innen der Geistes-, Sozial- und Rechtswissenschaften und Fachreferent*innen der MINT-Fachbereiche.

In dem projektinternen, gestaffelten Evaluationsverfahren werden innerhalb einer ersten Evaluationsrunde – direkt nach dem Workshop – erste Rückmeldungen zu der Veranstaltung, den

Veranstaltungsinhalten und der Organisation eingeholt. Vor allem die Workshop-Atmosphäre mit wiederholten Gruppenarbeiten zum gegenseitigen Austausch wurde als sehr gewinnbringend wahrgenommen. Für die eigene Veranstaltungspraxis empfanden es viele Teilnehmende als hilfreich, didaktische Grundlagen zur Ausrichtung kennenzulernen sowie verschiedene Online-Tools anzuwenden. Auch Listen mit URLs zu Tools, aktuellen Open-Access-Projekten und weiterführenden Internetseiten wurden als Hilfe zur Selbsthilfe wahrgenommen. Die Teilnehmenden stellten fest, dass das „Schlüsselerlebnis“ das Aufdecken ähnlich gelagerter Probleme sei, das den Workshop so wertvoll machte. „Der Austausch von Ideen und Erfahrungen stärkt das Wir-Gefühl einer Open-Access-Community ungemein“⁹. Als verbesserungswürdig sahen die Teilnehmenden das etwas knappe Zeit- und Pausenmanagement an und äußerten den Wunsch, – bspw. in einer Fortsetzung der Veranstaltungsreihe – stärker in einzelne Themenfelder einsteigen zu wollen.

Im Abstand von mindestens vier Monaten erfolgt jeweils eine zweite Evaluationsrunde, um den langfristigen Erfolg des Veranstaltungsformats in der Praxis zu eruieren. Dabei wurden u.a. die Ausrichtung eigener Veranstaltungen inkl. adressierter Zielgruppen, die Nachnutzung von Materialien, der Ausbau von Kontakten und die Beurteilung des Formats „Train-the-Trainer“ erfragt. Bei einem insgesamt zurückhaltenden Rücklauf gaben knapp die Hälfte der Antwortenden an, zwischenzeitlich eigene Veranstaltungen wie z.B. Online-Vorträge, -Seminare und -Workshops für die Zielgruppen Wissenschaftler*innen verschiedener Karrierestufen, Bibliothekspersonal und die Öffentlichkeit ausgerichtet zu haben. Mehr als die Hälfte der Befragten hatte demnach von dem Ausbau fachlicher Kontakte im Nachgang des Workshops und den erlernten didaktischen Methoden profitiert. Alle an der zweiten Evaluation Teilnehmenden beantworten die Frage, ob der Workshop ihnen insgesamt geholfen habe, bessere Trainer*innen zu werden, mit „ja“ oder „eher ja“.

4. Train-the-Trainer – ein nachhaltiges Format?

Nach bisher zehn durchgeführten Train-the-Trainer-Workshops im Rahmen des Projekts open-access.network stellt sich die Frage: Wie nachhaltig ist das Format und wie kann Nachhaltigkeit in diesem Kontext aussehen? Die Train-the-Trainer-Workshops dienen in ihrem interaktiven Format dazu, ein Miteinander zwischen den Teilnehmenden und die Identifikation ihrer selbst mit der Rolle als Trainer*in zu erzeugen. Durch die aktive Einbindung der Teilnehmenden und den Erfahrungsaustausch wird die Motivation für die Arbeit im Themenfeld Open Access an der eigenen Institution und darüber hinaus gefördert. Die Trainings von open-access.network unterstützen Multiplikator*innen, bilden sie weiter und tragen somit zu einer Stärkung der Open-Access-Community im deutschsprachigen Raum bei. Dem Wunsch nach einer Fortsetzung der Reihe und einer thematischen Vertiefung entsprechend, wäre die Weiterentwicklung des Formats hin zu „Train-the-Expert“¹⁰ überlegenswert.

9 Martin, Linda: Dritter Workshop von open-access.network zum Schwerpunkt „Open Access-Beauftragte an Universitäten und Hochschulen – Multiplikator*innen für eine offene Wissenschaft“, 2021, Online: <<https://open-access.network/services/news/artikel/open-access-staerken-wege-der-vermittlung-fuer-die-eigene-einrichtung-gemeinsam-entdecken>>, Stand: 16.05.2022.

10 Biernacka, Katarzyna u.a.: Open Access und wissenschaftliches Publizieren: Train-the-Trainer-Konzept, Zenodo, 2022, <<https://doi.org/10.5281/zenodo.6034407>>, S. IV.

Sabrina Stockhusen, Universität Bielefeld

Linda Martin, Open-Access-Büro Berlin, Freie Universität Berlin

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5829>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#).

Bericht aus der 82. Sitzung der Arbeitsgemeinschaft der Verbundsysteme am 28. April 2022

Aufgrund der Corona-Pandemie musste auch diese Sitzung als Videokonferenz stattfinden, die Tagesordnung wurde entsprechend angepasst. Der folgende Bericht hat den Stand April 2022.

Alma

OBV

Im OBV ist mit der Umstellung des letzten Lokalsystems der Umstellungsprozess nach Alma abgeschlossen. Die Produktionsaufnahme erfolgte Anfang Dezember 2021. Mit der Umstellung auf Alma sind nun nach beinahe 35 Jahren alle bundesstaatlichen Universitäten unter einem technologischen Dach mit den entsprechenden Vorteilen und den Möglichkeiten der Zusammenarbeit vereint. Bis zum Projektende wurden auch die letzten bisher auf Aleph bzw. das MAB-Format angewiesenen Anwendungen auf Alma und MARC 21 umgestellt. Der OBV und SLSP sind in engem Austausch zu Alma, da der Betrieb des Produkts „swisscovery“ unter Verwendung von Alma und Primo läuft.

SLSP

Der Betrieb des Produkts „swisscovery“ (Service Platform mit Alma und Primo) verläuft erwartungsgemäß gut. Die Mitarbeitenden in den Bibliotheken sind mittlerweile mit den neuen, zentralisierten Workflows vertraut. Neben der Konsolidierung von Workflows und Services hat die SLSP damit begonnen, die Weiterentwicklung von Services und weiteren Innovationen gezielt voranzutreiben. Es sind bereits neue Bibliotheken hinzugekommen und ab April 2022 nimmt auch der Bibliotheksverbund Alexandria, ein Verbund der Bibliotheken der schweizerischen Bundesbehörden, mit seinen 15 Bibliotheken an swisscovery teil.

hbz

Pünktlich zum Jahresbeginn 2022 wechselten die Bibliotheken der ersten Welle in den Alma-First-Level-Support des hbz. Bevor es weiter in die Cutover-Phase (Datenübernahme aus einem Altsystem in ein Produktivsystem) geht, werden von April bis Mai Workshops stattfinden, um dort die anstehenden wichtigen Fragen zu besprechen. Ab September 2022 wird die zweite Welle vorbereitet. Der Abschluss des NRW-Projektes ist nach einer weiteren Welle in 2023 vorgesehen.

bwLastCopies

Das gemeinsame Projekt bwLastCopies von BSZ und KIM Konstanz konnte Ende Dezember 2021 erfolgreich beendet werden. Der im Projekt entwickelte Workflow für die maschinelle Ermittlung seltener Bestände in den wissenschaftlichen Bibliotheken Baden-Württembergs über den K10plus konnte erfolgreich in die Routine übernommen werden. In diesem Workflow wird halbjährlich der aktuelle Stand der Bestände ermittelt und die so identifizierten, potenziell seltenen Titel werden anschließend im K10plus entsprechend gekennzeichnet bzw. die bestehenden Kennzeichnungen werden ggf. an die neue Situation angepasst (z.B. bei geänderter Anzahl).

Die Bibliotheken können auf Anfrage Listen ihrer seltenen Bestände erhalten und außerdem über ein eigens neu eingerichtetes Web-Tool Listen auszusondernder oder zu überprüfender Bestände hinsichtlich potenzieller Seltenheit prüfen lassen. Aktuell konnten so bereits über 2 Mio. Titel retrospektiv mit Archivierungszusagen zu beispielsweise Pflichtexemplaren, Dissertationen oder Titeln aus besonderen, bibliothekseigenen Sammlungen versehen werden.

Culturegraph und ORCID

Weitergeführt wird die Auswertung der aus Culturegraph erstellten Werkbündel zur Anreicherung von DNB-Datensätzen mit verschiedenen Metadatenelementen. Die Ergebnisse der Culturegraph-Werkbündelung sind jetzt öffentlich verfügbar.¹ Es handelt sich um eine Aggregatdatei der Werkbündel in MARCXML-ähnlichem Format, die monatlich bereitgestellt wird.

Seit Februar 2022 werden täglich aus ORCID-Claimings Vorschläge für neue Personennormdatensätze erzeugt und von einzelnen Kolleg*innen in DNB bearbeitet. Zu Beginn wurde außerdem eine Retro-Bearbeitung bereits geclimter DNB-Datensätze vorgenommen, bei der ca. 3.400 Vorschlagssätze erzeugt wurden. In Zukunft sollen Vorschlagssätze routinemäßig in der DNB-Formalerschließung weiterbearbeitet werden und perspektivisch ebenfalls von externen GND-Redakteur*innen. Eine Ausweitung des Verfahren auf andere Quellen für GND-Vorschlagssätze ist in Planung.

DeepGreen

Anfang des Jahres hat das DeepGreen-Projektteam den DFG-Abschlussbericht zur zweiten Förderphase von DeepGreen veröffentlicht.² Durch DeepGreen wurde eine technische und organisatorische Lösung zur automatisierten Verteilung von Artikeldaten wissenschaftlicher Verlage an institutionelle und fachliche Repositorien entwickelt. In der zweiten Projektphase lag der Fokus auf der Erprobung der Datendrehscheibe in der Praxis und der Ausweitung auf weitere Datenabnehmer und Verlage.

Eine neue Webseite³ informiert über DeepGreen. Alle Neuerungen sind in einem Screencast erklärt. Im März dieses Jahres ist zudem der Artikel „DeepGreen. Eine Infrastruktur für die Open-Access-Transformation“ in o-bib erschienen.⁴

1 Culturegraph-Werkbündelung, <<https://data.dnb.de/culturegraph/>>, Stand: 26.07.2022.

2 Bertelmann, Roland; Koch, Thorsten; Ceynowa, Klaus u.a.: DeepGreen: Etablierung und Weiterentwicklung rechtssicherer Workflows zur effizienten Umsetzung von Open-Access-Komponenten in Lizenzvereinbarungen für wissenschaftliche Publikationen. Abschlussbericht, Januar 2022. Online: <<https://nbn-resolving.org/urn:nbn:de:0297-zib-85420>>.

3 DeepGreen, <<https://info.aa-deepgreen.de/>>, Stand: 26.07.2022.

4 Boltze, Jana; Höllerl, Annika; Kuberek, Monika: DeepGreen. Eine Infrastruktur für die Open-Access-Transformation, in: o-bib 9 (1), 2022, S. 1–13. Online: <<https://doi.org/10.5282/o-bib/5764>>.

Deutsche Digitale Bibliothek

Für das Gemeinschaftsprojekt „Deutsche Digitale Bibliothek“ (DDB) sind an der Deutschen Nationalbibliothek (DNB) die Geschäftsbereiche Technik, Entwicklung und Service der gemeinsam mit der Stiftung Preußischer Kulturbesitz (SPK) wahrgenommenen Geschäftsführung angesiedelt.

Bis Ende Februar 2022 ist die Anzahl der Datenpartner auf 584 angestiegen, der Gesamtbestand der Objektnachweise belief sich auf rund 40 Mio. DDB-Objekte (davon ca. 14 Mio. mit Digitalisat). Der positive Trend bei der Zahl der Besucher*innen hat sich weiter fortgesetzt. Zwischen Oktober 2021 und Februar 2022 verzeichnete das DDB-Portal im Durchschnitt 15.600 eindeutige Besucher*innen pro Tag – ein Plus von 6.400. Auch die Nutzung des Archivportal-D hat weiter zugenommen. Durchschnittlich 3.010 Nutzer*innen pro Tag besuchten das Portal.

Am 28. Oktober 2021 wurde das Frontend-Release 7.2 veröffentlicht, mit dem neben dem Zeitungsportal auch die neue Startseite des DDB-Portals freigeschaltet wurde. Die neue Startseite ermöglicht es den Nutzer*innen des DDB-Portals, über explorative Einstiege einfach und komfortabel auf die vielfältigen Inhalte der DDB zuzugreifen. Redaktionelle Beiträge, virtuelle Ausstellungen und das Kalenderblatt werden auf der Startseite angeteasert und bekommen so deutlich mehr Sichtbarkeit, als dies mit der bisherigen Startseite möglich war. Auch die anderen Portale der DDB – Archivportal-D, das Deutsche Zeitungsportal und das Datengeber-Portal DDBpro – werden auf der neuen Startseite prominent verlinkt.

Digitaler Assistent

Der Digitale Assistent läuft im K10plus in kooperativer Betreuung durch die beiden Verbundzentralen und je einer Ansprechpartnerin aus Teilnehmerbibliotheken aus dem SWB und dem GBV im Routinebetrieb und wird von den teilnehmenden Einrichtungen rege genutzt. Ein Webauftritt für den Digitalen Assistenten ist in Vorbereitung und soll in den nächsten Monaten live gehen.

EBM-Tool

Das EBM-Tool, ein Dienstleistungsportfolio rund um E-Book-Metadaten, ist seit Juli 2020 für die Nutzung durch Bibliotheken und Verbundzentralen außerhalb des Bereichs des BSZ und des GBV freigegeben. Die Routineeinspielungen in den E-Book-Pool sowie die automatische Generierung von Bestandssätzen laufen stabil. Im ersten Quartal 2022 wurden ca. 500 EBM-Tool-Bestellungen der Bibliotheken für Metadaten erfolgreich vom BSZ bearbeitet. Die von Proquest und De Gruyter kurzfristig angebotenen Themenpakete zum Krieg in der Ukraine konnten dank der guten Zusammenarbeit von BSZ, GBV und den Datenlieferanten kurzfristig allen daran interessierten K10plus-Bibliotheken zur Verfügung gestellt werden.

GND-Dienste

Der Fokus der bisherigen Entwicklungsarbeiten für den GND-Explorer lag in der Darstellung und Konfigurierbarkeit von Grundinformationen zu einem GND-Datensatz in einem Faktenblatt sowie der Umsetzung einer Relationenvisualisierung. In 2022 sollen Suche und Trefferdarstellung optimiert sowie eine Hierarchievisualisierung umgesetzt werden. Eine erste Version mit Trefferliste, Relationen- und Hierarchievisualisierung soll im ersten Halbjahr 2022 präsentiert werden.

Bei den Arbeiten zum Thema Wikibase wurde der Fokus zunächst stärker auf die Dokumentationsplattform verlagert. In Zusammenarbeit mit dem beauftragten Dienstleister wurden Erweiterungen implementiert, die der besseren Darstellbarkeit der Datenbank-Inhalte dienen.

GND4C – GND für Kulturdaten

Die Projektpartner befinden sich in der zweiten Projektphase. In dieser zweiten Projektphase wird auf die Verstetigung und Erweiterung der Organisation und Gremienstrukturen sowie die Implementierung der entwickelten Konzepte und Werkzeug für den produktiven Einsatz fokussiert.

GOKb

Die Global Open Knowledgebase (GOKb) wird in Zusammenarbeit von hbz, VZG und der ZDB betrieben und weiterentwickelt. Die GOKb wird überregional von Bibliotheken für die kooperative Verwaltung von Metadaten zu Lizenzpaketen für E-Ressourcen genutzt. Die dort verzeichneten Daten sind für den maschinellen Austausch optimiert und frei zugänglich verfügbar, was eine Nachnutzung in ERM-Systemen unterstützt. Die GOKb dient beispielsweise den ERM-Komponenten von FOLIO als Quellsystem für Pakete und zuverlässig referenzierte Titelinformationen.

Als gemeinsames Projekt von hbz, VZG und ZDB ist es ein Ziel der GOKb, institutionell möglichst breit aufgestellt zu sein. Daher ist es eine große Bereicherung, dass seit Oktober 2021 auch die Elektronische Zeitschriftenbibliothek (EZB) vertreten ist.

Sowohl die Zahl der Nutzenden der GOKb als auch die in ihr enthaltenen Daten wachsen stetig. Die monatlich angebotenen GOKb-Infostunden werden gut angenommen und erweisen sich als geeignetes Forum, um themenorientiert über einzelne Aspekte zu informieren oder aufgetretene Probleme effizient zu klären.

Im November 2021 erschien die GOKb-Version 1.1. Der Einspielprozess für die KBART-Dateien ist dieser Version deutlich umgestaltet und läuft effizienter und stabiler. Auf Basis der farblich überarbeiteten Oberfläche können seitdem Produktiv- und Testsystem besser unterschieden werden. Die derzeit aktuelle GOKb-Version 1.2 ermöglicht die zentrale Anreicherung von Journaltiteln mit der EZB-ID.

OLE/FOLIO

Das Bibliotheksmanagementsystem FOLIO (The Future of Libraries is Open) ist eine cloudfähige Opensource-Softwarebasis für ein Bibliotheksmanagementsystem, das einfach über Zusatzservices erweitert werden kann. Mitglieder im deutschsprachigen Raum sind die Bibliotheksverbände GBV, hbz, hebis und BVB. Pilotbibliotheken sind die ZBW Kiel/Hamburg und die SuUB Bremen.

Das Ende 2021 veröffentlichte FOLIO Kiwi Release umfasste Neuerungen im Bereich des sogenannten Backends (z.B. für den Daten-Import) sowie Verbesserungen in der FOLIO-Bedienungsfläche. Zu den eingebauten neuen Funktionalitäten gehört die Löschfunktion eines Benutzenden-Datensatzes.

Das erste FOLIO-Lokalsystem wurde im Dezember 2021 an den WDR ausgeliefert und wird vom hbz gehostet. Vom hebis-Verbundrat wurden im Dezember 2021 weitere Entscheidungen für das FOLIO-Migrationsprojekt getroffen. Der Bundrat des BVB hat im September 2020 die Einrichtung einer BVB Task Force FOLIO (TFF) beschlossen, die bis 2022 ermitteln soll, ob ein Einsatz von FOLIO im BVB zu empfehlen ist. Erste Versuche, aus SISIS oder dem b3Kat exportierte Katalog- und Exemplaraten in FOLIO zu importieren, sehen vielversprechend aus.

Standardisierungsarbeit im deutschsprachigen Raum

RDA und das 3R-Projekt

Das von der DNB geleitete kooperative Projekt „3R für DACH-Bibliotheken“, das zum Ziel hat, ein Erschließungshandbuch für Bibliotheken des DACH-Raums unter Berücksichtigung der veränderten RDA-Regelungen zur erarbeiten, liegt im Zeitplan. Das Projektende ist für Dezember 2022 geplant. Eine Arbeitsgruppe befasst sich mit der Ausarbeitung eines Schulungskonzepts für die geplanten Anpassungsschulungen, die voraussichtlich im ersten Halbjahr 2023 stattfinden werden.

Als erste Communitys werden Ende des Jahres die Library of Congress gemeinsam mit dem Program for Cooperative Cataloging (PCC) in Nordamerika und die British Library den Umstieg auf das neue Toolkit in der Praxis vollziehen.

Die nächste Sitzung der Arbeitsgemeinschaft der Verbundsysteme findet am 17. November 2022 per Videokonferenz statt.

Edith Röschlau, Deutsche Nationalbibliothek

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5881>

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).

Aus der Deutschen Forschungsgemeinschaft

Die Sitzung des Ausschusses für Wissenschaftliche Bibliotheken und Informationssysteme (AWBI) am 24./25. Mai 2022 fand in hybrider Form statt. In Mittelpunkt der Diskussionen standen folgende Themen:

Klausurtagung des AWBI

Die letzte Klausurtagung des AWBI hat im Mai 2017 stattgefunden. Die Ergebnisse dieser Klausurtagung sind in dem Positionspapier „Förderung von Informationsinfrastrukturen für die Wissenschaft“ (März 2018)¹ festgehalten worden. Für Herbst 2023 plant der AWBI eine weitere Klausurtagung, in der er sich als übergeordnetem Thema mit den Wechselwirkungen bzw. der Verzahnung der DFG-Förderung mit anderen Akteuren im Bereich der wissenschaftlichen Informationsversorgung befassen wird. Hintergrund ist, dass die Wirksamkeit der DFG-Förderung auch von Voraussetzungen und Rahmenbedingungen abhängig ist, die jenseits der Reichweite der DFG-Förderung liegen. Dabei ist – wie bei allen DFG-Aktivitäten – zu berücksichtigen, wie die Perspektive von Wissenschaft und Forschung in die (Weiter-)Entwicklung von Informationsinfrastrukturen einfließen kann.

Fachinformationsdienste für die Wissenschaft (FID)

Entwicklung einer FID-Gesamtstruktur

Mit der Neuakzentuierung des Förderprogramms „Fachinformationsdienste für die Wissenschaft“ (FID) im Jahr 2020 ist die Etablierung einer FID-Governance-Struktur als weiteres Förderziel hinzugekommen. Das Förderziel „Aufbau und Entwicklung einer FID-Gesamtstruktur“ und der Beitrag der einzelnen FID dazu führt auch zu Anpassungen im Begutachtungs- und Bewertungsprozess der DFG. Entsprechende Aspekte in den Anträgen werden in allen Begutachtungssitzungen verstärkt in den Blick genommen und in der Bewertung vergleichend aufgerufen. Angesichts der Konsolidierungsphase, in der sich das Programm mittlerweile befindet, werden auch zentrale Indikatoren (wie Nutzungszahlen) verstärkt berücksichtigt. Eine Kontextualisierung der Nutzungszahlen wird von jedem Antrag erwartet.

Längerfristige Finanzierung der Fachinformationsdienste

Die bisherige Förderung der Fachinformationsdienste ist projektförmig mit einer jeweiligen Laufzeit von drei Jahren angelegt, wobei der maximale Förderzeitraum zwölf Jahre beträgt. Da infrastrukturelle Maßnahmen der Fachinformationsdienste häufig nicht innerhalb einer zwölfjährigen Laufzeit abgeschlossen werden können, soll eine Perspektive für eine längerfristige Finanzierung entwickelt werden, die ebenfalls in Projektform, aber mit längeren Laufzeiten organisiert sein soll. Dafür müssen einige grundlegende Fragen geklärt werden, wie beispielsweise: Für welche Aufgabenbereiche ist eine längerfristige Finanzierung erforderlich? Welche Kriterien und Indikatoren werden zur Auswahl

1 DFG: Förderung von Informationsinfrastrukturen für die Wissenschaft. Ein Positionspapier der Deutschen Forschungsgemeinschaft, März 2018, <https://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_informationsinfrastrukturen.pdf>, Stand: 04.08.2022.

dieser Aufgabenbereiche herangezogen? Wie und in welchen Zeiträumen soll eine qualitätssichernde Evaluierung erfolgen? Die Klärung dieser Fragen wird von einer vom AWBI eingesetzten Kommission begleitet. Zur inhaltlichen Ausarbeitung der einzelnen Themenfelder sind verschiedene Workshops und Rundgespräche geplant. Neben Mitgliedern des AWBI gehören der Kommission auch externe Expertinnen und Experten an, deren Kompetenzen in der Entwicklung von quantitativen und qualitativen Evaluierungskriterien und der Definition von Kennzahlen sowie bibliothekarischen Dienstleistungsportfolios der FID liegen. Ergänzt wird die Kommission durch Vertreterinnen bzw. Vertreter der AG FID sowie des FID-Lenkungsgremiums, die jeweils als Gäste an den Sitzungen teilnehmen.

Vernetzung der FID mit der Nationalen Forschungsdateninfrastruktur (NFDI)

Die Initiative des FID-Lenkungsgremiums, im November 2021 ein Rundgespräch zur Vernetzung von Fachinformationsdiensten mit den Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI) durchzuführen, hat der AWBI begrüßt. Aus Sicht des AWBI können sowohl die einzelnen Konsortien der NFDI als auch die Gesamtstruktur der NFDI von den Erfahrungen der FID profitieren. Angeregt wurde, den Rückkoppelungen mit den jeweiligen Fachcommunities, die ein zentrales Element der FID-Förderung darstellen, in den Abstimmungsprozessen mit der NFDI verstärkte Aufmerksamkeit zu widmen. Der AWBI versteht die Förderung von Fachinformationsdiensten und die Förderung von NFDI-Konsortien grundsätzlich als komplementär. Vernetzungs- und Abstimmungsprozesse zwischen einzelnen Fachinformationsdiensten und NFDI-Konsortien sind insbesondere dann erforderlich, wenn Fachinformationsdienste auch ausgewählte Angebote aus dem Bereich Forschungsdatenmanagement (z.B. Beratungsangebote, Nachweis von Forschungsdaten in FID-Portalen) in ihr Serviceportfolio integrieren.

Digitalisierung und Erschließung

DFG-Rundgespräch zu Inkunabeln

Auf Bitte des AWBI hat die Bayerische Staatsbibliothek München im Februar 2022 ein DFG-Rundgespräch zum Thema „Forschungsorientierte Inkunabelerschließung und -digitalisierung in deutschen Bibliotheken: Bedarfe und Perspektiven“ durchgeführt. Im Rahmen des Rundgesprächs wurde der aktuelle Sachstand der Erschließung, Digitalisierung und Präsentation von Inkunabelbeständen in Deutschland ausführlich erörtert und zusammengestellt. Damit liegt eine fundierte Basis vor, auf der zukunftsfähige Konzepte für Inkunabelbestände in Deutschland entwickelt werden können. Der AWBI regt an, in einem zweiten Schritt Überlegungen anzustellen, wie das Potenzial der Inkunabelbestände noch besser ausgeschöpft und neue Angebote für die Forschung gemacht werden können. Bedarf wird auch hinsichtlich der Konsolidierung und Aggregation vorhandener Daten gesehen. Dass im Nachgang des Rundgesprächs sowohl Handreichungen zur exemplarspezifischen Erschließung von Inkunabeln als auch Festlegungen erarbeitet werden sollen, welche Exemplare mehrfach digitalisiert werden sollen, wird vom AWB begrüßt.

Digitalisierung und Bereitstellung rechtebewehrter Objekte

Entsprechend dem Beschluss des AWBI aus dem September 2021 lud die DFG interessierte Einrichtungen im Dezember 2021 dazu ein, Interessensbekundungen für eine Beteiligung an einem

Koordinierungsprojekt zur Vorbereitung einer Pilotphase zur Digitalisierung rechtebewehrter Materialien einzureichen. Der AWBI zeigte sich erfreut darüber, dass der Aufruf auf eine hohe Resonanz gestoßen ist. Die Personen und Einrichtungen, die ihr Interesse an einer Beteiligung an einem Koordinierungsprojekt formuliert hatten, sind auf Einladung der DFG-Geschäftsstelle Anfang Mai 2022 zu einem Gespräch zusammengekommen, um die Konzeption eines Koordinierungsprojektes gemeinsam zu erörtern. Weitere Absprachen und Rollenverteilungen erfolgen nun im Vorfeld der geplanten Antragstellung für ein Koordinierungsprojekt in Eigenregie der Beteiligten.

DFG-Positionspapier „Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung“

Der AWBI hat das Mitte Mai 2022 veröffentlichte DFG-Positionspapier „Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung“² ausdrücklich begrüßt. Mit der dargelegten Auseinandersetzung zu gegenwärtigen Entwicklungen im Publikationswesen und der Darstellung der verschiedenen Herausforderungen und Handlungsfelder stellt es eine hervorragende Zusammenfassung der Entwicklungen der letzten Jahre dar. Es leistet einen Beitrag zur Stärkung wissenschaftsadäquater Rahmenbedingungen im Wissenschaftssystem und zu nationalen wie internationalen Diskussionen um die Fortentwicklung von Publikations- und Bewertungssystemen. Ziel ist es, einen Kulturwandel anzustoßen hin zu einer verbreiteten Bewertungsgrundlage sowie den Fokus auf eine inhaltliche Bewertung von Forschungsleistungen zu legen. Für die Arbeit des AWBI ist die Positionierung der DFG hinsichtlich der Auffindbarkeit von wissenschaftlichen Informationen und der Verankerung von Open Access von hoher Relevanz.

Action Plan for Diamond Open Access

Die DFG hat im März 2022 den von Science Europe, ANR (Agence Nationale de la Recherche), OPERAS (Open Scholarly Communication in the European Research Area for Social Sciences and Humanities) und der cOAlition S initiierten „Action Plan for Diamond Open Access“³ unterzeichnet. Die DFG ist dadurch Teil einer internationalen Community geworden, die sich der Umsetzung des „Action Plan“ widmet. Bisher zählen vorrangig Infrastrukturen zu den Unterstützern, weniger Fachgesellschaften. Daher hat der AWBI auf die wesentliche Bedeutung der Fachgesellschaften für das wissenschaftliche Publizieren hingewiesen. Die DFG kann hierbei aus Sicht des AWBI neben finanzieller Förderung auch koordinierend wirken und entsprechende Akteure zusammenbringen. Für den Herbst 2022 ist ein Rundgespräch zum Thema „Diamond Open Access“ geplant, bei dem vor allem aus der Perspektive von Infrastrukturen Bedarfe für die Förderung eruiert werden sollen. Grundsätzlich wird die Weiterentwicklung von Diamond-Open-Access-Infrastrukturen als langfristig relevantes Thema angesehen.

2 DFG: Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung. Herausforderungen und Handlungsfelder. Positionspapier, Mai 2022: <https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/publikationswesen/positionspapier_publikationswesen.pdf>, Stand: 04.08.2022.

3 Ancion, Zoé; Borrell-Damián, Lidia; Mounier, Pierre u.a.: Action Plan for Diamond Open Access. March 2022, <<https://doi.org/10.5281/zenodo.6282402>>.

Ulrike Hintze, Deutsche Forschungsgemeinschaft

Gruppe ‚Wissenschaftliche Literaturversorgungs- und Informationssysteme‘ (LIS)

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5882>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#).

Zuschrift von OCLC zum Beitrag „Das Lesen der Anderen“ von Renke Siems

Stellungnahme der Herausgebenden

Viel Aufmerksamkeit erregt hat der in Heft 1/2022 von o-bib veröffentlichte Aufsatz „Das Lesen der Anderen. Die Auswirkungen von User Tracking auf Bibliotheken“ von Renke Siems.¹ An einigen Stellen wird darin auch die Rolle von OCLC thematisiert, einem im internationalen Bibliothekskontext agierenden Anbieter von Daten, Software und Dienstleistungen. Vor kurzem erreichte uns ein Schreiben von OCLC mit der Bitte, der Fachöffentlichkeit eine „Richtigstellung“ zu diesem Beitrag zur Verfügung zu stellen.

Die Herausgebenden möchten zunächst auf das Qualitätssicherungsverfahren von o-bib hinweisen, bei dem die eingereichten Aufsätze systematisch einem fachwissenschaftlichen, sprachlichen und formalen Review unterzogen werden. Alle Fachbeiträge durchlaufen ein Peer Review durch zwei Begutachtende, die im Idealfall umfangreiche Erfahrungen im Themenfeld des Beitrags aufweisen. Diese beurteilen die Beiträge nicht nur aus fachlicher Sicht, sondern geben auch konstruktive Hinweise zu eventuellen Unklarheiten, Widersprüchen oder gar Fehlern. Ggf. ergänzt um Hinweise der geschäftsführenden Herausgebenden werden diese Gutachten den Autor*innen mit der Bitte um Berücksichtigung zur Verfügung gestellt. Nach der Einreichung eines auf dieser Grundlage überarbeiteten Textes erfolgt eine Prüfung seitens der geschäftsführenden Herausgebenden, inwieweit die Hinweise aufgegriffen wurden. Zusätzlich wird auf eventuell verbliebene Unklarheiten oder Inkonsistenzen sowie ggf. noch nicht korrekte Formalia geachtet. Erst nachdem alles verbessert ist, geht der Fachbeitrag ins Lektorat. Dort erfolgt durch eine*n Redakteur*in in Abstimmung mit der Chefredakteurin ein intensives sprachliches und formales Lektorat, das im Dialog mit den Autor*innen zu einer weiteren Optimierung des Texts führt.

Dieses u.U. mühsame und Zeit in Anspruch nehmende Verfahren hat sich in vielfacher Hinsicht bewährt und führt zu einer hohen Qualität der in o-bib veröffentlichten Beiträge. Dennoch können natürlich auch in einem solchen Verfahren Fehler übersehen werden oder sich womöglich sogar einschleichen. Vor diesem Hintergrund sind Hinweise durch Externe auf mögliche Ungereimtheiten in einem publizierten Beitrag immer ernst zu nehmen. Seitens der Herausgebenden von o-bib wurden deshalb die im Schreiben von OCLC kritisierten Punkte eingehend geprüft.

Wir, die geschäftsführenden Herausgebenden, sind zu dem Ergebnis gekommen, dass der Fachbeitrag nicht gegen die gute wissenschaftliche Praxis verstößt und keine Aussagen beinhaltet, die einer Richtigstellung bedürfen. Eine Verfälschung von Zitaten konnten wir nicht erkennen. Der Autor hat beispielsweise auch nicht behauptet, dass OCLC selbst für die University of Gloucestershire Daten aggregiert und analysiert habe. Fraglos sind manche Aussagen im Aufsatz sehr pointiert formuliert, werden aber wo nötig plausibel belegt. Subjektive Einschätzungen und Überlegungen des Autors, die sich auf mögliche zukünftige Szenarien beziehen, sind als solche erkennbar und als von der

1 Siems, Renke: Das Lesen der Anderen. Die Auswirkungen von User Tracking auf Bibliotheken, in: o-bib 9 (1), 2022, S. 1-25. Online: <<https://doi.org/10.5282/o-bib/5797>>.

Meinungsfreiheit gedeckt zu bewerten – zumal sich Positionen, die der des Autors ähneln, durchaus in der internationalen Fachdiskussion wiederfinden. So ist die Frage, wem eigentlich die in den WorldCat eingebrachten Katalogisate „gehören“ und was Bibliotheken damit machen oder nicht machen dürfen, seit vielen Jahren virulent und aktuell durch den Rechtsstreit zwischen OCLC und Clarivate wieder in den Vordergrund getreten. Und eine Primärkatalogisierung im WorldCat wurde auch für Deutschland schon diskutiert.

OCLC als Player im Ökosystem hat seinerseits selbstverständlich das Recht, seine von der Haltung des Autors abweichende Sichtweise zu artikulieren. Wir stellen deshalb das Dokument von OCLC, verfasst von Bénédict Ripperger, im vollen Wortlaut zur Verfügung. Sie können den Text von der Frontdoor-Seite dieses Beitrags herunterladen. Ebenso haben wir dem Autor die Möglichkeit gegeben, zu der von OCLC geäußerten Kritik Stellung zu nehmen. Auch diesen Text finden Sie auf der Frontdoor-Seite. Damit schaffen wir maximale Transparenz und geben allen Leser*innen die Möglichkeit, sich ein eigenes Bild von der Diskussion zu machen.

Stuttgart/Köln, Juli 2022

Heidrun Wiesenmüller und Achim Oßwald
geschäftsführende Herausgebende von o-bib

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5879>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

Facetten eines Missverständnisses

Ein Debattenbeitrag zum Terminus „Diamond Open Access“

Einleitung

In der letzten Zeit nehmen wir eine zunehmende Verwendung der Bezeichnung „Diamond Open Access“ (im Folgenden „Diamond OA“) in der Bibliothekswelt wahr. Dabei fällt auf, dass mit „Diamond OA“ unterschiedliche Konzepte und Sachverhalte bezeichnet werden, die nicht immer dasselbe meinen. Der folgende Beitrag lädt zu einer Auseinandersetzung mit dem Terminus „Diamond OA“ ein. Wie wir darstellen werden, sehen wir die Verwendung der Bezeichnung mit Skepsis: Das liegt vor allem an den vagen und unterschiedlichen Definitionen, von denen sich manche sogar gegenseitig widersprechen. Dennoch scheint sich die Bezeichnung zunehmend zu institutionalisieren – es werden sogar Fördergelder für „Diamond OA“-Projekte bereitgestellt, was eine Klärung des Terminus umso dringlicher macht. Wir möchten ein Bewusstsein für die unscharfe Verwendung der Bezeichnung wecken und anregen, „Diamond OA“ präziser zu definieren, um dadurch zielgerichteter über das jeweilige Anliegen diskutieren zu können. Um es vorweg deutlich zu sagen: Die verschiedenen Konzepte, die mit „Diamond OA“ beschrieben werden – Author Processing Charges (APC)-freies OA und seine Finanzierung, Publizieren und Publikationsinfrastrukturen in den Händen der Wissenschaft – finden wir überaus unterstützenswert. Unsere Kritik richtet sich explizit nicht gegen solche Publikationsmodelle. Auch begrüßen wir nachdrücklich das verstärkte Interesse in Bibliothekswelt und Forschungsförderung an Open-Access-Geschäftsmodellen und Infrastrukturen des wissenschaftlichen Publizierens. Die Bezeichnung „Diamond OA“ ist unserer Ansicht jedoch nur bedingt dazu geeignet, Klarheit über die zu erreichenden Ziele zu schaffen, insbesondere wenn sie mit ebenso mehrdeutigen Bezeichnungen wie „community-owned“ oder „scholar-led“ kombiniert wird.

Im Folgenden werden wir Beispiele zur Verwendung des Terminus anführen und die Felder identifizieren, in denen der Terminus verwendet wird. Hier kombinieren wir Beobachtungen mit unseren Einschätzungen. Am Ende plädieren wir dafür, an der Unterscheidung von „Gold“ und „Grün“ für Open-Access-Publikationen festzuhalten und schlagen vor, verschiedene Attribute eines Publikationsangebots präziser herauszuarbeiten. Hilfreicher als die Einführung neuer Sammelbezeichnungen ist unseres Erachtens die Entwicklung einer strukturierten Beschreibung verschiedener Eigenschaften eines Publikationsmodells: Konkrete Angebote können mit klaren Kriterien genauer beschrieben und beurteilt werden; Publikationsoptionen für die Zielgruppe Wissenschaftler*innen werden so deutlicher umschrieben und einfacher vermittelbar.

Neue Aufmerksamkeit für „Diamond OA“

Allen uns bekannten Definitionen ist gemein, dass sie ein Publikationsmodell bezeichnen, welches Open-Access-Erstveröffentlichungen ohne Kosten für Autor*innen und Leser*innen meint. Neue Bezeichnungen werden in der Regel eingeführt, wenn etwas Neues auf den Plan tritt. Sie definieren damit nicht nur das Neue, sondern grenzen es auch vom Alten ab. Im Fall der aktuellen Aufmerksamkeit für „Diamond OA“ sind allerdings weder das Phänomen noch die bibliothekarische Beschäftigung damit

neu, sondern gehen bereits mindestens 20 Jahre zurück.¹ Es ist sicher nicht ganz zufällig, dass „Diamond OA“ zu einem Zeitpunkt an Popularität gewinnt, an dem die Skalierbarkeit, Nachhaltigkeit und Finanzierbarkeit der großen Transformationsverträge durch führende Wissenschaftsvertreter*innen bezweifelt wird, die Preise von APCs auch finanzstarke Einrichtungen vor Probleme stellen² und die weite Verbreitung des APC-Modells als eine Barriere für Autor*innen aus finanzschwachen Ländern und Einrichtungen wahrnehmbarer diskutiert wird.³ Die weitere Zunahme der Marktkonzentration und die selbstverschuldete Abhängigkeit des Wissenschaftsbetriebs von kommerziellen (Publikations-) Infrastrukturen als Bedrohung der Wissenschaftsfreiheit wird ebenfalls breiter diskutiert, während ein Kulturwandel zur Öffnung der Wissenschaft weiter voranschreitet.⁴ Diese Unzufriedenheit hat sicherlich einen Anteil an der weiten Beachtung der von Science Europe und cOAlition S beauftragten Studie *The OA Diamond Journal Study. Exploring collaborative community-driven publishing models for Open Access*.⁵ Die Studie brachte die „Diamond“-Bezeichnung weiter in Umlauf und scheint derzeit mitunter als hoffnungsvolle Perspektive einer Open-Access-Welt zu dienen, die mit positiven Attributen wie „nicht-kommerziell“, „scholar-led“, „verlagsunabhängig“ oder „wissenschaftsgetrieben“ die Frustration über den derzeitigen Entwicklungsstand auflöst. Die Frustration mit dem Status Quo teilen wir. Wenn wir aus dieser Unzufriedenheit Kriterien für Ziele der Open-Access-Transformation ableiten, wären das: mehr Open Access – für alle finanzierbar, in den Händen der Wissenschaft. Mit der bisherigen Unbestimmtheit der Bezeichnung „Diamond OA“ und der Diversität der Interpretationen ist die Auseinandersetzung mit diesen Zielen schwerfällig, wodurch sie (noch?) kein geeignetes Instrument zur Erreichung der Ziele darstellt. Um das zu verdeutlichen, diskutieren wir im Folgenden einige beispielhaft zusammengestellte Definitionen der Bezeichnung.

- 1 Haschak, Paul G.: The 'platinum route' to open access. A case study of E-JASL: The Electronic Journal of Academic and Special Librarianship, in: *InformationResearch* 12 (4), 2007, <<http://informationr.net/ir/12-4/paper321.html>>, Stand 08.07.2022.
- 2 Dies trat beispielweise in der vom VDB organisierten Panel-Diskussion „Wie finanzieren wir die Open Access Transformation?“ auf dem Bremer Bibliothekartag 2021, 17.06.2021, deutlich zutage. Siehe <<https://dbt2021.abstractserver.com/program/#/details/presentations/379>>, Stand 08.07.2022.
- 3 Siehe hierzu Czerniewicz, Laura: Open access and social justice (Vortragsfolien). Zenodo, <<https://doi.org/10.5281/zenodo.5543440>>; Demeter, Márton; Istratii, Ronina: Scrutinising What Open Access Journals Mean for Global Inequalities, in: *Publishing Research Quarterly* 36 (4), 2020, S. 505–522, <<https://doi.org/10.1007/s12109-020-09771-9>>; Al Hamzy, Murad; de Villiers, Dominique; Banner, Megan; Lamprecht, Hein; Bruijns, Stevan R.: Access to Top-Cited Emergency Care Articles (Published Between 2012 and 2016) Without Subscription, in: *Western Journal of Emergency Medicine* 20 (3), 2019, S. 460–465, <<https://doi.org/10.5811/westjem.2019.2.40957>>; Asare, Samuel; Mitchell, Rafael; Rose, Pauline: How Accessible Are Journal Articles on Education Written by Sub-Saharan Africa-Based Researchers?, in: *Development and Change* 52 (3), 2021, S. 661–669, <<https://doi.org/10.1111/dech.12639>>; Smith, Audrey C., Merz, Leandra ; Borden, Jesse B.; Gulick, Chris; Kshirsagar, Akhil R.; Bruna, Emilio Miguel: Assessing the Effect of Article Processing Charges on the Geographic Diversity of Authors Using Elsevier's 'Mirror Journal System', in: *MetaArXiv*, <<https://doi.org/10.31222/osf.io/s7cx4>>; Burchardt, Jørgen: Researchers Outside APC-Financed Open Access, in: *SAGE Open* 4 (4), 2014. Online: <<https://doi.org/10.1177/2158244014551714>>.
- 4 Georg Fischer: Interview mit Björn Brems. Großverlage arbeiten daran, „den wissenschaftlichen Workflow zu monopolisieren“, in: *irights.info* (08.10.2021), <<https://irights.info/artikel/bjoern-brems-grossverlage-arbeiten-daran-den-wissenschaftlichen-workflow-zu-monopolisieren/31142>>, Zugriff 08.07.2022; Siems, Renke: Das Lesen der Anderen, in: *o-bib* 9 (1), 2022, <<https://doi.org/10.5282/o-bib/5797>>; Deutsche Forschungsgemeinschaft: Datentracking in der Wissenschaft. Aggregation und Verwendung bzw. Verkauf von Nutzungsdaten durch Wissenschaftsverlage. Bonn 2021, <https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_de.pdf>, Stand: 08.07.2022.
- 5 Bosman, Jeroen; Frantsovåg, Jan Erik; Kramer, Bianca; Langlais, Pierre-Carl; Proudman, Vanessa: OA Diamond Journals Study. Part 1: Findings. Zenodo 2021, <<https://doi.org/10.5281/zenodo.4558704>>; Becerril, Arianna; Bosman, Jeroen; Bjørnshauge, Lars; Frantsovåg, Jan Erik; Kramer, Bianca; Langlais, Pierre-Carl; Mounier, Pierre; Proudman, Vanessa; Redhead, Claire; Torny, Didier: OA Diamond Journals Study. Part 2: Recommendations. Zenodo. 2021, <<https://doi.org/10.5281/zenodo.4562790>>.

Diamond und Platin

In allen uns bekannten Definitionen werden „Diamond OA“ und „Platin OA“ synonym verwendet, wobei „Platin“ die ältere Bezeichnung zu sein scheint. 2007 schrieb Haschak über seine Erfahrungen der Gründung von *E-JASL: The Electronic Journal of Academic and Special Librarianship* im Jahr 2002: „It is my firm belief that all scholarly journal articles should be free and freely accessible. There never should be a charge to the readers, the authors, or the institutions for access. The editors of E-JASL are proud of what they have accomplished and they will continue to take a leadership role in the *platinum route* to open-access. Hopefully, the patterns of scholarly publishing will change to embrace the platinum route. Each of us has to do what we can to make this a reality.“⁶

Die (englischsprachige) Wikipedia und das im Januar 2022 gestartete Projekt „Platinum Open Access Funding (PLATO)“ setzen Diamond und Platin OA gleich: „Diamond open access refers to scientific articles published without charging authors for article processing charges. Alternative labels include platinum open access (...).“⁷

„PLATO ist ein Projekt, das gemeinsam von sechs Schweizer Universitäten initiiert wurde, um ein nachhaltiges Finanzierungsmodell zu entwickeln, das qualitativ hochstehendes Open-Access-Publizieren in der Schweiz ermöglicht. Platinum Open Access (auch bekannt als Diamond Open Access) steht für die Veröffentlichung und den Zugang zu wissenschaftlichen Publikationen ohne Kosten für Autor:innen und Leser:innen.“⁸

„Platin“ und „Diamond“ werden also synonym verwendet, aber ebenso wie Diamond OA ist Platin OA über die Kostenfreiheit für Leser*innen und Autor*innen hinaus nicht eindeutig definiert.

Diamond OA und „scholar-led“, „academic-owned“, „community-driven“, „non-commercial“ etc.

Unterschiedliche Ansichten gibt es zu der Frage, ob „Diamond OA“ intrinsisch nicht-kommerziell oder „wissenschaftsgeleitet“ ist. In diesem Aspekt werden Fragen des Geschäftsmodells in Verbindung mit Lizenzen und dem Copyright gebracht. So definiert die englischsprachige Wikipedia „Diamond OA“ als ein nicht-kommerzielles Geschäftsmodell das, durch die Gleichsetzung mit „commons“, mindestens implizit freie Lizenzen und bestimmte Eigentümerschaft am Journal beinhaltet: „Alternative labels include non-commercial open access, cooperative open access or, more recently, open access commons.“⁹ Ähnlich Fuchs und Sandoval in 2013: „In the Diamond Open Access Model, not-for-profit, non-commercial organizations, associations or networks publish material that is made available

6 Haschak, The 'platinum route' to open access.

7 Diamond Open Access, Wikipedia, <https://en.wikipedia.org/wiki/Diamond_open_access>, Stand: 08.07.2022.

8 Platinum Open Access Funding (PLATO), Universität Zürich, Online: <<https://www.openscience.uzh.ch/de/openaccess/plato.html>>, Stand: 08.07.2022.

9 Diamond Open Access, Wikipedia.

online in digital format, is free of charge for readers and authors and does not allow commercial and for-profit re-use.“¹⁰

Hier wird „Diamond OA“ als Gegensatz zu kommerziellem OA definiert und eine kommerzielle Nachnutzung sogar ausgeschlossen. Dies widerspricht nicht nur der „Diamond“-Definition der Beilstein-Zeitschriften, die „Diamond OA“ zwingend mit einer CC BY-Lizenz verbinden (siehe unten), sondern auch der Definition von Open Access aus der Berliner Open-Access-Erklärung:¹¹ „...grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship ...“¹²

Im 2022 erschienenen, von cOAlition S und Science Europe herausgegebenen „Action Plan for Diamond OA“ heißt es: „‘Diamond’ Open Access refers to a scholarly publication model in which journals and platforms do not charge fees to either authors or readers. Diamond Open Access journals represent community-driven, academic-led and -owned publishing initiatives. Serving a fine-grained variety of generally small-scale, multilingual, and multicultural scholarly communities, these journals and platforms embody the concept of bibliodiversity. For all these reasons, Diamond Open Access journals and platforms are equitable by nature and design.“¹³ Hier werden die Attribute „community-driven“, „academic-owned“ und „bibliodivers“ zusammengebracht. Dies ist aus operativer Sicht schwierig, da sich diese Facetten auf die Gesamtheit des „Diamond OA“-Ökosystems beziehen und sich nicht alle zur Beschreibung einzelner Zeitschriften nutzen lassen. Die Autor*innen leiten davon ab, dass „Diamond OA“-Zeitschriften inhärent „fair und gerecht“ seien. Wie wenig diese Attribute allerdings beim tatsächlichen Betrieb der Zeitschriften helfen, zeigt die Anmerkung unter der Überschrift „Nachhaltigkeit“ auf der nächsten Seite: „Although Diamond Open Access journals and platforms are scholar-owned and -led, their legal status and governance is often unspecified.“¹⁴ Diese Erkenntnis widerspricht dem Ideal, wie es von der Fair Open Access Alliance ausgeführt wird.¹⁵ Was also kann dann „scholar-led“ und „scholar-owned“ bedeuten, wenn es keinen klaren rechtlichen Status oder eine Rechtsform definiert? Provokativ gesagt, könnten auch klassische Subskriptionszeitschriften (und Open-Access-Zeitschriften, die den Autor*innen APCs in Rechnung stellen) bei kommerziellen Verlagen als „scholar-led“ und „wissenschaftsgeleitet“ bezeichnet werden, da meist Redaktion und inhaltliche Begutachtung durch aktive Wissenschaftler*innen geschehen und die Gründung der Zeitschrift oftmals auf eine Initiative von Wissenschaftler*innen zurückgeht. Welche

10 Fuchs, Christian; Sandoval, Marisol: The Diamond Model of Open Access Publishing. Why Policy Makers, Scholars, Universities, Libraries, Labour Unions and the Publishing World Need to Take Non-Commercial, Non-Profit Open Access Serious, in: triple C – Communication, Capitalism & Critique. Journal for a Global Sustainable Information Society 11 (2), 2013. Online: <<https://doi.org/10.31269/triplec.v11i2.502>>.

11 Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen, 22.10.2003, Max-Planck-Gesellschaft, <<https://openaccess.mpg.de/Berliner-Erklaerung>>, Stand: 07.07.2022.

12 Ebd.

13 Ancion, Zoé; Borrell-Damián, Lidia; Mounier, Pierre; Rooryck, Johan; Saenen, Bregt: Action Plan for Diamond Open Access. Zenodo. 2022, <<https://doi.org/10.5281/zenodo.6282403>>, S. 4.

14 Ebd., S. 5.

15 „The journal has a transparent ownership structure, and is controlled by and responsive to the scholarly community.“ Fair Open Access Alliance FOAA: The Fair Open Access Principles, <<https://www.fairopenaccess.org/the-fair-open-access-principles/>>, Stand 08.07.2022.

wissenschaftliche Zeitschrift, welcher wissenschaftliche Verlag wird das Attribut „scholar-led“ nicht für sich beanspruchen (wollen), zumal wenn dies Voraussetzungen für Zuwendungen aus Förderprogrammen und für Compliance mit Publikationsauflagen werden? Wenn wir Bezeichnungen wie „scholar-led“, „community-driven“ und „wissenschaftsgeleitet“ verwenden, sollten wir untereinander konsensfähige, klare Definitionen mit ihnen verbinden und sie nicht als Buzzwords in die Debatte einbringen. Nur dann können wir eine Vereinnahmung durch kommerzielle Großverlage erschweren und vielleicht sogar verhindern.

Es wäre wünschenswert, wenn übergeordnete Zusammenhänge wie z.B. Knowledge Exchange, cOAlition S oder Science Europe eine trennscharfe Definition vorschlagen und zur Diskussion stellen, die Orientierung durch die Vielfalt an Bezeichnungen bietet („scholar-led“, „scholar-owned“, „academic-led“, „academic-owned“, „wissenschaftsgeleitet“, „verlagsunabhängig“, „community-driven“, „non-commercial“). Es ist (uns) beispielsweise nicht klar, wo die Grenze zwischen „community-owned“, „verlagsunabhängig“ und „mit Verlag“ liegt. Sind nicht-kommerziell ausgerichtete Universitätsverlage per se „community-owned“ oder fallen sie bereits in die Kategorie „Verlag“? Sind Zeitschriften einer Fachgesellschaft, wie z.B. der Optical Society of America unterstützenswert oder nicht, weil sie die Einnahmen aus den Zeitschriften teilweise zur Querfinanzierung anderer Vereinsaktivitäten verwenden? Was ist der Status von großen Gold-OA-Verlagen, wie PLOS und Copernicus Publications in dieser Nomenklatur, die als Non-Profit organisiert sind, aber als Rechtsform z. B. eine GmbH sind? Gelten sie als „non-commercial“ und käme ein APC-freies Journal für „Diamond OA“-Fördergelder in Betracht oder wären sie ebenso ausgeschlossen wie offensichtlich profitorientierte Verlage wie z. B. MDPI? Sehen wir es als unsere Aufgabe in Bibliotheken, herauszubekommen, wem die GmbH gehört, wem sie verpflichtet ist und wer wirtschaftlich Berechtigte im Sinne des Geldwäschegesetzes sind? Müssen wir uns künftig durch Aktiengesellschaften und ihre häufig verschachtelte Unternehmenshierarchien hangeln, um herauszufinden, ob sie „scholar-led“ ist oder halten wir eine solche Rechtsform für inkompatibel mit „scholar-led“-Ansätzen? Gibt es eine Grenze dessen, was Angestellte eines Non-Profit-Verlags verdienen dürfen, ohne den moralischen Status „inhärent fair und gerecht“ aus dem Action Plan zu verlieren?

Auch „scholar-led“-Initiativen ist vermutlich mehr geholfen, wenn Finanzierung und Geschäftsmodell von anderen Aspekten in der Diskussion getrennt werden würden: Eine Umstellung der Finanzierung von publikationsbezogenen Ausgaben (APCs, Book Processing Charges, im Folgenden BPC) hin zu medienbezogener Finanzierung für APC/BPC-freies Open Access wird mit Änderungen des Konsortialgeschäfts einhergehen. Das ist gut und wichtig, liefert jedoch nicht unbedingt auch eine Lösung für (weitere) Bedarfe von „scholar-led“ oder „verlagsunabhängigen“ Publikationsinitiativen, die Unterstützung für den Betrieb der technischen Systeme, das Zeitschriften- und Metadatenmanagement, Hosting, verlegerische Dienstleistungen sowie Rechtsberatung benötigen.

Diamond und Publikationsinfrastruktur

Die DFG kündigt ihre Unterstützung für den *Action Plan für Diamond Open Access* wie folgt an: „Diamond Open Access“ ist ein Open-Access-Geschäftsmodell, bei dem weder für Publizierende noch für Lesende Gebühren anfallen. Die Publikationsinfrastruktur für die frei zugänglichen Erstpublikationen

wird von wissenschaftlichen Einrichtungen zur Verfügung gestellt oder durch fachlich organisierte Wissenschaftsverbände, wie zum Beispiel Fachgesellschaften, finanziert.“¹⁶ Und weiter: „Mit der Unterstützung des Aktionsplans wird die Koordination internationaler Aktivitäten für ein bedarfsgerechtes und wissenschaftsgeleitetes Publizieren im Open Access ohne Profitinteressen verstärkt.“¹⁷ Hier wird neben der Kostenfreiheit für das Lesen und Publizieren eine weitere entscheidende Eigenschaft mit „Diamond OA“ verknüpft, nämlich dass die Publikationsinfrastruktur durch wissenschaftliche Einrichtungen oder Wissenschaftsverbände zur Verfügung gestellt wird und geförderte Maßnahmen keine Profitinteressen vertreten. Was genau „bedarfsgerecht“ bedeutet, darum wird sicherlich noch gerungen werden. Die Tür für kommerzielle Produkte des Publizierens (Workflows, Plattformen, Journalmanagementsysteme) scheint jedoch nicht geschlossen, da es nur um das „zur Verfügung stellen“ geht, nicht um das „Besitzen“. Könnte sich also eine Universität einfach ein „Rundum-Sorglos-Paket“ eines kommerziellen Anbieters anschaffen oder viel Geld für den Imprint eines Verlags bezahlen, sich also weiterhin von kommerziellen Drittanbietern für das eigene Publizieren abhängig machen, und trotzdem mit der hier formulierten Diamond-Bezeichnung konform sein? Vermutlich würden in einem solchen Szenario die Kommunikation mit Herausgebenden, die Erklärung der Nutzung für neue Autor*innen und der First-Level-Support trotzdem auf Seiten der Wissenschaft liegen. Die Erfahrungen der Vergangenheit geben Grund zur Vermutung, dass für solche Angebote steigende Preise zu erwarten sind und der Finanzfluss von Wissenschaft zu kommerziellen Großverlagen auch unter „Diamond OA“ nicht durch die Verwendung der Bezeichnung allein ausgeschlossen ist. Wie einfach aus der Community gegründete Publikationsinfrastrukturen von profitorientierten Großverlagen übernommen werden können, zeigen beispielsweise die Übernahmen von Bepress durch RELX (dem Mutterkonzern von Elsevier) sowie von Knowledge Unlatched durch Wiley-VCH. Auch Dienste im Publikationsprozess wie Mendeley, Publons, Altmetric und Overleaf sollen nicht unerwähnt bleiben.¹⁸ Wenn die DFG erreichen möchte, dass Betrieb und Weiterentwicklung von Open-Source-Infrastrukturen des Publizierens durch öffentlich geförderte Wissenschaftseinrichtungen geschieht, müssten die Bedingungen expliziter benannt werden.

„Diamond“/Gold: Teilmenge oder andere Kategorie?

Das Verhältnis von „Diamond OA“ zu „Gold OA“ wird unterschiedlich bestimmt, und einige dieser Bestimmungen widersprechen sich. Im Groben lassen sich zwei Sichtweisen ausmachen: Einige sehen „Diamond OA“ als eine Teilmenge von „Gold OA“, andere nutzen es als eine autonome Kategorie zur Bezeichnung von für Autor*innen kostenfreien OA-Erstveröffentlichungen mit weiteren, wechselnd zugeschriebenen oder ausgeschlossenen Attributen. Leider sind auch innerhalb der oben zitierten *Open Access Diamond Journal Study* die Definitionen nicht konsistent. Im Ankündigungs-Post zur Studie heißt es auf der Website der cOAlition S: „community-driven open access (OA) journals across the

16 Deutsche Forschungsgemeinschaft: „DFG unterstützt ‘Action Plan for Diamond Open Access‘“, in: Ausschreibung – Information für die Wissenschaft (26), <http://www.dfg.de/foerderung/info_wissenschaft/2022/info_wissenschaft_22_26>, Stand 08.07.2022.

17 Ebd.

18 Siehe hierzu auch den Blog *Innovations in Scholarly Communication* von Bianca Kramer und Jeroen Bosman, <<https://101innovations.wordpress.com/tag/open-infrastructure/>>, Stand: 08.07.2022.

world that are free for readers and authors, usually referred to as 'OA diamond journals'.¹⁹ Hier wird „community-driven“ als eine definierende Eigenschaft von OA Diamond angeführt. Im Glossar der Studie heißt es: „OA diamond journals: Journals that publish without charging authors and readers, in contrast to APC Gold OA or subscription journal“²⁰ Hier wird OA Diamond nicht als Teilmenge, sondern als Kontrast zu „APC Gold OA“ definiert, Gold OA wird – zumindest implizit – mit APC gleichgesetzt. Im Vorwort der Studie definieren die Autor*innen das Anliegen: „... an analysis and overview of collaborative, community-driven open access journals and platforms (aka „OA diamond“)“.²¹ Hier wiederum fehlt der Kontrast zu APC Gold OA.

Noch 2017 definierte Hartington „Diamond OA“ als eine Teilmenge von Gold-OA: „Essentially, Diamond OA is a form of Gold OA that does not include a requirement for authors to pay article processing charges (APCs).“²²

In der Webinar-Einladung für die Veranstaltung „The Diamond Open Access Model: what impact on research?“, ausgerichtet von Academia Europaea Cardiff, KU Leuven Libraries and the Young Academy of Europe (28.03.2022) heißt es: „With escalating Article Processing Charges (APCs) under the Gold Open Access Model, attention has been turning to the Diamond Model, where scientific articles are not subject to APCs. Diamond journals represent a large percentage of open access publishing output and are mostly owned and managed by universities, learned societies and other not-for-profit organisations.“²³ Auch hier wird „Diamond OA“ per Definition von Gold OA abgesetzt und Gold OA auf ein Geschäftsmodell, nämlich mit APCs, beschränkt.

Die bereits oben erwähnten Zeitschriften des Beilstein Instituts (*Journal of Organic Chemistry* und *Journal of Nanotechnology*) definieren „Diamond OA“ in Abgrenzung zu Gold OA wie folgt: „What is diamond open access? Diamond open access means permanent and free access to published scientific works for readers and no publication fees for the authors – it is 100% free. All articles are published under the most flexible reuse standard – the CC BY license. Authors are not charged article processing fees or publication fees – no fees whatsoever. Importantly, authors retain copyright of their work and allow it to be shared and reused, provided that it is correctly cited. Readers anywhere in the world may download, share or use the work, free of charge. Besides diamond open access, there is the gold open access model, where the author pays a fee (article processing charge – APC) to the publisher to cover production costs, and the copyright holder (usually the author) grants readers permanent and royalty-free access to the final version of the published article. Unfortunately, some publishers place limits on the reuse of the articles, restrict text and data mining, hold the copyright (or only transfer it

19 cOAlition S: Diamond unearthed. Shining light on community-driven Open Access publishing, 09.03.2021), <<https://www.coalition-s.org/diamond-unearthed-shining-light-on-community-driven-open-access-publishing/>>, Stand: 08.07.2022.

20 OA Diamond Journals Study. Part 1: Findings, S. 12.

21 Ebd., S. 6.

22 Hartington, Robert: Diamond Open Access, Societies and Mission. The Scholarly Kitchen, 01.06.2017, <<https://scholarlykitchen.sspnet.org/2017/06/01/diamond-open-access-societies-mission/>>, Stand: 08.07.2022.

23 <https://cardiff.zoom.us/webinar/register/WN_6qYh3UoRSBKmLud8kkm1ow>, Stand: 23.03.2022, inzwischen offline.

upon additional payment), and even adopt the practice of allowing only 'temporary open access'.²⁴ Hier wird Diamond OA explizit von Gold OA abgesetzt und zusätzlich mit einer bestimmten Lizenz (CC BY) sowie dem Umstand gleichgesetzt, dass Autor*innen ihr Copyright behalten (welches nicht im deutschen, aber in anderen Rechtsräumen übertragbar ist). Gold OA hingegen wird als grundsätzlich mit APC verbunden und als potenziell weniger autor*innenfreundlich dargestellt.

Der Action Plan for Diamond OA zieht keine Vergleiche zu OA Gold, die Bezeichnung „Gold“ kommt in dem achtseitigen Dokument nicht vor. Die Tendenz, „Diamond OA“ unabhängig von Gold OA zu definieren, so scheint uns, nimmt zu. Wenn dem so ist, kann dies als Zeichen für seine Etablierung als eine eigenständige Kategorie gesehen werden, die sich nicht mehr als Teilmenge von Gold OA begreift.

Eine aktuelle Gegenposition zum „Diamond-Separatismus“ nimmt der Wissenschaftsrat ein. In seinen 2022 veröffentlichten „Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access“ empfiehlt der Wissenschaftsrat, die Bezeichnung „Gold OA“ zu behalten. Diese Empfehlung wird schon in der Kurzfassung sehr explizit formuliert: „Dieser sogenannte „Goldene Weg“ zum offenen Zugang (Gold Open Access) ist mit verschiedenen Geschäftsmodellen vereinbar. So können Publikationsdienstleister Einnahmen aus einer auf die einzelne Publikation bezogenen Gebühr beziehen. Aber auch Modelle für eine institutionelle Finanzierung von Reihen, Zeitschriften oder größeren Portfolios über einen längeren Zeitraum sind erfolgreich etabliert worden.“²⁵ Die Fußnote erläutert: „Teilweise werden solche Modelle als „Diamond OA“ von Geschäftsmodellen abgegrenzt, die auf Publikationsgebühren beruhen. Aus Sicht des Wissenschaftsrats sollten Zugangsregime und Geschäftsmodelle jedoch analytisch getrennt werden. Insbesondere ist „Gold OA“ nicht mit einer Finanzierung über Article Processing Charges (APC) gleichzusetzen.“²⁶ Diese Unterscheidung ist unserer Auffassung nach zentral: Wenn wir über Open Access sprechen, sollten wir zwischen Zugang und Geschäftsmodell differenzieren. Im Glossar des zitierten Papers heißt es zu Diamond OA: „Gängige Bezeichnung für OA-Publikationsorgane, die auf Publikationsgebühren bzw. Article Processing Charges (APC) verzichten. Äquivalent dazu ist die Kombination von Goldenem Weg mit einem medienbezogenen Geschäftsmodell.“²⁷ Gold OA bzw. Goldener Weg wird ebenda definiert als „Sofortige, unbefristet freie Zugänglichkeit der Version of Record am ursprünglichen Publikationsort unter einer offenen Lizenz.“²⁸

Für eine präzisere Diskussion

Ohne einen zusätzlichen Verwendungshinweis ist die Bezeichnung „Diamond OA“ derzeit nur sehr bedingt dazu geeignet, einen Sachverhalt präzise zu beschreiben. Analytisch ist die Bezeichnung „Diamond OA“ unscharf, da sie Erscheinungsweise bzw. Zugang (Erstveröffentlichung) und

24 Beilstein Journal of Organic Chemistry, Diamond Open Access, <<https://www.beilstein-journals.org/bjoc/openAccess>>, Stand: 08.07.2022, Hervorhebungen im Original entfernt. Wortgleich auf der Webseite: Beilstein Journal of Nanotechnology, <<https://www.beilstein-journals.org/bjnano/openAccess>>, Stand: 08.07.2022.

25 Wissenschaftsrat: Empfehlungen zur Transformation des wissenschaftlichen Publizierens zu Open Access; Köln, Januar 2022, <<https://doi.org/10.57674/fyrc-vb61>>, Stand 08.07.2022, S. 23.

26 Ebd., S. 8 (Fußnote 4).

27 Ebd., S. 104.

28 Ebd.

Geschäftsmodell (APC-frei) und in manchen Fällen auch weitere Eigenschaften zur Eigentümerschaft der Zeitschrift (nicht-kommerzieller Verlag / Verlag oder Publikationsinfrastruktur in Händen der scholarly community) kombiniert. Die Kombinationen variieren stark, widersprechen sich teilweise und werden oft mit ebenso vage definierten Aspekten ergänzt, weshalb außer sofortige, unbefristet freie Zugänglichkeit der Version of Record am ursprünglichen Publikationsort unter einer offenen Lizenz (= die Definition von Gold OA in der Berliner Erklärung) ohne Publikationskosten für Autor*innen nicht viel an Gemeinsamkeit übrig bleibt.

Wir erachten es als zielführend, der Empfehlung des Wissenschaftsrates, „Gold OA“ nicht mit einer Finanzierung über Article Processing Charges (APC) gleichzusetzen, zu folgen, um nicht zur (weiteren) Verwirrung beizutragen. Definitionen, die man nicht operationalisieren und für Entscheidungen anwenden kann, sind schließlich nutzlos. Um für alle Beteiligten deutlicher zu machen, was in den Diskussionen, die unter „Diamond OA“ stattfinden, genau gemeint ist, schlagen wir vor, mit der Definition von Gold OA zu beginnen – „Sofortige, unbefristet freie Zugänglichkeit der Version of Record am ursprünglichen Publikationsort unter einer offenen Lizenz“. Es spricht nichts dagegen, für die jeweiligen Zwecke Merkmale zu definieren, die für die eigenen Entscheidungsprozesse relevant sind. Dafür wäre unseres Erachtens eine Liste weiterer spezifizierender Kriterien mit weiteren Attributen besser geeignet als eine allgemeine Kategorie, die je nach Sachlage unterschiedlich ausgelegt oder neu definiert wird (und deshalb Vergleiche erschwert). Hierfür müssen wir nicht die mühsam etablierten Begriffe „Gold OA“ und „Grün OA“ verändern oder außen vor lassen. Weitere Facetten und Merkmale könnten sein:

Juristisches (Eigentum, Governance, Rechtsform)

- Copyright/Verwertungsrechte liegen bei den Autor*innen.
- Das Eigentum am Titel der Zeitschrift oder Schriftenreihe liegt bei den Herausgeber*innen oder einer Fachgesellschaft mit einer not-for-Profit Rechtsform.²⁹
- Transparentes Reporting über Ausgaben ist in der Satzung festgehalten.

Infrastrukturen

- Genutzte Systeme für Journal Hosting und redaktionelle Workflows sind im Open Source verfügbar, mindestens aber mit offenen Schnittstellen.
- Betrieb von Infrastrukturen, Hosting, Zugang und Bereitstellung werden von Wissenschaftseinrichtungen (Hochschulen oder Fachgesellschaften) oder bei nonprofit-Verlagen bzw. nonprofit-Drittanbietern geleistet.

Finanzierung

- Das Geschäftsmodell nutzt zur Finanzierung der Betriebskosten Open-Access-Konsortien, Crowdfunding, Querfinanzierung (z.B. durch Mitgliedschaftsbeiträge), ehrenamtliches Engagement (ggfs. mit Vergabe von Credit Points) oder eine Mischung dieser Optionen. Alternativ kann die Finanzierung durch eine Hochschule, ein Forschungsinstitut oder eine Fachgesellschaft geschehen.

²⁹ Fair Open Access Alliance, Principle 1.

- Die Organisation und die Abwicklung der Finanzierung erfolgt durch die Wissenschaftscommunity (Konsortialstellen in Bibliotheken, KOALA, Lyrisis, ...).

Datenschutz

- Zeitschriften verzichten auf Trackingsoftware; es besteht die Möglichkeit, technisch nicht notwendige Cookies abzulehnen.

Je nach Vorhaben kann diese Liste um weitere Facetten und Merkmale erweitert werden – beispielsweise zu technischen Anforderungen (wie etwa von Plan S definiert³⁰), zur Organisation der Qualitätssicherung oder zu ethischen Verpflichtungen, wie sie zum Beispiel von COPE vorgeschlagen werden.³¹ Ebenfalls relevant aber schwieriger messbar ist die Frage, wie eine Zeitschrift zur Bibliodiversität beiträgt und welche Maßnahmen Redaktionen, Verlage und Herausgeber*innen ergreifen, um sich für Diversität, Inklusion und Gleichheit in ihrer Community einzusetzen.

Weitere Probleme mit der Verwendung von „Diamond“ in Abgrenzung zu „Gold OA“

Es ist nicht nur eine Frage des eigenen Idealismus, ob man sich der Diskursmacht kommerzieller Verlage widersetzt, die die Gleichung „Erstveröffentlichung im OA = Gold OA = APC“ etabliert haben und darauf insistiert, dass Gold OA auch APC-frei möglich ist. Die Alternative wäre das Ausweichen in eine nicht-etablierte Kategorie: „Diamond OA“ wird zwar verstärkt in der Bibliothekswelt genutzt, aber es ist keine gängige Selbstbezeichnung von Zeitschriften.³² Wie wahrscheinlich ist es, dass unsere primäre Zielgruppe – Wissenschaftler*innen in ihrer Rolle als Autor*innen und Herausgeber*innen – die Bezeichnung „Diamond OA“ dankbar annehmen? Zusätzlich zu erwartbaren Kommunikationsproblemen in der Publikationsberatung kann das Streiten über die Interpretationshoheit der Kategorie „Diamond OA“ zu weiteren Problemen führen, die sich letztlich gegen die Intention der Diamond-Befürwortenden richten: Das jahrelange Lobbyieren für Open-Access-Policies an Hochschulen hat die Begriffe „Gold OA“ und „Grün OA“ etabliert. Keine uns bekannte Open-Access-Policy an deutschen Hochschulen nutzt derzeit die Bezeichnung „Diamond Open Access“, auch wenn Passagen ein Engagement beschreiben, welches unter der Bezeichnung „Diamond OA“ fallen könnte.³³ Auch in Förderrichtlinien sind die Bezeichnungen „Gold“ und „Grün“ etabliert.³⁴ Grenzt man „Diamond“ von „Gold“ ab, wird die Anerkennung zur Publikation in solchen Zeitschriften zumindest nicht explizit

30 cOAlition S: Principles & Implementation, Part III: Technical Guidance and Requirements, <https://www.coalition-s.org/technical-guidance_and_requirements/>, Stand: 08.07.2022.

31 Committee on Publication Ethics, <<https://publicationethics.org/>>, Stand: 08.07.2022.

32 „Finally, it is relevant to note that of all the OA diamond journals encountered in the sample, next to zero self-identify as being a diamond or no-APC journal.“ OA Diamond Journals Study. Part 1: Findings, S. 27.

33 Siehe zum Beispiel die OA Policy der TU Berlin: „Das Engagement für nicht-kommerzielle Angebote wird besonders befürwortet.“ TU Berlin: Open-Access-Policy der TU Berlin, 06.12.2017, <<https://www.tu-berlin.de/?191164>>, Stand: 08.07.2022.

34 Siehe z.B. die Richtlinien des EU Programms Horizon 2020: European Research Council (ERC): Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020, Version 1.1, 21.04. 2017, <https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf>, Stand: 08.07.2022, S. 7. „Beneficiaries can choose one of two main ways to meet this requirement: 1. Self-archiving / 'green' OA: (...) 2. Open access publishing / 'gold' OA: (...)“.

durch die Policy ermuntert. Eine Abgrenzung zwischen Diamond und Gold auf Ebene von Publikationen stellt keine Bereicherung für Diskussionen zu Umsetzungsstrategien von Open Access dar. Die Gefahr, dass „Diamond“-Publikationen aus dem Raster fallen, besteht vor allem in Einrichtungen, die „Gold OA“ in ihren Open-Access-Policies explizit erwähnen und als Beitrag zu Open Science besonders bewerten. Zumindest hier sollte die bibliothekarische Community dafür sorgen, dass „Diamond OA“ als Untermenge von „Gold OA“ konzipiert, etabliert und auch definiert wird.

Zweitens verstellt das Zusammendenken von „Diamond OA“ als „APC-frei und deswegen nicht-kommerziell“ den Blick dafür, dass auch Projekte mit kommerziellen Partnern APC-frei funktionieren können etwa (SCOAP³)³⁵ oder Verträge wie DEAL für die teilnehmenden Einrichtungen. Für die konsortiale Finanzierung im KOALA-Projekt³⁶ ist es keine Bedingung, dass sich Redaktionen von ihren Verlagen trennen, egal ob es sich um Großverlage oder Kleinverlage handelt.

In der „Documentation“ zur im Mai 2022 veröffentlichten Liste „Diamond Open Access Journals, Germany (DOAG)“ wird für die bibliometrische Analyse „Diamond OA“ als Teilmenge von Gold-OA definiert – als APC-freies Gold OA: „In order to verify the open access status of the journals of the list, it was matched against the ISSN-Gold-OA list 5.0, which is a curated evidence source for Gold OA journals. In a final step, the journals for which payment information were found in OpenAPC for 2020 and 2021 were removed from the list.“³⁷ „Diamond OA“ anders zu verstehen wäre allein aus forschungspragmatischen Gründen wenig zielführend gewesen.

Fazit

Die Bibliothekswelt hat lange daran gearbeitet, die Bezeichnungen „Gold“ für Open-Access-Erstveröffentlichungen und „Grün“ für Open-Access-Zweitveröffentlichungen in der Wissenschaftswelt zu etablieren. Alle anderen „Farbschattierungen“ (hybrid, bronze, schwarz, platin, usw.) sind außerhalb der Bibliothekswelt kaum bekannt. In der Kommunikation innerhalb der Bibliothekscommunity wird die Bezeichnung für eine Vielzahl unterschiedlicher Aspekte verwendet. Das scheint eine Verständigung auf gemeinsame Ziele zu erleichtern, da alle meinen über das zu reden, was ihnen selbst gerade vorschwebt. Etwaige Missverständnisse offenbaren sich dann aber spätestens in der Umsetzung. Wir ermutigen deshalb, die Bezeichnung „Gold OA“ in der Kommunikation für frei zugängliche Erstveröffentlichungen unabhängig vom Finanzierungsmodell des Publikationsorgans zu verwenden.

Um es noch einmal deutlich zu sagen: Mit der Formulierung unserer Bedenken hinsichtlich der Bezeichnung „Diamond OA“ verwerfen wir nicht die Idee APC-freier Erstveröffentlichungen oder des Publizierens in den Händen der Wissenschaft, im Gegenteil! Wir sind der festen Überzeugung, dass wissenschaftliche Publikationen dauerhaft offen und frei zugänglich sein sollen – für Leser*innen und für Autor*innen. Wenn finanzierbare Open-Access-Erstveröffentlichungen ohne Kosten für

35 Sponsoring Consortium for Open Access Publishing in Particle Physics - SCOAP³, <<https://scoap3.org/>>, Stand 28.09.2022.

36 Konsortiale Open-Access-Lösungen aufbauen - KOALA, <<https://projects.tib.eu/koala>>, Stand 08.07.2022.

37 Bruns, Andre; Taubert, Niels Christian; Cakir, Yusuf; Kaya, Sibel; Beidaghi, Samaneh: Diamond Open Access Journals Germany (DOAG), <<https://doi.org/10.4119/unibi/2963331>>, Stand: 08.07.2022. Bielefeld University 2022, S. 3.

Autor*innen unser Ziel sind, sind verschiedene Geschäftsmodelle denkbar. Wenn wir wollen, dass Eigentum und Betrieb von Infrastrukturen in die Hände der Wissenschaft gehören, müssen wir uns mehr mit Vergabe- und Vertragsrecht, mit Softwarelizenzen, der nachhaltigen Finanzierung von Infrastrukturen und der Anerkennung des Engagements in Infrastrukturen wissenschaftlichen Publizierens für akademische Karrieren beschäftigen. Lasst uns in konkreten, operationalisierbaren Bezeichnungen darüber sprechen, wie wir diese Ziele erreichen.

S. Dellmann, X. van Edig, J. Rücknagel, S. Schmeja, *Technische Informationsbibliothek (TIB), Hannover*

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5849>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#).

Qualität in der Inhaltserschließung / herausgegeben von Michael Franke-Maier, Anna Kasprzik, Andreas Ledl und Hans Schürmann. – Berlin, Boston: De Gruyter Saur, 2021. – VI, 420 Seiten : Illustrationen. – (Bibliotheks- und Informationspraxis ; Band 70). – ISBN 978-3-11-069149-8 : EUR 69.95 (auch als E-Book im Open Access verfügbar unter <https://doi.org/10.1515/9783110691597>)

Seit dem Jahr 2017 ist die Frage, welchen Qualitätskriterien die inhaltliche Erschließung von Informationsressourcen verpflichtet sein soll, eines der großen Themen in der bibliothekarischen Community des deutschsprachigen Raums. Damals hatte sich Klaus Ceynowa, der Generaldirektor der Bayerischen Staatsbibliothek, in einem Artikel, der in der Frankfurter Allgemeinen Zeitung erschienen war und schon allein dadurch ein hohes Maß an Aufmerksamkeit und Wirkmächtigkeit erreichte, kritisch mit der Inhaltserschließungspraxis der Deutschen Nationalbibliothek (DNB) auseinandergesetzt.¹ Im Fokus der Kritik standen die Resultate maschineller Inhaltserschließung, die nun nicht mehr nur in Bezug auf elektronische Publikationen, sondern auch auf die Reihen B und H der Deutschen Nationalbibliografie Anwendung fand und perspektivisch auf die Reihe A, also Verlagspublikationen, ausgeweitet werden sollte. Im Gleichklang mit Ceynowa äußerten sich zahlreiche Stimmen aus dem Bibliothekswesen kritisch zu den Ergebnissen des von der DNB zur Verschlagwortung eingesetzten maschinellen Verfahrens, wobei sich die eine oder andere Stellungnahme nicht nur durch einen hohen Erkenntniswert auszeichnete, sondern auch – eher ungewöhnlich für bibliothekarische Themen – wegen der integrierten Beispielsammlung besonders misslungener Erschließungsergebnisse einen beträchtlichen Unterhaltungswert aufwies.² Die anfangs gelegentlich etwas polemisch geratene Gegenüberstellung von intellektueller und maschineller Inhaltserschließung ist in den letzten Jahren dem Bestreben gewichen, beide Erschließungsvarianten miteinander zu harmonisieren. Grundlegend bleibt dabei aber die Frage nach der Qualität der inhaltserschließenden Daten, seien sie intellektuell oder maschinell generiert. Die Festlegung verbindlicher Maßstäbe, Standards und Kriterien, die eine qualitativ hochwertige Inhaltserschließung definieren, hatte Michael Franke-Maier bereits in einem 2018 erschienenen Artikel eingefordert, in dem er verschiedene Dimensionen von Qualität unterschied.³ Nun greifen er und seine Mitherausgeber*innen Anna Kasprzik, Andreas Ledl und Hans Schürmann dieses Thema in dem vorliegenden Band wieder auf.

- 1 Ceynowa, Klaus: In Frankfurt lesen jetzt zuerst Maschinen, in: Frankfurter Allgemeine Zeitung, 31.07.2017. Online: <https://www.faz.net/aktuell/feuilleton/buecher/maschinen-lesen-buecher-deutsche-nationalbibliothek-setzt-auf-technik-15128954.html>, Stand: 15.07.2022.
- 2 Vgl. Wiesenmüller, Heidrun: Maschinelle Indexierung am Beispiel der DNB. Analyse und Entwicklungsmöglichkeiten, in: o-bib 5 (4), 2018, S. 141–153, hier S. 143–149. Online: <https://doi.org/10.5282/o-bib/2018H4S141-153>, sowie dies.: RSWK reloaded – Verbale Sacherschließung im Jahr 2018, in: BuB 70 (1), 2018, S. 26–29, hier S. 28. Online: <https://nbn-resolving.org/urn:nbn:de:0290-opus4-160857>.
- 3 Franke-Maier, Michael: Anforderungen an die Qualität der Inhaltserschließung im Spannungsfeld von intellektuell und automatisch erzeugten Metadaten, in: ABI Technik 38 (4), 2018, S. 327–331. Online: <https://doi.org/10.1515/abitech-2018-4005>.

Insgesamt liefern die 18 hier versammelten Beiträge in eindrucksvoller thematischer Breite Vorschläge und Beispiele, wie Qualität in verschiedenen Bereichen bibliothekarischer Inhaltserschließung umzusetzen und zu gewährleisten ist. Angesichts der Vielfalt des behandelten Themenspektrums muss an dieser Stelle darauf verzichtet werden, jeden einzelnen der durchwegs höchst aufschlussreichen Artikel in seiner inhaltlichen Ausprägung darzustellen. Hierfür sei auf das Editorial des Bandes verwiesen, das alle Beiträge in der gebotenen Kürze präsentiert (S. 8–15) und – wie der gesamte Band – erfreulicherweise im Open Access zugänglich ist. Stattdessen sollen im Folgenden thematische Cluster gebildet und dabei ausgewählte Artikel vorgestellt werden, deren Inhalte das jeweilige Themengebiet repräsentieren.

Während das Gros der Beiträge gegenwartsorientiert Probleme der Qualitätssicherung aufzeigt, Lösungsansätze darlegt und Entwicklungsperspektiven vorschlägt, wählen zwei Aufsätze eine historische Perspektive. So bietet Andreas Ledl mit seinem Überblick über das Qualitätsstreben der bibliothekarischen Inhaltserschließung in den Jahren 1970 bis 2020 eine hilfreiche Hinführung zum Thema, wohingegen Ulrike Junger und Frank Scholze ihren Fokus enger fassen und die Entwicklung der Inhaltserschließungspolitik an der DNB seit 2010 referieren. Ihr Hauptaugenmerk gilt dabei dem Einsatz maschineller Erschließungsverfahren ebenso wie der Frage nach der Qualität der damit generierten Erschließungsergebnisse. Mit dem Verweis auf die „Strategischen Prioritäten 2021–2024“ der DNB geben sie jedoch auch einen Ausblick auf die Weiterentwicklung der Inhaltserschließungspolitik in den kommenden Jahren.⁴

Einen multiperspektivischen Zugang zu der Frage, was Qualität in der Inhaltserschließung ausmacht und wie diese zu gewährleisten sei, bietet das Expertenteam RDA-Anwendungsprofil für die verbale Inhaltserschließung (ET-RAVI), ein Gremium des Standardisierungsausschusses, in einem kurzen, aber ungemein gehaltvollen Beitrag. Hinsichtlich der Kriterien, die einer qualitativ hochwertigen Inhaltserschließung zugrunde liegen sollen, werden dabei vier Qualitätsdimensionen unterschieden: die Erfassungsregeln für das Generieren von Normdaten, die Verwendungsregeln für die inhaltliche Erschließung von Informationsressourcen, die transparente Auswertung inhaltserschließender Metadaten durch Retrievalsysteme und die Öffnung von Datenkorpora für die Nachnutzung in wissenschaftlichen Kontexten. Jede dieser vier Qualitätsdimensionen wird zudem durch eine differenzierte Auflistung konkreter Anforderungen spezifiziert.

Zwei Beiträge widmen sich dem Themenkreis Wissensbasen und Normdaten und ihrer Nutzung für eine effektive und nutzungsorientierte verbale Inhaltserschließung. Esther Scheven beschäftigt sich in diesem Zusammenhang mit der Gemeinsamen Normdatei (GND). Nach einem Überblick über deren Entstehung stellt sie verschiedene Maßnahmen zur Qualitätssicherung in der GND dar, der hinsichtlich der Öffnung der GND für nichtbibliothekarische Anwendergruppen und der sich daraus ergebenden Einspielung von Fremddaten eine hohe Bedeutung zukommt. Neben der GND findet verstärkt auch Wikidata als Wissensbasis Verwendung in inhaltserschließenden Projekten. Lydia Pintscher, Peter Bourgonje, Julián Moreno Schneider, Malte Ostendorff und Georg Rehm befassen sich in ihrem Beitrag mit der Nutzbarmachung von Wissensbasen im Rahmen automatisierter

4 Deutsche Nationalbibliothek: Strategische Prioritäten 2021–2024. Online: <<https://d-nb.info/1224705858/34>>.

Erschließungsverfahren und setzen sich konkret mit der Qualität der Daten in Wikidata auseinander. Sicherlich tragen das Grundprinzip der Offenheit und der Community-Charakter, denen Wikidata verpflichtet ist, zum weltweiten Erfolg und zur vermehrten Nutzung als Wissensbasis bei. Zu Recht wird jedoch auf die Gefahr hingewiesen, dass Wikidata aus diesem Grund auch verstärkt zum Ziel von Kampagnen werden könne, die im Rahmen von gezielten Sabotageakten Falschinformationen in Wikidata verankern. Dies stelle eine umso größere Herausforderung an die Community dar, welche die Qualität der Daten dauerhaft sicherstellen müsse.

Mit den Möglichkeiten des Datenaustauschs und der Anreicherung mit inhaltserschließenden Metadaten beschäftigt sich eine weitere Gruppe von Beiträgen. Während sich Cyrus Beck mit der Qualität der Fremddatenanreicherung im Kontext des Projekts FRED befasst, das 2017 von der Zentralbibliothek Zürich und den Universitätsbibliotheken Basel und Bern umgesetzt wurde, stellen Uma Balakrishnan, Stefan Peters und Jakob Voß in einem gemeinsamen Beitrag coli-conc vor. Coli-conc, von der Verbundzentrale des GBV entwickelt, bietet die technische Infrastruktur für das Erstellen von Mappings zwischen Wissensorganisationssystemen und unterstützt Datenaustausch- und Anreicherungsprozesse. Eine zentrale Anwendung stellt das Mapping-Tool Cocoda dar, mit dessen Hilfe auf einfache Weise intellektuelle Mappings zwischen Notationen verschiedener bibliothekarischer Klassifikationen generiert werden können. Auch der von Rita Albrecht, Barbara Block, Mathias Kratzer und Peter Thiessen gemeinsam verfasste Beitrag, der die Rolle der Bibliotheksverbünde bei der Optimierung der Inhaltserschließung in den Blick nimmt, fokussiert sich auf das Thema Datenanreicherung. Überzeugend bewerten die Verfasser*innen in diesem Zusammenhang Quantität und Qualität nicht etwa als einander gegenüberstehende Kategorien. Stattdessen verweisen sie mit Recht auf den Umstand, dass sich aus Perspektive der Nutzer*innen die Qualität bibliothekarischer Rechercheinstrumentarien in einem Information Retrieval ausdrückt, das einen hohen Recall liefert. Entscheidend sei demnach neben der inhaltlichen und formalen Qualität der inhaltserschließenden Metadaten auch ihre Quantität. Diese könne man auf unterschiedliche Arten in den Verbundsystemen steigern. Als Beispiele werden der Einsatz effizienter Sacherschließungstools wie etwa des Digitalen Assistenten (DA-3) oder von hebis-SET zur Unterstützung der intellektuellen Inhaltserschließung genannt, vor allem aber die Nachnutzung der Inhaltserschließung der DNB, der Datenaustausch zwischen den Verbänden und andere Anreicherungsprojekte. Besonders eindrucksvoll und leistungsstark erscheint in dieser Hinsicht das Projekt Culturegraph der DNB. Dabei handelt es sich um ein Werkclusterverfahren, das in einem ersten Schritt aus den Katalogdaten eines Verbundes oder aller deutschsprachigen Verbände die Metadaten der gesamten Manifestationen eines Werks zu einem Cluster zusammenführt. Anschließend werden die vorhandenen inhaltserschließenden Metadaten einer Manifestation auf alle anderen Manifestationen in diesem Cluster übertragen. Ausgaben und Auflagen ein und desselben Werks, die bisher nicht inhaltlich erschlossen waren, können auf diesem Wege effizient mit GND-Schlagwörtern, Schlagwortfolgen und Notationen, die von anderen Manifestationen aus demselben Cluster stammen, angereichert werden.

Eine Studie zur automatisierten Anreicherung von digitalisierten Volltexten mit inhaltserschließenden Metadaten aus der GND und aus Wikidata präsentiert ein aus Sina Menzel, Hannes Schnaitter, Josefine Zinck, Vivien Petras, Clemens Neudecker, Kai Labusch, Elena Leitner und Georg Rehm bestehendes Team von Autor*innen. Grundlage ihrer Untersuchung ist das Projekt SoNAR, in dessen

Rahmen Daten von Gedächtnisinstitutionen für die Bedarfe der Digital Humanities – konkret für die Historische Netzwerkanalyse – aufbereitet werden. Über maschinelle Verfahren werden dabei in einem Textkorpus zunächst Personen, Orte, Körperschaften, Ereignisse und andere Named Entities identifiziert (= Named Entity Recognition) und ausgezeichnet. Weiterhin werden die Named Entities mit dem entsprechenden Schlagwort in einer Normdatei verknüpft (= Named Entity Linking) und im Rahmen einer Volltextsuche recherchierbar gemacht. Der Qualitätsgewinn für Nutzer*innen liegt dabei in dem Mehrwert, den eine semantische Suche gegenüber einer Stichwortsuche im Volltext bietet. Analog zu einer Schlagwortsuche im Bibliothekskatalog kann man so gezielt im Textkorpus nach bestimmten Entitäten recherchieren und dabei etwa irrelevante Homonyme ausschließen. Außerdem enthält die Treffermenge neben dem gesuchten Begriff auch die in der betreffenden Normdatei aufgeführten Synonyme.⁵ Darüber hinaus kann auch nach weiteren Begriffen, die in Beziehung zu dem Suchbegriff stehen und in der Normdatei verzeichnet sind, recherchiert und somit der Suchfokus erweitert werden. Zudem lässt sich der Suchraum auf andere Datenbanken ausdehnen, die ebenfalls mit der betreffenden Normdatei verlinkt sind. Ein interessantes Ergebnis der vorgelegten Studie ist die Feststellung, dass sowohl die GND als auch Wikidata nur etwa die Hälfte der Entitäten enthalten, die für die Auszeichnung der im Rahmen einer Stichprobe ausgewerteten Volltexte notwendig gewesen wären. Vor allem zahlreiche (historische) Personen, Körperschaften und Ereignisse waren in den beiden Wissensbasen nicht verzeichnet.

Grundlage für eine effektive Nutzung digitalisierter Volltexte, die mit inhaltserschließenden Metadaten angereichert sind, ist dabei eine hohe Qualität der Optical Character Recognition (OCR), durch die der Text (retro-)digitalisierter Dokumente erst maschinell lesbar und verarbeitbar wird. Am Beispiel retrodigitalisierter Zeitungen zeigen Clemens Neudecker, Karolina Zaczynska, Konstantin Baierer, Georg Rehm, Mike Gerber und Julián Moreno Schneider in einem gemeinsamen Beitrag die Herausforderungen bei der OCR-Bearbeitung in Bezug auf Texterkennung, Layoutanalyse und die Erkennung von Strukturen und Textabschnitten auf. Anwendungsbezogen haben sie dabei ebenfalls vor allem die Bedarfe der Digital Humanities im Blick, deren Forschungsergebnisse in hohem Maße von einer qualitativ hochwertigen OCR-Bearbeitung der untersuchten digitalen Textkorpora abhängig sind.

Nicht mit dem bedeutungstragenden Inhalt, sondern mit der Form inhaltserschließender Daten, deren Qualität sich auch in ihrer Lesbarkeit und Verarbeitbarkeit ausdrücke, beschäftigt sich Jakob Voß. Um die Nachnutzung und den Austausch vorhandener Inhaltserschließungsdaten zu gewährleisten, seien nicht nur Regeln für die Befüllung der einzelnen Datenfelder festzulegen (und einzuhalten), sondern diese seien auch angemessen zu dokumentieren.

Die vom ET-RAVI geforderte Optimierung des Information Retrievals durch eine Verbesserung der Recherchewerkzeuge, um eine effektive Auswertung aller inhaltserschließenden Metadaten zu ermöglichen sowie den Endnutzer*innen erweiterte Suchmöglichkeiten zu bieten, stellt einen weiteren Qualitätsaspekt dar, der in mehreren Beiträgen aufgegriffen wird. Heidrun Wiesenmüller, die sich in ihrem Artikel mit den Funktionen inhaltlicher Erschließung befasst, unterscheidet bezüglich

5 Zu den Vorteilen einer semantischen Suche siehe im selben Band auch Sack, Harald: Hybride Künstliche Intelligenz in der automatisierten Inhaltserschließung, S. 387–405, hier S. 398f.

der Funktion, Zugang zu Informationsressourcen zu bieten, zwei Recherchearten, die durch ein adäquates Information Retrieval unterstützt werden sollen: Die „präzise Recherche“, d.h. die konkrete Suche nach einem bestimmten Thema, und die „Überblicksrecherche“, also eine weiter gefasste Suche nach einem größeren Themenfeld. Ergänzt würden beide Suchverfahren durch die Option, mithilfe einer differenzierten Filterfunktion Treffermengen thematisch einzugrenzen. Neben dem Zweck, das Auffinden von Medien im Zuge einer thematischen Recherche zu ermöglichen, haben inhaltserschließende Angaben zudem die Aufgabe, durch die in den Katalogisaten transportierten Informationen den Endnutzer*innen „Orientierung“ zu bieten, d.h. sie darüber zu informieren, welches Thema in einem Werk repräsentiert ist. Darüber hinaus solle im Rahmen der Explorationsfunktion auch ermöglicht werden, durch ein freies, nicht zielgerichtetes Navigieren in Katalog oder Discovery-System inhaltlich nützliche Ressourcen zu entdecken. Hierfür bedürfe es adäquater Suchwerkzeuge, die im Idealfall intuitiv bedienbar sein und den Endnutzer*innen vielfältige Möglichkeiten zur Interaktion bieten sollten. Die Option, in einem Online-Katalog oder Discovery-System unmittelbar nach GND-Schlagwörtern und den hierarchisch über- und untergeordneten sowie den thematisch verwandten Begriffen suchen zu können, würde die Rechercheoptionen der Endnutzer*innen beträchtlich erweitern.⁶ Ergänzend dazu bieten Rudolf Ungváry und Péter Király in ihrem Beitrag zur Qualitätsbewertung von MARC-21-Datensätzen einen weiteren Optimierungsvorschlag: Sie regen an, Bibliothekskataloge so zu ertüchtigen, dass die mithilfe einer Schlagwortsuche gewonnene Treffermenge um solche Titel erweitert werden könne, die mit einer Notation aus einem Klassifikationssystem erschlossen sind, deren Benennung dem Inhalt der Schlagwortsuche entspricht.

Jan Frederik Maas fordert in seinem Beitrag, dass nicht nur die bibliothekarische Inhaltserschließung, sondern auch die Entwicklung von Discovery-Systemen nutzungsorientiert anhand spezifischer Anwendungsfälle zu gestalten sei. Um eine effektive Umsetzung unterschiedlicher Rechercheoptionen mit nutzungsfreundlicher Funktionalität zu erreichen, plädiert er für die konsequente Einbindung von Expert*innen der bibliothekarischen Inhaltserschließung in die Entwicklung von Discovery-Systemen. Neben den Metadaten, ihrer Erfassung, Verwendung, Kuratierung, Darstellung, Auswertung und Anreicherung widmet sich der Band also noch einer weiteren Qualitätsdimension, die ebenfalls eine wichtige Grundlage für eine qualitativ hochwertige Inhaltserschließung darstellt: der Dimension „Mensch“. Menschen spielen in Bezug auf die Frage nach einer qualitätvollen Inhaltserschließung eine zweifache Rolle: Einmal als Anwender*innen bibliothekarischer Regelwerke und aktive Nutzer*innen von Normdaten im Prozess der intellektuellen Erschließung, zum anderen als Endnutzer*innen mit spezifischen Suchmethoden und Erwartungshaltungen an die Leistungsfähigkeit von Rechercheinstrumenten. Bei den Endnutzer*innen lassen sich mit der von Heidrun Wiesenmüller vorgeschlagenen Differenzierung „Normalnutzer*innen“ von „Expert*innen“ unterscheiden – mithin zwei Zielgruppen, die sich in ihrer Sachkompetenz wie in ihren Recherchegewohnheiten voneinander abgrenzen lassen. Die Ergebnisse einer anwenderspezifischen Inhaltserschließung müssten daher mit Blick auf verschiedene Zielgruppen, ihre Bedarfe und Erwartungen unterschiedlich ausfallen.⁷ Neben der von Jan Frederik Maas angeregten Konzeption von „Personas“ oder dem Rückgriff auf die

6 Vgl. im selben Band auch Maas, Jan Frederik: Inhaltserschließung für Discovery-Systeme gestalten, S. 303–324, hier S. 307.

7 Vgl. auch Wartena, Christian; Golub, Koraljka: Evaluierung von Verschlagwortung im Kontext des Information Retrievals, S. 325–348, hier S. 328.

Methode des „Scenario Based Design“⁸ stellt der in mehreren Beiträgen formulierte Vorschlag, den Endnutzer*innen ein Feedback zur Bewertung der Nützlichkeit bibliothekarischer Suchwerkzeuge und inhaltserschließender Daten zu ermöglichen,⁹ eine wünschenswerte Option dar, um Bedarfe verschiedener Gruppen von Endnutzer*innen zu erheben.

Mit Blick auf die Menschen, die sich aktiv in der intellektuellen Inhaltserschließung, in Projekten zur maschinellen Erschließung und in regionalen oder überregionalen Gremien der Inhaltserschließung engagieren, wäre es durchaus interessant, auch die strategischen und institutionellen Rahmenbedingungen etwas genauer zu betrachten, unter denen Sacherschließung in Bibliotheken betrieben wird. Da die im Prozess der intellektuellen Inhaltserschließung gewonnenen Metadaten als Trainingsdaten für maschinelle Verfahren herangezogen werden und die Ergebnisse maschineller Inhaltserschließung – zumindest in Stichproben – wiederum intellektuell überprüft werden, um ihre Qualität zu bestimmen, erscheint eine hochwertige intellektuelle Inhaltserschließung auch als Voraussetzung dafür, eine höhere Akzeptanz von maschinell generierten, inhaltserschließenden Daten zu erreichen.¹⁰ Die Frage nach bibliothekspolitischen und organisatorischen Kriterien, die sich qualitätsfördernd oder qualitätshemmend auf die Inhaltserschließung auswirken können, erscheint auch vor diesem Hintergrund als berechtigt.

Die hohe Bedeutung, die der inhaltlichen Erschließung von Informationsressourcen und ihren qualitativen Ansprüchen in der bibliothekarischen Öffentlichkeit beigemessen wird und die sich in zahlreichen Arbeitsgruppen, Tagungen, Workshops und Publikationen zu diesem Themenkreis widerspiegelt, sollte nicht darüber hinwegtäuschen, dass die Inhaltserschließung mit anderen strategischen Handlungsfeldern des Bibliothekswesens konkurriert. Bibliothekspolitisch ist daher auch generell die Frage zu stellen, wo innerhalb des „Magischen Dreiecks“ des (Projekt-)Managements, das durch die drei Eckpunkte (niedrige) Kosten, (geringer) Zeitaufwand und (hohe) Qualität gebildet wird, die intellektuelle Inhaltserschließung, die (Weiter-)Entwicklung (semi-)automatisierter Inhaltserschließung und das Engagement in Sacherschließungsgremien verortet werden. Freilich haben alle diese Bereiche keinen Projektcharakter, sondern sind vielmehr als eine Daueraufgabe der betreffenden Institutionen zu betrachten. Gerade bei der intellektuell vollzogenen Inhaltserschließung hängen Qualität und Quantität unmittelbar von der dafür verfügbaren Arbeitszeit und der Bereitschaft ab, die sich daraus ergebenden (Personal-)Kosten zu tragen. Die im Editorial dieses Bandes formulierte Befürchtung, der Einsatz maschineller Erschließungsverfahren könne angesichts knapper personeller und finanzieller Ressourcen primär als ein Mittel zur Effizienzsteigerung und zur Kostenreduktion interpretiert werden (S. 3), zielt in eben diese Richtung. Eine solche Sichtweise birgt jedoch die Gefahr, Quantität über Qualität zu stellen und damit gleichermaßen den Anspruch wie die Kernkompetenz

8 Vgl. Maas: Inhaltserschließung für Discovery-Systeme gestalten, S. 320-322.

9 Vgl. im selben Band Laczny, Joachim: Fit for Purpose – Standardisierung von inhaltserschließenden Informationen durch Richtlinien für Metadaten, S. 35–54, hier S. 44 und 47. Ebenso Pintscher, Lydia; Bourgonje, Peter; Moreno Schneider, Julián u.a.: Wissensbasen für die automatische Erschließung und ihre Qualität am Beispiel von Wikidata, S. 71–92, hier S. 83, und auch Wiesenmüller, Heidrun: Verbale Erschließung in Katalogen und Discovery-Systemen – Überlegungen zur Qualität, S. 279–301, hier S. 291, sowie Wartena; Golub: Evaluierung von Vorschlagwortung, S. 338.

10 Auf das Problem der mangelnden Akzeptanz maschinell generierter Inhaltserschließung verweist auch Sack, Harald: Hybride Künstliche Intelligenz in der automatisierten Inhaltserschließung, S. 402.

der wissenschaftlichen Bibliotheken zu konterkarieren, mittels einer an den Bedarfen der Kund*innen orientierten Inhalterschließung ein zeiteffizientes und zielgenaues Auffinden möglichst aller relevanten Titel zu einem recherchierten Thema zu ermöglichen. Dem „Appell an die strategische Managementebene“ (S. 4), die Inhalterschließung generell in ihrer Relevanz anzuerkennen und sich an entsprechenden Qualitätsstandards zu orientieren, ist daher schon im Sinne des Servicegedankens bibliothekarischer Dienstleistungen unbedingt beizupflichten. Welchen Standards der Qualitätsaspekt in verschiedenen Teilbereichen der Inhalterschließung verpflichtet sein sollte, dafür liefert der vorliegende Band eine Fülle an spannenden Anregungen und überzeugenden Beispielen.

Martin Völkl, Universitätsbibliothek Augsburg

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5843>

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).

Bedrohte Bücher : eine Geschichte der Zerstörung und Bewahrung des Wissens / Richard Ovenden ; aus dem Englischen von Ulrike Bischoff. – Erste Auflage. – Berlin: Suhrkamp, 2021. – 416 Seiten : Illustrationen. – Originaltitel: Burning the books. – ISBN 978-3-518-43007-1 : EUR 28.00 (auch als E-Book verfügbar)

Die gezielte Zerstörung von verkehrs- und kommunikationstechnischer Infrastruktur, aber auch von Orten und Symbolen kultureller Identität ist allgegenwärtig – nicht nur in kriegerischen Auseinandersetzungen. Seit Jahrtausenden haben macht- und wirtschaftspolitisch, religiös, ethnisch oder kulturell motivierte Eiferer ganz gezielt nicht zuletzt Bücher, Bilder, Archivbestände und sonstige Kulturgüter angegriffen, geplündert, geraubt oder häufig durch Feuer zerstört. Mit der Beseitigung oder Auslöschung der physischen Datenträger sollten auch die in ihnen enthaltenen Informationen und Erkenntnisse als Belege für das Wissen und die Kultur von Gemeinschaften – im Extremfall auch für deren gesamte Existenz! – ausgelöscht werden. Neben anderen haben Bibliothekar*innen und Archivar*innen sowie ihre Unterstützer*innen immer wieder versucht, dies durch ihre Arbeit zu verhindern und kodifiziertes Wissen zu bewahren. Von beidem, vom Zerstören, aber auch vom Sichern und Bewahren des Wissens, handelt das engagiert geschriebene Buch des britischen Bibliothekars Richard Ovenden.

Eingerahmt von einer Einleitung, die das Grundanliegen des Buches darstellt, und einem Schlusskapitel mit fünf Thesen zur Frage „Warum wir immer Bibliotheken und Archive brauchen werden“ (s.u.), behandelt Ovenden das Thema in 14 Kapiteln.¹ Ergänzend stellt der Autor einen nach Kapiteln gegliederten wissenschaftlichen Apparat mit Literaturangaben und Erläuterungen, eine umfangliche Bibliografie der genutzten Literatur sowie ein Register der erwähnten Namen und Institutionen bereit. Das Thema der Zerstörung greift er unter vielen Aspekten auf. Unter den Beispielen sind bibliotheksgeschichtliche „Klassiker“ wie die mesopotamischen Tontafeln des Königs Assurbanipal und deren Zerstörung, die allmähliche Zerstörung der antiken Bibliothek von Alexandria durch Vernachlässigung und Brände, die religiös motivierte Vernichtung von Buchbeständen und Archiven (z.B. im Kontext der Reformation) und die zweimalige Vernichtung der Universitätsbibliothek von Leuven durch deutsches Militär im 20. Jahrhundert. Ovenden schreibt auch über die nationalsozialistisch motivierte Vernichtung von hebräischer und jiddischer Literatur und Überlieferung und über den z.T. erfolgreichen Versuch der sogenannten Papierbrigade in Litauen, dies zu verhindern. Weitere Beispiele sind die gezielte Vernichtung von muslimischer Kultur und Wissensbeständen, oder auch die von Archivdaten über den Alltag der muslimischen Bevölkerung in Bosnien-Herzegowina durch Serben in den 1990er-Jahren. Ein weiterer Fall systematischer Bücherzerstörung war das Niederbrennen der Library of Congress 1814 durch britische Truppen im Britisch-Amerikanischen Krieg.

Bei seiner Darstellung der Bedrohungen und Zerstörungen von Wissensspeichern betrachtet Ovenden Bibliotheken und Archive als prinzipiell zusammengehörig. Er nimmt hierbei das Gemeinsame und

1 Inhaltsverzeichnis unter <<https://d-nb.info/1230568212/04>>, Stand: 10.08.2022.

nicht – wie das in der deutschen Diskussion zumeist der Fall ist – das Trennende in den Blick. Bei Archiven allerdings kommt die Bedrohung häufig nicht nur von außen, sondern auch von innen: von den Betreibern, insbesondere von jenen, deren Handeln durch die Zerstörung oder Beseitigung von Archivmaterial vertuscht oder nicht mehr nachweisbar sein soll. Ovenden vertieft diesen Aspekt an den mehr als 100 Regalkilometer einnehmenden Stasi-Akten und ihrer Rettung durch Bürger*innen der DDR ebenso wie an der Sicherung von Akten der irakischen Baath-Partei oder am Beispiel des südafrikanischen Umgangs mit Akten des Apartheid-Regimes.² Die gezielte Vernichtung bzw. die Nicht-Übergabe von Akten der Kolonialmächte an unabhängig gewordene Staaten prangert er in diesem Zusammenhang als einen Versuch an, die Aufarbeitung von staatlichem Unrechtshandeln und damit u.a. auch eine gesellschaftliche Versöhnung zu verhindern. Ovenden formuliert: „Wenn Gemeinschaften keinen Zugang zu ihrer eigenen Geschichte haben, lässt sich die Darstellung ihrer Vergangenheit kontrollieren und manipulieren und die kulturelle und politische Identität ernstlich untergraben.“ (S. 247)

An der Schnittstelle von Bibliothek und Archiv, der Literatur und den Literaturarchiven ist der dritte Aspekt der von Ovenden thematisierten Bedrohungen angesiedelt: Die von Autor*innen bzw. Urheber*innen gewünschte, erbetene oder befohlene Vernichtung von Unterlagen aus ihrem jeweiligen Vor- bzw. Nachlass. Die mit einer erwünschten Vernichtung verbundene grundsätzliche ethische Frage, ob ein solcher Auftrag umgesetzt oder verweigert werden sollte, diskutiert Ovenden u.a. an dem von Franz Kafka zur Vernichtung bestimmten literarischen Nachlass und der Weigerung des von ihm beauftragten Freundes Max Brod, seinen Wunsch auch umzusetzen. An weiteren Beispielen wie der Autobiografie von Lord Byron oder dem Nachlass der US-Autorin Sylvia Plath verdeutlicht er das ethische Dilemma sowie die möglichen Motive, sich dem Wunsch eines Nachlassgebers oder einer Nachlassgeberin zu widersetzen. Auch wenn es heutzutage bei Schriftsteller*innen und anderen Nachlassgebenden nur noch selten um die gewollte Vernichtung von Notiz- und Tagebüchern geht, sondern eher um Fragen des digitalen Nachlasses: Das grundsätzliche Dilemma bleibt bestehen. Allerdings ist heute die endgültige Löschung von Facebook- oder anderen Social-Media-Spuren deutlich schwieriger zu erreichen, als dies bei physischen Objekten der Fall ist.

Als vierte Variante der Bedrohung von Wissen stellt Ovenden – mit einer Referenz auf die allmähliche Zerstörung der antiken Bibliothek von Alexandria – die Frage der Vernichtung durch Vernachlässigung oder Nichtstun im Umgang mit digital gespeichertem Wissen in den Mittelpunkt seiner Betrachtungen. Durch die Nicht-Sicherung und Nicht-Archivierung unserer digitalen Aktivitäten, Materialien und Spuren im Netz sieht er die Chance gefährdet, heutige oder zukünftige Entwicklungen und Einflussnahmen auf Politik und Gesellschaft transparent zu machen bzw. aufzudecken. Hierbei spricht er die Notwendigkeit einer umfassenden Archivierung sowohl des Web als auch von Kommunikation in Social Media an. Beide seien die unabdingbaren Grundlagen einer ggf. später erwünschten Analyse

2 Die bewusste oder unbewusste, auf jeden Fall aber fahrlässige Vernichtung von Landkarten der sog. „Windrush-Generation“ durch das britische Innenministerium war nach Aussage von Ovenden der Auslöser für das Schreiben des vorliegenden Buches. Hierbei waren diese Dokumente über das „Anlanden“ von Migrant*innen in Großbritannien in den 1940ern und 1950ern, die diesen als Beleg für ihre Einwanderung und damit auch für die Erlangung der ihnen zugesagten Staatsbürgerschaft hätten dienen können, 2010 unter der Verantwortung von Innenministerin Theresa May vernichtet worden. In der Konsequenz wurden Einwanderer und Einwanderinnen mangels anderer Belege für ihre legale Einwanderung unrechtmäßig aus Großbritannien ausgewiesen.

von Kommunikationsverhalten und einer möglicherweise erfolgten Einflussnahme auf gesellschaftliches Handeln und Geschehnisse, wie sie – unter Missachtung von Persönlichkeitsrechten – von großen Technologiekonzernen, aber auch von Firmen wie Cambridge Analytica bekannt geworden sind. Bislang aber sind – so sein berechtigter Hinweis – diesbezüglich weder die gesellschaftsinternen Zuständigkeiten für solche Sicherungsmaßnahmen abgeklärt noch deren Finanzierung. Ohne solche Maßnahmen zur Sicherung und Transparenz von digitalen Aktivitäten sei man jedoch, so Ovenden, den Technologiekonzernen ebenso ausgeliefert wie autoritären Regimen, die in ihren Archiven solche Spuren dokumentierten. Sie hätten ganz offensichtlich die von Jacques Derrida formulierte Regel „Keine politische Macht ohne Kontrolle des Archivs“³ besser verstanden als die von ihnen kontrollierten Gesellschaften.

Die abschließenden Betrachtungen zur Frage „Warum wir immer Bibliotheken und Archive brauchen werden“ unterstreicht der Autor durch fünf Funktionen, die sich auch als Quintessenz seiner Sicht verstehen lassen:

„*Erstens* unterstützen Bibliotheken und Archive die Bildung der gesamten Gesellschaft und spezifischer Gemeinschaften.

Zweitens stellen sie eine große Bandbreite an Wissen und Ideen zur Verfügung.

Drittens tragen sie zum Wohl der Bevölkerung bei und stützen die Prinzipien der offenen Gesellschaft, indem sie zentrale Rechte bewahren und die Integrität von Entscheidungsprozessen fördern.

Viertens liefern sie einen festen Bezugspunkt, der es erlaubt, mittels Verfahren des Transparentmachens und Verifizierens, des Zitierens und Reproduzierens zwischen Wahrheit und Unwahrheit zu unterscheiden.

Und *fünftens* helfen sie, Gesellschaften in ihrer kulturellen und historischen Identität zu verwurzeln, indem sie die Schriftzeugnisse dieser Gesellschaften und Kulturen bewahren.“
(S.309; Hervorhebungen im Original)

Zu Recht ist dieses Buch nicht nur in der Fach-, sondern auch in der überregionalen Presse besprochen und gewürdigt worden. Es hat aktuelle Bezüge und steht im Juli 2022 vielleicht auch deshalb auf der Longlist für den Preis des Wissenschaftsbuches des Jahres 2022. Schon 2021 stand es auf der Shortlist für den Wolfson History Prize 2021.

Die Zerstörungen von materiellem wie immateriellem Kulturerbe in der Ukraine belegen Ovendens Sicht und Thesen genauso wie all die anderen Vernichtungsmaßnahmen gegenüber Kulturen, Völkern, Ethnien, religiösen oder nach anderen Aspekten definierten Gruppen und ihren Wissensbeständen. Wir haben dies in den letzten Jahren und Jahrzehnten weltweit erlebt und werden es wohl auch in Zukunft weiter erleben müssen – trotz der Haager Landkriegsordnung von 1907 sowie der Haager Konvention zum Schutz von Kulturgut bei bewaffneten Konflikten von 1954.

3 Hier zitiert nach Ovenden, S. 299; zuerst in Derrida, Jacques: Dem Archiv verschrieben. Eine Freudsche Impression, Berlin 1997, S. 14.

Die latente Enteignung unserer Daten und Kommunikationsspuren, die zunehmend in der Verfügungsgewalt von privaten, gewinnorientierten Firmen liegen und einer gesellschaftlichen Kontrolle entzogen sind, birgt Gefahren, die weit über solche zerstörerischen Maßnahmen hinaus gehen. George Orwell hat sie schon in seinem Roman „1984“ eindrucksvoll dargestellt. Das bislang unkontrollierte bzw. kaum kontrollierbare Geschäft mit der Sammlung und Analyse von Daten, die ggf. ex post aus unterschiedlichen Kontexten zusammengeführt werden (data analytics), gerät allerdings erst allmählich in den Blick. Der Wissenschaftsbereich, in dem dies um sich zu greifen droht, mag als Beispiel für die zunehmende Macht der im und mit dem Internet agierenden Konzerne dienen.⁴

Die Alternative zu solchen Zuständen sieht Ovenden in einer gesamtgesellschaftlichen, z.B. an Bibliotheken und Archive zu übertragenden, finanziell und strukturell abgesicherten Aufgabe: Die „Bewahrung des in digitaler Form erzeugten Wissens [ist] eine wichtige Voraussetzung für offene, demokratische Gesellschaften“ (S. 306). Sie ist also nicht etwa Selbstzweck, sondern eine Voraussetzung für ein verantwortliches und selbstbestimmtes gesellschaftliches Miteinander.

Die titelgebenden „bedrohten Bücher“ sind für Ovenden heute demzufolge auch bedrohte Daten und Kommunikationsmöglichkeiten als Grundlage und Ergebnis der Generierung gesellschaftlich relevanten Wissens. Schade, dass sich diese aktuelle Dimension des Themas nicht im Titel des Buches niedergeschlagen hat!

Richard Ovenden war von 2009 bis 2013 Vorsitzender der Digital Preservation Coalition⁵ und dürfte insofern mit den Fragen rund um die Sicherung digitaler Objekte und Daten schon lange vertraut sein. Dennoch entsteht der – vermutlich falsche – Eindruck, als habe Ovenden erst bei der intensiven Beschäftigung mit dem Thema die weitergehenden, digitalen Implikationen seiner „Geschichte der Zerstörung und Bewahrung des Wissens“ – so der Untertitel – zunehmend realisiert. Allerdings hat er sie kompetent thematisiert und auf grundsätzliche Weise auch aufgearbeitet.

Seine Vertrautheit und Verbundenheit mit bibliothekshistorischen Themen verleitet ihn aus Sicht des Rezensenten manchmal zu etwas zu ausführlich beschriebenen Details zur Entstehung jener Bibliotheken, deren Zerstörung er beschreibt, zur englischen Bibliotheksgeschichte sowie insbesondere zur Bodleian Library in Oxford, deren 25. Bibliothekar der Autor seit 2014 ist. All dies kann Lesende ermüden, die vorzugsweise am Kern des Themas, nämlich an den Gründen und Konsequenzen der Gefährdung und Zerstörung von Medien sowie der in ihnen enthaltenen Erkenntnisse interessiert sind.

Diese Längen sind aber verkraftbar und verzeihlich, gerade bei einem Autor, der nicht in der Bibliotheksgeschichte verharrt, sondern ganz offensichtlich willens und in der Lage ist, eine grundsätzlichere Perspektive einzunehmen. Die Bewahrung unseres Kulturerbes schließt heutzutage eben unabdingbar auch all jene digitalen Daten, Dokumente und die digitale Kommunikation über sie mit ein, die unser digitales Ökosystem konstituieren. Dies wie Richard Ovenden als eine

4 Vgl. Siems, Renke: Das Lesen der Anderen. Die Auswirkungen von User Tracking auf Bibliotheken, in: o-bib 9 (1), 2022, S. 1–25. Online: <<https://doi.org/10.5282/o-bib/5797>>.

5 Digital Preservation Coalition: About the Digital Preservation Coalition, <<https://www.dpconline.org/about>>, Stand: 10.08.2022: „The Digital Preservation Coalition exists to secure our digital legacy“.

gemeinsame, gesellschaftlich abzusichernde Aufgabe zu betrachten, die ganz wesentlich auch von Bibliothekar*innen und Archivar*innen zu bearbeiten ist, zeugt von einer guten Mischung aus Weit-sicht und Selbstbewusstsein, die noch nicht verbreitet genug ist.

Achim Oßwald, Technische Hochschule Köln

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5880>

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#).

Gemeinsame Kommission Informationskompetenz von VDB und dbv

„Alles IK oder was? Ideen und Beispiele für die Vernetzung von Literacies in Bibliotheken gesucht!“

Best-Practice-Wettbewerb Informationskompetenz 2022

Dieser Beitrag der gemeinsamen Kommission Informationskompetenz von VDB und dbv möchte der Fachcommunity die vier prämierten Einreichungen des Best-Practice-Wettbewerbs Informationskompetenz 2022 vorstellen und damit insbesondere Konzepte bekannt machen, welche über den Kern der IK-Förderung hinausdenken und die Förderung vielfältiger Literacies in den Blick nehmen.

Bibliotheken sprechen oft von Informationskompetenz, wenn es um die Angebote geht, mit denen sie die Teilhabe an der Informationsgesellschaft fördern – sei es im Privaten oder in der Wissenschaft. Um den wachsenden Ansprüchen einer zunehmend digitalen Welt gerecht zu werden, müssen aber heute etablierte Aufgabenfelder erweitert und mit neuen Themen vernetzt werden: Digital- und Medienkompetenz, Daten- und Schreibkompetenz, Kommunikations- und Lernkompetenzen stehen nicht nur für sich, sondern überschneiden und ergänzen sich. Hier entsteht ein Bedarf an neuem Wissen und Fertigkeiten, um die Potentiale der digital geprägten Gesellschaft und Wissenschaft effektiv nutzen zu können – ein Bedarf, der auch neue Vermittlungsangebote, -methoden und -formate fordert. Bibliotheken stehen damit vor didaktischen, organisatorischen und fachspezifischen Herausforderungen: Modelle und Konzepte sind nötig, die die Berührungspunkte dieser Kompetenzen oder „Literacies“ ausloten und in gemeinsame Formate übertragen. Dazu sind auch neue Kooperationsformen, Partner und Möglichkeiten der Vernetzung gefragt.

Der diesjährige Best-Practice-Wettbewerb Informationskompetenz rief dazu auf, insbesondere Beispiele einzureichen, die aufzeigen wie öffentliche und wissenschaftliche Bibliotheken sich in der Praxis in diesem Spannungsfeld unterschiedlicher Literacies positionieren - sei es allein oder gemeinsam mit externen Partnern.

Ziel des Wettbewerbs war es, vorbildliche Konzepte und Umsetzungen zu fördern, Best-Practice-Beispiele bekannt zu machen und zum Erfahrungsaustausch und zur Nachahmung sowie ggf. zur Nachnutzung anzuregen. Teilnahmeberechtigt waren Bibliotheken, Bibliothekarinnen und Bibliothekare aller Sparten.

Eingereicht wurden insgesamt 19 Beiträge von wissenschaftlichen und öffentlichen Bibliotheken sowie deren Kooperationspartnern. Die hohe Anzahl an Einreichungen verdeutlicht die Aktualität des gewählten Themas und zeigt damit auch auf, dass viele Bibliotheken längst Aktivitäten entfaltet haben, um ihre Angebote auf die Anforderungen einer zunehmend daten- und digitalorientierten Wissenschaft und Gesellschaft anzupassen. Anders und plakativer formuliert: Die Zeiten, in denen sich Bibliotheken ausschließlich auf Bibliotheksführungen und Katalogschulungen konzentrieren konnten, gehören endgültig der Vergangenheit an. Die Einreichungen belegen eine enorme thematische Breite der Aktivitäten und zeigen das Innovationspotential der Literacies auf: Das Spektrum reichte

von Data Literacy und neuen digitalen Lehr-Lern-Konzepten über neue Formen der Schreibberatung bis zu digitaler Quellenkritik und Informationsbewertung.

Ermittelt wurden die Preisträger*innen in einem mehrstufigen Verfahren durch eine vierköpfige Fachjury (Charlotte Becker, Beratungsstelle für Öffentliche Bibliotheken Südniedersachsen; Dagmar Knorr, Leiterin des Schreibzentrums der Leuphana Universität Lüneburg; Raphaela Müller, JFF-Institut für Medienpädagogik; Ulrike Wuttke, FH Potsdam, Professurvertretung Bibliotheks- und Informationstechnologien und Digitale Services).

Aufgrund gleicher Punktzahl sind dabei gleich zwei der bestbewerteten Einreichungen auf dem zweiten Platz gelandet.

Die Preisträger lauten:

1. Preis: Uwe Johnson-Bibliothek Güstrow
2. Preis: Universitätsbibliothek Wuppertal
2. Preis: ZBW – Leibniz-Informationszentrum Wirtschaft
3. Preis: Bibliothek der Hochschule für Technik und Wirtschaft Dresden

1. Preis: Uwe Johnson-Bibliothek Güstrow (Tilmann Wesolowski)

Die Uwe Johnson-Bibliothek Güstrow präsentiert mit ihrer Einreichung ein Projekt, das sich der Vermittlung von Digital- und Technikkompetenz an einer öffentlichen Bibliothek widmet. Darin werden digitale Angebote mit haptischen Stationen und bibliothekspädagogischen Bausteinen kombiniert. Die Umsetzung erfolgt in unterschiedlichen Formaten und verteilt in den Räumlichkeiten der Bibliothek. Zu den Angeboten zählen u.a. Makerspaces, „Probier’s aus-Boxen“, Medienstationen und flankierenden Workshops. Aus diesen Komponenten entsteht ein Gesamtkonzept, das kreative Elemente der Digitalität, der technischen Bildung sowie der Bibliotheks- bzw. Museumspädagogik miteinander verbindet. Das Konzept ist dabei modular aufgebaut, damit auch andere Bibliotheken einzelne Bausteine unabhängig von vorhandenen Ressourcen, wie z.B. dem zur Verfügung stehenden Raum, adaptieren können. Dieser Aufbau fördert auch die Ansprache unterschiedlicher Zielgruppen. So werden z.B. Kurse zur Robotik für Grundschulen angeboten und ein Workshop Photogrammetrie für die 8. und 9. Klasse. Begleitet wird das Angebot zudem jeweils durch ausgewählte Literatur zu den einzelnen Bausteinen.

Die Uwe Johnson-Bibliothek hat mit diesem Projekt ein Angebot für den MINT-Bereich an einer öffentlichen Bibliothek geschaffen, das prototypisch für die Möglichkeiten steht, die auch kleineren Bibliotheken zur Verfügung stehen.

Für die Jury war vor allem die Nachnutzbarkeit des Konzepts auf der Ebene der öffentlichen Bibliotheken ein besonders positives Merkmal des Angebots. Hinzu kamen die überzeugende Kombination unterschiedlicher Literacies und die Verbindung von analogen und digitalen Komponenten bei der Förderung derselben. Der Vorbildcharakter des Angebotes wird durch die aktuelle Nutzung, eine positive Evaluierung sowie die bereits stattfindende Nachnutzung einzelner Komponenten durch zahlreiche Einrichtungen bestätigt.

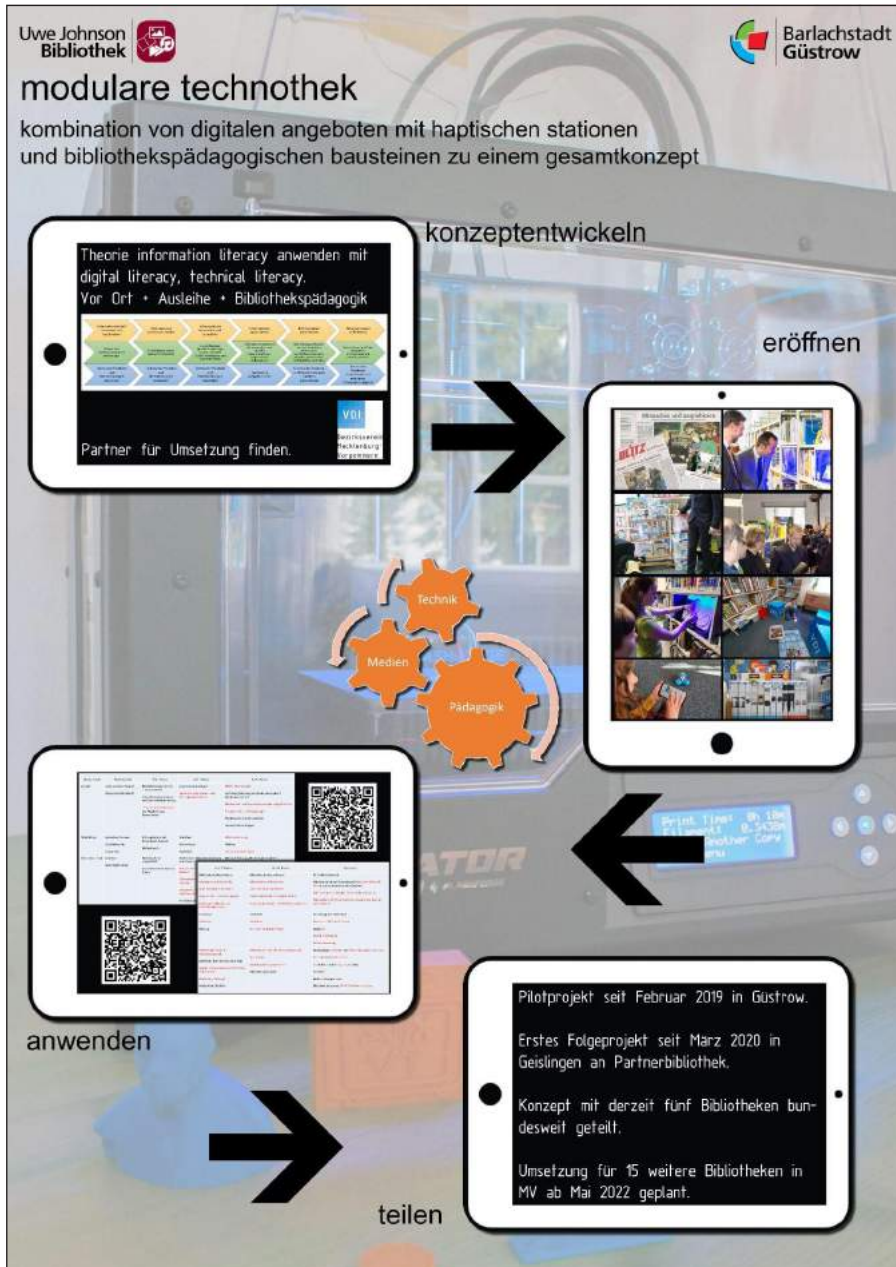


Abb. 1: Modulare Technotheek. Plakat der Uwe Johnson-Bibliothek Güstrow im Rahmen des Best-Practice-Wettbewerbs 2022 der Gemeinsamen Kommission für Informationskompetenz von VDB und dbv. Copyright: Uwe Johnson-Bibliothek Güstrow.

Platz 2: Universitätsbibliothek Wuppertal (Jan Fritze, Kerstin Kaiser, Fred Demmer)

Die Universitätsbibliothek Wuppertal reichte das Vorhaben *digi-komp.nrw*¹ für den Wettbewerb ein. Das Kooperationsprojekt, an dem mehrere Hochschulbibliotheken und wissenschaftliche Einrichtungen aus Nordrhein-Westfalen beteiligt sind, wird seit 2019 vom Ministerium für Kultur und Wissenschaft des Bundeslandes gefördert. Ziel ist eine Online-Plattform für interdisziplinäre und modular aufgebaute E-Learning-Selbstlernkurse, die zur Förderung der Medien-, Informations- und Digitalkompetenz von Studierenden – vor allem der ersten Semester – beiträgt. Thematische Schwerpunkte der Plattform sind: „Informationsrecherche und -bewertung“, „Medienproduktion und Präsentation“, „Prinzipien der digitalen Welt“ (hier werden v. a. technische Themen wie Daten, KI und Algorithmen aufbereitet), „Rechtliche Grundlagen“ sowie ein übergeordneter Einstieg „Information, Wissenschaft und Kommunikation“. Bei der Themenwahl orientierten sich die Projektpartner an den sogenannten Future Skills, die die zentralen Schlüsselqualifikationen für die Zukunft umfassen. Bei der Vermittlung dieser Themen wird auf eine mediale Vielfalt aus Text, Bild, Videos und weiterem audiovisuellem Material gesetzt.

Ein Schwerpunkt liegt im modularen Aufbau des Lernangebots, durch den die Einheiten leicht in individuelle Lernkontexte integriert werden können. Außerdem fördert das Konzept die eigenverantwortliche Wissensaneignung durch die Studierenden und die bedarfsgerechte Integration z. B. in Schulungsangebote von Bibliotheken als Teil von Blended-Learning-Szenarien. Die dadurch erzielte Flexibilität des Lernangebots wurde von den Jurymitgliedern besonders hervorgehoben.

Ebenso überzeugte die Jury die Bereitstellung der Materialien als Open Educational Resources (OER). So werden die digitalen Lernmaterialien landesweit unter der freien Lizenz CC-BY-SA 4.0 verfügbar gemacht.

Als besonders innovativ wurde die Verknüpfung verschiedener Literacies betont, durch die aktuelle Bedarfe der Wissensvermittlung eingelöst werden.

1 Vgl. <<http://digi-komp.uni-wuppertal.de/>>, Stand: 19.08.2022

digi-komp.nrw E-Learning für Studierende

DAS PROJEKT

Ziel: Erstellung von Lernmaterialien zu Themen der „Future Skills“, also Informations-, Medien- und Digitalkompetenzen


Zielgruppe: Studierende der Studienleistungsphase (1. - 3. Semester) in NRW

Format: Umsetzung der Inhalte in Moodle zum Selbstlernen und/oder zur modularen Einbindung in die Lehre

VERNETZUNG

Konsortium: Mitglieder aus universitären Einrichtungen, v. a. Bibliotheken, in ganz NRW

Austausch: Vernetzung mit Projekten der Digitalen Hochschule NRW wie z. B. SoC@ware.nrw, kmh.nrw, soaccess.nrw, modc@.nrw und weiteren Projekten aus dem Förderprogramm Data Literacy Education.nrw



SCHWERPUNKTTHEMEN


Wissenschaft, Information, Kommunikation	Informationsrecherche und -bewertung	Medienproduktion und Präsentation	Prinzipien der digitalen Welt	Rechtliche Grundlagen
--	--------------------------------------	-----------------------------------	-------------------------------	-----------------------

Orientierung: KMK-Strategie „Bildung in der digitalen Welt“ und Medienkompetenzrahmen NRW

UMSETZUNG

Moodle-Kurs: Entwicklung eines Kurses zum Themenbereich Informationsrecherche und -bewertung für Studierende und Mitarbeiter*innen der RWUW im Mai 2020

Einführungskurse: Erstellung fachspezifischer Kurse zum Themenbereich Informationsrecherche und -bewertung für die Germanistik (WS 2021), Anglistik (WS 2021) und Romanistik (WS 2022) an der BUW




EVALUATION

Feedback: Review von Kurs „Uni@Home – Bibliothek“ durch das Konsortium und zusätzlicher anonymer Feedbackbögen für weitere Hochschullangezeitige

Forschungsgruppen: Evaluation mit Studierenden verschiedener Fachrichtungen

Umfrage: Digitale und anonyme Bewertung des Fachkurses Germanistik von Studierenden am Semesterende (Rücklauf ca. 20%, n=173)

Ergebnisse: Stärkung der inhaltlichen Ausrichtung, Optimierung der Usability, Wunsch nach mehr multimedialen Inhalten und mehr alternativen Beispielen





MULTIMEDIALE AUFBEREITUNG

Text und Grafiken: Produktion von audiovisuellen Materialien zur Anreicherung der Kursmaterialien

Erklärvideos: Realisation als Legetechnik-Animation und mit Sprecher*innen im Kurzformat (~ 1,5-2,5 Min.)

OER: Veröffentlichung aller Materialien unter CC BY-SA 4.0







VERÖFFENTLICHUNG UND AUSBLICK

Selbstlernkurse: Zukünftige Veröffentlichung der Kurse auf dem Landesportal ORCA (Open Resources Campus NRW)

Best Practices: Bereitstellung von Anleitungen und Dokumentationen für die Verwendung der Materialien in der Lehre

Einbindung: Mögliche Integration in Bewältigungswege der Bibliotheken und in die Lehre an den Hochschulen des Konsortiums





100 Ausführende der Logos des Bibliothekswirtd-Kolleg und Wuppertal sind im Landes-REB, dem Organ der Hochschulbibliothek, der Landesbibliothek für Bildung und Lehre (LBB) NRW und der Landesbibliothek (LBB) NRW als gleichberechtigter Partner an diesem Platz und dem Landes-REB NRW, und sind verfügbar unter www.wuppertal.komp.nrw

Gefördert durch:




Abb. 2: digi-komp.nrw. E-Learning für Studierende. Plakat der Universitätsbibliothek Wuppertal im Rahmen des Best-Practice-Wettbewerbs 2022 der Gemeinsamen Kommission für Informationskompetenz von VDB und dbv. Copyright: Universitätsbibliothek Wuppertal.

Platz 2: ZBW – Leibniz-Informationszentrum Wirtschaft (Ellen Heidelberger, Tamara Pianos)

Mehrere wissenschaftliche Beobachtungen belegten, dass schriftliche Arbeiten von Studierenden der Wirtschaftswissenschaften häufig nicht den Qualitätsansprüchen der Lehrenden genügen. Die Studierenden tun sich mit den fachspezifischen Konventionen schwer und sind nicht mit den grundlegenden Techniken des wissenschaftlichen Arbeitens vertraut. Ausgehend von diesem Befund entwickelte das Team der ZBW einen Guide zum Thema "Schreibe Deine beste Hausarbeit in Wirtschaftswissenschaften".

Bei der Konzeption des Guides beschränkte sich das Projektteam nicht auf den Bereich der Informationskompetenz, sondern berücksichtigte weitere Literacies und deren Schnittmengen. Dazu gehören u.a.: Die recherchierten Informationen in ihrer Qualität zu bewerten, Informationen aus unterschiedlichen Quellen miteinander in Bezug zu setzen, aus den Informationen Argumente zu formulieren und diese Argumente in wissenschaftlicher Sprache auszudrücken. Natürlich spielen auch die "traditionellen" Skills des korrekten Zitierens und der effizienten Recherche eine Rolle im Guide. Daneben unterstützt der Guide die Studierenden beim Zeit- und Projektmanagement, das als eine besondere Herausforderung identifiziert wurde.

Gerade das ganzheitliche Konzept des Guides, das auf die individuelle Lebenssituation der Studierenden eingeht, überzeugte die Jurymitglieder. Auch wurde gelobt, dass der Guide die Studierenden mit den Spezifika der Wirtschaftswissenschaften vertraut macht, wie die Bedeutung von Working Papers.

Schließlich wurde die attraktive Gestaltung des Guides als besonders positiv bewertet. Die Broschüre, deren erste Auflage 2019 fertiggestellt worden ist, liegt mittlerweile auch in einer englischen Ausgabe vor. Dies sowie die hohe Verbreitung des Guides zeugen von der positiven Resonanz und dem Nachnutzungspotential: Bislang wurden über 75.000 gedruckte Exemplare von über 250 Institutionen bestellt. Zusätzlich wurde eine digitale PDF-Version auf EconBiz, dem Suchportal der ZBW, veröffentlicht und dort bislang über 3.000 Mal heruntergeladen. Eine weitere PDF-Version ohne lizenzpflichtige Abbildungen unter einer CC-BY-Lizenz ermöglicht es interessierten Institutionen, den Guide in eigene Angebote einzubinden.²

2 Vgl. <https://www.econbiz.de/eb/fileadmin/ik_broschuere/schreibe_deine_beste_hausarbeit_booklet.pdf>, Stand: 19.08.2022

SCHREIBE DEINE BESTE HAUSARBEIT

Ein Guide zum wissenschaftlichen Schreiben für Studierende der Wirtschaftswissenschaften

Wie sind wir vorgegangen?

6. Nachnutzung

- Nachnutzung im Rahmen von drei halbtägigen Online-Seminaren in deutscher und in englischer Sprache
- Einführung von ergänzenden interaktiven Online-Trainings (aktuell in Arbeit)

3. Layout, Druck, Versand

- Auflage: mehr als 70.000 Broschüre und 5.000 englische gedruckte Exemplare
- Versand: kostenlose Versand an über 250 Universitäten, Hochschulen, öffentliche und wissenschaftliche Bibliotheken, Archive, etc. in Deutschland, Großbritannien und der Schweiz
- PDF-Download: außerdem mehr als 3.000 Downloads

4. Co-Kreation gemeinsam mit unserer Community

- Erstellen mit ZBW-Kolleginnen aus verschiedenen Abteilungen
- Feedbackschreiben mit Professorinnen und Studierenden der Wirtschaftswissenschaften (FH und Uni) sowie mit Informationsmanager:innen und Bibliothekar:innen

1. Ziel

- Unser Ziel: Studienzeile beim Erstellen von Haus- und Abschlussarbeiten unterstützen
- Angangspunkt des Projektes: Historie von Lehrenden, dass die wissenschaftlichen Arbeiten von Studierenden in den Wirtschaftswissenschaften oftmals nicht den Ansprüchen genüge
- Reaktion: Änderung: gezielte und strukturierte Informationen zu allen Arbeitsschritten beim Schreiben von Haus- und Abschlussarbeiten bereitstellen

2. Konzept für den Guide

- Erklärung verschiedener Kompetenzen entlang des Schreibprozesses einer Hausarbeit in einem Guide: Recherchen, Lesen, Schreiben, Zitiere, Korrektur
- bestmögliche Entscheidung für eine Print-Broschüre, die Studierende parallel zu ihrer Arbeit am Bildschirm mit an und in der die-bibliotek können (evtl. als eBook mit Kindle und QR-Code)
- ein gedruckter Guide, der auch die explizite Wertschöpfung von Studierenden und Lehrenden

3. Erarbeiten der Lerninhalte zu verschiedenen Kompetenzen

- Informationskompetenz: Informationen recherchieren, Quellen bewerten
- Schreibkompetenz: wissenschaftlich schreiben, korrekt zitieren
- Zeitmangement: zentrale Arbeitsschritte erkennen, Zeit im Blick behalten, Pausen einplanen

So sieht der Guide aus:

Checklisten

QR Codes

praktische Tipps

schlüsselwörter Zitiervorlagen für ein besseres Zitiermanagement

Platz für den Screenshot der Tabelle, der Abbildung etc.

Ein Angebot der:

Leibniz-Informationszentrum
Wirtschaft
Leibniz Information Centre
for Economics

Kontakt: info@econbiz.de
Web: www.econbiz.de

zum PDF-Guide
(Download)

zum Best-Practice-Material
für Print-Guides

Abb. 3: Schreibe Deine beste Hausarbeit in Wirtschaftswissenschaften. ZBW – Leibniz-Informationszentrum Wirtschaftswissenschaften im Rahmen des Best-Practice-Wettbewerbs 2022 der Gemeinsamen Kommission für Informationskompetenz von VDB und dbv. Copyright: ZBW – Leibniz-Informationszentrum Wirtschaftswissenschaften.

3. Preis: Bibliothek der Hochschule für Technik und Wirtschaft Dresden (Petra-Sibylle Stenzel)

Gerade für kleinere wissenschaftliche Bibliotheken stellt die Umsetzung eines zeitgemäßen und für unterschiedliche Zielgruppen geeigneten Schulungs- und Workshopangebots eine große Herausforderung dar. Die Einreichung der Bibliothek für Technik und Wirtschaft Dresden (HTW) hat dafür eine überzeugende Lösung in Form von Kooperationen und eines breit angelegten Netzwerks entwickelt. Grundlegend für das Konzept ist es, dass die Schlüsselqualifikation Informationskompetenz als Teil der ganzheitlichen Hochschulbildung verstanden wird und das Vermittlungsangebot deshalb nicht nur Studierende adressiert, sondern auch auf Lehrende, Forschende und universitäre MitarbeiterInnen ausgedehnt wird. Gerade durch die Verbindung der bibliothekarischen Expertise mit den Kenntnissen von inner- und außeruniversitären Akteuren ist es möglich, auch den neuen Anforderungen der sich stetig wandelnden Informationstechnologie gerecht zu werden. Die Kooperationen ermöglichen es, ressourcenschonend neue Themenfelder, wie z.B. die Gute Wissenschaftliche Praxis, in das Schulungsangebot aufzunehmen. Gleichzeitig können so die Qualität und die Reputation der Informationskompetenzvermittlung gesteigert werden. Formate wie die „Lange Nacht der aufgeschobenen Hausarbeiten“ zeugen von der erfolgreichen Umsetzung des skizzierten Ansatzes und dem Gelingen des Vorhabens.

Die Jurymitglieder lobten die durch den interdisziplinären Ansatz geschaffenen Synergieeffekte und die praxisnahe Umsetzung, die zur Nachnutzung einlädt. Auch die Berücksichtigung und Verschränkung unterschiedlicher Literacies sowie deren zielgruppengerechte Vermittlung überzeugte die Jury, die die Einreichung mit dem dritten Platz belohnte.



Abb. 4: Zusammen recherchiert man weniger allein – kooperative Informationsvermittlung an der Hochschule für Technik und Wirtschaft Dresden (HTW). Bibliothek der Hochschule für Technik und Wirtschaft Dresden im Rahmen des Best-Practice-Wettbewerbs 2022 der Gemeinsamen Kommission für Informationskompetenz von VDB und dbv. Copyright: Bibliothek der Hochschule für Technik und Wirtschaft Dresden.

Fazit zum Best-Practice-Wettbewerb Informationskompetenz 2022

Das Ziel des diesjährigen Wettbewerbs Informationskompetenz war es, vorbildhafte Best-Practice-Beispiele aus der Community vorzustellen und zu würdigen, die aufzeigen, wie das wechselseitige Zusammenspiel von unterschiedlichen Literacies innovativ umgesetzt werden kann. Damit sollten auch Lösungen für die Herausforderungen der Digitalität für Gesellschaft und Wissenschaften gezeigt werden und die Rolle von Bibliotheken in diesem Kontext beleuchtet werden.

Die prämierten Beiträge machen deutlich, wie unterschiedlich die Antworten auf diese Herausforderungen sein können. Vor allem belegten die Einreichungen eine bestehende Experimentierfreude, neue Inhalte und Konzepte auszuprobieren bzw. alte Konzepte neu zu denken. Die Bandbreite reicht dabei von der Kombination physischer und digitaler Elemente bis zu digital-only-Angeboten. Dabei werden etablierte Konzepte, wie z.B. die Lese- oder Schreibkompetenzförderung, neu gedacht und gänzlich neue Angebote, wie z.B. zu Data oder Digital Literacy, entwickelt. Deutlich wird, dass sich Bibliotheken aktiv in diesem Feld beteiligen. Dabei profitieren sie von ihrer vorhandenen Erfahrung in der Informationskompetenzvermittlung und gehen gleichzeitig aktiv die Chance an, die eine Kombination der "neuen" Literacies bietet. Als treibende Kraft lässt sich zudem der Wille beobachten, zielgruppenspezifische und moderne Angebote zu schaffen.

Die gemeinsame Kommission Informationskompetenz von VDB und dbv möchte an dieser Stelle allen Teilnehmenden am Wettbewerb danken, sie gratuliert insbesondere den vier Preisträgern und hofft, dass die Beiträge anderen Bibliotheken Impulse geben und Mut machen, das eigene Vermittlungsangebot breiter aufzustellen. Gerade die modular und kooperativ aufgestellten Projekte zeigen auf, wie auch mit kleinen Maßnahmen und geringen Budgets attraktive Vermittlungsprogramme entwickelt werden können.

Christina Schmitz, Staatsbibliothek zu Berlin

Erik Senst, Leuphana Universität Lüneburg

Timo Steyer, UB Braunschweig

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5876>

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](#).