

Aufbau einer offenen Bibliometriedatenbank innerhalb des Kompetenznetzwerks Bibliometrie (KBOPENBIB): Teilprojekt Kuration Zitationsdaten und Open-Science-Evidenzen

BMFTR-Abschlussbericht

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt unter dem Förderkennzeichen 16WIK2301E gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor:innen.

Verfasst von

Nick Haupka, Dr. Najko Jahn (SUB Göttingen)

Zuwendungsempfänger: Georg-August-Universität Göttingen
Förderkennzeichen: 16WIK2301E
Vorhabenbeschreibung: Aufbau einer offenen Bibliometriedatenbank innerhalb des Kompetenznetzwerks Bibliometrie (KBOPENBIB): Teilprojekt Kuration Zitationsdaten und Open-Science-Evidenzen
Laufzeit des Vorhabens: 01.05.2023–31.12.2025

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

Teil I

Kurzbericht

1 Ursprüngliche Aufgabenstellung und Stand der Forschung

Ziel des Vorhabens war der Aufbau einer offenen Bibliometriedatenbank innerhalb des Kompetenznetzwerks Bibliometrie (KB). Das Teilprojekt der SUB Göttingen hatte dabei einen besonderen Fokus auf die Analyse, Kuration und Bereitstellung von Dokumenttypen, Open-Science-Evidenzen und Zitationsdaten. Grundlage der Kuration war die Datenbank OpenAlex, die einen umfangreichen Datenbestand und einen kostenfreien Zugang bietet. Die mit dem Vorhaben verbundene Öffnung des KB hin zu offenen Bibliometriedaten sollte zugleich eine Alternative zu den proprietären Datenbanken Web of Science und Scopus schaffen, deren Nutzung mit erheblichen Lizenzkosten einhergeht.

Zu Projektbeginn stützte sich das KB ausschließlich auf Web of Science und Scopus. Während OpenAlex sich als grundsätzlich vielversprechende, inklusive Alternative abzeichnete, war seine Eignung für bibliometrische Analysen noch nicht systematisch geprüft; zudem ließ die hohe Entwicklungsdynamik der Datenbank einen erheblichen Bedarf an gezielter Kuration und kontinuierlicher Qualitätssicherung erkennen. Das Vorhaben knüpfte an eine wachsende Forschung und eigene Vorarbeiten an, die Abdeckung, Vollständigkeit und Datenqualität offener Quellen im Vergleich zu den etablierten Datenbanken untersuchten.

2 Ablauf des Vorhabens

Das Vorhaben gliederte sich in vier Arbeitsbereiche: Im ersten Arbeitsbereich wurde das FIZ Karlsruhe unterstützt, OpenAlex in die Dateninfrastruktur des KB zu integrieren. Die Unterstützung aus dem Teilprojekt bezog sich auf die Entwicklung von Importroutinen, die Erstellung von Datenschemata sowie ein kontinuierliches Monitoring von Schemaänderungen. Zugleich koordinierte das Teilprojekt die Open-Data-Veröffentlichungen des Projekts und setzte sie über das Repositorium Zenodo sowie über das Open Scholarly Data Warehouse der SUB Göttingen auf Grundlage von Google BigQuery um.

Der zweite Arbeitsbereich verglich die Datenbasis von OpenAlex mit den im KB etablierten Datenbanken Scopus und Web of Science. Untersucht wurden insbesondere Publikationsaufkommen, Zitationsdaten, Dokumenttypen sowie weitere spezifische Metadaten wie Open-Access-Evidenzen.

Im dritten Arbeitsbereich wurden die Erkenntnisse aus dem Datenbankvergleich aufgegriffen und in Kurationsverfahren überführt, mit denen die Datenbasis von OpenAlex im KB optimiert wurde. Im Vordergrund stand die Entwicklung eines Verfahrens zur verbesserten Dokumenttypenklassifikation. Die kuratierten Ergebnisse wurden in Open-Data-Releases gebündelt und über Zenodo bereitgestellt. Der vierte Arbeitsbereich diente der Vernetzung nach innen und außen, von der verbundinternen Abstimmung bis zum Transfer der Ergebnisse in die Fach- und Anwendungscommunity. Begleitend wurden die Resultate durch Blogposts und Zeitschriftenartikel dokumentiert.

3 Wesentliche Ergebnisse

Zu den wesentlichen Ergebnissen zählen:

- **Offene Bibliometriedatenbank** OpenAlex wurde als zusätzliche Datenquelle im KB etabliert. Ergänzend wird im Scholarly Data Warehouse der SUB Göttingen regelmäßig eine OpenAlex-Version des KBs gemeinsam mit weiteren offenen Datenquellen frei angeboten.
- **Anreicherung von Zitationsdaten.** Auf Basis offener Quellen wie Semantic Scholar wurde geprüft und erprobt, wie sich die Referenzdaten in OpenAlex ergänzen lassen.
- **Erschließung von Open-Science-Evidenzen.** Durch den Abgleich von cOAlition-S-, Crossref- und OpenAlex-Daten wurden Artikel Open-Access-Transformationsverträgen automatisch zugeordnet und ein entsprechender Datensatz erstellt und validiert.
- **Kuration von Dokumenttypen.** Mithilfe eines eigens entwickelten, quelloffenen Machine-Learning-Klassifikators lassen sich rund 10 % aller Zeitschriftenartikel in OpenAlex als nicht-originäre Beiträge (etwa Book Reviews, Abstracts oder Editorials) zusätzlich kennzeichnen.
- **Datenpublikation.** Die im Vorhaben erstellten Kurationen wurden als Teil des verbundübergreifenden OPENBIB-Data-Releases, dessen Veröffentlichung die SUB Göttingen koordinierte, über Zenodo offen bereitgestellt, darunter ein Datensatz zu Dokumenttypen und ein Datensatz zu Open-Access-Transformationsverträgen.
- **Publikationen und Vernetzung.** Zwischen- und Endergebnisse wurden in mehreren Blogposts und Zeitschriftenartikeln dokumentiert, allen voran die Abdeckungsanalysen, die OpenAlex mit Scopus und Web of Science vergleichen. Über die Beteiligung an der Initiative ORION-DBs und die Zusammenarbeit mit dem französischen CNRS wurden Verfahren und Daten des Vorhabens auch international anschlussfähig gemacht.

Teil II

Ausführlicher Bericht

1 Ausgangslage

Offene Bibliometriedaten sind zunehmend Gegenstand der Wissenschaftsforschung und -politik. Prominentes Beispiel für diese Entwicklung ist OpenAlex (Priem, Piwowar & Orr, 2022), eine quelloffene Datenbank, die aus dem Microsoft Academic Graph hervorgegangen ist und verschiedene Datenquellen unter Verwendung persistenter Identifier (DOI, ROR, ORCID) zusammenführt. Angesichts der wachsenden wissenschaftspolitischen Bedeutung besteht aus bibliometrischer Sicht allerdings ein großer Bedarf an einer prüfenden Betrachtung von Abdeckung, Vollständigkeit und Datenqualität offener Quellen gegenüber den etablierten proprietären Datenbanken Web of Science und Scopus. Zudem stellt die dynamische Weiterentwicklung von OpenAlex eine Herausforderung dar, da neue Versionen sowohl zusätzliche Features als auch grundlegende Änderungen (*breaking changes*) enthalten können.

Das Kompetenznetzwerk Bibliometrie (KB) stützte sich zu Projektbeginn ausschließlich auf die proprietären Datenbanken Web of Science und Scopus. Eine Erweiterung um offene Bibliometriedaten wie OpenAlex ermöglicht, was geschlossene bibliometrische Dateninfrastrukturen bislang verwehrten: die freie Nutzung kuratierter Daten und Methoden, deren Nachnutzung durch Dritte und damit eine Erweiterung des Nutzerkreises über die etablierte bibliometrische Forschung hinaus. Zudem erfassen nicht-selektive Datenquellen auch Publikationsarten und -weisen, die in Web of Science und Scopus unberücksichtigt bleiben, und erweitern so die Analysemöglichkeiten.

Vor diesem Hintergrund zielte das vorliegende Projekt auf den wissenschaftsgeleiteten Aufbau einer offenen Bibliometriedatenbank innerhalb des KB. Während in den ersten beiden Jahren die initiale Bereitstellung, Analyse und Kuration offener Daten mit Fokus auf die deutsche Wissenschaftslandschaft im Mittelpunkt stand, konzentrierten sich die Arbeiten im Jahr 2025 darauf, automatisierte Verfahren zur kontinuierlichen Bereitstellung offener Bibliometriedaten zu implementieren und zu evaluieren. Damit wurde der Betrieb einer offenen Bibliometriedatenbank innerhalb des KB ab 2026 vorbereitet, sodass künftig nur noch eine der beiden proprietären Datenbanken genutzt werden muss.

Der Aufbau der offenen Bibliometriedatenbank ist als Verbundvorhaben angelegt, in dem jede Partnereinrichtung ihre fachlichen Stärken und Erfahrungen in die Kuration der jeweiligen Entitäten einbringt (siehe Tabelle 1). Den Schwerpunkt des Teilprojekts der SUB Göttingen bildeten dabei die Prüfung, Kuration und Bereitstellung von Daten zu Dokumenttypen sowie zu Open-Science-Evidenzen. Darüber hinaus wirkte die SUB Göttingen an der vergleichenden Analyse von OpenAlex hinsichtlich Zitationsdaten und -abdeckung mit. Als koordinierende Einrichtung übernahm sie zudem die Organisation regelmäßiger Verbundtreffen, um die Kommunikation zwischen den Partnern aufrechtzuerhalten. Schließlich unterstützte die SUB Göttingen das FIZ Karlsruhe bei der Bereitstellung der Datenbank, alternativ über Google BigQuery.

Entität	Verbundpartner
Items (nur Datenbankvergleich)	SUB Göttingen DZHW
Zitationen	SUB Göttingen DZHW FZ Jülich GESIS
Adress-Informationen	Universität Bielefeld
Journale	FZ Jülich
Verlagsbezeichnungen	FZ Jülich
Autoren	GESIS
Förderinformationen	DZHW
Dokumenttypen	SUB Göttingen
Open-Science-Evidenzen	SUB Göttingen GESIS

Tabelle 1: Fachliche Arbeitsteilung im Projektverbund

2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Mit der Hinwendung zu Open Science sind in den vergangenen Jahren mehrere offene Quellen für bibliometrische Untersuchungen entstanden, darunter Crossref, Semantic Scholar und OpenAlex. OpenAlex wurde Anfang 2022 von OurResearch als Nachfolger des eingestellten Microsoft Academic Graph veröffentlicht, auf dem es im Wesentlichen aufsetzt (Scheidsteger & Haunschild, 2022), und gewann durch seine Offenheit, Inklusivität und breite Abdeckung rasch an Aufmerksamkeit. Da die Daten frei verfügbar sind, eröffnet OpenAlex die Möglichkeit reproduzierbarer bibliometrischer Studien, die mit den proprietären Quellen Web of Science und Scopus aufgrund ihrer Lizenzbedingungen kaum möglich sind.

Diese Perspektive schlug sich zunehmend in der Wissenschaftspolitik nieder. Spätestens mit der Barcelona Declaration on Open Research Information (Kramer, Neylon & Waltman, 2024) im Jahr 2024 gewann die Nachnutzung offener Forschungsinformationen, zu denen auch Daten über Publikationsaktivitäten zählen, an Gewicht, und Einrichtungen begannen, Lizenzen für proprietäre Datenquellen zu hinterfragen und verstärkt auf offene Alternativen zu setzen. Der Hintergrund ist ein doppelter: Zum einen lässt sich mit offenen Daten ein hohes Maß an Datenhoheit erreichen, wodurch die wissenschaftspolitisch kritisierte Abhängigkeit von kommerziellen Anbietern verringert wird (Aspesi & Brand, 2020); zum anderen verbessern offene Daten die Transparenz und Nachvollziehbarkeit bibliometrischer Ergebnisse und Evaluierungen (Wilkinson et al., 2016). Weil Web of Science und Scopus zudem selektiv indexieren, versprechen inklusive Quellen wie OpenAlex eine Diversifizierung von Wissenschaftsforschung und Forschungsbewertung, indem bislang unberücksichtigte Publikationsarten einbezogen werden können (Alperin, Portenoy, Demes, Larivière & Haustein, 2024).

Diese Entwicklung verändert auch die Anforderungen an bibliometrische Dateninfrastrukturen. Ihre Zukunftsfähigkeit bemisst sich zunehmend daran, wie weit sie zusätzliche Datenquellen über vielfältige Aspekte des wissenschaftlichen Publizierens einbeziehen (Sugimoto & Larivière, 2018; Waltman & Larivière, 2020). Das Kompetenznetzwerk Bibliometrie hatte diesbezüglich erste Erfahrungen bei der Integration von Crossref und Unpaywall in ihre Dateninfrastruktur gesammelt. Darüber hinaus betreibt die SUB Göttingen mit dem Open Scholarly Data Warehouse eine leistungsfähige Analyseplattform auf Grundlage von Google BigQuery, die ausschließlich offene Quellen wie Crossref, OpenAlex

und Unpaywall einbindet (Hauptka, Jahn & Hobert, 2022). Aufgrund seiner Performance und transparenten Preisstruktur rückt Google BigQuery zunehmend in den Fokus weiterer Akteure für die Bereitstellung von offenen Bibliometriedaten. So hat das brasilianisch-niederländische Projekt InSysPo im Sommer 2024 seinen gesamten Datenbestand öffentlich zugänglich gemacht (Mazoni & Costas, 2024). Diese Vorarbeiten bildeten einen geeigneten Ausgangspunkt, um OpenAlex unter kontrollierten Bedingungen in eine bestehende Dateninfrastruktur zu überführen.

Parallel dazu entwickelte sich eine wachsende Forschung, die die Eignung von OpenAlex für bibliometrische Analysen prüfte (Céspedes et al., 2025). Einzelne Arbeiten untersuchten die Abdeckung von Zeitschriften und Open-Access-Publikationen (Simard, Basson, Hare, Larivière & Mongeon, 2024) Genauigkeit von Affiliationsmetadaten (Akbaritabar, Theile & Zagheni, 2024) sowie den Vergleich von Zitations- und Referenzdaten verschiedener Datenquellen (Visser, van Eck & Waltman, 2021). Wiederkehrende Befunde waren, dass OpenAlex umfangreicher, aber heterogener als die etablierten Datenbanken ist, dass seine Datenqualität stark schwankt und sich mit jeder Version ändert und dass ein allgemeiner Goldstandard für die Bewertung der Abdeckung fehlt. Für die Klassifikation von Dokumenttypen war bereits aus der Forschung zu Web of Science und Scopus bekannt, dass uneinheitliche Typologien und Kurationsstrategien zu erheblichen Diskrepanzen führen und unmittelbar in bibliometrische Indikatoren und Rankings einfließen (Donner, 2017). Bereits kurz nach der Veröffentlichung widmeten sich daher erste Arbeiten gezielt der Abdeckung und Datenqualität von OpenAlex im Vergleich zu Scopus und Web of Science.

An diesen Forschungsstand knüpfte das Vorhaben an. OpenAlex erschien als grundsätzlich vielversprechende Alternative zu Web of Science und Scopus, deren Eignung für bibliometrische Analysen jedoch noch nicht systematisch geprüft war und die aufgrund ihrer hohen Dynamik einen erheblichen Bedarf an gezielter Kuration und kontinuierlicher Qualitätssicherung erkennen ließ.

3 Planung und Ablauf des Vorhabens

Arbeitsbereich 1: Datenbankbereitstellung

Das Ziel dieses Arbeitsbereichs bestand darin, die Datenbank OpenAlex in die Dateninfrastruktur des KB zu integrieren. Die Bereitstellung der Datenbank erfolgte durch das FIZ Karlsruhe. Das Teilprojekt unterstützte den Prozess innerhalb der AG Technik. Im Zuge der Diskussion über die Öffnung offener Bibliometrie-Datenbanken über Google BigQuery hatte die SUB Göttingen ihr Scholarly Data Warehouse öffentlich verfügbar gemacht und damit die ersten Datenveröffentlichungen des KB über diese Cloud-Computing-Umgebung ermöglicht. Zugleich war es Aufgabe der SUB Göttingen, das FIZ Karlsruhe technisch sowie organisatorisch dabei zu unterstützen, eine Google-Cloud-Computing-Umgebung für die offenen KB-Datenbestände mittels des GEANT OCRE 2024 Framework anzubieten.

Arbeitsbereich 2: Datenbankvergleich

Ziel dieses Arbeitsbereiches war es, OpenAlex hinsichtlich der Qualität und Quantität der enthaltenen Daten zu prüfen und mit den proprietären Datenbanken im KB Scopus und Web of Science abzugleichen. Der Vergleich hielt außerdem den Ausgangspunkt der anschließenden Datenkuration fest. Die Datenbankvergleiche erfolgten auf Grundlage der Dateninfrastruktur des FIZ Karlsruhe. Im Sinne der Arbeitsteilung im Verbund lag der Fokus des Teilprojekts der SUB Göttingen auf Items, Zitationsdaten, Dokumenttypen und Open-Science-Evidenzen.

Arbeitsbereich 3: Datenkuratierung

In diesem Arbeitsbereich erfolgten die geplanten Datenkurationen, wobei das Hauptaugenmerk auf den Entitäten Zitationen, Open-Science-Evidenzen und Dokumenttypen lag. Die Datenkuration erfolgte durch automatische Verfahren, welche zu diesem Zweck implementiert wurden:

- **Zitationen:** Ziel war es, Zitationsdaten aus weiteren Quellen anzureichern. Hierfür sollten Daten aus offenen Datenquellen wie Crossref und Semantic Scholar nachgenutzt werden.
- **Dokumenttypen:** Für die Erstellung eines semi-automatischen Verfahrens zur Identifizierung von nicht-originären Forschungsartikeln wurde ein Datenset mit Dokumenttypen aus der medizinischen Datenbank PubMed zusammengestellt und mit weiteren Metadaten aus Crossref und OpenAlex angereichert. Der Trainingsdatensatz wurde genutzt, um einen Machine-Learning-Klassifikator zu trainieren.
- **Open-Science-Evidenzen:** Der Schwerpunkt der Arbeiten lag auf der Zuordnung von Artikeln zu Open-Access-Transformationsverträgen. Zu diesem Zweck wurden Daten des Tools „cOAlition S Journal Checker“ mit den Datenbanken Crossref und OpenAlex abgeglichen. Dabei wurde auch auf Affiliationen geachtet, um Vertragsbeziehungen zu berücksichtigen. Um der wachsenden Bedeutung des wissenschaftsgeleiteten Open Access gerecht zu werden, wurde der Diamond Open Access Discovery Hub integriert, um den Nachweis zu optimieren.

Arbeitsbereich 4: Vernetzung

Aufgabe des Arbeitsbereichs Vernetzung war es, den Informationsfluss sowohl innerhalb des Verbunds als auch nach außen sicherzustellen. Hierfür wurde von Beginn an eine kontinuierliche, niedrigschwellige Kommunikationsstruktur etabliert.

Die verbundinterne Abstimmung erfolgte über zweiwöchentliche Calls der Projektpartner, organisiert durch die SUB Göttingen, in denen der Stand der arbeitsteilig durchgeführten Kurations- und Analysearbeiten besprochen und offene Fragen geklärt wurden. Ergänzend wurde eine User Group mit bis zu dreißig Mitgliedern betreut, die als Schnittstelle zwischen dem Projekt und der weiteren Anwendungscommunity diente. Über regelmäßige Rückmeldungen im Rahmen einschlägiger Fachtagungen wurde der Austausch über Zwischenergebnisse und Entwicklungen zusätzlich verstetigt.

Die Außenvernetzung war auf den Transfer der Projektergebnisse in die Fach- und Anwendungscommunity ausgerichtet, mit einem Schwerpunkt auf der Bibliothekscommunity. Geplant und umgesetzt wurden hierfür insbesondere Fachvorträge auf einschlägigen Konferenzen sowie der gezielte Aufbau von Kooperationen mit nationalen und internationalen Akteuren im Bereich offener Forschungsinformationen.

4 Inhaltliche Ergebnisse

Arbeitsbereich 1: Datenbankbereitstellung

Erweiterung KB-Dateninfrastruktur. Im Zuge des Vorhabens wurden Daten aus OpenAlex innerhalb der Dateninfrastruktur des Kompetenznetzwerks Bibliometrie bereitgestellt. Parallel integrierte die SUB Göttingen weitere offene Quellen in die KB-Infrastruktur, insbesondere Crossref und Semantic Scholar, sodass OpenAlex nicht isoliert, sondern als Teil eines breiteren Bestands offener Forschungsinformationen nutzbar wurde.

Datenpublikation. Die SUB Göttingen hat die Datenpublikation der beiden Versionen des KB in Zenodo koordiniert und umgesetzt. Der Schwerpunkt lag dabei auf der Harmonisierung der Schemata der Datentabellen und der Erstellung der Dokumentation.

Haupka, N., Culbert, J., Donner, P., Jahn, N., Lenke, C., Mayr, P., Meier, A., Mittermaier, B., Scheidt, B., Stahlschmidt, S., & Taubert, N. (2025). *OPENBIB: Selected curated open metadata based on OpenAlex* [Data set]. Kompetenznetzwerk Bibliometrie. <https://doi.org/10.5281/zenodo.15308680>

Haupka, N., Culbert, J., Donner, P., Jahn, N., Lenke, C., Mayr, P., Meier, A., Mittermaier, B., Scheidt, B., Stahlschmidt, S., & Taubert, N. (2026). *OPENBIB: Selected curated open metadata based on OpenAlex* [Data set]. Kompetenznetzwerk Bibliometrie. <https://doi.org/10.5281/zenodo.18429476>

Die Daten wurden zudem über das Scholarly Data Warehouse der SUB Göttingen verfügbar gemacht, um deren Nachnutzung mit Google BigQuery zu befördern:

<https://orion-dbs.community/collections/subugoe/#openbib>

Vernetzung der Datenbestände durch ORION-DBs. Über die Bereitstellung hinaus vernetzte sich die SUB Göttingen mit internationalen Akteuren, um die offene Bereitstellung von Forschungsinformationen über Google BigQuery zu standardisieren. Google BigQuery hat sich in der Fachcommunity als leistungsfähige Umgebung etabliert, um große offene Datenbestände kombinierbar und auswertbar zu machen; mehrere Einrichtungen, unter ihnen das CWTS und Digital Science, stellen dort eigenständig öffentliche Versionen einzelner Quellen bereit. Ein erstes Ergebnis dieser Zusammenarbeit war die Öffnung des Open Scholarly Data Warehouse der SUB Göttingen für die Allgemeinheit.

Die Nutzung von Google BigQuery des Open Scholarly Data Warehouses der SUB Göttingen basiert auf dem GÉANT IaaS OCRE-Framework (Open Clouds for Research Environments). Dieses ermöglicht eine rechtskonforme Nutzung kommerzieller Cloud-Dienste für Forschungszwecke. Die Kosten für die Aufbereitung und Speicherung der Datenbestände werden durch die SUB Göttingen finanziell getragen und durch das Rechenzentrum der Universität Göttingen, der GWGD, ihr in Rechnung gestellt. Die Kosten für Abfragen tragen Dritte selbst, wobei es hierfür großzügige Freikontingente gibt.

Darüber hinaus war die SUB Göttingen Mitbegründerin der Initiative Open Research Information Online Databases (ORION-DBs), die diese bislang verteilt und unabhängig voneinander betriebenen Datenbereitstellungen bündelt und auf gemeinsame Standards für Aufbereitung, Dokumentation und Zugang hinarbeitet. Zu diesem Zweck programmierte die SUB Göttingen eine automatisierte Dokumentationsplattform, die die zuvor verteilten Datenbestände auf Grundlage von Quarto und GitHub Actions in Form einer Webseite nachweist und auffindbar macht.

Neylon, C., Kramer, B., Fernandes Mazoni, A., Costas, R., Jahn, N., & van Eck, N. J. (2026). *Sharing the load: Building a collective to support open research information online*. <https://doi.org/10.54900/2pnyq-nhx95>

Mit Stand Juni 2026 umfasst ORION-DBs 7 Partner mit 58 offen zugänglichen Datensätzen im Gesamtumfang von 18 TB (<https://orion-dbs.community/>).

Arbeitsbereich 2: Datenbankvergleich

Abdeckungsanalyse von Zitationen und weiteren Metadatenelementen. In diesem Arbeitspaket wurde die Datenbank OpenAlex mit den proprietären Datenbanken Scopus und Web of Science durch eine Vollerhebung hinsichtlich des Publikationsvolumens, der Anzahl der Referenzen sowie weiterer Kriterien verglichen. Hierbei zeigte sich, dass OpenAlex eine vergleichbare Anzahl an *source*-Referenzen wie die etablierten Datenbanken Scopus und Web of Science aufweist. *Source*-Referenzen sind Referenzen, bei denen die Referenz selbst auch in der jeweiligen Datenbank als Publikation enthalten ist. Der Vergleich zeigte auch, dass die Abdeckung von Förderinformationen in den klassischen Datenbanken ausgeprägter ist als bei OpenAlex. Die Abdeckung von ORCID-Identifiern ist wiederum ausgeprägter in OpenAlex.

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*, 130(4), 2475–2492. <https://doi.org/10.1007/s11192-025-05293-3>

Dokumenttypen. In einer weiteren Abdeckungsanalyse zu den Publikations- und Dokumenttypen in OpenAlex wurden auf Basis eines gemeinsamen Korpus von rund 10 Millionen Artikel große Diskrepanzen bei der Klassifikation jener Entitäten innerhalb verschiedener Datenbanken festgestellt. Konkret wurde hier OpenAlex mit den Datenbanken Web of Science, Scopus, PubMed und Semantic Scholar verglichen. Dabei zeigte sich, dass bei den Publikationstypen in OpenAlex, Web of Science und Scopus zwar weitgehende Überschneidungen bestehen, bei der Dokumenttypenklassifikation zwischen

den Datenbanken jedoch deutliche Unterschiede auftreten, da OpenAlex fast durchgängig Dokumente als Artikel klassifizierte, während andere Datenbanken wie Web of Science oder PubMed differenzierter vorgehen und auch Reviews und klinische Studien zuordnen.

Haupka, N., Culbert, J. H., Schniedermann, A., Jahn, N., & Mayr, P. (2026). Analysis of the publication and document types in OpenAlex, Web of Science, Scopus, PubMed and Semantic Scholar. *Quantitative Science Studies*, 7, 179–194. <https://doi.org/10.1162/qss.a.406>

Open-Access-Transformationsverträge. Ein weiterer Analyseschwerpunkt war die Validierung eines offenen Ansatzes zur Schätzung von Artikeln, die durch Open-Access-Transformationsverträge ermöglicht wurden. Trotz der massiven finanziellen Kosten, die mit diesem Geschäftsmodell verbunden sind, fehlt es häufig an Transparenz über diese Ausgaben. Die Vollerhebung von über 13.000 hybriden Zeitschriften mit fast zehn Millionen Artikeln zwischen 2019 und 2023 und der anschließende Vergleich des offenen Verfahrens mit Scopus und Web of Science zeigen ein erhebliches, durch diese Verträge bedingtes Wachstum des Open Access, wobei die Ergebnisse über alle drei Datenquellen hinweg konsistent waren und trotz Unterschieden in Zeitschriftenabdeckung und Metadatenverfügbarkeit hohe Korrelationen auf Länderebene aufwiesen. Der vergleichende Ansatz stützt damit die Nutzung offener Metadaten für großskalige Analysen des hybriden Open Access und macht zugleich deutlich, dass die Kombination mehrerer Datenquellen ein robusteres Bild liefert als jede einzelne Datenbank für sich.

Jahn, N. (2025). Estimating transformative agreement impact on hybrid open access: a comparative large-scale study using Scopus, Web of Science and open metadata. *Scientometrics*, 131(2), 901–924. <https://doi.org/10.1007/s11192-025-05390-3>

Im Zuge der Erstellung von Open-Access-Analysen für die internationale Initiative OA2020 fiel zudem ein deutlicher Rückgang der Affiliationsmetadaten in OpenAlex auf. Bei der Aufbereitung von Länderdaten zu Zeitschriften in Transformationsverträgen zeigten sich zahlreiche Datensätze, die zwar Autorenangaben, aber keine Affiliationen mehr enthielten. In einer umfassenden Analyse von über 13 Millionen Artikeln großer kommerzieller Verlage (2018–2025) wurde dieser Befund quantifiziert: Während 2023 noch rund 95 % der untersuchten Datensätze institutionelle Affiliationen aufwiesen, sank dieser Anteil bis 2025 auf etwa die Hälfte, wobei Elsevier als größter Verlag besonders betroffen war. Als Ursachen ließen sich zum einen fehlende Affiliationsmetadaten ausmachen, die einige Verlage nicht über Crossref bereitstellen, zum anderen technische Probleme im Übergang zur neuen OpenAlex-Version „Walden“.

Die Analyse legte damit eine strukturelle Schwäche des derzeitigen OpenAlex-Ansatzes offen: Da Affiliationsangaben von vielen Verlagen nicht offen über Crossref geliefert werden, greift OpenAlex für einen erheblichen Teil der Publikationen auf das Auslesen von Verlagswebseiten (Web Scraping) zurück. Diese Datengrundlage erwies sich als instabil, insbesondere seit Verlage angesichts zunehmender KI-gestützter Datensammlung verstärkt Schutzmaßnahmen gegen automatisierte Zugriffe einsetzen.

Die Befunde wurden im Austausch mit OpenAlex und in der Community offen diskutiert und fanden auch über das Vorhaben hinaus Beachtung, etwa in einem Beitrag des kommerziellen Anbieters Digital Science.¹ Dieser griff die Analyse auf und nutzte sie als Beleg dafür, dass aggregierte offene Quellen nur dann eine tragfähige Grundlage bilden, wenn die zentralen Metadaten, insbesondere Affiliationen mit ROR-IDs, bereits an der Quelle über Crossref offen bereitgestellt werden, statt nachträglich durch Web Scraping ergänzt zu werden.

Jahn, N. (2025). Decreasing affiliation metadata coverage in OpenAlex. <https://doi.org/10.59350/z3c5x-bfk63>

Arbeitsbereich 3: Datenkuratierung

Zitationen. Im Laufe des Projektes gab es vielfältige Änderungen in der OpenAlex-Datenbank, wodurch die Kuration von Zitationen in OpenAlex erschwert wurde. Für eine mögliche Ergänzung von Zitationsdaten wurde Semantic Scholar herangezogen, da diese Datenquelle ein einfacheres Datenschema im Vergleich zu OpenCitations bietet. Ein Vergleich der Referenzen in OpenAlex und Semantic Scholar hat daraufhin gezeigt, dass sich Semantic Scholar eignen würde, um Referenzen in OpenAlex anzureichern. Um Referenzen aus Semantic Scholar nachnutzen zu können, wurde ein entsprechender Datenbanksnapshot im Scholarly Data Warehouse der SUB Göttingen veröffentlicht, welcher knapp 3 Milliarden Referenzen enthält. Mit dem Update auf Walden im November 2025 sind allerdings in OpenAlex etliche neue Referenzen in die Datenbank hinzugekommen, weshalb die Eignung von Semantic Scholar zur Anreicherung von Zitationsdaten erneut geprüft werden müsste. Das Ergebnis der Referenzanalyse von Semantic Scholar und OpenAlex wurde als Blogpost veröffentlicht.

Hauptka, N. (2025). Is Semantic Scholar suitable for enriching references in OpenAlex? <https://doi.org/10.59350/8t2s7-vtw86>

Dokumenttypen. In diesem Arbeitspaket wurde ein Verfahren entwickelt, welches die Identifizierung von nicht-originiären Forschungsartikeln in OpenAlex unterstützt. Der Ansatz basiert auf einem Machine-Learning-Klassifikator, welcher fehlerhaft klassifizierte Forschungsartikel und Reviews erkennt und entsprechend kennzeichnet. Von etwa 42,7 Millionen Zeitschriftenartikeln in OpenAlex, die zwischen 2014 und 2023 veröffentlicht wurden, wurden etwa 4,5 Millionen Artikel identifiziert, die womöglich fälschlicherweise dem Dokumenttyp Artikel zugeordnet worden sind (etwa 10 %). Das heißt, es sind in diesen Fällen keine Forschungsartikel, sondern etwa Book Reviews, Abstracts oder Editorials. Ein Datensatz mit sämtlichen klassifizierten Zeitschriftenartikeln in OpenAlex (Publikationsjahre 2014 bis 2025) wurde im Rahmen des Open-Data-Releases dieses Projektes auf Zenodo veröffentlicht. Zusätzlich wurde ein Artikel über das Verfahren in der Zeitschrift Scientometrics eingereicht und veröffentlicht. Das French National Centre for Scientific Research (CNRS) bietet außerdem den Klassifikator auf seiner Webseite für Forschende in Frankreich an.² Im Rahmen der Arbeit mit Doku-

¹<https://www.digital-science.com/blog/no-shortcuts-to-research-information-citizenship/>

²<https://services.istex.fr/classification-de-documents-openalex-par-type-de-document/>

menttypen konnten darüber hinaus Bugs im Quellcode von OpenAlex gefunden und in Absprache mit den Entwicklern behoben werden.

Haupka, N. (2026). Presenting a classifier to improve the identification of research journal publications in OpenAlex. *Scientometrics*. <https://doi.org/10.1007/s11192-025-05524-7>

Open-Science-Evidenzen. Ein weiteres Ergebnis war die Bereitstellung eines kuratierten Datensatzes zu Open-Access-Transformationsverträgen, der im Rahmen des ersten OPENBIB-Datenrelease des Kompetenznetzwerks Bibliometrie an der SUB Göttingen entstand. Transformationsverträge haben sich seit ihrem ursprünglichen Vorschlag zu einem vorherrschenden Modell zur Finanzierung von Open Access in wissenschaftlichen Zeitschriften entwickelt, doch die Messung ihrer Wirkung bleibt schwierig, weil die zugehörigen Daten über verschiedene Quellen verstreut sind, bestehende Nachweise wie das cOAlition S Journal Checker Tool nur laufende Verträge abbilden und gängige Bibliometriedatenbanken keine Informationen zu Transformationsverträgen integrieren. Um diese Lücke zu schließen, verbindet der unter CC0 lizenzierte Datensatz die Daten der cOAlition S mit OpenAlex und umfasst 1.043 Transformationsverträge mit 17.199 erfassten Zeitschriften und 5.355 teilnehmenden Einrichtungen sowie 1.362.221 zuordenbare Artikel von Erstautorinnen und Erstautoren dieser Einrichtungen. Vorversionen des Datensatzes flossen bereits in das Hybrid Open Access Dashboard³ der SUB Göttingen sowie in Studien zur Wirkung von Transformationsverträgen auf das hybride Open-Access-Publizieren ein; eine unabhängige Validierung anhand von NWO-geförderten Publikationen bestätigte, dass sich auf Grundlage der Transformationsvertragsdaten und OpenAlex die Mehrheit der unter diesen Verträgen erschienenen Artikel zuverlässig identifizieren lässt (de Jonge, Kramer & Sondervan, 2025).

Jahn, N. (2025). Introducing Open Metadata about Transformative Agreements. https://subugoe.github.io/scholcomm_analytics/posts/openbib_ta_release/

Arbeitsbereich 4: Vernetzung

Im Berichtszeitraum wurde die verbundinterne Kommunikation erfolgreich aufrechterhalten; die zweiwöchentlichen Treffen und die User Group trugen wesentlich dazu bei, die Arbeiten der Partner aufeinander abzustimmen und Anforderungen aus der Praxis frühzeitig aufzunehmen.

Nach außen wurden die Ergebnisse des Vorhabens über zahlreiche Fachvorträge sichtbar gemacht, unter anderem auf der BiblioCon als der größten deutschsprachigen Fachkonferenz des Bibliothekswesens sowie innerhalb der internationalen OA2020-Community. Über diesen Austausch hinaus entstanden konkrete Anschlüsse an weitere Initiativen: Mit dem französischen CNRS wurde eine Zusammenarbeit zur Weiternutzung des im Vorhaben entwickelten Klassifikators aufgenommen, und über die Beteiligung an der Initiative Open Research Information Online Databases (ORION-DBs) wurden offene Datenbestände gemeinsam bereitgestellt und für Dritte nachnutzbar gemacht.

³<https://subugoe.github.io/hoaddash/>

Insgesamt trug der Arbeitsbereich damit dazu bei, die im Projekt entwickelten Verfahren und Daten in der Fach- und Anwendungscommunity zu verankern, Rückmeldungen aus der Praxis systematisch einzubeziehen und die Ergebnisse über den Verbund hinaus anschlussfähig zu machen.

5 Wichtigste Positionen des zahlenmäßigen Nachweises

Die Personalkosten dienten der Umsetzung des Arbeitsplans durch qualifiziertes Personal, das nach TV-L 13 vergütet wurde. Sämtliche Ausgaben erfolgten in Übereinstimmung mit den Richtlinien der Georg-August-Universität Göttingen.

6 Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Das Vorhaben ist von hoher Relevanz für die bibliometrische Landschaft in Deutschland, aber auch international. Die Bereitstellung von offenen Daten im KB erleichtert den Zugang und die Arbeit mit Bibliometriedaten, wodurch Analysen und Entscheidungsprozesse gestärkt werden können. Gleichzeitig bieten die erstellten Abdeckungsanalysen Transparenz beim Umgang mit und Umfang der neuen Datenquelle OpenAlex.

Die im Projekt entwickelten Verfahren zur Datenkuratierung sind öffentlich verfügbar, ermöglichen weiterführende Studien und können zudem für spezifische Anwendungen nachgenutzt werden.

7 Voraussichtlicher Nutzen, Verwertbarkeit der Ergebnisse und zukünftige Planungen im Sinne des Verwertungsplans

Im Vorhaben konnte OpenAlex als zusätzliche, offene Datenquelle im Kompetenznetzwerk Bibliometrie etabliert werden. Damit steht erstmals eine Alternative zu den kostenpflichtigen Datenbanken Scopus und Web of Science bereit, die ab 2026 den Regelbetrieb ausschließlich einer offenen Bibliometriedatenbank im KB ermöglicht. Die im Vorhaben erstellten Abdeckungsanalysen legen Stärken und Schwächen dieser neuen Datengrundlage offen und versetzen Nutzende in die Lage, fundiert zwischen OpenAlex und proprietären Quellen abzuwägen. Die entwickelten Kurationsverfahren ergänzen das Angebot von OpenAlex gezielt um Aspekte, die bislang nicht oder nur unzureichend abgedeckt sind.

Die Verwertbarkeit der Ergebnisse zeigt sich bereits an ihrer Nachnutzung und Resonanz. Die im Open-Data-Release bereitgestellten kuratierten Daten wurden auf Zenodo rund 2.000-mal aufgerufen und etwa 10.000-mal heruntergeladen. Der entwickelte Klassifikator zur Identifizierung nicht-originärer Forschungsartikel wird über das French National Centre for Scientific Research (CNRS) Forschenden in Frankreich angeboten und damit über den Verbund hinaus produktiv eingesetzt. Die im Vorhaben aufbereiteten Daten zu Open-Access-Transformationsverträgen finden zudem Eingang in die Arbeit von Verhandlungscommunities und wurden von Dritten aufgegriffen, etwa vom niederländischen

Förderer NWO, der das offene Verfahren anhand eigener Daten validiert hat (de Jonge et al., 2025). Schließlich greift OpenAlex selbst die im Vorhaben durchgeführten Abdeckungsanalysen auf, um die eigene Datenbasis zu verbessern. Die zugrunde liegenden Verfahren und Daten sind durchgängig offen lizenziert und ermöglichen sowohl weiterführende Studien als auch die Anpassung an spezifische Anwendungen.

Auch die wissenschaftliche Verwertung in international referierten Zeitschriften ist erfolgreich verlaufen: Die zentralen Ergebnisse wurden in den einschlägigen, hochrangigen Fachzeitschriften *Scientometrics* und *Quantitative Science Studies* veröffentlicht und werden bereits intensiv zitiert. Ergänzend dokumentieren mehrere Blogposts Zwischen- und Endergebnisse und machen sie für die Praxis zugänglich.

Über die unmittelbare Nachnutzung hinaus hat das Vorhaben auch einen strategischen Nutzen. Die Integration offener Daten in das Kompetenznetzwerk Bibliometrie stärkt die Verhandlungsposition gegenüber den proprietären Anbietern, da mit OpenAlex eine tragfähige Alternative zur Verfügung steht und damit die Abhängigkeit von einzelnen kommerziellen Datenquellen verringert wird. Mit der Bereitstellung offener Bibliometriedaten über Google BigQuery und der Mitgründung der Initiative Open Research Information Online Databases (ORION-DBs) wurde zugleich eine Grundlage für die gemeinsame, arbeitsteilige Pflege offener Forschungsinformationen geschaffen, die auch von Dritten nachgenutzt und weiterentwickelt werden kann. Für die Zukunft ist vorgesehen, OpenAlex im KB dauerhaft zu betreiben, die Kurationsroutinen kontinuierlich fortzuführen und die begonnene Vernetzung mit nationalen und internationalen Akteuren, etwa dem CNRS und den ORION-DBs-Partnern, weiter auszubauen.

8 Fortschritt auf dem Gebiet des Vorhabens während der Durchführung des Vorhabens bei anderen Stellen

Dokumenttypen. Während der Projektlaufzeit gab es umfangreiche Änderungen am Datenbestand seitens OpenAlex. Insbesondere die Klassifikation von Dokumenttypen wurde durch OpenAlex erheblich optimiert. Unter anderem wurden Dokumenttypen aus PubMed in OpenAlex integriert, was zu einer Verbesserung bei der Identifizierung von Editorials und Letters in OpenAlex geführt hat. Des Weiteren hat OpenAlex weitere Dokumenttypen implementiert, darunter Reviews. Diese Entwicklungen wurden im Projekt durch Blogposts aufgegriffen und dokumentiert. Auch von externer Seite wurden Analyseergebnisse bezüglich der Dokumenttypen in OpenAlex veröffentlicht. Mongeon et al. (2025) hat Diskrepanzen bei der Klassifikation von Dokumenttypen zwischen OpenAlex und Web of Science untersucht und dabei festgestellt, dass OpenAlex bei der Klassifikation von Forschungsartikeln und Reviews zu Missklassifikationen neigt. Ein Vergleich der Dokumenttypen in OpenAlex und Scopus wurde von Alperin et al. (2024) durchgeführt, wobei hier auf die unterschiedlichen Klassifikationssysteme in beiden Datenbanken hingewiesen wird. So benutzt OpenAlex den Dokumenttyp Artikel als Aufgangsbegriff für alle Dokumente, die sich keinem speziellen Dokumenttyp im Klassifikationssystem von OpenAlex zuordnen lassen. Auch Maisano, Mastrogiacomio, Ferrara und Franceschini (2025) weisen

auf Genauigkeitsprobleme bei der Klassifikation von Dokumenttypen in bibliometrischen Datenbanken hin. Gleichzeitig stellen sie ein Verfahren vor, mit dem sich fehlerhafte Dokumenttypenklassifikationen identifizieren lassen.

Open-Science-Evidenzen. Während der Projektlaufzeit waren im Bereich offener Open-Science-Evidenzen und Monitoring erhebliche Fortschritte zu verzeichnen, insbesondere durch Partner der Open Science Monitoring Initiative (OSMI).⁴ Auch OpenAlex hat sein Spektrum an erfassten Entitäten kontinuierlich erweitert, sodass zahlreiche für die Open-Science-Analyse relevante Informationen mittlerweile aus dieser Quelle bezogen werden können. Das Vorhaben hat sich daher bewusst auf die Ergänzung jener Aspekte konzentriert, die in OpenAlex bislang nicht oder nur unzureichend abgedeckt sind, allen voran Open-Access-Transformationsverträge.

Ein verwandtes Feld, das sich während der Laufzeit dynamisch entwickelt hat, betrifft die Erfassung von Diamond Open Access, also von Zeitschriften, die weder für das Lesen noch für das Publizieren Gebühren erheben. Hier besteht derzeit noch kein einheitliches Verständnis; vielmehr konkurrieren verschiedene Definitionsvorschläge, deren Tragfähigkeit gegenwärtig geprüft wird. Das Vorhaben hat begonnen, diese Diskussion zu begleiten, indem es Daten aus dem Diamond Open Access Discovery Hub⁵ in die eigene Dateninfrastruktur einspielt und so zur bibliometrischen Fundierung der Debatte beiträgt.

Offene Forschungsinformation. Ein weiterer Fortschritt betrifft die Bereitstellung offener Forschungsinformationen über Google BigQuery. Diese Umgebung hat sich als leistungsfähiges Werkzeug etabliert, um große offene Datenbestände kombinierbar und auswertbar zu machen, ohne dass auf Seiten der Nutzenden spezielle Hard- oder Software erforderlich ist. Mehrere Einrichtungen, darunter die SUB Göttingen, das niederländische CWTS, die brasilianischen Teams InSySPo/MultiObs sowie Sesame Open Science, haben dort unabhängig voneinander öffentliche Versionen einzelner offener Quellen bereitgestellt, bei denen der Anbieter die Speicherkosten trägt und Nutzende lediglich für ihre eigenen Abfragen aufkommen.

Mit der Gründung der Initiative Open Research Information Online Databases (ORION-DBs), an der die SUB Göttingen beteiligt ist, wurden diese zuvor verteilten Aktivitäten gebündelt. Ziel ist es, die Lasten der Bereitstellung gemeinsam zu tragen, doppelte Arbeit zu vermeiden und schrittweise gemeinsame Standards für Aufbereitung, Dokumentation und Zugang offener Forschungsinformationen zu entwickeln. Über eine zentrale Website werden die verfügbaren Datensätze einschließlich Bezugsquelle und Aktualisierungsstand nachgewiesen, sodass sie sowohl für gelegentliche Ad-hoc-Analysen als auch für umfangreiche, kontinuierliche Auswertungen zugänglich sind.

Diese Entwicklung wirkt unmittelbar auf die Kurationsstrategie des Vorhabens zurück. Da kein einzelner Datenbestand alle Anforderungen abdeckt, gewinnt die Kombination mehrerer offener Quellen für die Kuration zunehmend an Bedeutung: Lücken einer Quelle lassen sich durch andere ausgleichen, und die gemeinsame, auswertbare Bereitstellung erlaubt vergleichende Qualitätsprüfungen über Datenbankgrenzen hinweg. Entscheidend ist dabei, dass diese Form der Kuration nicht auf das Konsor-

⁴<https://open-science-monitoring.org/>

⁵<https://ddh.edch.eu/en>

tium beschränkt bleibt, sondern dank der offenen und niedrigschwelligen Bereitstellung auch von Dritten nachvollzogen, nachgenutzt und weiterentwickelt werden kann.

9 Erfolgte oder geplante Veröffentlichungen der Projektergebnisse

Zeitschriftenartikel

Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*, 130(4), 2475–2492. <https://doi.org/10.1007/s11192-025-05293-3>

Haupka, N., Culbert, J. H., Schniedermann, A., Jahn, N., & Mayr, P. (2026). Analysis of the publication and document types in OpenAlex, Web of Science, Scopus, PubMed and Semantic Scholar. *Quantitative Science Studies*, 7, 179–194. <https://doi.org/10.1162/qss.a.406>

Haupka, N. (2026). Presenting a classifier to improve the identification of research journal publications in OpenAlex. *Scientometrics*. <https://doi.org/10.1007/s11192-025-05524-7>

Jahn, N. (2025). Estimating transformative agreement impact on hybrid open access: a comparative large-scale study using Scopus, Web of Science and open metadata. *Scientometrics*, 131(2), 901–924. <https://doi.org/10.1007/s11192-025-05390-3>

Blogposts

Haupka, N., Dörner, S., & Jahn, N. (2024). Recent Changes in Document type classification in OpenAlex compared to Web of Science and Scopus. https://subugoe.github.io/scholcomm_analytics/posts/openalex_document_types/

Haupka, N. (2024). Identifying journal article types in OpenAlex. https://subugoe.github.io/scholcomm_analytics/posts/oal_document_types_classifier/

Haupka, N. (2025). Identifying journal article types in OpenAlex using OPENBIB data. <https://www.open-bibliometrics.de/posts/20250717-DocumentTypeClassifier/>

Haupka, N. (2025). Is Semantic Scholar suitable for enriching references in OpenAlex? https://subugoe.github.io/scholcomm_analytics/posts/s2_reference_analysis/s2_reference_analysis.html

Jahn, N. (2025). Decreasing affiliation metadata coverage in OpenAlex. <https://doi.org/10.59350/z3c5x-bfk63>

Jahn, N. (2025). Introducing Open Metadata about Transformative Agreements. https://subugoe.github.io/scholcomm_analytics/posts/openbib_ta_release/

Datenpublikation

Haupka, N., Culbert, J., Donner, P., Jahn, N., Lenke, C., Mayr, P., Meier, A., Mittermaier, B., Scheidt, B., Stahlschmidt, S., & Taubert, N. (2026). *OPENBIB: Selected curated open metadata based on OpenAlex (0.3)* [Data set]. Kompetenznetzwerk Bibliometrie. <https://doi.org/10.5281/zenodo.18429476>

Haupka, N., Culbert, J., Donner, P., Jahn, N., Lenke, C., Mayr, P., Meier, A., Mittermaier, B., Scheidt, B., Stahlschmidt, S., & Taubert, N. (2025). *OPENBIB: Selected curated open metadata based on OpenAlex (0.1)* [Data set]. Kompetenznetzwerk Bibliometrie. <https://doi.org/10.5281/zenodo.15308680>

Literatur

- Akbaritabar, A., Theile, T. & Zagheni, E. (2024). Bilateral flows and rates of international migration of scholars for 210 countries for the period 1998-2020. *Scientific Data*, 11 (1). doi: 10.1038/s41597-024-03655-9
- Alperin, J. P., Portenoy, J., Demes, K., Larivière, V. & Haustein, S. (2024). *An analysis of the suitability of OpenAlex for bibliometric analyses*. Zugriff auf <https://arxiv.org/abs/2404.17663>
- Aspesi, C. & Brand, A. (2020). In pursuit of open science, open access is not enough. *Science*, 368, 574 - 577. doi: 10.1126/science.aba3763
- Céspedes, L., Kozłowski, D., Pradier, C., Sainte-Marie, M. H., Shokida, N. S., Benz, P., ... Larivière, V. (2025, Januar). Evaluating the linguistic coverage of OpenAlex: An assessment of metadata accuracy and completeness. *Journal of the Association for Information Science and Technology*, 76 (6), 884–895. doi: 10.1002/asi.24979
- de Jonge, H., Kramer, B. & Sondervan, J. (2025). Tracking transformative agreements through open metadata: Method and validation using Dutch Research Council NWO funded papers. *Quantitative Science Studies*, 6, 1215–1227. doi: 10.1162/qss.a.24
- Donner, P. (2017, August). Document type assignment accuracy in the journal citation index data of web of science. *Scientometrics*, 113 (1), 219–236. doi: 10.1007/s11192-017-2483-y
- Haupka, N., Jahn, N. & Hobert, A. (2022). Praxisbericht Big Scholarly Data an der SUB Göttingen. *LIBREAS. Library Ideas* (41). doi: 10.18452/24800
- Kramer, B., Neylon, C. & Waltman, L. (2024, April). *Barcelona Declaration on Open Research Information*. Zenodo. doi: 10.5281/zenodo.10958522
- Maisano, D. A., Mastrogiovanni, L., Ferrara, L. & Franceschini, F. (2025). A large-scale semi-automated approach for assessing document-type classification errors in bibliometric databases. *Scientometrics*. doi: 10.1007/s11192-025-05244-y
- Mazoni, A. & Costas, R. (2024). Towards the democratisation of open research information for scientometrics and science policy: the Campinas experience. doi: 10.59350/eqmfk-82y98

- Mongeon, P., Hare, M., Riddle, P., Wilson, S., Krause, G., Marjoram, R. & Toupin, R. (2025). *Investigating document type, language, publication year, and author count discrepancies between openalex and web of science*. Zugriff auf <https://arxiv.org/abs/2508.18620>
- Priem, J., Piwowar, H. & Orr, R. (2022). *Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. Zugriff auf <https://arxiv.org/abs/2205.01833>
- Scheidsteger, T. & Haunschild, R. (2022). *Comparison of metadata with relevance for bibliometrics between Microsoft Academic Graph and OpenAlex until 2020*. Zugriff auf <https://arxiv.org/abs/2206.14168>
- Simard, M.-A., Basson, I., Hare, M., Larivière, V. & Mongeon, P. (2024). *The open access coverage of OpenAlex, Scopus and Web of Science*. Zugriff auf <https://arxiv.org/abs/2404.01985>
- Sugimoto, C. R. & Larivière, V. (2018). *Measuring research: what everyone needs to know*. Oxford University Press.
- Visser, M., van Eck, N. J. & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2 (1), 20–41. doi: 10.1162/qss_a_00112
- Waltman, L. & Larivière, V. (2020). Special issue on bibliographic data sources. *Quantitative Science Studies*, 1 (1), 360–362. doi: 10.1162/qss_e_00026
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3. doi: 10.1038/sdata.2016.18