

„Data is King“

Strategische Überlegungen zum Sammelauftrag der Bibliothek im digitalen Zeitalter

„Wie kann es sein, dass ein Besitz, verteilt / Auf mehr Besitzer, diese reicher macht /
Als wenn er nur von wenigen besessen?“ (Dante Alighieri)

Die bibliothekarische wissenschaftliche Sammlung, so das Credo von Publikationen der letzten Jahre,¹ verliert angesichts ubiquitär verfügbarer digitaler Werke an Bedeutung oder wandelt doch stark ihr Gesicht. Dabei spielen verschiedene Faktoren eine Rolle, die letztlich alle auf die seit Jahren schnell voranschreitende Digitalisierung der Literatur- und Informationsversorgung zurückzuführen sind. Einige Zahlen belegen den Wandel: Die Ausleihen gedruckter Werke gingen in den letzten fünf Jahren um fast 50 % zurück.

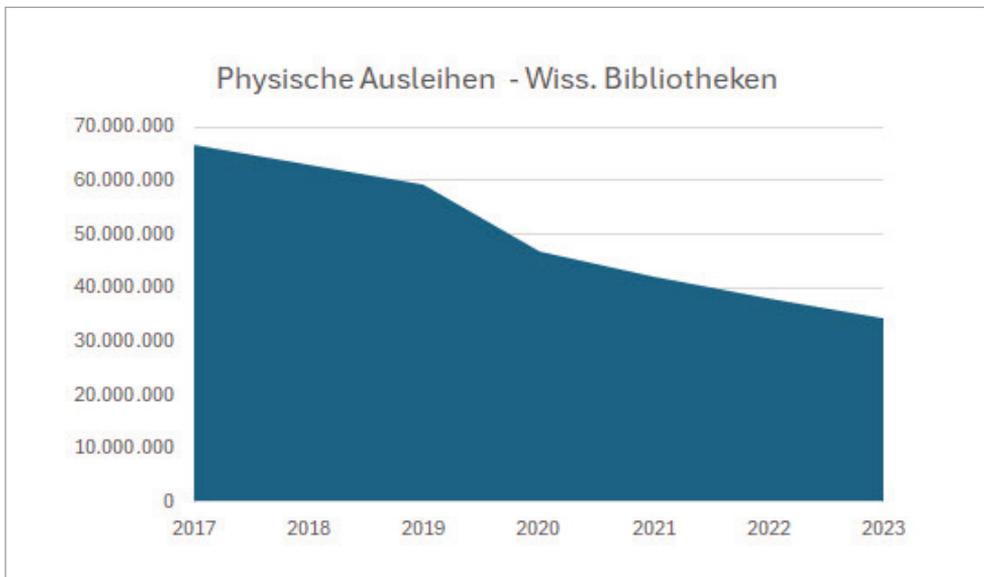


Abb. 1: Physische Ausleihen – Wiss. Bibliotheken, Quelle: Deutsche Bibliotheksstatistik (DBS)

- 1 Anderson, Rick: Collections 2021. The future of the library collection is not a collection, in: *Serials* 24 (3), 2011, S. 211–215, <https://doi.org/10.1629/24211>; Bonte, Achim; Ceynowa, Klaus: Bibliothek und Internet. Die Identitätskrise einer Institution im digitalen Informationszeitalter, in: *Lette International* 100 (Frühjahr 2013), 2013, S. 115–117; Degkwitz, Andreas: Digitale Sammlungen. Vision eines Neubeginns, in: *Bibliothek Forschung und Praxis* 38 (3), 2014, S. 411–416, <https://doi.org/10.1515/bfp-2014-0064>; Kempf, Klaus: Bibliotheken ohne Bestand?, in: *Bibliothek Forschung und Praxis* 38 (3), 2014, <https://doi.org/10.1515/bfp-2014-0057>; Altenhöner, Reinhard: Aspekte des Bestandsaufbaus im digitalen Zeitalter. Eine Herausforderung für sammlungsorientierte Bibliotheken, in: Hauke, Petra; Kaufmann, Andrea; Petras, Vivien (Hg.): *Bibliothek – Forschung für die Praxis. Festschrift für Konrad Umlauf zum 65. Geburtstag*, Berlin 2017, S. 321–337, <https://doi.org/10.1515/9783110522334-028>; Kempf, Klaus: Die Sammlung: Ein Auslaufmodell? Bestandsaufbau in digitalen Zeiten, in: Alker-Windbichler, Stefan; Kuhn, Axel; Lodes, Benedikt u.a. (Hg.): *Akademisches Lesen*, Göttingen 20221, S. 237–264, <https://doi.org/10.14220/9783737013970.237>.

Die Anzahl der Open-Access-Beiträge von wissenschaftlicher Literatur liegt mittlerweile bei gut 70 %. Das ist eine beachtliche Steigerung gegenüber 2014, als der Anteil noch bei 50 % lag.² Dass Open Access (OA) die Zukunft gehört, dürfte unstrittig sein, auch wenn der Anteil nach den Zahlen des OA-Monitors³ in den letzten Jahren nicht mehr so stark wächst (71–75 % in den Jahren 2020 bis 2023). Interessant ist, dass sich auch die Nutzung lizenzierter elektronischer Werke nach einer Coronadelle nicht wieder erholt hat bzw. stagniert.

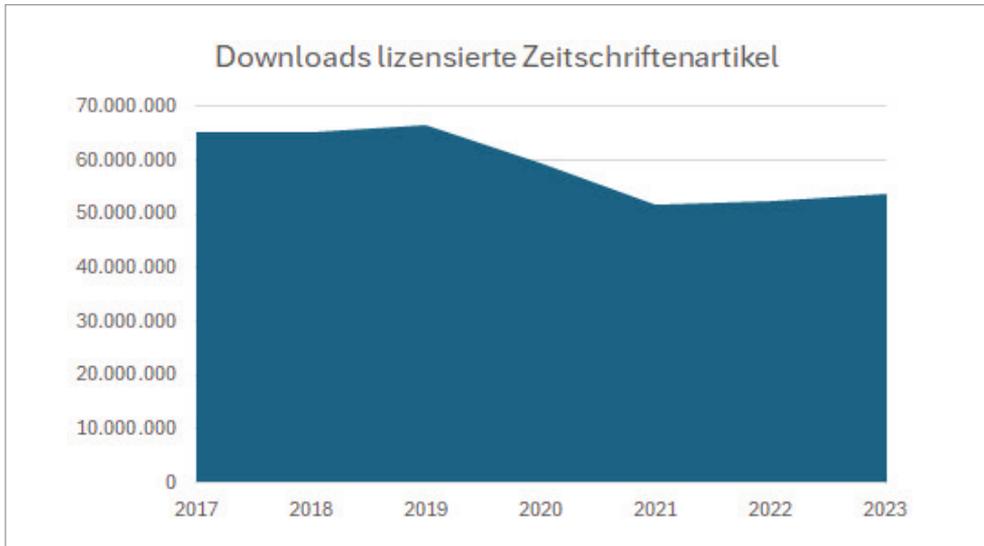


Abb. 2: Downloads lizenzierte Zeitschriftenartikel, Quelle: DBS

Die Nutzung von elektronischen Ressourcen verschiebt sich also nicht nur entlang der Achse von analog zu digital, sondern auch von lizenziert zu offen, ein Phänomen, das in der Logik des Digitalen liegt und auch in der Literatur vorhergesagt wurde: “In der langfristigen Perspektive wird der Informationsraum durch das Netz abgelöst werden; der Content steht im OA bereit; eines Zwischenhändlers für die Handhabung von Lizenzierungen bedarf es dann nicht mehr”⁴. Hinzu kommt, dass die erfolgreichen Retrodigitalisierungsmaßnahmen der letzten Jahre – man denke an die Förderung der Digitalisierung der VDs (Verzeichnis der im deutschen Sprachbereich erschienenen Drucke) durch die DFG – zunehmend die Sondersammlungslesesäle entvölkert haben⁵ und urheberrechtsfreie Materialien

2 Piwowar, Heather; Priem, Jason; Larivière, Vincent u. a.: The state of OA. A large-scale analysis of the prevalence and impact of Open Access articles, in: PeerJ 6, 13.02.2018, S. 4, <https://doi.org/10.7717/peerj.4375>.

3 <https://www.fz-juelich.de/de/zb/open-science/open-access/open-access-monitor>, Stand: 13.09.2024.

4 Altenhöner, Reinhard: Aspekte des Bestandsaufbaus im digitalen Zeitalter. Eine Herausforderung für sammlungsorientierte Bibliotheken, in: Hauke, Petra; Kaufmann, Andrea; Petras, Vivien (Hg.): Bibliothek – Forschung für die Praxis. Festschrift für Konrad Umlauf zum 65. Geburtstag, Berlin 2017, S. 321–337, <https://doi.org/10.1515/9783110522334-028>.

5 Dazu gibt es keine gesicherten Zahlen, allerdings ist es eine breit geteilte Wahrnehmung, dass die Nutzung von Sonderlesesälen stark zurückgegangen ist. Frühere Hoffnungen, dass mit der Digitalisierung eher das Gegenteil eintreten würde und Digitalisate Nutzer zum Original führten, haben sich nicht erfüllt.

durch effiziente Prozesse auf Nachfrage (*Books on Demand*) digitalisiert werden. Die Verschiebung des Erwerbungssetats hin zu einem Informationsbudget flankiert diese Prozesse, und es ist absehbar, dass der Erwerbungssetat in Zukunft immer weniger für Käufe und Lizenzen, als für die Finanzierung von Publikationen aufgewendet wird.⁶ Für die bibliothekarische Sammlung blieben allenfalls eng abgezielte landesbibliothekarische Aufgaben der Pflichtablieferung zur Sicherung des kulturellen Erbes oder die retrospektive Nationalbibliothek der Sammlung Deutscher Drucke,⁷ die womöglich in Zukunft als Kopiergrundlage dient oder als Möglichkeit, die Authentizität und Integrität einer Kopie zu überprüfen. Die lokale Sammlung würde demnach nur noch für urheberrechtsbewehrte Werke, Werke in nicht-digitaler Form, Pflichtexemplare oder Originale der kulturellen Überlieferung benötigt. Die Aufgabe der Bibliothek reduzierte sich auf ihre lokale Archivfunktion und mit der schleichenden Musealisierung ihrer Bestände schwände auch ihre Relevanz für Lehre und Forschung.⁸

Offensichtlich sind die Konsequenzen aus diesem Wandel für die Bibliotheken gravierend, denn warum sollte eine Bibliothek noch außerhalb der besagten engen Aufgaben sammeln, wenn alles Open Access ist? Die Frage ist deswegen kritisch für die Bibliothek, weil die Sammlung trotz gegenteiliger Aussagen⁹ ein wesentliches Konstitutiv von Bibliothek ist und sie sich über die Jahrhunderte hinweg über die Trias Sammeln, Erschließen, Vermitteln in immer neuen Formen und Gestalten definiert hat. Hat die Bibliothek mit der forcierten Unterstützung von Open Access, wie Jochum und andere einst warnten,¹⁰ womöglich an ihrem eigenen Ast gesägt? Tatsache ist, dass man die Bibliothek, wenn alle

6 Schimmer, Ralf: Open Access und die Re-Kontextualisierung des Bibliothekserwerbungssetats, in: *Bibliothek Forschung und Praxis* 36 (3), 2012, <https://dx.doi.org/10.1515/bfp-2012-0038>.

7 <https://www.ag-sdd.de>

8 Vgl. Rat für Informationsinfrastrukturen (RfII): Sammlungen als multimodale Infrastrukturen. Analog und digital für die verknüpfte Nutzung erschließen, Göttingen 2024, <https://nbn-resolving.org/urn:nbn:de:101:1-2023021766>. Der RfII betont zwar mit Recht die weitere Bedeutung des Analoges für *wissenschaftliche* Sammlungen, etwa durch Aufbewahrung medizinischer Proben, allerdings bestehen bibliothekarische Sammlungen in ihrem Kern aus *Publikationen*, zielen also traditionell auf schriftkulturelles und schriftbasiertes Wissen ab, das Gegenstand eines Kommunikationsprozesses ist und war. In jüngerer Zeit kommen Daten und Software hinzu, die aber ebenfalls eher als Publikationen betrachtet werden und insofern keinen unikatlen Charakter tragen, so dass die Argumentation des RfII zugunsten des Analoges oder der Verschränkung von Analogem und Digitalem für bibliothekarische Sammlungen nur in der historischen Dimension für materielle Informationsträger greift und für moderne Medien allenfalls Übergangsprozesse umfasst. Die analoge bibliothekarische Sammlung behält ihr Recht, weil die Materialität (das Analoge) das Verständnis der Schrift- bzw. des Textes modifiziert. Schrift auf einer mittelalterlichen Handschrift gibt dem Text eine weitere Bedeutungsdimension, die zwar forschungsrelevant, aber nicht unbedingt mit dem Text als solchem identisch ist, der auch unabhängig vom Träger, z.B. als verlustfrei kopierbare Datei auf flüchtigen Speichern existieren kann. Der digitale Speicher ist anders als der materielle Träger gegen den Text indifferent, wodurch keine durch die Materialität bedingte Unikalität mehr möglich ist. Das RfII-Papier verunklart diesen Sachverhalt, indem es von einer „Materialität des Digitalen“ (S. 14) spricht und ein Kontinuum unterstellt, das es so nicht gibt. Unglücklich ist auch, dass die von Thibodeau, K. „The State of Digital Preservation: An International Perspective, chapter Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years“, Report Number 31, CLIR, 2002, eingeführte Unterscheidung von *Physical*, *Logical* und *Conceptual Object* nicht berücksichtigt wird, was zu Fehlurteilen hinsichtlich der Bedeutung von „verlustfreiem Kopieren“ führt (gegen das Allianzpapier, das diese Unterscheidung berücksichtigt und eben nicht von verlustfreiem „Transformieren“ spricht, welches nicht auf der physischen Ebene stattfindet; s. RfII S. 13 und Anm. 20). Verlustfreiheit und Trägeragnostik im physischen Kopiervorgang ist ein wesentliches Charakteristikum des Digitalen und so im Analogen nicht möglich. Diese Verlustfreiheit und Disseminationseffizienz macht den entscheidenden Unterschied in der Frage des Sammlungsbaus. Tatsache bleibt, dass die OA-Logik in der Publikation (!) das Analoge aus der aktuellen bibliothekarischen Sammlung zurückdrängen wird. Anders formuliert, die Schallplatte wird zwar weiter für Liebhaber und die historische Forschung Relevanz haben, die Musik spielt aber anderswo.

9 Lankes, R. David: *The Atlas of New Librarianship*. Cambridge [Mass.] et al. 2016, <https://dx.doi.org/10.7551/mitpress/9780262529082.003.0008>.

10 Jochum, Uwe: Die Selbstabschaffung der Bibliotheken, in: Jochum, Uwe; Schlechter, Armin (Hg.): *Das Ende der Bibliothek? Vom Wert des Analoges*. Frankfurt am Main 2011, S. 11–25.

wissenschaftlichen Werke im Open Access zugänglich sind, nicht mehr als Vermittler braucht. Die Bibliothek ist als ein Ort, der Zugang zu Information und Literatur zum Zwecke des Lesens schafft, mit dieser Entwicklung an ein Ende gekommen. Wer einen Titel lesen will, findet ihn im Netz und nicht mehr in der Bibliothek. So formulierte es schon Anderson mit einer seinerzeit provozierenden These:

“I predict that within ten years, the days of wondering whether or not one’s library can give access to a desired resource will be a fading memory, in much the same way that reliance on a printed card catalog is a faint and fading memory in 2011.”¹¹

Doch ist damit auch die bibliothekarische Sammlung an ihr Ende gekommen? Die Digitalisierung verändert nicht nur die Bibliothek, sondern auch die Wissenschaft und Forschung. In den Natur- und Technikwissenschaften, in der Medizin, in den Rechts-, Wirtschafts- und Sozialwissenschaften und den Geisteswissenschaften – prominent im Bereich der Digital Humanities¹² – sind digitale Werkzeuge auf dem Vormarsch. Zugleich rücken mit der Digitalisierung des Publikationswesens und der Maschinenlesbarkeit der Medien digitale Auswertungsmethoden für Informationen und Literatur in den Fokus und lassen Forderungen nach einer *Datafication* der bibliothekarischen Sammlung laut werden.¹³ Die bibliothekarische Sammlung, so zeichnet sich ab, wird in Zukunft nicht nur oder vielleicht sogar nicht einmal mehr maßgeblich dem Lesen dienen (als Literatur), sondern der algorithmischen Auswertung, dem „Rechnen“ von digitalisierter Literatur (als Daten), wie es Lauer einst formulierte.¹⁴ Data Science, ein Begriff, den es schon seit 40 Jahren gibt, erhält spürbar neuen Auftrieb, was die wachsende Anzahl von Publikationen in diesem Bereich belegt,¹⁵ und es ist nicht zu übersehen, dass wir es mit grundsätzlichen Veränderungen zu tun haben. Mit einem gewissen Recht kann man daher mit Hey von einem *Fourth Paradigm* der *Data-Intensive Science*¹⁶ sprechen, das sich nach den Paradigmata der Empirie, der Theorie, und der Computersimulation jetzt Bahn bricht. Data Science aber braucht Daten. Nun könnte man annehmen, dass Bibliotheken mit ihren Sammlungen prädestiniert für die Lieferung dieser Daten sind. Doch überraschenderweise haben Bibliotheken in der Fläche keine oder nur sehr selektiv Daten gesammelt, geschweige denn in einer Form, dass man damit wissenschaftlich arbeiten könnte. Digitale Medien oder elektronische Literatur und damit potentielle Daten wurden trotz der E-Only-Devisen in den letzten zwanzig Jahren überwiegend nicht gesammelt, sondern nur lizenziert. Die für Data Science notwendigen Dateien (Files) befinden sich im Regelfall auf Verlagsservern. Ggf. vorhandene Archivkopien (vgl. Nationales Hosting elektronischer Ressourcen (NatHosting) u.ä.) oder auch Kopien von National- und Allianzlizenzen stehen in der Regel der

11 Anderson, Rick: Collections 2021. The future of the library collection is not a collection, in: *Serials. The Journal for the Serials Community* 24 (3) 2011, S. 211-215, <http://dx.doi.org/10.1629/24211>, S. 211.

12 Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte (Hg.): *Digital Humanities*, Stuttgart 2017, <https://doi.org/10.1007/978-3-476-05446-3>.

13 Vgl. z.B. das Kapitel 1.2 zum Digitalen Zwilling in: Rat für Informationsinfrastrukturen (RfII): *Sammlungen als multimodale Infrastrukturen. Analog und digital für die verknüpfte Nutzung erschließen*, Göttingen 2024, <https://nbn-resolving.org/urn:nbn:de:101:1-2023021766>.

14 Lauer, Gerhard: Literatur rechnen. Lektüre im Computerzeitalter, in: *FAZ* vom 26.08.2009, Online: <https://www.faz.net/aktuell/feuilleton/geisteswissenschaften/literatur-rechnen-lecture-im-computerzeitalter-1840973.html>, Stand: 13.09.2024.

15 Eine Recherche im GBV nach “data science” ergibt: Erscheinungsjahr 2000:117, 2010:312; 2015: 394, 2020: 782 Treffer.

16 Hey, Tony: *The fourth paradigm. Data-intensive scientific discovery*, Redmond, Washington 2009.

Forschung nicht unmittelbar zur Verfügung und werden selbst berechtigten Einrichtungen wegen schwieriger lizenzrechtlicher Rahmenbedingungen nicht ohne weiteres zur Nutzung überlassen. Für das Lesen waren die Lizenzen ausreichend, nicht aber für die Data Science. Frühe Warnungen, wie die von Doina Oehlmann¹⁷ blieben seinerzeit ungehört, mit der Folge, dass Bibliotheken heute praktisch mit leeren Händen dastehen. Jetzt könnte man im Sinne des Zitats von Anderson sagen: Das ist doch alles im Netz! Die analogen oder lizenzierten Objekte, die es heute noch gibt, werden digitalisiert sein oder in Zukunft weitgehend frei im Netz zur Verfügung stehen. Diese naive Sicht verkennt die Anforderungen, die die Data Science an die Daten stellt und übersieht den Unterschied zwischen Lesen und Auswerten. Für das Lesen ist es sicher richtig, aber für das Auswerten sind andere Anforderungen an die Sammlung zu stellen. Ein einfacher Versuch möchte es verdeutlichen. Wenn ich z.B. alle Artikel über den Klimawandel, die im Directory of Open Access Journals (DOAJ) nachgewiesen sind, auswerten möchte, stehe ich vor beträchtlichen Hürden. Meine Anfrage nach *Climate Change* führt zu 95.338 Treffern. Wenn ich mir den ersten Treffer ansehe, erhalte ich einen Link zu einer HTML-Seite, von dort führt ein Link auf ein PDF, das aber eine eigene Applikation von Wiley öffnet. Erst ein Download führt zu einem PDF, das ich theoretisch herunterladen und auswerten kann, auch wenn PDF anders als z.B. JATS nicht optimal für Text Mining bzw. Data Science ist. Wie aber komme ich an die anderen 95.337 Treffer bzw. Texte? Kann ich mit vernünftigen Zeitaufwand ein Skript schreiben, das die Besonderheiten verlegerischer Angebote und Schnittstellen abbildet? Wohl kaum. Oder lade ich jedes einzelne Dokument per Hand herunter? Wie aufwändig wäre das für 95.000 Treffer? Wie komme ich an gut strukturierte Texte, die es erlauben, Fußnoten, Inhaltsverzeichnisse, Stichwörter u.ä. zu filtern? Es wird schnell klar, dass die Angebote im Netz alles andere als bequem zu benutzen und sozusagen Data-Science-Ready sind. Eine Nutzung der im DOAJ zugänglichen OA-Materialien für Zwecke des Text- und Data-Mining (TDM) ist so gut wie unmöglich oder so aufwändig, dass die meisten an dieser Aufgabe scheitern werden. Das gilt nicht nur für das DOAJ, sondern für die überwiegende Zahl der offenen Portale im Netz. Sie sind Zugänge zu Lesefassungen und deren Metadaten, nicht aber zu den Daten selbst. Was also hat Kempf gemeint, wenn er schreibt: „... die überwiegende Zahl der wissenschaftlichen Bibliotheken ... [wird] ihren eigenen Bestand im Sinne einer ganz auf Service abstellenden Gebrauchssammlung, eine ›online core collection‹ für das Tagesgeschäft, ausrichten und letztlich nur noch ›just in time‹ »sammelnd« aktiv werden“?¹⁸ *Just in Time* braucht im Bereich der Daten die Aggregation „*Just in Case*“ bzw. vorgeschaltete Prozesse des *Data Cleansing* und der Standardisierung. Sammeln ist Ordnen innerhalb eines Ordnungsrahmens. Ohne Ordnung sind Daten nicht oder nur eingeschränkt nutzbar. Wie die Bibliothek als bloße Ansammlung von Literatur, die ungeordnet auf einem Haufen liegt, nicht nutzbar ist, selbst wenn alles “da” wäre, so sind auch ungeordnete Daten nicht nutzbar. Entsprechend ist schon früh bemerkt worden, dass das Internet keine Bibliothek ist.¹⁹ Erst die Aufbereitung zu kuratierten digitalen Sammlungen mit einheitlichen Dateiformaten schafft die Voraussetzungen, die für eine wissenschaftliche Nutzung nötig

17 Oehlmann, Doina: Lizenzen oder Texte, Nutzung oder Hosting: können Bibliotheken ihren Auftrag in Zeiten elektronischer Texte weiterhin ausreichend erfüllen?, in: Zeitschrift für Bibliothekswesen und Bibliographie 59, 2012, S. 231–235, <http://dx.doi.org/10.3196/186429501259511>.

18 Kempf, Klaus: Die Sammlung: Ein Auslaufmodell? Bestandsaufbau in digitalen Zeiten, in: Alker-Windbichler, Stefan; Kuhn, Axel; Lodes, Benedikt u. a. (Hg.): Akademisches Lesen, Göttingen 20221, S. 254, <https://doi.org/10.14220/9783737013970.237>.

19 Zimmer, Dieter E.: Die Bibliothek der Zukunft: Text und Schrift in den Zeiten des Internet, Hamburg 2001.

sind. Zur Ordnung gehört die Frage der Reproduzierbarkeit und Provenienz von Daten und Analysen ebenso wie die der Datensouveränität. Schon heute gibt es Angebote im Netz, die das Aggregieren und Sammeln von eigentlich offenen Texten und Daten vermarkten²⁰ und damit auch Möglichkeiten der ungehinderten Nutzung von freien Ressourcen in Frage stellen. Die Entwicklungen der Vergangenheit sollten Bibliotheken eine Warnung sein. Nachdem das “Lesen”, um einen Ausdruck von Marx zu verwenden, expropriert wurde, darf man nicht zulassen, dass auch das “Auswerten” expropriert wird. Nur das systematische eigene Sammeln von digitalen Publikationen und deren Aufbereitung zu Daten für Zwecke des TDM oder Künstliche Intelligenz (KI) stellt für die Zukunft freie Lehre und Forschung wirklich sicher und ist der logische Schritt der Weiterentwicklung der Bibliothek als zentraler Ort der öffentlichen Informationsversorgung. Erforderlich dafür ist nichts weniger als eine vollständige Umgestaltung der bisherigen Geschäftsgänge und Arbeitsabläufe, die sich strategisch an folgendem Dekalog von Leitgedanken orientieren kann:

1. Jede erworbene (lizenzierte oder OA) Publikation wird als Datei (*File*) erworben. Retrodigitalisierung ist ein Erwerbungsfall.
2. Jede Publikation wird in Originalform gespeichert und zusätzlich in mindestens ein Standardformat konvertiert.
3. Erworbene Dateien werden in der Regel dauerhaft nach OAIS-Prinzipien²¹ archiviert.
4. Alle relevanten Metadaten werden aus den Daten extrahiert und (semi-)automatisch standardisiert.
5. Standards sind webtaugliche Standards und berücksichtigen die Maschinenlesbarkeit der Publikationen.
6. Publikationen werden nach qualitativen Kriterien (Erwerbungsrichtlinien) gesammelt (gefiltert).
7. Digitale Sammlungen lassen sich als ganze oder in Teilen nach inhaltlichen, formalen und transformativen²² Kriterien zu frei konfigurierbaren Datensets zusammenstellen und abrufen.
8. Datensets von OA-Publikationen sind auch OA.
9. Datensets können zum Zwecke der Reproduzierbarkeit persistent adressiert werden.
10. Digitale Sammlungen können als ganze nachgenutzt und ihrerseits aggregiert werden.

Gerade der letzte Punkt zeigt, dass die digitale Sammlung im Sinne von *Collections as Data*²³ um einen Himmel verschieden ist von der früheren analogen Sammlung. Sammlungen in diesem Verständnis sind selbst nichts anderes als Bücher im Regal einer Bibliothek und bilden ihrerseits virtuelle Bibliotheken. Die Sammlung hat darin ihre aus dem langjährigen Fleiß der Bibliothekarinnen und Bibliothekare und aus dem Sammlerglück der Zeiten erwachsenen Besonderheit und Einzigartigkeit eingebüßt. Im Gegenteil, die Qualität und der Wert einer Sammlung zeigt sich weniger in ihrer Besonderheit als

20 Vgl. <https://core.ac.uk/>, Stand: 13.09.2024.

21 OAIS = Open Archival Information System, vgl. <https://www.iso.org/standard/57284.html>, Stand: 13.09.2024.

22 Gemeint ist, dass Sammlungen oder Teilsammlungen in verschiedene Formate überführt werden können, z.B. plain text, JATS/XML, TEI/XML usw.

23 Vgl. das Santa-Barbara-Statement zu Collections as Data: <https://zenodo.org/records/3066209>, Stand: 13.09.2024.

in ihrer Offenheit und Austauschbarkeit.²⁴ Das Digitale kennt keine Ortsgebundenheit und keinen Besitzerstolz.²⁵ Original und Kopie sind ununterscheidbar. Der Mehrwert des Digitalen liegt in seiner Teilbarkeit.²⁶ Und doch bleibt der Ort als Schnittstelle und Drehscheibe, als Integrationspunkt und Filter, als Garant von Qualität, Zuverlässigkeit und Authentizität von größter Relevanz für die Sammlung. Hier fließen die Daten nach Maßgabe der jeweiligen Anforderungen zusammen und von hier werden sie abgerufen. Die Bibliothek sorgt für das Herstellen von FAIRen Publikationen qua Daten. Sie aggregiert und vereinheitlicht (*Data Cleansing*) interne und erworbene externe Sammlungen. Sie schafft Möglichkeiten der Selektion *Just in Time*, der Rekombination oder Neuordnung von Sammlungsteilen oder der Informations- und Wissensextraktion. Sie transformiert Daten und stellt Derivate zur Verfügung, um TDM- oder KI-Anwendungen zu unterstützen.

Nun könnte man der Meinung sein, dass es ja nur einer einzigen großen digitalen Weltbibliothek bedürfe, um von dort nach dem Wort von Kempf *Just in Time* die Daten zu holen, die gerade benötigt werden. Abgesehen davon, dass diese Phantasie nicht neu ist und schon die Gründer der Bibliothek von Alexandria bewegte, mag der Gedanke theoretisch richtig sein. Aber weder ist eine solche Bibliothek in Sicht, noch gäbe es wohl eine einzige Einrichtung, die in der Lage wäre, eine solche allumfassende Bibliothek aufzubauen und zuverlässig für die Welt zu hosten. Die praktischen Hürden wären enorm und die politischen Risiken in einer zunehmend unsicherer werdenden Welt groß. Was geschähe mit dieser Bibliothek, wenn das Sitzland sich entschiede, die Sammlungen zu sperren oder kriegerische Einwirkungen ihren Fortbestand bedrohten? Der eine Ort ist nicht nur Segen, sondern auch Fluch.²⁷ Umgekehrt ist Redundanz im Digitalen anders als im Analogen kein Makel. Der Platzbedarf ist perspektivisch, was den Text anlangt, vernachlässigbar. Was spricht also dagegen, dass alle alles haben, wenn ohnehin die Auswertung in den meisten Fällen vor Ort geschehen wird? Oder wie Hey formulierte: „But fundamentally, you can stick a document in any of these archives and it will get replicated to all the other archives“.²⁸

Dass eine Weltbibliothek weniger nach dem Top-Down- als nach dem Bottom-Up-Modell denkbar ist, ist dagegen eine ernstzunehmende These: „Eine Weltbibliothek im digitalen Zeitalter ist nicht mehr so strikt an die räumlichen Bedingungen gebunden, wie noch ihre Vorgängerin. Um daher ihr besonderes Potential zu heben, muss sie es ermöglichen, dass ihre Bestände mit denen anderer Bibliotheken, Archive und Sammlungen zu Korpora verknüpft werden können, digital und konzeptuell.“²⁹ Die Arbeitsteilung einer einheitlichen Aufbereitung ist eine vielversprechende, ja, notwendige

24 Stäcker, Thomas: Die Sammlung ist tot, es lebe die Sammlung!, in: Bibliothek Forschung und Praxis 43 (2), 2019, S. 304–310, <https://doi.org/doi:10.1515/bfp-2019-2066>.

25 Der RfII spricht davon, dass „das Besitzprinzip, den jeweiligen Bestand betreffend, [...] relativiert“ wird, vgl.: Rat für Informationsinfrastrukturen (RfII): Sammlungen als multimodale Infrastrukturen. Analog und digital für die verknüpfte Nutzung erschließen, Göttingen 2024, <https://nbn-resolving.org/urn:nbn:de:101:1-2023021766>, S. 33.

26 Der RfII bringt „Sammlungsbestände als eine Art Allmende“ ins Spiel, vgl.: Rat für Informationsinfrastrukturen (RfII): Sammlungen als multimodale Infrastrukturen. Analog und digital für die verknüpfte Nutzung erschließen, Göttingen 2024, <https://nbn-resolving.org/urn:nbn:de:101:1-2023021766>, S. 34.

27 Ovenden, Richard: *Burning the books. A history of the deliberate destruction of knowledge*, Cambridge, Massachusetts 2022.

28 Hey, Tony: *The fourth paradigm. Data-intensive scientific discovery*, Redmond, Wash. 2009, S. XXVII.

29 Lauer, Gerhard: Die Weltbibliothek und ihre Korpora. Bibliothekarische Infrastruktur als Bedingung der Möglichkeit zur Korpusbildung, in: Zeitschrift für Bibliothekswesen und Bibliographie 61 (4–5) 2014, S. 251–53, <https://doi.org/10.3196/18642950146145149>, S.252.

Perspektive und wie heute Bibliotheken gemeinsam an Metadaten in Bibliotheksverbänden arbeiten, so arbeiten sie vielleicht in Zukunft gemeinsam an Daten in Datenverbänden und erzeugen einheitlich kodierte Publikationen, die zu interoperablen Korpora zusammenwachsen. Die Zukunft wird zeigen, welches Prinzip sich durchsetzt. Ganz sicher wäre es aber falsch, darauf warten zu wollen. Die Zeit zum Sammeln ist jetzt.

Thomas Stäcker, Universitäts- und Landesbibliothek Darmstadt / Fachhochschule Potsdam;
ORCID: <https://orcid.org/0000-0002-1509-6960>

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/6101>

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