

Editorial #41: Big Scholarly Data

Redaktion LIBREAS

Diese Ausgabe ist Zeugnis einer Schwerpunktverlagerung – und zwar vom Konkreten zum Allgemeinen und zurück zu einem anderen Konkreten. War der Ausgangspunkt unseres Call for Paper noch der Verlust von Microsoft Academic Services und damit das Einfrieren des Microsoft Academic Graph (MAG), einer großen und frei verfügbaren Datensammlung über wissenschaftliche Literatur, so scheinen mittlerweile eher Fragen nach dem Umfang und der Vielfalt der Landschaft von Big Scholarly Data von hoher Relevanz für Bibliotheken und Informationseinrichtungen. Insbesondere das Ende 2021 öffentlichkeitswirksam gestartete OpenAlex, vom Anspruch her eine Art Weltkatalog des wissenschaftlichen Publikationsaufkommens,¹ wird in Forschung und Praxis als offene Alternative zu den kommerziellen Bibliometriedatenbanken wie dem Web of Science, Scopus oder Dimensions erörtert.

Zugleich schlägt sich etwa im Informationspapier „Datentracking in der Wissenschaft“ des Ausschusses für Wissenschaftliche Bibliotheken und Informationssysteme der Deutschen Forschungsgemeinschaft² ein gestiegenes Bewusstsein über negative Folgen des Wandels der Geschäftsmodelle wissenschaftlicher Verlage in Richtung Data Analytics für Wissenschaft und Bibliotheken nieder. Paradoxerweise würde gerade das ursprünglich als Gegenmodell zur übermäßigen Kommodifizierung der Wissenschaft angedachte Konzept des Open Access heute zu neuen verlagsspezifischen Plattformlösungen führen. Großverlage bereiten Daten über wissenschaftliche Informationsprozesse auf und bieten diese wissenschaftlichen Einrichtungen für die Unterstützung komplexer Entscheidungsfindungen an, etwa im Kontext der Anbahnung und Evaluation von Open-Access-Transformationsverträgen.³

Dahingegen glauben wir, dass der souveräne, also anbieterunabhängige Umgang mit Big Scholarly Data in der bibliothekarischen Berufspraxis ein Schlüssel ist, um diesen Trend der Kommodifizierung von Daten über wissenschaftliche Publikationen zu begegnen. Wir freuen uns daher, in der Ausgabe #41 vielfältige Anwendungsbeispiele von Big Scholarly Data für die Beantwortung von bibliothekarischen Informationsbedürfnissen versammeln zu dürfen.

In ihrem Werkstattbericht stellen Franziska Stanzel, Irene Barbers, Philipp Pollack und Barbara Lindstrot von der Bibliothek des Forschungszentrums Jülich ihre Arbeit mit kommerziellen als

¹Singh Chawla, D. (2022). Massive open index of scholarly papers launches. Nature.

<https://doi.org/10.1038/d41586-022-00138-y>

²Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme der Deutschen Forschungsgemeinschaft (2021). Datentracking in der Wissenschaft: Aggregation und Verwendung bzw. Verkauf von Nutzungsdaten durch Wissenschaftsverlage.

https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_de.pdf

³Scholl, D. (2021). Datentracking: Die schöne neue Welt der Wissenschaftsverlage. Wikimedia Deutschland.

<https://blog.wikimedia.de/2021/12/07/datentracking-die-schoene-neue-welt-der-wissenschaftsverlage/>

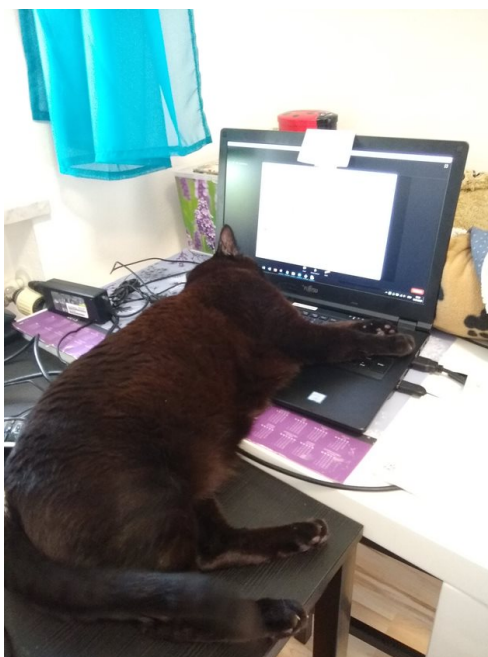


Abbildung 1: Redaktionsorte XX: Arbeitstreffen online.

auch offenen Datensätzen im Kontext des Open Access Monitors vor. Sie gehen dabei insbesondere auf ihre Normierungsaufwände ein, die entstehen, wenn unterschiedliche Datenquellen miteinander kombiniert werden.

Die Beraterin und Projektmanagerin Anne Christensen erläutert die Frage nach Gemeinsamkeiten und Unterschieden zwischen akademischen Suchmaschinen und den in Bibliotheken eingesetzten Discovery-Systemen. Sie sieht sowohl bezogen auf die Usability als auch der Öffnung der Datenbestände ein großes Potenzial für Discovery-Systeme, damit sie stärker als bisher mit ihren zugrunde liegenden Erschließungsverfahren als Suchraum für wissenschaftliche Literatur sowie als Grundlage für Datenanalysen wahrgenommen werden.

Philipp Zumstein, Universitätsbibliothek Mannheim, nutzt offene Daten, um die Publikationspraxis des Mega-Journals Academia Letters zu beschreiben. Insbesondere die Vernetzung mit dem sozialen Netzwerk Academia und ein Einreichungs- und Begutachtungsworkflow, der auf kurze Fristen setzt, sind mögliche Erklärungen für die gestiegene Anzahl an Artikeln, die in den letzten Monaten im Journal veröffentlicht wurden. Er beleuchtet zudem Aspekte der (fehlenden) Qualitätssicherung und Transparenz des Publikationsservice.

Nick Haupka, Najko Jahn und Anne Hobert von der SUB Göttingen illustrieren abschließend, wie sie dank Cloud-Computing-Umgebung auch ohne eigene Datenbank- und Serveradministration Analysen großer Datenmengen bewerkstelligen. Dabei fokussieren sie auf die Datenquellen Unpaywall, Crossref und OpenAlex, die etwa für die Beantwortung von Fragestellungen im Kontext des hybriden Open Access genutzt werden.

Neben diesen Schwerpunktbeiträgen untersucht der Bibliothekswissenschaftler Karsten Schuldt die Einführung von Minitel und Btx an Bibliotheken in den 1980er Jahren als ein zeithistorisches Beispiel, wie Bibliotheken neue Technologien aufgriffen. Und selbstverständlich fasst auch in

dieser Ausgabe die Kolumne DLDL in der Tradition einer Referatezeitschrift fachliche Beiträge zusammen. Obwohl man heute vielleicht noch überzeugender wirken könnte, wenn man von einem Small-Scholarly-Data-Graphen spricht. Gern weisen wir zudem darauf hin, dass die bibliografischen Daten der in DLDL besprochenen Beiträge nun in einer öffentlich zugänglichen Zotero-Gruppe zu finden sind: https://www.zotero.org/groups/4620604/libreas_dldl/library

Viel Vergnügen und Anregung wünscht die LIBREAS Redaktion!

(Berlin, Hannover, Göttingen, Lausanne, München, Potsdam)