

Komparative Analyse und Kuratierung Deutscher Metadaten in Offenen Bibliometriedaten (OPENBIB)

BMFTR-Abschlussbericht

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung, Technologie und Raumfahrt unter dem Förderkennzeichen 16WIK2301C gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor:innen.

Verfasst von

Barbara Scheidt und Bernhard Mittermaier (FZJ)

Zuwendungsempfänger: Forschungszentrum Jülich GmbH

Förderkennzeichen: 16WIK2301C

Vorhabenbeschreibung: Komparative Analyse und Kuratierung Deutscher Metadaten in Offenen Bibliometriedaten. Teilprojekt: Normierung von Daten zu Verlagen und Zeitschriften.

Laufzeit des Vorhabens: 01.05.2023-31.12.2025

Gefördert durch:



**Bundesministerium
für Forschung, Technologie
und Raumfahrt**

Teil 1

Kurzbericht

1 Ursprüngliche Aufgabenstellung und Stand der Forschung

Ziel des Vorhabens war der Aufbau einer offenen Bibliometriedatenbank innerhalb des Kompetenznetzwerks Bibliometrie (KB). Bei diesem Teilprojekt lag ein gesonderte Fokus auf der Analyse und Bereitstellung von Verlagen und Zeitschriften. Die Grundlage der Kuration bildete die Datenbank OpenAlex, das einen umfangreichen Datenbestand und einen kostenfreien Zugang bietet. Die mit diesem Projekt verbundene Öffnung des KBs hin zu offenen Bibliometriedaten sollte zudem eine Alternative zu proprietären Datenbanken wie Web of Science und Scopus unterstützen, deren Nutzung mit erheblichen Lizenzkosten einhergeht.

Zunächst wurden im Projekt Abdeckungsanalysen durchgeführt, die den gegenwärtigen Stand in OpenAlex mit den aktuell genutzten Datenquellen im KB, Scopus und Web of Science, vergleichen. In einem nächsten Schritt sollte OpenAlex im KB vollumfänglich zur Verfügung gestellt und mit kuratierten Daten angereichert werden. Gleichzeitig sollten die kuratierten Daten einem breiteren Publikum im Sinne des Open Data Prinzips zugänglich gemacht werden. Das KB konnte dabei auf jahrelange Erfahrung mit bibliometrischen Datenbanken zurückgreifen.

2 Ablauf des Vorhabens

Das Projekt ist in drei Arbeitsbereiche unterteilt, das FZJ hat Arbeitspakete in den Bereichen 2, 3 und 4 durchgeführt.

Nach der Bereitstellung von OpenAlex in der Dateninfrastruktur des KBs sowie über die Google Cloud Analytics Plattform BigQuery der SUB Göttingen durch die Projektpartner (Arbeitspaket 1), verfolgte der zweite Arbeitsbereich das Ziel, die Datenbasis von OpenAlex mit denen im KB etablierten Bibliometriedatenbanken Scopus und Web of Science zu vergleichen. Hierbei sollten insbesondere Verlage und Zeitschriften analysiert und verglichen werden.

Die Erkenntnisse aus dem zweiten Arbeitsbereich wurden im dritten Arbeitsbereich aufgegriffen und zur Entwicklung von Kurationsverfahren herangezogen mit denen die Datenbasis von OpenAlex im KB optimiert werden soll. Für Verlage und Zeitschriften wurden Möglichkeiten und Verfahren zur Datenkuratierung ermittelt, mit dem Ziel, die Datenqualität für das KB zu verbessern.

Erste Ergebnisse der Verlagsnormierung wurden in einem Open Data Release gebündelt und über das Datenrepositorium Zenodo zur Verfügung gestellt.¹ Zusätzlich wurden die Ergebnisse durch einen Blogpost dokumentiert.²

3 Wesentliche Ergebnisse

- **Vergleichsanalyse der Verlags- und Zeitschriftendaten von OpenAlex, Web of Science und Scopus.** Während des Projektes wurden die Verlags- und Zeitschriftendaten zwischen OpenAlex, Web of Science und Scopus verglichen.
- **Entwicklung eines Kurationsverfahrens zur Verbesserung der Verlags- und Zeitschriftendaten im KB.** Im Laufe des Projektes wurden Verlagsdaten normiert und ein mögliches Beziehungssystem der Verlage sowie eine Historisierung entwickelt. Für die

¹ <https://zenodo.org/records/15308680> , <https://zenodo.org/records/18429476>.

² <https://open-bibliometrics.de/posts/20250903-PublisherNormalization/>

Zeitschriftendaten wurden offene Quellen als Goldstandard zum Aufbau einer normierten Zeitschriftendatenbank evaluiert und erste Kuratierungsverfahren geprüft.

- **Datenbereitstellung.** Die in diesem Projekt erstellten Datenkurationen wurden über Zenodo einem breiteren Publikum zur Verfügung gestellt. In diesem Projekt wurde eine konkrete Datenkuratierung für Verlage beigesteuert.
- **Vergleichsanalyse.** Um die Auswirkung der Verwendung von normierten Daten in Analysen zu überprüfen, wurden auf Basis der OpenAlex-Daten Analysen vorgenommen und mit Analysen unter Verwendung kuratierter Daten durchgeführt.
- **Publikationen.** Das Vorhaben wurde mit einem Blogpost zu Ergebnissen aus der vergleichenden Analyse der Verlage und dem Kuratierungsverfahren begleitet.

Teil 2

Ausführlicher Bericht

1 Ausgangslage

Die Nachnutzung offener Bibliometriedaten ist in den letzten Jahren verstärkt in den Fokus vieler Einrichtungen gerückt. Neben steigenden Lizenzkosten für proprietäre Datenquellen stellen auch Restriktionen hinsichtlich der Weitergabe von Daten aus diesen Datenquellen ein Problem bei der Nutzung klassischer Bibliometriedatenbanken dar, was auch Auswirkungen auf die Nutzung von Scopus und Web of Science im Kompetenznetzwerk Bibliometrie (KB) hat. Aus diesem Grund verfolgt das KB eine strategische Neuausrichtung, bei der auch offene Datenquellen eine zunehmend wichtige Rolle einnehmen.

Im Zentrum der Kurationen steht der kostenlose Datenservice OpenAlex, das aus dem Microsoft Academic Graph hervorgegangen ist. Die Nutzung von OpenAlex im KB ist besonders interessant, da die Daten unter einer CC0-Lizenz veröffentlicht werden und daher uneingeschränkt nachgenutzt werden können. Darüber hinaus bietet OpenAlex eine sehr große Anzahl an Items, wodurch Vorteile gegenüber anderen Bibliometriedatenbanken entstehen können.

Nichtsdestotrotz waren mit der Datenbank OpenAlex zu Projektbeginn einige Probleme in Bezug auf bibliometrische Qualitätsstandards verbunden, weshalb es qualitätssichernder Analysen und Kurationen bedurfte.

Zu Beginn des Projekts war unklar, in welcher Vollständigkeit und Qualität Angaben zu Verlagen und Zeitschriften in den Publikationsdaten von OpenAlex vorliegen. Die Bibliometriedatenbanken Web of Science und Scopus bieten für diesen Datenbestand zwar nicht die gewünschte Qualität, vor allem in Bezug auf normierten Daten, stellen bisher aber die Datengrundlagen für derartige Analysen, wenn auch verbunden mit weiteren Bearbeitungsschritten zur Vereinheitlichung von Verlagen und Zeitschriften. Um die Notwendigkeit des letztgenannten Arbeitsschritts bei Verwendung einer offenen Datenbasis zukünftig zu vermeiden, galt es zunächst, den Istzustand der Daten in OpenAlex zu erfassen.

Im nächsten Schritt stand die Prüfung weiterer freier Datenquellen an, um die in OpenAlex indexierten Publikationen mit qualitativ hochwertigen und normierten Daten über Verlage und Zeitschriften optimieren zu können. Die Entwicklung erster Kuratierungsverfahren sollte diese Arbeit verstetigen und die Daten in die Infrastruktur des KBs zu integrieren.

2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Spätestens seit der Einführung der Barcelona Declaration im Jahr 2024 hat die Nachnutzung von offenen Forschungsinformationen wie etwa aus OpenAlex an Bedeutung gewonnen. Das hat zur Folge, dass Einrichtungen Lizenzen für proprietäre Datenquellen kündigen und verstärkt auf offene Alternativen setzen. Das KB konnte in Bezug auf offene Forschungsinformationen bereits Erfahrungen sammeln. So wurde unter anderem schon mit dem kostenlosen Open Access-Tool Unpaywall eine frei zugängliche Datenquelle im KB integriert. Die SUB Göttingen betreibt mit dem Open Scholarly Data Warehouse zusätzlich eine Analyse-Plattform, bei der ausschließlich offene Datenquellen wie Crossref, OpenAlex und Unpaywall integriert und genutzt werden (Hauptka, Jahn & Hobert, 2022).

Im Vergleich zu den proprietären Datenquellen Scopus und Web of Science, welche seit Jahrzehnten durch zahlreiche Studien untersucht werden, gibt es zu den neuen offenen Datenbanken vergleichsweise wenigen Studien. Eine genauere Untersuchung von kostenfreien Datenquellen könnte hierbei Aufschluss geben, inwiefern sich die Datenqualität zwischen offenen und proprietären Datenquellen unterscheidet.

Mit Blick auf den Datenumfang zeigt eine Studie von Simard, Basson, Hare, Larivière und Mongeon (2025), dass OpenAlex z.B. bei der Anzahl der (Diamond OA-) Zeitschriften inklusiver ist als Scopus und Web of Science, da die Aufnahme von wissenschaftlichen Journalen in den beiden proprietären Datenbanken von einem restriktiveren Aufnahmeprozess begleitet wird. Das hat zur Folge, dass OpenAlex zwar einen größeren Datenumfang bietet, was einerseits zu einer höheren Sichtbarkeit spezieller Disziplinen (Maddi, Maisonneuve & Boukacem-Zeghmouri, 2025) und Sprachen (Céspedes et al., 2025) führen kann, andererseits aber die Aufnahme von Datensätzen mit geringer Qualität begünstigen kann (Thelwall & Jiang, 2025). Im Vergleich zu anderen offenen Datenbanken wie etwa Semantic Scholar und dem etwas restriktiveren Service Google Scholar wird OpenAlex jedoch eine gute Metadatenqualität attestiert (Delgado-Quirós & Ortega, 2024).

Untersuchungen, die besonders die Angaben über Verlage oder Zeitschriften im Fokus haben, gibt es aber kaum. Und wenn doch, beziehen sie sich auf das Vorhandensein von ISSN und Informationen, die zur Identifikation einzelner Beiträge dienen, wie Volumes und Issues (Delgado-Quirós & Ortega, 2024). Zum Projektstart gab es also kaum Erkenntnisse über die Abdeckung und Datenqualität dieser Metadaten.

3 Planung und Ablauf des Vorhabens

3.1 Arbeitsbereich 2: Datenbankvergleich

Ziel dieses Arbeitsbereiches war es, die Datenbank 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 soll außerdem den Ausgangspunkt der anschließenden Datenkuration festhalten.

3.2 Arbeitsbereich 3: Datenkuratierung

In diesen Arbeitsbereich fallen die geplanten Datenkurationen der Verlags- und Zeitschrifteninformationen.

3.3 Arbeitspaket 3.3: Journale

In diesem Arbeitsbereich wurden Möglichkeiten einer Kuratierung von Zeitschriftendatensätzen mit Daten aus der Zeitschriftendatenbank ZDB vorgenommen.

3.4 Arbeitspaket 3.4: Verlagsbezeichnungen

Zur Verbesserung der Informationen über Verlage wurden Verlagsnamen vereinheitlicht und Vorgehensweisen zur Aktualisierung der Disambiguierung und Historisierung der Verlage geprüft.

3.5 Arbeitspaket 4.2: Use Cases

Konkrete Anwendungsbeispiele zeigen auf, welche Auswirkungen nicht normierte Daten auf bestimmte Ergebnisse haben können.

4 Inhaltliche Ergebnisse

4.1 Arbeitsbereich 2: Datenbankvergleich

In diesem Arbeitspaket wurde die Datenbank OpenAlex mit den proprietären Datenbanken Scopus und Web of Science hinsichtlich Verlage und Zeitschriften verglichen. Konkretes Ziel war zu ermitteln, ob die frei nutzbaren Daten aus OpenAlex mindestens die gleiche Abdeckung wie die beiden proprietären Datenbanken erreichen und qualitativ deren Standard für bibliometrische Analysen entsprechen. Darüber hinaus sollten die Ergebnisse Ansätze für eine notwendige Datenkuratierung liefern, indem Lücken oder Fehler identifiziert werden, die der Möglichkeit guter Analysen entgegenstehen.

Untersuchungseinheiten

Für den direkten Vergleich der Verlags- und Zeitschriftenangaben wurden die rund 25 Millionen Core-Publikationen untersucht, die in allen drei Datenbanken erfasst sind. Alle diese Publikationen haben eine DOI zur eindeutigen Identifikation, so dass es sich jeweils um Untermengen der drei Datenbankbestände handelt. Basierend auf der Annahme, dass die Angaben über die Quelle einer bestimmten Publikation, im konkreten Fall also die Angaben über Verlage und Zeitschriften, unabhängig von der Bereitstellung über eine Datenbank identisch sein müssten, können Abweichungen von dieser Annahme auf Fehler, oder zumindest Alternativen hinweisen.

Etwas anders verhält sich die Untersuchung von exklusiven Publikationen, die also zum Beispiel nur in OpenAlex, nicht aber in Web of Science oder Scopus enthalten sind. An diese zusätzlichen Publikationen knüpft sich die Hoffnung, Analysen auch für bisher unterrepräsentierte Fachgebiete oder Regionen zu ermöglichen. Die Untersuchung auf diese Publikationen auszuweiten, auch ohne Einzelvergleichsmöglichkeit auf Publikationsbasis, dient dem Erkenntnisgewinn über Vollständigkeit der gewünschten Daten und mögliche Probleme.

Ergebnisse für die Core-Publikationen

Der Vergleich zwischen den Datenbanken hat Unterschiede in den konkreten Angaben von Informationen offenbart. Im Fokus stehen die Serientitel wie Zeitschriften oder Veröffentlichungsreihen von Konferenzen sowie die Angaben über die herausgebenden Verlage.

Zeitschriften können über den Serientitel, aber auch die ISSN als "International Standard Serial Number" identifiziert werden. Beide bieten sich als Vergleichsmöglichkeit zwischen den Datenbanken an. Allerdings gibt es für viele Titel mehrere ISSN, z.B. für Druckausgaben andere als für die elektronische Form der Veröffentlichung, oder durch Veränderungen des Titels oder der Verlagszugehörigkeit. Dadurch und auch aufgrund des Formats erhöhen sich die Chancen auf Übereinstimmung zwischen den drei Datenbanken, so dass zunächst geprüft wurde, in welcher Vollständigkeit diese Information vorhanden ist. Ein Vergleich der Titel selber wurde ebenfalls vorgenommen, ist aber mit mehr Unsicherheiten behaftet: das liegt an unterschiedlichen Schreibweisen oder anderer Darstellungen von Zeichen als Bestandteil eines Titels und erfordert im Vorfeld eine Bearbeitung der Titel, um eine höhere Übereinstimmung zu erzielen. Zudem sind Titel nicht unbedingt exklusiv.

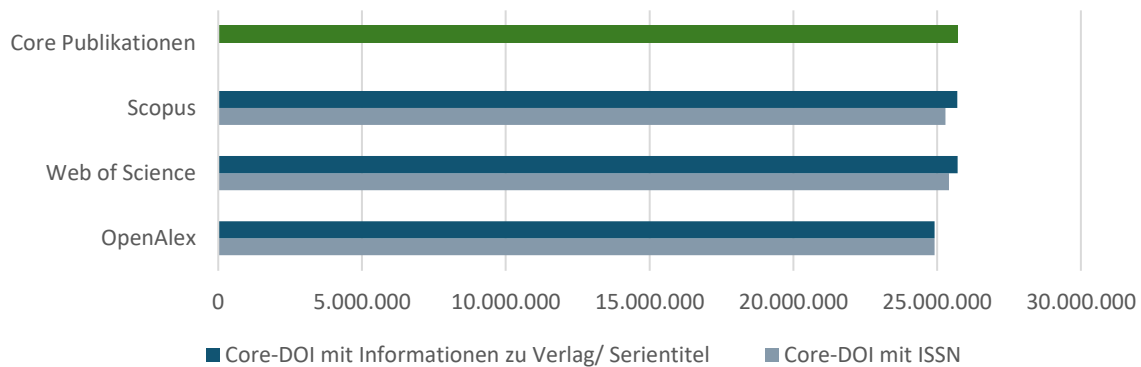


Abbildung 1: Anzahl der Datensätze (DOI-Zählung) mit Verlags- und Zeitschrifteninformationen bzw. mit ISSN im Vergleich zur Anzahl der Core-Publikationen je Datenbank

Eine erste Erkenntnis des Vergleichs ist, dass nicht alle Publikationsdatensätze die notwendigen bzw. wünschenswerten Informationen über Verlage, Zeitschriften und ISSN beinhalten: damit reduzieren sich die in OpenAlex auswertbaren Publikationen um rund 3 Prozent während es bei WoS und Scopus mit 1 bzw. 2 Prozent etwas weniger sind.

Anders als OpenAlex verzeichnen sowohl WoS als auch Scopus auch Konferenzbände mit einer Buchnummer (ISBN) als Kennung. Da Konferenztitel sehr unterschiedlich dargestellt werden und nur mit Mühe vergleichbar sind, werden in den folgenden Auszählungen nur Datensätze berücksichtigt, die eine ISSN enthalten, wodurch sich die Anzahl der Datensätze in WoS und Scopus um rund 1,1 bzw. 1,6 Prozent verringert hat.

Web of Science mit mehr Zeitschriftentiteln als ISSN, bei Scopus und WoS umgekehrt

Im folgenden Abschnitt geht es um die Anzahl distinkter Serientitel und ISSN als Hinweis auf Unterschiede in der konkreten Darstellung dieser Informationen in den Datenbanken. Im Idealfall wäre von einer hohen Übereinstimmung auszugehen.

Die in Tabelle 1 dargestellten Zahlen belegen jedoch eine andere Realität: während die jeweilige Anzahl der ISSN noch relativ ähnlich ist, unterscheidet sich die Anzahl der Serientitel teilweise sehr stark. Mehrere ISSN je Zeitschrift sind nicht ungewöhnlich,³ aber es gibt auch den umgekehrten Fall, der in der Regel darauf zurückzuführen ist, dass ein Titel in leicht unterschiedlichen Schreibweisen existiert. Die deutlich größere Zahl von Zeitschriftentiteln (39.987) zu ISSN (27.487) in WoS deutet jedoch auf andere Gründe hin. Und das Zahlenverhältnis ist sowohl bei Scopus als auch OpenAlex genau umgekehrt: die Anzahl der ISSN liegt über der Anzahl der Titel.

Tabelle 1: Anzahl der Core-Publikationen (DOI), Quellentitel und ISSN je Datenbank in distinkter Zählung.

	Anzahl DOI	Anzahl Sourcetitel	Anzahl ISSN
OpenAlex	24.915.717	14.360	27.932
Web of Science	25.416.116	39.987	27.487
Scopus	25.294.707	21.714	26.176

³ OpenAlex verzeichnet bis zu 11 ISSN für einige Titel, am häufigsten jedoch 2.

Die geringste Anzahl unterschiedlicher Quellentitel verzeichnet OpenAlex, während diese Zahl bei Scopus 1,5 mal so groß ist und bei Web of Science sogar den 2,8-fachen Wert erreicht.

Als Folge von abweichenden Schreibweisen, z.B. durch zusätzliche Kommas, Spiegelstriche, Leerzeichen oder '&'-Zeichen statt 'and' oder 'und' kommt es im direkten Titelvergleich zu weniger Übereinstimmungen bei gleichen Titeln. Selbst innerhalb einer Datenbank kommt es dadurch zu Mehrfach-Nennungen eines Titels.

Genau dieser Sachverhalt ist für bibliometrische Analysen immer wieder problematisch: Zeitschriftentitel erscheinen im Ergebnis mehrfach, meistens auch mit unterschiedlichen Publikationszahlen, obwohl diese eigentlich zusammengehören. Die voneinander abweichenden Schreibweisen sind oft unbekannt, werden nicht berücksichtigt und führen so zu fehlerhaften Analyseergebnissen.

Beispiel aus OpenAlex

<i>Chemical engineering research & design</i>	5.127
<i>Chemical Engineering Research & Design</i>	4.432

Abbildung 2: Zeitschriftentitel aus OpenAlex in unterschiedlicher Schreibweise und Item-Zählung

Welche Unterschiede gibt es?

OpenAlex und Web of Science indexieren eher Zeitschriftentitel, auch, wenn es sich um ein Special Issue mit einem eigenen Namen handelt. Dieser wiederum wird in Scopus häufiger als Quelle genannt anstelle des Serientitels (vgl. Abbildung 4).

Beispiel aus Web of Science

<i>acoustics & vibration of mechanical structures</i>	58
<i>acoustics and vibration of mechanical structures (avms-2017)</i>	49

Abbildung 3: Zeitschriftenbeispiel aus dem Web of Science – Titel durch Kurzform und Jahresangabe

Beispiel Journal of Proteomics, ISSN 1874-3919

<i>opal</i>	<i>Journal of Proteomics</i>	4.095
<i>wos</i>	<i>JOURNAL OF PROTEOMICS</i>	4.092
	<i>Journal of Proteomics</i>	3
<i>scp</i>	<i>Journal of Proteomics</i>	4.039
	<i>V ICAP Proteogenomics</i>	22
	<i>10 Year Anniversary of Proteomics</i>	21
	<i>The Central and Eastern European Proteomics Conferences: 10 Years Anniversary</i>	6
	<i>Food and Crop Proteomics</i>	3

Abbildung 4: Ergebnisse zur ISSN 1874-3919 in den drei Datenbanken mit einer größeren Titelvarianz bei Scopus, die auf Titeln von Special-Issues zurückzuführen ist.

Wie kommt es aber zu so einer großen Anzahl von Titeln in Web of Science bzw. warum ist die Anzahl der Titel in Scopus und vor allem in OpenAlex so viel niedriger?

Web of Science-Titel enthalten sehr häufig auch die Angaben eines bestimmten Bandes (Vol oder Volume mit der entsprechenden Zählung, vgl. Abbildung 3), während dies bei OpenAlex gar nicht und bei Scopus-Datensätzen nur in geringem Maße der Fall ist. Insgesamt wurden im Core-Bestand in Web of Science 34.445 DOI mit einer solchen Angabe gezählt, in Scopus nur 736.

Und wo Scopus oder OpenAlex den Buchserientitel als Quelle angeben, nennt Web of Science die Unterreihe. Beispiel: *Advances in Experimental Medicine and Biology* ist die Serie, während das 2-bändige Werk den Titel "Hypertension: from basic research to clinical practice" heißt und so auch in WoS als Quellentitel erfasst ist (<https://link.springer.com/book/10.1007/978-3-319-44251-8>). Um vergleichbare Auswertungen über die Datenbanken erstellen zu können, müssten die Titel vor allem in Web of Science also zunächst entsprechend bereinigt bzw. angeglichen werden.

Insgesamt gehen durch diese Praxis in Scopus und OpenAlex relevante Informationen nicht verloren. Die Angaben über die Quellen sind sogar präziser. Für die Analysen der Quellentitel können die sehr detaillierten Angaben aber von Nachteil sein. Die Ausgangslage in OpenAlex hinsichtlich Analysen von Zeitschriftentiteln ist also vergleichsweise gut. Dennoch würden eine Prüfung und Vereinheitlichung dieses Datenbereiches die Qualität und Zuverlässigkeit von Auswertungen weiter verbessern.

Welche Erkenntnisse gibt es über die Verlage in OpenAlex-Daten im Vergleich zu WoS und Scopus?

Ein Abgleich der Verlagsinformationen der den knapp 25 Millionen Core-Publikationen hat unterschiedliche Ergebnisse in allen drei Datenbanken erbracht. Während die Abdeckung von Core-Publikationen mit Verlags-Angaben bei OpenAlex bei durchschnittlich 97 Prozent liegt, ist sie bei den WoS-Daten vollständig, liegt bei Scopus aber nur bei 81 Prozent. Die Schnittmenge von Publikationen, zu denen in jeder Datenbank Verlags-Angaben vorliegen, umfasst 20.104.320 Publikationen.

Bereits die Auswertung nach Anzahl der unterschiedlichen Verlagsnamen weist auf Unterschiede in der Datenpraxis hin: die geringste Anzahl gibt es in OpenAlex (2.008), gefolgt von Web of Science (3.487) und Scopus (3.699). Die Gründe für diese Ergebnisse sind vielfältig. Auch hat die vergleichende Auswertung der Verlage mit den höchsten Publikationszahlen selbst nach groben manuellen Zusammenführungen von unterschiedlichen Schreibweisen sehr deutlich voneinander abweichende Ergebnisse erbracht, wie Abbildung 5 zeigt.

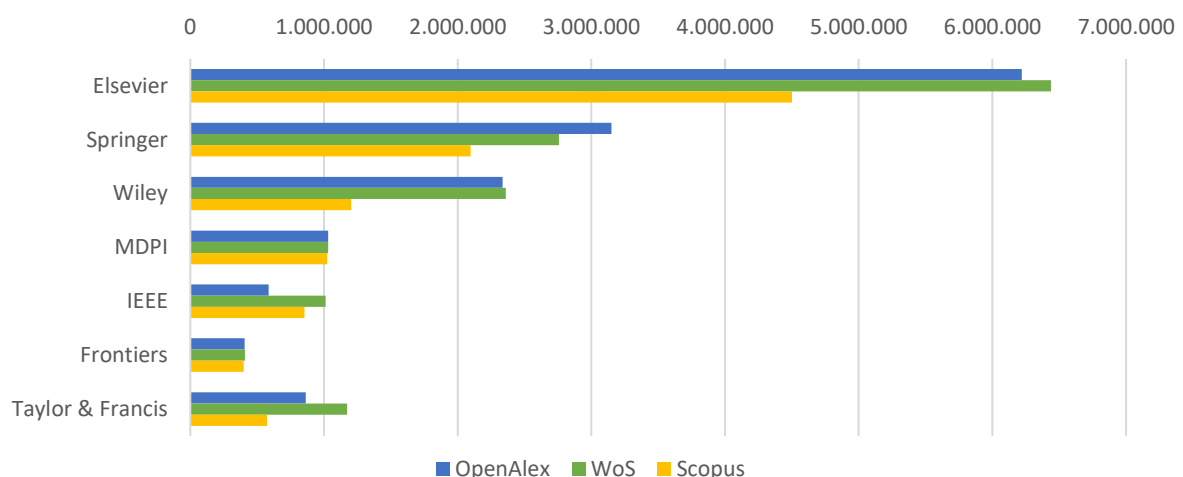


Abbildung 5: Anzahl Publikationen der größten Verlage in OpenAlex, Web of Science und Scopus 2010-2022

Nähere Prüfungen haben gezeigt, dass nicht nur Schreibweisen, sondern auch die Berücksichtigung oder Nicht-Berücksichtigung von Verlags-Hierarchien bzw. Zugehörigkeiten eine Rolle spielen. So nennt z.B. Scopus viel häufiger die Namen der zu Elsevier gehörenden Unterverlage als Web of Science oder OpenAlex: Dies erklärt die so stark voneinander abweichenden Zahlen zu Elsevier (Abbildung 5).

Zu einzelnen Publikationen gibt es also in den verschiedenen Datenbanken nicht selten unterschiedliche Verlage. Das Beispiel einer Auswahl von Publikationen aus dem Journal 'Advances in Nutrition' illustriert die Problematik wechselnder Herausgeberschaften und die Rolle von

Wissenschaftlichen Gesellschaften als Besitzer von Zeitschriften. Die Zeitschrift gehört der American Society for Nutrition, wurde und wird aber zu unterschiedlichen Zeiten durch unterschiedliche Verlage herausgegeben. Während OpenAlex ausschließlich diese Society als Verlag nennt, führen Web of Science ebenso korrekt zusätzlich Oxford University Press und Elsevier und Scopus wiederum zusätzlich (aber nicht korrekt) Medline als Verlage auf. Elsevier hat die Zeitschrift eigentlich erst nach dem Untersuchungszeitraum übernommen, und in Scopus fehlen für 16 % der Publikationen komplett die Verlagsangaben.

Der Vergleich der Verlagsangaben zwischen den drei Datenbanken zeigt durch dieses Beispiel, aber auch generell, für OpenAlex eine vergleichsweise hohe Vollständigkeit und keine schlechtere Datenqualität. Durch eine nicht einheitliche Zuordnungspraxis und fehlende Normierung ist diese als Grundlage für Auswertungen jedoch ebenso wie die in WoS und Scopus nur bedingt geeignet und erfordert umfangreiche Überprüfungen und Datenkuratierungen, um gute Ergebnisse liefern zu können.

Margin Publikationen

OpenAlex hat über die gemeinsamen Publikationen mit WoS und Scopus hinaus die absolut größte Menge an zusätzlichen Publikationen, die hier als 'margin' bezeichnet werden. Der Anteil von Datensätzen ohne die Angabe eines Verlags oder einer Quelle ist allerdings deutlich höher als bei den Core-Publikationen.

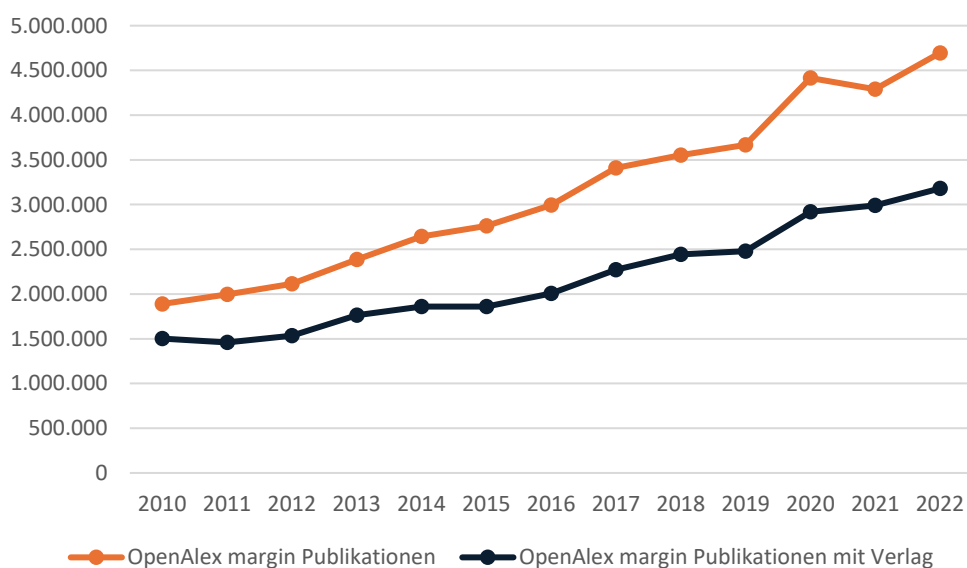


Abbildung 6: Anzahl Margin-Publikationen im Zeitverlauf und Publikationen mit Verlagsangaben

Rund 19 % der rund 44 Millionen exklusiven Publikationen in OpenAlex enthalten keine Informationen über die Veröffentlichungsquellen. In der Auswertung der vorhandenen Verlage sind aber durchaus alle großen Verlage vertreten, darüber hinaus enthalten die Verlagsnamen aber häufig Angaben wie 'Society', 'University' oder 'Association'.

Aufschluss über die Art des zusätzlichen Contents gegenüber WoS und Scopus gibt auch eine Analyse der Beitragstitel: so werden in nicht geringem Anteil Inhaltsverzeichnisse und Editorial Boards als Beiträge indexiert und leider auch häufig als Dokumenttyp 'article' bezeichnet. Folglich wäre eine Korrektur der Bezeichnung der Publikationsarten in OpenAlex unabdingbar zur Identifizierung von wissenschaftlichen Veröffentlichungen.

3.5 Arbeitspaket 4.2: Use Cases

Verschiedene Nutzungsszenarien haben die Auswirkungen von nicht standardisierten Daten auf Analysen untersucht. In einem konkreten Anwendungsfall für das Forschungszentrum Jülich sollte die Frage beantwortet werden, bei welchen Verlagen Mitarbeitende des Zentrums am häufigsten publizieren. Ausgehend von den Core-Publikationen ist ein Publikationsset erstellt worden, das den Zeitraum 2010 bis 2022 auf Basis des Publikationsjahres im WoS abdeckt sowie durch die FZJ-Adressen gemäß der Institutionenkodierung ebenfalls aus dem WoS identifiziert wurde. Wie unterschiedlich die Ergebnisse einer Verlagsauswertung mit den in den Angaben der jeweiligen Datenbanken sein können, zeigen die nachfolgenden Tabellen.

VERLAG OPENALEX		ANZAHL PUBLIKATIONEN
1	Elsevier BV	6.077
2	American Physical Society	2.044
3	American Chemical Society	1.618
4	Springer Science+Business Media	1.612
5	IOP Publishing	1.464
6	Wiley	1.313
7	Wiley-Blackwell	1.176
8	American Institute of Physics	1.019
9	Royal Society of Chemistry	972
10	Nature Portfolio	871
VERLAG WEB OF SCIENCE		ANZAHL PUBLIKATIONEN
1	AMER PHYSICAL SOC	2.126
2	AMER CHEMICAL SOC	1.610
3	ELSEVIER	1.600
4	IOP PUBLISHING LTD	1.247
5	SPRINGER	1.216
6	WILEY-V C H VERLAG GMBH	1.170
7	PERGAMON-ELSEVIER SCIENCE LTD	1.099
8	WILEY	1.057
9	ELSEVIER SCIENCE BV	1.046
10	ROYAL SOC CHEMISTRY	980
VERLAG SCOPUS		ANZAHL PUBLIKATIONEN
1	Elsevier Ltd	1.971
2	Elsevier B.V.	1.671
3	American Physical Society	1.568
4	American Chemical Society	1.360
5	Institute of Physics Publishing	936
6	Royal Society of Chemistry	833
7	Wiley-VCH Verlag	705
8	American Institute of Physics Inc.	602
9	John Wiley and Sons Inc	569
10	Copernicus GmbH	511

Abbildung 7: Analyse der Verlage in den drei Datenbanken für Publikationen des FZJ 2010-2022

Alle Ergebnisse sind sehr unterschiedlich und in allen Tabellen kommt bereits unter den Top 10 mindestens ein Verlag doppelt vor, in verschiedenen Varianten und mit unterschiedlichen Publikationszahlen.

Besser sieht es bei der Auswertung der Zeitschriften aus, wie die nächsten Tabellen mit den Top 10-Auswertungen der Quellen zeigen.

ZEITSCHRIFTEN OPENALEX		ANZAHL PUBLIKATIONEN
1	Nuclear fusion	476
2	Physical review. B./ Physical review. B	437
3	Physical review. B, Condensed matter and materials physics	408
4	Atmospheric chemistry and physics	407
5	Physical review letters	387
6	Scientific reports	326
7	Journal of nuclear materials	268
8	Fusion engineering and design	258
9	Applied physics letters	243
10	Nature communications	226
ZEITSCHRIFTEN WEB OF SCIENCE		ANZAHL PUBLIKATIONEN
1	PHYSICAL REVIEW B	845
2	NUCLEAR FUSION	476
3	ATMOSPHERIC CHEMISTRY AND PHYSICS	408
4	PHYSICAL REVIEW LETTERS	387
5	SCIENTIFIC REPORTS	326
6	PHYSICAL REVIEW D	277
7	JOURNAL OF NUCLEAR MATERIALS	268
8	FUSION ENGINEERING AND DESIGN	258
9	APPLIED PHYSICS LETTERS	243
10	NATURE COMMUNICATIONS	226
ZEITSCHRIFTEN SCOPUS		ANZAHL PUBLIKATIONEN
1	Nuclear Fusion	476
2	Physical Review B	436
3	Physical Review B - Condensed Matter and Materials Physics	408
4	Atmospheric Chemistry and Physics	408
5	Physical Review Letters	386
6	Scientific Reports	325
7	Journal of Nuclear Materials	268
8	Fusion Engineering and Design	258
9	Applied Physics Letters	243
10	Nature Communications	226

Abbildung 8: Auswertung der Zeitschriften in den drei Datenbanken für Publikationen des FZJ 2010-2022

Oft stimmen Titel und Anzahl der Publikationen überein, aber die Varianten der Zeitschrift *Physical Review B* zeigen die Auswirkungen auf die Analyse und Notwendigkeit von einheitlichen Benennungen auf. Um die Frage zu klären, ob die Zusammenführung der Titel *Physical Review B* und *Physical Review B - Condensed Matter and Materials Physics* zu einem Titel, wie in den WoS-Daten, korrekt ist, sind zusätzliche Quellen wie die ZDB hilfreich und notwendig. Ist diese Information in einem kuratierten Datensatz vorhanden, spart man sich bei Analysen nicht nur Bearbeitungsprozesse zum Bereinigen der Daten, sondern kann je nach Fragestellung auch Titelveränderungen berücksichtigen, die z.B. im Zusammenhang mit Lizenzmanagementanforderungen stehen.

4.3 Arbeitsbereich 3: Datenkuratierung

4.3.1 Arbeitsbereich 3: Datenkuratierung

Ergebnisse der Vergleichsanalyse und erneute Datenprüfungen im Projektverlauf haben die Notwendigkeit der Datenkuratierung sowohl für die Zeitschriften als auch die Verlage in OpenAlex aufgezeigt.

Das bedeutet konkret, dass ergänzend zu den in OpenAlex bereitgestellten Zeitschriftentiteln und Verlagen zu den Publikationen jeweils standardisierte Bezeichnungen dieser beiden Datenbereiche angeboten werden sollen. Mit Hinblick auf die Erkenntnisse aus den Vergleichsanalysen werden damit Schreibvarianten oder sogar falsche Titel oder Verlage korrigiert und ggf. auch Identifikatoren.

Für die Zeitschriften waren folgende Teilinformationen für die Kuratierung von Bedeutung:

1. Titel: Schreibweise standardisieren (ggf. sind leichte Änderungen, z.B. des Untertitels eine Herausforderung)
2. Identifikatoren: ISSN erfassen (sie dienen zum Abgleich mit den ISSN in den Publikationsdatensätzen von OpenAlex)
3. Gültigkeit: Beginn und ggf. Enddatum für den Erscheinungsverlauf einer Zeitschrift.
4. Beziehungen: gibt es Vorgänger- oder Nachfolgetitel?
5. Bei welchem Verlag erscheint eine Zeitschrift? (ebenfalls mit Gültigkeitsdauer)

Damit notwendigerweise verbunden sind Verlagsangaben in ebenfalls standardisierter Form:

1. Name einheitlich (herausfordernd bei großen Verlagsstrukturen mit z.B. vielen Standorten)
2. Verlinkungen zu den Verlagswebseiten
3. Identifikatoren oder Verknüpfungen zu weiteren Informationssystemen (z.B. Research Organisation Registry ROR, das auch Verlage indexiert, oder Wikidata)
4. Gültigkeit: Gründung und ggf. Ende des Verlags
5. Beziehungen/ Relationen: viele Verlage sind nicht (mehr) eigenständig und in größere Verlagsstrukturen integriert. Das ermöglicht in Analysen verschiedene Perspektiven. Gültigkeitszeiträume für Beziehungen gehören dazu und ermöglichen es nicht nur, Verlagsentwicklungen nachzuvollziehen sondern auch die damit einhergehenden Wechsel der Zeitschriften dieser Verlage.

4.3.2 Arbeitspaket 3.3: Journale

Die Kuratierung der Zeitschriftendaten ist im Projektzeitraum ausführlich geprüft worden, die Umsetzung steht in der nachfolgenden Förderperiode an. Das Hauptaugenmerk lag auf der Analyse der Voraussetzungen in OpenAlex und der Überprüfung von Datenquellen für den Aufbau einer Informationsbasis mit kuratierten Zeitschrifteninformationen, wie oben beschrieben. Im Folgenden werden diese möglichen Quellen vorgestellt, mit der ZDB und der EZB zunächst zwei Angebote aus Deutschland.

Die Zeitschriftendatenbank (ZDB)⁴ ist ein wichtiges Angebot für die aufwändige Verwaltung von Zeitschriften in den Bibliotheken, die Daten ihres Bestandes an die ZDB liefern und umgekehrt nachnutzen können. Durch die stetig wachsende Menge an Serientiteln, die Veränderungen der Titel und Verlagszugehörigkeiten sowie das Hinzukommen neuer Medienarten ist zwar eine unverzichtbare und wertvolle Wissensbasis gewachsen, aber auch eine mit uneinheitlichen Teilinformationen. Die Daten der ZDB werden kooperativ durch über 4.000 Bibliotheken eingepflegt. Neben ihrer Bedeutung als Nachweisdatenbank für deutsche und österreichische Bibliotheken von Zeitschriften weltweit und mit umfangreichen Informationen über Titel und Titelveränderungen, Erscheinungsverläufen, Verlagsinformationen, ISSN und weiteren IDs war die ZDB zunächst als Quelle zum Anreichern der Zeitschrifteninformationen in der offenen Bibliometriedatenbank in Betracht gezogen worden. Zudem sind die Daten frei verfügbar und nachnutzbar.

Nach einer eingehenden Prüfung wurde aber dieses Vorgehen verworfen. Viele Felder enthalten keine normierten Daten sondern bringen als freie Textfelder einen sehr großen Variantenreichtum an Einträgen mit sich, so dass erneut ein größerer Aufwand erforderlich wäre, bevor diese Daten nachgenutzt werden könnten. Die Informationen sind wertvoll im Rahmen gezielter Einzelrecherchen, erscheinen aber angesichts fehlender Normierung und Klarheit für die Datenkuratierung als zu problematisch.

Eine weitere relevante Quelle für wissenschaftliche Zeitschriften ist die Elektronische Zeitschriftendatenbank EZB⁵, die als Bibliotheksservice an der Universität Regensburg betrieben wird, die ebenso wie die ZDB die Daten zur freien Nachnutzung zur Verfügung stellt. Durch eine enge Kooperation mit der ZDB gibt es über die ZDB-ID eine Verlinkung zwischen den elektronischen Medien der EZB und den entsprechenden Nachweisen in der ZDB sowie gemeinsame Projekte zur Datenqualitätsverbesserung und Verfügbarkeitsrecherchen. Die enthaltenen Informationen über die Verlage bieten sich zwar an, um im Vorhaben der Zeitschriften- und Verlagskuratierung die jeweiligen Besitzverhältnisse gleich mit zu erfassen, werden aber ebenso wie in der ZDB in Freitextfeldern vorgehalten und sind nicht vereinheitlicht bezüglich Schreibvariante oder Gültigkeit.

Daher wurde für das Vorhaben nach einer weiteren umfangreichen Prüfung auf Daten zurückgegriffen, die unmittelbar durch die Verlage bereitgestellt werden. Im Format einer KBART-Datei können Verlage standardisierte Informationen ihrer Zeitschriften in vorgegebenen Feldern verfügbar machen— z.B. für Bibliotheken zu Nachnutzung in ihren Managementsystemen für Elektronische Ressourcen. KBART steht für Knowledge Base And Related Tools und "ist eine Initiative zum Austausch von Metadaten".⁶ Wichtige Informationen, die dadurch zur Verfügung gestellt werden, sind die Titel und Identifikatoren wie ISSN (oder im Falle von Büchern ISBN), Gültigkeitszeiträume, Medienarten und Zugriffsmöglichkeiten. Da die KBART-Dateien im tsv-Format direkt von den Anbietern kommen, ist die Verlagszugehörigkeit der Zeitschriften damit ebenfalls dokumentiert. In den Dateien großer Verlage sind sogar häufig für einzelne Zeitschriftentitel andere Verlage aufgeführt, bei denen es sich i.d.R. um Teilverlage des Hauptverlags handelt. Damit liefern sie wichtige Informationen zu einem der geplanten Teilbereiche im Projekt, nämlich zu den Beziehungen zwischen Eltern- und Kind-Verlag. Ein großer Vorteil ist auch, dass regelmäßige Aktualisierungen angeboten werden. Ändern sich also Zeitschriftentitel, wechseln zu einem anderen Verlag oder enden ganz, so können diese Informationen aus den KBART-Listen entnommen werden und auch im KB-Schema regelmäßig angepasst werden.

Es gibt jedoch auch Nachteile: die KBART-Listen müssen von jedem Verlag einzeln eingesammelt werden und nicht alle Verlage stellen diese bereit. Zudem werden nicht alle Standards der KBART-Initiative zuverlässig eingehalten bzw. nicht alle notwendigen Datenfelder mitgeliefert.

⁴ <https://zeitschriftendatenbank.de/startseite> (26.06.2026).

⁵ <https://ezb.ur.de/>

⁶ <https://format.gbv.de/kbart>; <https://www.niso.org/standards-committees/kbart>;

Obwohl dies Einschränkungen einen nicht unerheblichen Aufwand beim Aufbau der kuratierten Zeitschriften- und auch Verlagsdaten erfordert, stehen im Vergleich zu den anderen Möglichkeiten die bestmöglichen Informationen durch die KBART-Dateien zur Verfügung. Vor allem der Direktbezug von den Verlagen, die Aussicht auf regelmäßige Aktualisierungen und die wachsende Bedeutung von KBART-Listen für andere Nachweissysteme⁷ sprechen für die Verwendung dieser Datenquelle im Rahmen des Vorhabens. Ein weiterer Vorteil ist, dass sie Daten für beide Vorhaben liefern, sowohl für die Zeitschriften- als auch die Verlagsnormierungen.

4.3.3 Arbeitspaket 3.4: Verlagsbezeichnungen

Die Verlagsbezeichnungen sind im Projektverlauf bereits bearbeitet und veröffentlicht worden. Die Herangehensweise ist eine andere gewesen, als die geplante und oben im Zusammenhang mit dem Aufbau einer kuratierten Datenbasis sowohl der Zeitschriften als auch der Verlage beschriebene. Die erzielten Ergebnisse und gewonnen Erkenntnisse sind für das Anschlussvorhaben aber sehr nützlich.

Ausgangspunkt für die Standardisierung waren die Verlagsangaben durch OpenAlex. Diese waren bereits erkennbar vereinheitlicht, aber weder fehlerfrei noch ohne Varianten. Ausgehend von den Verlagen mit den höchsten Publikationszahlen wurden weitere Schreibvarianten dieser Verlage – auch durch Teilsting-Abgleiche – ermittelt, jedoch ausschließlich solche in lateinischer Schreibweise. Recherchen auf den Verlagswebseiten, in Wikidata oder auf der ROR-Webseite dienten der Überprüfung der vorhandenen Angaben, haben zusätzliche Informationen über die Verlage geliefert und zum Aufbau von zwei Tabellen beigetragen: Standardbezeichnungen und Relationen. Auch OpenAlex hat in den Rohdaten vereinzelt Links oder Angaben über Verlagsbeziehungen dokumentiert, die im Projekt ggf. nach weiterer Prüfung nachgenutzt worden sind.

Die Ergebnisse wurden in einem Blogbeitrag⁸ beschrieben und im Projektbereich von Zenodo⁹ veröffentlicht.

Tabelle 2: Tabelle Verlage mit kuratierten Daten, Beispiel 'Nature Publishing Group'.

table field	data
publisher_id	34297
publisher_id_orig	P4366736682
publisher_name	Nature Publishing Group
standard_name	Nature Portfolio
unit_pk	182
wikidata	Q180419
ror	03dsk4d59
url	https://www.nature.com/nature-portfolio

Tabelle 2 illustriert das Ergebnis der Datenkuratierung für den Verlag Nature Publishing Group: die Publisher-IDs und der Publisher_Name stammen aus OpenAlex, ebenso wie Identifikatoren von Wikidata und ROR. Letztere wurden manuell auf Richtigkeit überprüft und in anderen Fällen selber recherchiert, ebenso wie der Verlagslink (URL). Mit dem primären Schlüssel (Unit-PK) in der Tabelle des KB-Schemas ist diesem Datensatz eine eindeutige Kennung mit der Möglichkeit zur Weiterverknüpfung mit anderen Einträgen und Tabellen zugewiesen worden.

Diese Verknüpfung wird in Tabelle 3 angewendet, die das Beispiel des Verlags in Nature Portfolio in einer Beziehung zu einem übergeordneten Verlag abbildet. Die Relation zu Macmillan Publishers als

⁷ <https://gokb.org/kbart-listen-besorgen-und-pruefen/>; https://www.bib-bvb.de/documents/d/ker/4_ker_gokb_weisheit

⁸ <https://open-bibliometrics.de/posts/20250903-PublisherNormalization/>

⁹ <https://zenodo.org/records/18429476>

Elternverlag bestand von 1999 bis 2015, wie durch den Gültigkeitszeitraum dargestellt ist. Der Child_Name ist der Standardname für alle Schreibvarianten aus der vorherigen Tabelle, im Feld Child_Unit ist die eindeutige Kennung übernommen worden. Dasselbe gilt für die Einträge in Parent_Name und Parent_Unit, sie sind so ebenfalls in der Tabelle mit den Standardnamen erfasst.

Tabelle 3: Tabelle mit kuratierten Daten für die Verlagsbeziehungen, Beispiel 'Nature Portfolio' als Kind- und 'Macmillan Publishers' als Eltern-Verlag mit einem Gültigkeitszeitraum von 1999 bis 2015.

table field	data
p_relation_id	337
child_name	Nature Portfolio
child_id	P4310319908
child_unit	182
parent_name	Macmillan Publishers
parent_id	P4310319964
parent_unit	38
first_date	01.01.1999
last_date	30.04.2015

5 Wichtigste Positionen des zahlenmäßigen Nachweises

Die Personalkosten dienen der Umsetzung des Arbeitsplans durch qualifiziertes Personal, es wurde nach TVÖD-Bund, Entgeltgruppe 13 vergütet. Die Ausgaben erfolgten in Übereinstimmung mit den Richtlinien des Forschungszentrums Jülich.

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 und Umfang der neuen Datenquelle OpenAlex.

Die im Projekt entwickelten Verfahren zur Datenkuratierung sind öffentlich verfügbar und 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 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. OpenAlex greift selbst die im Vorhaben durchgeführten

Abdeckungsanalysen auf, um die eigene Datenbasis zu verbessern. Die Standardisierung der Verlags- und Zeitschrifteninformationen wird im Open Access Monitor¹⁰ bereits nachgenutzt; die EZB hat Interesse an den kuratierten Verlagsdaten gezeigt. 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 ist erfolgreich verlaufen. Ein Blogpost der Endergebnisse macht 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 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

An den Daten der Verlage und Zeitschriften hat es in OpenAlex keine wesentlichen Veränderungen gegeben. In aktuelleren Datenständen wurde überprüft, ob sich die Anzahl der Publikationen mit und ohne Verlagsinformationen geändert hat und die Aktualisierung der Standardisierung der Verlage Ende 2025 wurde ebenfalls genutzt, um mögliche Veränderungen zu erfassen. Generell ist die Vollständigkeit diesbezüglich in den jüngeren Jahren höher als für ältere Publikationsjahre, aber grundlegende Veränderungen konnten nicht erkannt werden.

Die Datenkuratierung sowohl für die Verlage als auch für die Zeitschriften wird zukünftig im Wesentlichen aus freien Quellen gespeist werden und wird insbesondere im Aufbau auch von manuellen Bearbeitungsschritten unterstützt. Die für die automatisierten Importschritte der Daten und für regelmäßige Aktualisierungen relevanten KBART-Dateien stehen von einer großen Anzahl von Verlagen zur Verfügung. Für zusätzliche Informationen können Daten der ZDB und EZB nachgenutzt werden, die in großem Stil kooperieren und sich um eigene Qualitätsverbesserungen bemühen. In persönlichen Gesprächen insbesondere mit der EZB ist bereits eine Nachnutzung der Verlagsinformationen aus dem KB-Projekt zu den Zeitschriften in der EZB angedacht worden, während die ZDB über die Datenerfassung Unterstützung eines weiteren Projekts aus der GOKB¹¹ anstrebt (z.B. die ISSN-Angaben). Die GOKB kann als Verwaltungsplattform für elektronische Medien genutzt werden und wird ebenso wie die ZDB und die EZB und unter deren Beteiligung kooperativ gepflegt. Zur Verbesserung der Datenqualität und Vervollständigung von Datensätzen erfolgt eine gegenseitige Nachnutzung der Daten. Die genannten Datenbanken verwalten also alle dieselben Datenbereiche mit jeweils etwas anderen Nutzungsanforderungen. Dass die Nutzung sich in erster Linie an deutsche Einrichtungen richtet, ändert nichts an der Tatsache, dass der größte Anteil der Serientitel internationaler Herkunft ist. Für die Verwendung in einer offenen Datenbasis im KB sind beide Aspekte, der Fokus auf Deutschland und die internationalen Titel, ideal. Dass hier mehrere Initiativen mit einer ähnlichen Datenbasis in ihren Vorhaben in dieselbe Richtung gehen ist ein positiver Effekt, der dem KB nutzt, die Nachnutzbarkeit seiner Daten aber auch wahrscheinlich macht.

¹⁰ <https://open-access-monitor.de/>

¹¹ <https://gokb.org/uber-die-gokb/>

9 Erfolgte oder geplante Veröffentlichungen der Projektergebnisse

9.1 Blogposts

Scheidt, Barbara. 2025. "Publisher Normalization for OpenAlex Data." September 3, 2025.
<https://www.open-bibliometrics.de/posts/20250903-PublisherNormalization/>.

9.3 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>

Literatur

Alperin, J. P., Portenoy, J., Demes, K., Larivière, V. & Haustein, S. (2024). An analysis of the suitability of OpenAlex for bibliometric analyses. arXiv. <https://10.48550/arXiv.2404.17663>

Céspedes, L., Kozłowski, D., Pradier, C., Sainte-Marie, M. H., Shokida, N. S., Benz, P., ... Larivière, V. (2025). 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.
<https://doi.org/10.1002/asi.24979>

Delgado-Quirós, L. & Ortega, J. L. (2024). Completeness degree of publication metadata in eight free-access scholarly databases. *Quantitative Science Studies*, 5 (1), 31–49.
https://doi.org/10.1162/qss_a_00286

Maddi, A., Maisonnobe, M. & Boukacem-Zeghmouri, C. (2025). Geographical and disciplinary coverage of open access journals: Openalex, scopus, and wos. *PLOS ONE*, 20 (4), 1-21.
<https://doi.org/10.1371/journal.pone.0320347>

Mongeon, P., Hare, M., Krause, G., Marjoram, R., Riddle, P., Toupin, R. & Wilson, S. (2025). Investigating Document Type Discrepancies between OpenAlex and the Web of Science. *Proceedings of the Annual Conference of CAIS / Actes du congrès annuel de l'ACSI*.
<https://doi.org/10.29173/cais1943>

Simard, M.-A., Basson, I., Hare, M., Larivière, V. & Mongeon, P. (2004). The open access coverage of openalex, socpus and web of science. <https://arxiv.org/abs/2404.01985>

Thelwall, M. & Jiang, X. (2025). Is openalex suitable for research quality evaluation and which citation indicator is best? *Journal of the Association for Information Science and Technology*, 76 (12), 1660-1681. <https://doi.org/10.1002/asi.70020>