

Projektbericht
Juli 2026

Stephan Stahlschmidt

Schlussbericht

„Komparative Analyse und Kuratierung von Metadaten
deutscher Forschungseinrichtungen in Offenen Biblio-
metriedaten: Förderinformationen und Vernetzung“
FKZ 16WIK2301D

Dieses Werk steht unter der Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Projektleitung

Dr. Stephan Stahlschmidt
Telefon +49 30 2064177-18
E-Mail: stahlschmidt@dzhw.eu

Projektmitarbeiter*innen

Nikita Sorgatz
Dimity Stephen
Matteo Ottaviani
Paul Donner
Alexander Schniedermann
Valeria Aman

Impressum

Herausgeber

Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)
Lange Laube 12 | 30159 Hannover | www.dzhw.eu
Postfach 2920 | 30029 Hannover
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

Geschäftsführung

Dr. Marcus Beiner
Regina Oelfke

Vorsitzender des Aufsichtsrats

Ministerialdirigent Peter Greisler

Registergericht

Amtsgericht Hannover | B 210251
Umsatzsteuer-Identifikationsnummer:
DE291239300

Juli 2026

Inhaltsverzeichnis

1	Kurzbericht	2
1.1	Aufgabenstellung	2
1.2	Planung und Ablauf des Vorhabens	2
1.3	Wesentliche Ergebnisse	3
2	Eingehende Darstellung	4
2.1	der Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen, mit Gegenüberstellung der vorgegebenen Ziele	4
2.1.1	Ausgangslage	4
2.1.2	Eingehende Darstellung der Projektaktivitäten	5
2.2	der wichtigsten Positionen des zahlenmäßigen Nachweises	8
2.3	der Notwendigkeit und Angemessenheit der geleisteten Arbeit	8
2.4	des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans	8
2.5	des während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen	9
2.6	der erfolgten oder geplanten Veröffentlichungen des Ergebnisses	9
2.7	Literaturverzeichnis	10

1 Kurzbericht

1.1 Aufgabenstellung

Ziel des Vorhabens war der Aufbau einer offenen Bibliometriedatenbank innerhalb des Kompetenznetzwerks Bibliometrie (KB). Das Teilprojekt am DZHW hatte dabei einen besonderen Fokus auf die Analyse, Kuration und Bereitstellung von Dokumenttypen, Zitationsdaten und Forschungsförderinformationen. 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 an, die Abdeckung, Vollständigkeit und Datenqualität offener Quellen im Vergleich zu den etablierten Datenbanken untersuchte, und konnte dabei auf die langjährige Erfahrung des KB mit bibliometrischen Datenbanken aufbauen.

1.2 Planung und Ablauf des Vorhabens

Das Vorhaben gliederte sich in vier Arbeitsbereiche.

Im ersten Arbeitsbereich wurde OpenAlex in die Dateninfrastruktur des KB integriert und über Google BigQuery öffentlich zugänglich gemacht. Dies umfasste die Entwicklung von Importroutinen, die Erstellung von Datenschemata sowie ein kontinuierliches Monitoring von Schemaänderungen.

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

Im dritten Arbeitsbereich wurden die Erkenntnisse aus dem Datenbankvergleich aufgegriffen und in Kurationsverfahren überführt, mit denen die Datenbasis von OpenAlex im KB optimiert wird. Die kuratierten Ergebnisse wurden in einem Open-Data-Release gebündelt und über das Repositorium 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.

1.3 Wesentliche Ergebnisse

Folgende zentrale Erkenntnisse konnten im Projekt erzielt werden:

- Bereitstellung einer offenen Bibliometriedatenbank im KB. OpenAlex wurde als zusätzliche Datenquelle im KB bereitgestellt und kontinuierlich aktualisiert.
- Datenpublikation. Die im Vorhaben erstellten Kurationen wurden als Teil des verbundübergreifenden OPENBIB-Data-Releases über Zenodo offen bereitgestellt.
- In einem datenbank-übergreifenden Vergleich zu den Dokumententypen und Referenzen konnten wichtige Qualitätsmerkmale von OpenAlex identifiziert werden, aber auch bestehende Desiderate aufgedeckt werden
- Die offenen Indexierungspolitik von OpenAlex erhöht die Sichtbarkeit von deutschen Publikationen um 40%, jedoch indexiert OpenAlex auch Fachzeitschriften von zweifelhafter Qualität (ca 3% predatory journals im Gesamtkatalog)
- Basierend auf ergänzten Zitationsindikatoren konnte mittels eines Cases die direkte Nutzbarkeit von OpenAlex für das bibliometrische Reporting geprüft werden. Für die geprüfte Forschungsinstitution konnten keine substanziellen Unterschiede im bibliometrischen Standardbericht des KB zwischen Web of Science und OpenAlex aufgefunden werden. Dies unterstreicht die bereits bestehende Datenqualität von OpenAlex.
- Im Vergleich zu den proprietären Daten sind Förderinformationen zu zentralen deutschen Forschungsförderern in OpenAlex noch stark verbesserungsfähig. Proprietäre Daten enthalten bis zu 24% mehr Förderinformationen als OpenAlex für zentrale deutsche Forschungsförderer.
- Mittels der Analyse von Schlussberichten zu BMFTR-geförderten Projekten und Prozessdaten der DFG Förderung konnten ergänzende Datenquellen für Förderinformationen erprobt werden.
- Die unternommenen Vernetzungsaktivitäten unterteilen sich in gezielte Ansprachen strategisch relevanter Akteure in Deutschland, der Vernetzung mit internationalen bibliometrischen Dateninfrastrukturen und Disseminationsaktivitäten der Projektergebnisse stießen auf ein großes Echo und zeigten die nationale und internationale Relevanz der Projektergebnisse auf.
- 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 verglichen.

2 Eingehende Darstellung

2.1 der Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen, mit Gegenüberstellung der vorgegebenen Ziele

2.1.1 Ausgangslage

Offene Bibliometriedaten werden zunehmend zum Gegenstand der Wissenschaftsforschung und -politik. Prominentes Beispiel für diese Entwicklung ist OpenAlex, 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 zu besseren Konditionen 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 des DZHW bildeten dabei die Prüfung, Kuration und Bereitstellung von Daten zu Förderinformationen. Darüber hinaus wirkte das DZHW an der vergleichenden Analyse von OpenAlex hinsichtlich Zitationsdaten und -abdeckung mit, die gemeinsam mit weiteren maßgeblich beteiligten Partnern durchgeführt wurde.

Tabelle: Fachliche Arbeitsteilung im Projektverbund

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

2.1.2 Eingehende Darstellung der Projektaktivitäten

Arbeitspaket 2.1.1: Datenbankweiter Vergleich Items/Works

Die komparative analytische Exploration von OpenAlex zielt darauf ab, die Abdeckung und die Datenqualität von OpenAlex zu Projektbeginn zu bestimmen und mit den beiden proprietären Datenbanken Web of Science und Scopus zu vergleichen. Ein solcher Vergleich ist notwendig, da es im Projekt darum geht, nach Möglichkeit eine Alternative zu proprietären Datenbanken zu entwickeln.

Zur Vereinheitlichung entsprechender Analysen durch die OPENBIB Kooperationspartner wurde im Teilprojekt des DZHW mit der SUB Göttingen Kern- und Randpublikationen in OpenAlex und den proprietären Datenbanken identifiziert. Basierend auf diesen Arbeiten wurde gemeinsam mit GESIS und der SUB Göttingen eine Publikation zur Datenqualität von Referenzen in der offenen Bibliometriedatenbank OpenAlex erstellt. Hierbei zeigt sich, dass OpenAlex über eine vergleichbare Anzahl an source-Referenzen wie die etablierten Datenbanken Scopus und Web of Science verfügt. source-Referenzen sind Referenzen, bei denen die Referenz selbst auch in der jeweiligen Datenbank als Publikation enthalten ist. Diese Arbeiten wurde auf einem Workshop des OST Paris unter Beteiligung des CWTS und OST Montréal vorgestellt und diskutiert, bzw. im Projektverlauf in der bibliometrischen Fachzeitschrift Scientometrics veröffentlicht.

Ebenfalls wurde mit den Projektpartnern GESIS und SUB Göttingen eine Analyse und Klassifikation von Dokumententypen in OpenAlex umgesetzt. Dabei zeigte sich, dass bei den Publikationstypen in OpenAlex, Web of Science und Scopus zwar weitgehende Überschneidungen bestehen, bei der Dokumententypenklassifikation 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. Die zugehörigen Daten wurden auf Zenodo publiziert, die zugehörige Publikation erschien in Quantitative Science Studies.

Weiterhin wurde im Teilprojekt des DZHW eine Analyse der im Vergleich zu Web of Science exklusiv in OpenAlex indextierten Publikationen mit einem Fokus auf in Deutschland erstellten Fachbeiträge umgesetzt. Weniger als die Hälfte aller in beiden Datenbanken indextierten Publikationen konnten auch in beiden Datenbanken aufgefunden werden und mehr als die Hälfte aller Publikationen ist nur in einer der Datenbanken indextiert und somit nur partikulär in Analysen sichtbar. Der Großteil dieser erhöhten Sichtbarkeit fällt dabei auf OpenAlex (40% exklusiv in OpenAlex vs. <1% exklusiv im Web of Science), ein substantieller Anteil an Publikationen konnte jedoch aufgrund fehlender Identifier nicht einbezogen werden und mehr als 5% der analysierten Publikationen verfügte nur im Web of Science über eine deutsche Affiliation. Diese Ergebnisse deuten auf eine verbesserte Sichtbarkeit bei gleichzeitig schlechterer Datenqualität für indextierte Publikationen in OpenAlex hin. Die Ergebnisse wurden als Blog-Beitrag auf der Projekt-Website veröffentlicht.

Auch wurden potenziell negative Implikationen der offenen Indexierungspolitik von OpenAlex untersucht und hierfür ein Abgleich der in OpenAlex indextierten Fachzeitschriften mit der proprietären Datenbank Cabells Predatory Reports umgesetzt. Cabells Predatory Report unternimmt Prüfungen von Zeitschriften in Bezug auf unterschiedlichen Kriterien und trifft Aussagen, ob Fachzeitschrift ggf. nicht den wissenschaftlichen Publikationskriterien der Fachdiskurse entsprechen. Mittels eines Stichprobenansatzes konnte die Schnittmenge laut Cabells zweifelhafter Fachzeitschriften auf 6300 bis 8200 in OpenAlex indextierter Zeitschriften abgeschätzt werden. In Bezug auf die Gesamtmenge aller in OpenAlex indextierter Fachzeitschriften entspricht dies einem Anteil von ca. 3%. Die Ergebnisse wurden als Blog-Beitrag auf der Projekt-Website veröffentlicht.

Arbeitspaket 3.1.2: Berechnung feldnormalisierter Zitationsindikatoren

Basierend auf dem durch das DZHW für das KB erstellten Berechnungsroutinen konnte der KB-Hoster FIZ Karlsruhe auch für die OpenAlex Datensnapshots feldnormalisierte Zitationsindikatoren bereitstellen. Diese wurden sowohl innerhalb des Kompetenznetzwerk Bibliometrie, als auch öffentlich über Google Big Query zur Nachnutzung zur Verfügung gestellt.

Basierend auf Bereitstellung der Zitationsindikatoren und den OpenAlex Datenbanksnapshot im KB konnte der im KB entwickelte Standardbericht auf die OpenAlex Daten angewendet werden. In einer vergleichenden Analyse der Standardberichte für OpenAlex und Web of Science für eine Institution der Leibniz-Gemeinschaft konnten relativ ähnliche Ergebnisse in den bibliometrischen Indikatoren abgelesen werden. Hieraus konnte geschlossen werden, dass OpenAlex bereits heute auf der Aggregationsebene von Forschungsinstitutionen ähnliche Informationen zum Web of Science anbietet, dass aber auch die ausgewählte Institution nicht von der erhöhten Sichtbarkeit in OpenAlex profitiert hat.

Arbeitsbereich 3: Förderinformationen

Die Forschungsförderung durch Drittmittel nimmt eine zentrale Funktion im Wissenschaftssystem ein und entsprechend relevant sind Förderinformationen zu Publikationen, da sie Outputs von Förderhandeln direkt für Analysen erfassbar machen.

Die Arbeiten im Arbeitsbereich Förderinformationen wurden durch komparative Analyse der Förderinformationen in der freien OpenAlex Datenbank im Vergleich zu den proprietären Datenbanken Scopus und Web of Science initiiert. Es konnte eine im Vergleich zu proprietären Daten inferiore Abdeckung von Förderkontexten der bekannten deutschen Forschungsförderer in OpenAlex festgestellt werden, welche ergänzende Kuratierung der offenen Daten durch alternative Datenquellen motiviert haben. Zunächst wurden die Förderkontextinformationen in der offenen auf einzelnen Publikationen

zentrierte Datenbank Crossref mit OpenAlex abgeglichen und anschließend die auf Wissenschaftler*innen ausgerichtete Datenbank ORCID als weitere Datenquelle geprüft. Insbesondere Crossref Daten können die Abdeckung von Förderinformationen in Crossref um weitere 5% erhöhen. Ein Blogbeitrag auf der Projektwebsite präsentiert diese Ergebnisse öffentlich.

Um die Abdeckung von Förderinformationen für das deutsche Wissenschaftssystem substanziell zu erhöhen zu können, wurden anschließend ergänzende Datenquellen für Förderkontextinformationen für die beiden wichtigen deutschen Forschungsförderer DFG und BMFTR erprobt. Zunächst wurden Schlussberichte einer Stichprobe BMFTR geförderter Projekte mittels Natural Language Processing Techniken analysiert, um die dort enthaltenen Metadaten zu Projektpublikationen zu extrahieren. Die Ergebnisse und bestehenden Herausforderungen wurden in einem Blogpost auf der Projektwebsite öffentlich geteilt.

Anschließend wurde in Kooperation mit dem Projektpartner FZ Jülich dort vorliegende Daten zu Förderkontexten der DFG extrahiert. Diese basieren auf Anträgen im Rahmen der DFG-Förderung Open Access Publizieren für DFG-geförderte Projekte. Diese Daten ergänzen die in OpenAlex vorhandenen Förderinformationen komplementär und konnten mittels der beiden Open Data Releases des OPENBIB Projektes auch öffentlich zur Nachnutzung bereitgestellt werden.

Arbeitsbereich 4: Vernetzung

Die unternommenen Vernetzungsaktivitäten unterteilen sich in gezielte Ansprachen strategisch relevanter Akteure in Deutschland, der Vernetzung mit internationalen bibliometrischen Dateninfrastrukturen und Disseminationsaktivitäten der Projektergebnisse.

Gezielt wurden dabei insbesondere Akteure der (Selbst-)Steuerung der Wissenschaft angesprochen, da diese über ein strategischen Datenbedarf und entsprechendes Interesse an den Projektaktivitäten verfügen. Zu nennen wären hierbei insbesondere persönliche Treffen mit der DFG und dem European Research Council. Im Bereich der Selbststeuerung wurde mit Fachvertretern in der Medizin und Rechtswissenschaft gesprochen. Für den Bereich Transfer wurde sich mit der universitären Start-Up Unterstützungsinitiative K.I.E.Z. in Berlin ausgetauscht. Durch diese Aktivitäten konnten die betreffenden Stakeholder für das Projekt sensibilisiert werden und der Nutzungsbedarf offener Bibliometriedaten unter diesen Stakeholdern genau erfasst werden.

Neben dem KB gibt es international noch weitere Gruppen, die bibliometrische Dateninfrastrukturen basierend auf proprietären und offenen Bibliometriedaten unterhalten. Zum Erfahrungsaustausch haben sich Projektmitarbeiter*innen des DZHW Teilprojektes daher an verschiedenen Austauschformaten beteiligt. Neben Korrespondenzen per E-Mail kam es hierbei zu einem Austausch mit einem französischen Experten zum politisch motivierten Umschwung weg von proprietären Anbietern und hin zu OpenAlex im französischen Wissenschaftssystem und einem Präsenztreffen mit einer Delegation eines taiwanesischen Projektes zu nationalen Institutionendisambiguierungen. Gemeinsam mit dem CWTS, der Curtin Open Knowledge Initiative, der Humboldt-Universität zu Berlin und Sesame Open Science wurden auch zwei Special Sessions zu Offenen Bibliometriedaten zur Konferenz STI2024 eingereicht. Diese mit anderen Infrastrukturen gemeinsam durchgeführten Sessions wurden auch zur Bekanntmachung der OPENBIB Projektergebnisse genutzt. Stellvertretend für das Gesamtprojekt hat das Teilprojekt am DZHW einen Input aus dem OPENBIB Projekt für eine dieser beiden Sessions gehalten. Weiterhin wurden die Projektergebnisse im KB, aber auch in weiteren Initiativen wie der DACH-Gruppe zur Barcelona Declaration vorgestellt. Für die Wissenschaftsöffentlichkeit wurde eine Website aufgesetzt (www.open-bibliometrics.de), die einerseits über die im Projekt generierten offenen Daten informiert und somit die Nachnutzung unterstützt und andererseits mittels des zuvor erwähnten Blogs einen Einblick in die Projektarbeit ermöglicht und somit auch die Weiterentwicklung des KB hin zu einer offenen Infrastruktur unterstützt.

2.2 der wichtigsten Positionen des zahlenmäßigen Nachweises

Im zahlenmäßigen Nachweis stellen die Personalausgaben die wichtigste Position dar. Diese betragen 90,9 % der Gesamtausgaben. Die Sachkosten betragen 9,1 % der Gesamtausgaben. Der größte Anteil der Sachkosten (100 %) entfällt auf Position 843 für Sonstige allgemeine Verwaltungsausgaben. Weitere Sachausgaben entstanden in Position 846 für Dienstreisen (44,2 %). Die Gesamtausgaben liegen bei 180,3 % der ursprünglich bzw. mit dem ersten Zuwendungsbescheid bewilligten Mittel.

2.3 der Notwendigkeit und Angemessenheit der geleisteten Arbeit

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.

2.4 des voraussichtlichen Nutzens, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen 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. 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: Die zentralen Ergebnisse wurden in einschlägigen, hochrangigen Fachzeitschriften wie *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.

2.5 des während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit gab es umfangreiche Änderungen am Datenbestand seitens OpenAlex. Insbesondere die Klassifikation von Dokumententypen 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, Portenoy, Demes, Larivière und Haustein (2024) durchgeführt, wobei hier auf die unterschiedlichen Klassifikationssysteme in beiden Datenbanken hingewiesen wird. So benutzt OpenAlex den Dokumenttyp Artikel als Auffangbegriff 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 stellten sie ein Verfahren vor, mit dessen fehlerhafte Dokumenttypenklassifikationen identifiziert werden können.

Förderinformationen lagen zur Beantragung des Projektes noch nicht vor (Priem et al., 2022), wurden aber zu Projektbeginn in OpenAlex ergänzt. Diese Entwicklung wurde im Projekt aufgegriffen und haben die vergleichenden Analysen zwischen OpenAlex und proprietären Daten motiviert, bzw. substantielle Desiderate in OpenAlex erkennen lassen.

2.6 der erfolgten oder geplanten Veröffentlichungen des Ergebnisses

Aman, V. (2025). *Comparison of German Publications in OpenAlex and Web of Science*. August 12, 2025. <http://www.open-bibliometrics.de/posts/20250812-GermanPublications/>

Culbert, J., 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, 2475-2492. <https://doi.org/10.1007/s11192-025-05293-3>

Donner, P. (2025). *Prevalence of Predatory Journals in OpenAlex*. June 3, 2025. <http://www.open-bibliometrics.de/posts/20250603-QuestionableJournals/>

Donner, P. (2025). *Towards Open Citation Data from Research Project Reports*. October 30, 2025. <http://www.open-bibliometrics.de/posts/20251030-TowardsOpenCitationData/>

Haupka, N., Culbert, J., Schniedermann, A., Jahn, N., Mayr, P. (2025). 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., 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>

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>

Sorgatz, N. & Schniedermann A. (2025). *Coverage and Quality of OpenAlex Funding Data*. February 25, 2025. <http://www.open-bibliometrics.de/posts/fundingdata/fundingdata-db-comp24.html>

2.7 Literaturverzeichnis

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>

Maisano, D. A., Mastrogiacomo, L., Ferrara, L. & Franceschini, F. (2025). A large-scale semi-automated approach for assessing document-type classification errors in bibliometric databases. *Scientometrics*. <https://doi.org/10.1007/s11192-025-05244-y>

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>

Priem, Jason, Heather Piwowar und Richard Orr (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv*, 16. Juni 2022. <https://doi.org/10.48550/arXiv.2205.01833>.