

Wissenschaftliche Literatur entdecken: Was bibliothekarische Discovery-Systeme von der Konkurrenz lernen und was sie ihr zeigen können

Anne Christensen

Zusammenfassung: In den letzten Jahren ist das Angebot an Academic Search Engines für die Recherche nach Fachliteratur zu allen Wissenschaftsgebieten stark angewachsen und ergänzt die beliebten kommerziellen Angebote wie Web of Science oder Scopus. Der Artikel zeigt die wesentlichen Unterschiede zwischen bibliothekarischen Discovery-Systemen und Academic Search Engines wie Base, Dimensions oder Open Alex auf und diskutiert Möglichkeiten, wie beide von einander profitieren können. Diese Entwicklungsperspektiven betreffen Aspekte wie die Kontextualisierung von Wissen, die Datenmodellierung, die automatische Datenanreicherung sowie den Zuschnitt von Suchräumen.

Dieser Text ist Dr. Dieter Müller (1948–2022) in Liebe und Dankbarkeit gewidmet.

Vorbemerkung: Dieser Text sollte eigentlich eine Rezension für OpenAlex werden. Diese Academic Search Engine wurde Ende 2021 angekündigt und erscheint hinsichtlich ihres Datenmodells und der grundsätzlichen Offenheit von Code und Daten als besonders vielversprechende Ergänzung des bestehenden Marktes. Da zum Zeitpunkt der Veröffentlichung jedoch nur die Schnittstelle, nicht aber eine eigene Benutzungsoberfläche zur Verfügung stand und ein Test ohne Heranziehung zusätzlicher Tools nicht erfolgen konnte, werden nun die Academic Search Engines insgesamt aus bibliothekarischer Perspektive betrachtet. In diese Betrachtung sind neben den zitierten Quellen vor allem auch meine persönlichen Erfahrungen und Bewertungen aus über 15-jähriger theoretischer und praktischer Beschäftigung mit Discovery-Systemen eingeflossen.

Über Jahrzehnte galt in Bibliotheken: Kataloge für Bestandsnachweise, Bibliografien für darüber hinausgehende, vor allem thematische Suchen. Mit Discovery-Systemen versuchen Bibliotheken, diese Dualität aufzuheben, um einen zentralen Einstiegspunkt für die Recherche anzubieten, der sowohl lokale Bestände als auch zentrale bibliografische Ressourcen recherchierbar macht. Die Vielzahl von Herausforderungen, die es dafür in den Bereichen Metadatenmanagement, Technik und Lizenzen zu bewältigen gibt, sind noch nicht überwunden, und die Systeme sind weit hinter den Erwartungen zurückgeblieben.¹

¹Christensen, Anne, und Matthias Finck. 2021. „Discovery-Systeme: Eine Analyse ihrer Geschichte und Gegenwart mit dem Hype-Zyklus“. *Bibliothek Forschung und Praxis* 45 (3): 497–508. <https://doi.org/10.1515/bfp-2021-0039>

Gleichzeitig sind neben dem klassischen, fachübergreifenden Web of Science in den vergangenen 20 Jahren weitere Plattformen für die Recherche nach wissenschaftlicher Literatur entstanden – die Academic Search Engines. Unter diesen war unter anderem Google mit dem Ableger Google Scholar ein Pionier, aber auch die Bielefeld Academic Search Engine (BASE). Diese geht genauso wie Google Scholar auf das Jahr 2004 zurück und ist inzwischen ein etabliertes und unabhängiges Instrument mit internationalem Renommee. Besonders in den letzten Jahren ist das Angebot solcher Suchmaschinen noch einmal weiter gewachsen: Es gibt sowohl kommerzielle Anbieter, bei denen wesentliche Teile des Angebotes kostenlos nutzbar sind, als auch Dienste, die ausschließlich auf öffentliche Finanzierung und freie Zugänglichkeit und Nachnutzung setzen.²

Die Rezeption der Academic Search Engines in Bibliothekskreisen ist erstaunlich zurückhaltend, obwohl diese eine Rolle als potenzielle Empfänger von Metadaten aus Repositorien spielen, und sie Beobachtungsgegenstand der bibliometrischen Forschung sind.³ In Erwerbsabteilungen werden mitunter Vergleiche zwischen den kommerziellen Diensten Scopus und Web of Science gezogen, um Entscheidungen über den (Weiter-)Bezug dieser Dienste zu treffen. Was jedoch fehlt, ist eine Auseinandersetzung mit den Funktionen, Technologien und Strategien von Academic Search Engines im Zusammenhang mit den bibliothekarischen Discovery-Systemen.

Kaum eine Universitätsbibliothek verzichtet inzwischen darauf, neben oder sogar statt des klassischen Bibliothekskataloges, ein Discovery-System anzubieten, das neben den OPAC-Funktionalitäten einen erweiterten Suchraum sowie eine zeitgemäße Benutzungsoberfläche umfasst. Eine Orientierung an Suchmaschinen, egal ob allgemeiner Natur oder eben Academic Search Engines, findet inzwischen kaum mehr statt – obgleich der Impuls zur Entwicklung dieser Systeme eben aus jener Orientierung stammte. Trotzdem haben sich nur wenige Prinzipien der Suchmaschinenteknologie in den Discovery-Systemen durchsetzen können. Vielmehr wurde das Augenmerk darauf gelegt, die Funktionalitäten des OPACs möglichst gut nachzubauen beziehungsweise zu integrieren. Dies hat zum Phänomen der „Reduktion auf einfachste Suchmöglichkeiten“⁴ geführt, das die ursprüngliche Idee der Discovery-Systeme ad absurdum führt. Möglicherweise kann ein Blick auf nicht-bibliothekarische Suchsysteme helfen, der in einer gewissen Stagnation befindlichen Entwicklung der Discovery-Systeme wieder mehr Dynamik zu verleihen.

Academic Search Engines – Der Stand der Dinge

Suchmaschinen für wissenschaftliche Literatur und anderer Fachinformationen gibt es wie erwähnt bereits seit fast 20 Jahren, und sie unterziehen sich auch schon so lange den Vergleichen mit bibliothekarischen Angeboten.⁵ Ebenso werden sie regelmäßig miteinander in Bezug auf

²Tay, Aaron. 2020. „The next generation discovery citation indexes - a review of the landscape in 2020“. <http://musingsaboutlibrarianship.blogspot.com/2020/11/the-next-generation-discovery-citation.html>

³Zum Beispiel durch das Kompetenzzentrum Bibliometrie, <https://bibliometrie.info/>

⁴Mattmann, Beat, und Noah Regenass. 2021. „Eine neue Form der Recherche in Bibliotheken“. *Bibliothek Forschung und Praxis* 45 (2): 304–16. <https://doi.org/10.1515/bfp-2021-0010>

⁵Kortekaas, Simone, und Bianca Kramer. 2014. „Thinking the unthinkable – doing away with the library catalogue“. *Insights: the UKSG journal* 27 (3): 244–48. <https://doi.org/10.1629/2048-7754.174>

ihren Abdeckungsgrad und der Eignung für einzelne Fragestellungen verglichen.⁶ Damit, dass Crossref als Serviceplattform für Wissenschaftsverlage im Jahr 2013 eine Schnittstelle zum kostenlosen Abruf von Metadaten bereitgestellt hat, konnten auch verlagsunabhängige Anbieter eine kritische Masse an Metadaten regelhaft abrufen und über eigene Dienste recherchierbar machen. Diesen Dienst nutzen auch einzelne bibliothekarische Discovery-Systeme.⁷

Academic Search Engines gibt es von kommerziellen wie auch gemeinnützigen Anbietern. Die Benutzung ist in der Regel kostenlos, zumindest in Bezug auf die Funktionalitäten der Literaturrecherche. Mit dem Microsoft Academic Graph stand zwischen 2015 und 2021 das Angebot eines kommerziellen Dienstleisters zur Verfügung, der mit über 120 Millionen Referenzen über die größte Datenbasis verfügte und diese auch zur Benutzung über eine Schnittstelle bereitstellte.⁸

Diese Datenbasis wurde von OpenAlex übernommen, einem Ende 2021 angekündigten Dienst, der bereits jetzt über eine Schnittstelle benutzt werden kann und in naher Zukunft auch mit einer eigenen Web-Schnittstelle ausgestattet werden soll. Allein wegen der Datenbasis ist OpenAlex ein besonders vielversprechendes Instrument. Hinzu kommt, dass OpenAlex in jeder Hinsicht den Prinzipien der Openness folgt, also nicht nur auf Nutzung freier Metadaten und freie Bereitstellung derselben setzt, sondern auch die Software, die für die Aggregation und Nachbearbeitung der Metadaten benutzt wird, quelloffen zur Nachnutzung zur Verfügung stellt.⁹ OpenAlex wird von dem Non-Profit-Unternehmen OurResearch betrieben, das unter anderem auch den Dienst Unpaywall betreibt. OurResearch hat für den Aufbau von OpenAlex nach eigenen Angaben Fördergelder in Höhe von 4,5 Millionen US-Dollar einer gemeinnützigen Stiftung bekommen.¹⁰

Inwieweit diese Gelder ausreichen, um den geplanten Dienst dauerhaft anbieten zu können, wird sich zeigen. Mitanbieter wie die ebenfalls gemeinnützige Plattform TheLens.org finanzieren sich über Stiftungsgelder, zum Teil aber auch unterstützt über Crowdfunding. Die kommerziellen Mitbewerber verfügen vermutlich über deutlich mehr Mittel. Die zur Verlagsgruppe Holtzbrinck gehörende Firma Digital Science erzielt mit ihren Diensten, zu denen die Academic Search Engine Dimensions gehört, beispielsweise einen jährlichen Umsatz von 76,1 Millionen US-Dollar.¹¹

Neben der Bereithaltung der technischen Infrastrukturen für die verarbeiteten Datenmengen investieren die Academic Search Engines auch stark in die konzeptionelle Entwicklung ihrer Dienste. Ein weiteres Auszeichnungsmerkmal von Academic Search Engines ist die Orientierung an Use Cases, also erwarteten Nutzungsszenarien. Für die prototypische geisteswissenschaftliche Suchmaschine SoNAR wurden diese Use Cases gemeinsam mit künftigen Nutzen-

⁶Visser, Martijn, Nees Jan van Eck, und Ludo Waltman. 2021. „Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic“. *Quantitative Science Studies* 2 (1): 20–41. https://doi.org/10.1162/qss_a_00112

⁷Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV). 2020. „KOBV-Portal mit Crossref-Index und neue ALBERT-Instanzen online“. <https://www.kobv.de/kobv-portal-mit-crossref-index-und-neue-albert-instanzen-online/>

⁸Microsoft Academic Graph: <https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/>

⁹GitHub-Repositorium „openalex-guts“: <https://github.com/ourresearch/openalex-guts>

¹⁰OpenAlex About-Webseite: <https://openalex.org/about>

¹¹Growjo. Digital Science Revenue and Competitors. https://growjo.com/company/Digital_Science

den des Systems entwickelt.¹² Das Team hinter OpenAlex publiziert regelmäßig „Tips of the Day“, in denen die Funktionen der Schnittstelle demonstriert werden.¹³

Diese beiden Dienste, SoNAR und OpenAlex, haben außerdem gemeinsam, dass sie bibliografische Daten nicht nur aggregieren und auf ein einheitliches Format bringen, sondern auch eine andere Form der Modellierung für die Daten wählen und anstelle von Suchmaschinentechnologie Graphdatenbanken verwenden, um die in der Regel vielfältigen Beziehungen zwischen einzelnen Elementen der bibliografischen Beschreibungen korrekt darzustellen und sinnvoll durchsuchbar und visualisierbar zu machen (vergleiche Abschnitt 1). Mit diesen Technologien dürften zumindest einige der Academic Search Engines den bibliothekarischen Erschließungsprinzipien näher kommen als die derzeitigen Discovery-Systeme, die bei der Vernetzung von Daten nach formalen oder inhaltlichen Kriterien hinter den Bibliothekskatalogen zurückbleiben.

Wegen ihrer vergleichsweise innovativen Ansätze bei der Datenmodellierung und -haltung, aber auch wegen des Zuschnitts ihrer Suchräume und der völlig anderen Herangehensweisen an nachträgliche Erschließung und Erschließungsqualität, verdienen die Academic Search Engines mehr Aufmerksamkeit durch Bibliothekar*innen.

Was können Discovery-Systeme heute von Academic Search Engines lernen?

Wissenszusammenhänge besser darstellen

Ein wesentliches Unterscheidungsmerkmal von Academic Search Engines und bibliothekarischen Recherchertools betrifft die Kontextualisierung von Wissen. Während sich bibliothekarische Angebote darauf beschränken, ähnliche Treffer auf Grundlage von ähnlicher Erschließung oder Klickzahlen zu ermitteln, ist die Heranziehung von Zitationen bei allen Academic Search Engines Standard. Auch wenn sich die Academic Search Engines meistens nicht explizit darauf beziehen, ist der Science Citation Index als erste interdisziplinäre „Suchmaschine“ diesbezüglich prägend für die gesamte neue Generation der hier betrachteten Instrumente.¹⁴

Da in bibliothekarischen Katalogen und Discovery-Systemen in der Regel sowohl der klassische Bibliotheksbestand als auch Artikeldaten enthalten sind, wäre eine Heranziehung der beispielsweise inzwischen über Crossref verfügbaren Zitationen nur für eine Teilmenge der Daten in diesen Systemen relevant. Es scheint seitens der Anbieter von bibliothekarischen Discovery-Systemen, zumindest derzeit, als zu große Herausforderung angesehen zu werden, eine solche Inkonsistenz verständlich zu vermitteln, vor allem an die bibliothekarische Zielgruppe. Allerdings bleibt damit ein großes Potenzial ungenutzt, um neben einer formalen und inhaltlichen Ordnung auch eine „Bedeutungseinordnung“ für Literatur zu erlauben. Grundlage dafür könnten neben den Zitationen einerseits auch die Heranziehung von Exemplarzahlen bei monografischer Literatur sein, wenn man annimmt, dass diese Zahl ein Indikator für hohe Nachfrage und breite Nutzung ist. Insgesamt sind die bibliothekarischen Suchsysteme, insbesondere die Discovery-Systeme, bei der Erprobung und Evaluierung solcher Ansätze sehr zurückhaltend.

¹²Balck, Sandra, Menzel, Sina, Petras, Vivien, Schnaitter, Hannes, and Zinck, Josefine. 2022. „Fluch und Segen der Visualisierung - Unterschiedliche Zielfunktionen im Forschungsprozess der historischen Netzwerkanalyse“. Zenodo. <https://doi.org/10.5281/zenodo.6328123>

¹³Siehe zum Beispiel: https://www.twitter.com/OpenAlex_org/status/1501583631873044488

¹⁴Small, Henry. 2018. „Citation Indexing Revisited: Garfield's Early Vision and Its Implications for the Future“. *Frontiers in Research Metrics and Analytics* 3. <https://doi.org/10.3389/frma.2018.00008>

Zu einer Bedeutungseinordnung von Treffermengen kann auch eine Visualisierung von Treffermengen und den darin enthaltenen Bezügen beitragen. Academic Search Engines benutzen Visualisierungen bereits jetzt in weitaus stärkerem Maße, als dies bibliothekarische Tools tun – allerdings häufig nur für eher eindimensionale Aspekte wie Erscheinungsjahre, aber eben auch für Zitationsnetzwerke. Es ist anzunehmen, dass neben Zitationen auch inhaltliche und ausgewählte formale Aspekte wie eine institutionelle Zugehörigkeit geeignet sind, um optisch aufbereitet dargestellt zu werden und auf diese Weise die Erforschung von Wissenszusammenhängen zu ermöglichen.

Um die visuelle Darstellung komplexer Wissenszusammenhänge zu ermöglichen, nutzen viele Academic Search Engines andere Datenmodelle und Datenbank-Technologien als klassische Suchmaschinen. So setzen beispielsweise OpenAlex, der Prototyp der fachspezifischen (und derzeit nur prototypisch verfügbaren) Suchmaschine SoNAR oder Open Knowledge Graph auf Graphdatenbanken und entsprechende Modellierungen der Metadaten, um sowohl textuelle als auch grafische Zusammenhänge aufzubereiten. Graphdatenbanken ermöglichen es, die einzelnen Bestandteile der bibliografischen Beschreibung als Entitäten zu definieren und die Zusammenhänge der Entitäten untereinander zu beschreiben. Mit der Fokussierung auf die Zusammenhänge zwischen Entitäten unterscheiden sie sich stark von den relationalen Datenbanken, die hinter Bibliothekskatalogen stehen, oder den Suchmaschinenindices wie Solr.

Zusammenhänge gibt es innerhalb bibliografischer Datensets in vielfältiger Form: Es handelt sich dabei um formale Zusammenhänge, zum Beispiel bei Artikeln, die Teile einer Zeitschrift sind, bei Zugehörigkeiten von Personen zu Institutionen oder bei einzelnen Teilen einer Schriftenreihe oder eines mehrbändigen Werkes. Vor allem sind Bibliotheken aber stolz auf ihre inhaltliche Erschließung, mit der über Schlagwörter oder Klassifikationen thematische Zusammenhänge erzeugt werden. Eine verbesserte Ausnutzung dieser Erschließungsdaten innerhalb von Discovery-Systemen wird immer wieder gefordert.¹⁵ Mit den gängigen Datenmodellierungen – in der Regel einer Konvertierung von Daten aus proprietären Bibliotheksmanagement-Systemen nach MARC – werden sich die Herausforderungen jedoch vermutlich nicht lösen lassen, da das Datenmodell keine hierarchischen Beziehungen zulässt und Verknüpfungen, zum Beispiel zu übergeordneten Begriffen, in der Regel über eigene Identifikatoren der Herkunftssysteme erstellt werden. Die konsequente Verwendung von Linked Open Data könnte die Verwendung der Erschließungsdaten in Discovery-Systemen verbessern.¹⁶

Zuschnitt der Suchräume überdenken

Einer der wesentlichen Unterschiede zwischen Academic Search Engines und klassischen Rechercheinstrumenten wie Bibliothekskataloge, Discovery Systeme und Fachdatenbanken besteht darin, dass die Academic Search Engines ihre Inhalte nicht selbst kuratieren, sondern automatisiert beziehen. Allerdings erfolgt die Beschaffung von Literatur in Bibliotheken auch nicht

¹⁵Heidrun Wiesenmüller. 2021. „Verbale Erschließung in Katalogen und Discovery-Systemen – Überlegungen zur Qualität“. In *Qualität in der Inhaltsererschließung*, herausgegeben von Michael Franke-Maier, Anna Kasprzik, Andreas Ledl, und Hans Schürmann, 279–302. De Gruyter Saur. <https://doi.org/10.1515/9783110691597-014>

¹⁶Pohl, Adrian, und Patrick Danowski. 2013. „Linked Open Data in der Bibliothekswelt: Grundlagen und Überblick“. In *(Open) Linked Data in Bibliotheken*, herausgegeben von Patrick Danowski und Adrian Pohl, 1–44. De Gruyter Saur. <https://doi.org/10.1515/9783110278736.1>

mehr ausschließlich auf Grundlage intellektuell getroffener Auswahlentscheidungen, sondern im E-Medien-Bereich gleichsam nach Marktlage (E-Book-Pakete, National- oder Allianzlizenzen).

Die meisten Academic Search Engines beziehen in irgendeiner Weise Daten aus Crossref ein, das sich als Non-Profit-Organisation der Auszeichnung von wissenschaftlicher Literatur mit eindeutigen Identifikatoren (DOIs) verschrieben hat. Die Mitgliedschaft bei Crossref ist Content-Anbietern im wissenschaftlichen Bereich vorbehalten, so dass der Datenbestand in Crossref auf diese Weise auch als kuratiert gelten kann, wenngleich natürlich auf einem anderen Niveau als es bei Erwerbungsentscheidungen in Bibliotheken der Fall ist, da diesen in der Regel weitaus differenziertere Qualitätskriterien zu Grunde liegen.

Die Datengrundlage von bibliothekarischen Discovery-Systemen bilden in der Regel die Katalogs- und Bestandsdaten der eigenen Bibliothek, die mit Suchmaschinentechologie aufbereitet werden. Da die wenigsten Bibliotheken diese Aufbereitung selbst erledigen können, nutzen sie entweder gemeinschaftlich betriebene Indices wie den K10Plus Zentral oder kommerzielle Angebote wie EDS.¹⁷ Die Möglichkeiten der Einflussnahme auf die Inhalte in diesen Indices sind für die einzelnen Bibliotheken ausgesprochen begrenzt. Die Aufnahme von individuell gewünschten zusätzlichen Metadatenkollektionen ist aus technischen und lizenzrechtlichen Gründen oftmals unmöglich oder mit großen Hürden verbunden. Das bedeutet in der Regel, dass lokal durchaus sehr bedeutsame Datenkollektionen nicht ohne Weiteres über die Discovery-Systeme zugänglich gemacht werden können. Für Bibliotheken mit nicht katalogisierten Altbeständen oder selbst erstellten Bibliografien stellt das einen erheblichen Nachteil dar.

Die Bielefeld Academic Search Engine hingegen erlaubt es, dass Content-Anbieter eigene Quellen liefern können, vorausgesetzt es handelt sich dabei um zumindest teilweise frei zugängliche Volltexte. Eine niedrigschwellige Öffnung der Discovery-Indices für das Einspielen von Metadaten und/oder Volltexten wäre auch mit Blick auf Forschungsinformationssysteme wünschenswert, die sich immer weiter in der universitären Informationslandschaft etablieren und die Discovery-Systeme um zwar nicht bibliografische, aber dennoch lokal bedeutsame Daten bereichern können.

Offenheit für automatisierte Erschließung erhöhen

Voraussetzung für die niedrigschwellige Öffnung der Discovery-Indices für weitere Inhalte ist eine weitgehend maschinelle Vorverarbeitung und Anreicherung der Daten. Allerdings ist bereits jetzt eine häufige Klage von Bibliotheksmitarbeitenden, dass die aktuell im Einsatz befindlichen Verarbeitungsmechanismen schlecht nachvollziehbar seien – obwohl zum Beispiel für den K10Plus einschlägige Dokumentationen offen bereitstehen. Auch erste Versuche der Deutschen Nationalbibliothek zur automatischen Erschließung wurden in der Bibliotheks-Community eher kritisch beurteilt.¹⁸

¹⁷Wikipedia. „EBSCO Discovery Service“. https://de.wikipedia.org/wiki/EBSCO_Discovery_Service

¹⁸Wiesenmüller, Heidrun. 2018. „Maschinelle Indexierung am Beispiel der DNB : Analyse und Entwicklungsmöglichkeiten“. In *o-bib. Das offene Bibliotheksjournal*. VDB, Bd. 5 Nr. 4 (2018), 141–153. <https://doi.org/10.5282/o-bib/2018H4S141-153>

Die Kriterien für die Bemessung der Datenqualität sind bei Academic Search Engines und Discovery-Systemen höchst unterschiedlich. Bibliothekarische Systeme gehorchen den einschlägigen Richtlinien für die formale Erfassung und setzen bei der inhaltlichen Erschließung nahezu ausschließlich auf intellektuelle Verfahren. Im Gegensatz dazu gilt beispielsweise bei Crossref, dass sich gute Datensätze vor allem durch die Verfügbarkeit von Zitationen, offenen IDs wie ORCID, Abstracts und Volltextlinks auszeichnen.¹⁹

Angesichts der Tatsache, dass es sich bei Academic Search Engines in der Regel um Volltext-Suchmaschinen handelt, spielt die inhaltliche Erschließung eine eher untergeordnete Rolle und wird hauptsächlich für die Zuordnung von Werken zu Wissenschaftsgebieten genutzt. Diese Zuordnung erfolgt über automatische Verfahren.²⁰ Bei OpenAlex erfolgt eine thematische Auszeichnung nach eigenen Konzepten und Wikidata.²¹ Die Bielefeld Academic Search Engine setzt bereits seit etwa 2011 auf die automatisierte Anreicherung der nachgewiesenen Daten durch Systemstellen der Dewey Decimal Classification.²²

Eine Offenheit dafür, automatisierte inhaltliche Erschließung zu betreiben, ist auch in einschlägigen Projekten großer wissenschaftlicher Bibliotheken erkennbar.²³ Für eine breite Adaption von maschinellen Erschließungsverfahren in Bibliotheken scheint jedoch noch ein erheblicher Vertrauens- und Kompetenzaufbau notwendig zu sein, der möglicherweise aber auch durch den Einsatz von Verfahrensweisen des maschinellen Lernens angetrieben werden könnte. Diese Verfahrensweisen erlauben eine kontinuierliche Weiterentwicklung der automatischen Erschließung und haben das Potenzial, eine hohe Erschließungskonsistenz zu erreichen – was wiederum ein essentieller Bestandteil der bibliothekarischen Qualitätskriterien darstellt. Gleichzeitig zeichnet sich jedoch auch ab, dass automatisierter Verfahren nicht ausreichen werden und es nach wie vor Bedarf an intellektueller Erschließung gibt.²⁴

Fazit

Die Zielsetzungen und strategischen Ausrichtungen von Academic Search Engines und bibliothekarischen Suchsystemen unterscheiden sich in vielen Punkten. Es gibt jedoch einige augenscheinliche Möglichkeiten, wie beide voneinander profitieren können.

¹⁹Meddings, Kirsty, und Anna Tolwinski. 2018. „How good is your metadata?“
<https://www.crossref.org/blog/how-good-is-your-metadata/>

²⁰Vergleiche zum Beispiel Hook, Daniel W., Simon J. Porter, und Christian Herzog. 2018. „Dimensions: Building Context for Search and Evaluation“. In *Frontiers in Research Metrics and Analytics* 3: 6.
<https://doi.org/10.3389/frma.2018.00023>

²¹OpenAlex documentation. „Concept“. <https://docs.openalex.org/about-the-data/concept>

²²Waltinger, Ulli, Alexander Mehler, Mathias Lösch, und Wolfram Horstmann. 2011. „Hierarchical Classification of OAI Metadata Using the DDC Taxonomy“. In *Advanced Language Technologies for Digital Libraries*, herausgegeben von Raffaella Bernardi, Sally Chambers, Björn Gottfried, Frédérique Segond, und Ilya Zaihrayeu, 6699:29–40. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-642-23160-5_3

²³Arndt, Susanne, Berit Genat, und Mila Runnwerth. 2020. „Evaluierung von annif an der TIB - ein Werkstattbericht“. Zenodo. <https://doi.org/10.5281/zenodo.4304303>

²⁴Auer, Sören, Allard Oelen, Muhammad Haris, Markus Stocker, Jennifer D'Souza, Kheir Eddine Farfar, Lars Vogt, Manuel Prinz, Vitalis Wiens, und Mohamad Yaser Jaradeh. 2020. „Improving Access to Scientific Literature with Knowledge Graphs“. In *Bibliothek Forschung und Praxis* 44 (3): 516–29. <https://doi.org/10.1515/bfp-2020-2042>

Grundsätzlich ist die Innovationskraft auf Seiten der Academic Search Engines höher als im Bereich der bibliothekarischen Discovery-Systeme. Der Einsatz von automatisierten Verfahren für die Erschließung und Anreicherung von Daten ist angesichts der großen Datenmengen bei den meisten Angeboten selbstverständlich. Auch die Modellierung von Daten als Wissensgraphen ist für den Bibliotheksbereich vielversprechend, insbesondere dann, wenn auch Visualisierungen von Beziehungen gewünscht sind. Entsprechende explorative Projekte gibt es mit SONAR oder dem Datendienst der SLUB Dresden bereits.²⁵ Diese haben das Potenzial, die derzeit eher evolutionär verlaufende Entwicklung der Discovery-Systemen zu beschleunigen. Voraussetzung hierfür ist allerdings die Bereitschaft von Bibliotheken, Technologien mit einem relativ geringen Reifegrad als Chance zu verstehen und sich auf iterative Prozesse bei der Weiterentwicklung einzulassen, die zudem zuvor erarbeitete Nutzungsszenarien nicht außer Acht lassen sollte.

Unabhängig davon haben einige der frei verfügbaren Academic Search Engines durchaus das Potenzial, an die Stelle von kommerziellen Lösungen wie Web of Science oder Scopus treten zu können und damit ein bibliotheksseitiges Bekenntnis zu Openness zu stärken. Die interdisziplinäre Ausrichtung der meisten Academic Search Engines macht sie für viele Fragestellungen in den zunehmend ebenfalls interdisziplinären wissenschaftlichen Diskursen relevant. Ihre offenen Schnittstellen ermöglichen es außerdem, Datensets herunterzuladen und eigene Analysen durchzuführen, beispielsweise um Übersichtsartikel zu identifizieren oder Zitationsnetzwerke zu erforschen.²⁶

Allerdings können Academic Search Engines ihrerseits auch von Bibliotheken und ihren Erschließungswerkzeugen profitieren, ganz konkret von den jahrzehntelang gewachsenen und als Linked Open Data vorliegenden Normdateien. Eine kleine, aber hoffentlich wachsende Anzahl von Klassifikationen wird als Linked Open Data veröffentlicht und könnte als zusätzliches System der Wissensorganisation herangezogen werden.²⁷ Es ist verwunderlich, dass die Academic Search Engines bislang kaum auf einschlägige Normdaten aus dem bibliothekarischen Bereich zurückgreifen, insbesondere da beispielsweise mit dem Virtual International Authority File²⁸ ein erstklassig nutzbares Instrument für die Disambiguierung von Personennamen verfügbar ist.

Die verschiedenen Ökosysteme im Internet – seien es allgemeine soziale Netzwerke, journalistische Angebote oder Plattformen für die Recherche und Publikation von wissenschaftlicher Literatur und wissenschaftlichen Daten – werden in Zukunft noch stärker als bisher auf Interoperabilität setzen müssen.²⁹ Auch kommerzielle Systeme wollen mehr untereinander agieren, wie die Vision des Metaversums zeigt. Bibliotheken und andere Informationsanbieter, insbesondere die gemeinnützigen Academic Search Engines haben die Möglichkeit, gemeinsam an einer offenen, den FAIR-Prinzipien genügenden Vision von wissenschaftlichen Rechercheplattformen zu arbeiten und damit einen Beitrag zu einer offenen digitalen Infrastruktur für die Wissenschaft zu leisten.³⁰ Beide sollten in einen Austausch kommen, um diese Möglichkeit zu

²⁵SLUB Dresden. „SLUB LOD API documentation“. <https://data.slub-dresden.de/>

²⁶Williams, Brett. 2020. „Dimensions & VOSviewer bibliometrics in the reference interview“. In *Code4Lib Journal*, Nr. 47. <https://journal.code4lib.org/articles/14964>

²⁷SkoHub: <https://skohub.io/>

²⁸Virtual International Authority File: <https://viaf.org>

²⁹FAIR-Prinzipien: <https://www.go-fair.org/fair-principles/>

³⁰Zuckerman, Ethan. 2020. „The Case for Digital Public Infrastructure“. Columbia University. <https://doi.org/10.7916/d8-chxd-jw34>

nutzen. In die bibliothekarischen Discovery-Systeme ist in den vergangenen Jahren viel Zeit und Geld geflossen, vor allem, um den Status quo des klassischen OPACs zu erhalten – was aber nicht der ursprünglichen Zielsetzung entspricht und innerhalb der Discovery-Community kritisch reflektiert wird.³¹ Die Academic Search Engines zeigen auf, wie Bibliotheken mit mehr Innovationsfreude, Pragmatismus und Orientierung an Use Cases bessere Discovery-Systeme schaffen können.

Anne Christensen ist Bibliothekarin und hat über 20 Jahre in wissenschaftlichen Bibliotheken gearbeitet, zuletzt als stellvertretende Direktorin der Bibliothek an der Leuphana Universität Lüneburg. Seit 2020 ist sie Beraterin und Projektmanagerin bei der Hamburger Firma effective WEBWORK, die Open-Source-Systeme für Bildungseinrichtungen und Bibliotheken implementiert. Anne Christensen ist außerdem als Lehrbeauftragte in verschiedenen bibliothekarischen Ausbildungsgängen tätig.

³¹Keßler, Kristof, Anne Christensen, Schrader Jarmo, und Jan F. Maas. 2019. „Discovery-Systeme, die ihrem Namen Ehre machen: Verbesserungspotenziale für bibliothekarische Suchmaschinen“. <https://nbn-resolving.org/urn:nbn:de:0290-opus4-162505>