

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 16WIK2301E gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor:innen.

Verfasst von:

PD Dr. Niels Taubert

Christopher Lenke

Zuwendungsempfänger: Universität Bielefeld, Universitätsstraße 25, 33615 Bielefeld

Förderkennzeichen: 16WIK2301A

Vorhabenbeschreibung: Komparative Analyse und Kuratierung Deutscher Metadaten in Offenen Bibliometriedaten – Teilprojekt Adressdaten

Laufzeit des Vorhabens: 01.07.2023-31.12.2025

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

Teil 1

Kurzbericht

1 Ursprüngliche Aufgabenstellung und Stand der Forschung

Im Rahmen des Kompetenznetzwerks Bibliometrie (KnB) wurde in der Förderphase 2022–2025 der Aufbau einer offenen Bibliometriedatenbank auf Basis von OpenAlex vorangetrieben. Ziel war es, eine transparente, frei nutzbare Alternative zu den proprietären Datenbanken Web of Science (WoS) und Scopus zu schaffen und gleichzeitig Verfahren zur Qualitätsverbesserung bibliometrischer Metadaten zu entwickeln. Ausgangspunkt des Vorhabens waren die Einschränkungen proprietärer Datenbanken hinsichtlich Datenhoheit, Transparenz, Nachvollziehbarkeit und Kosten. Durch die Nutzung offener Datenquellen wie OpenAlex sollten diese Restriktionen überwunden und eine nachhaltige Infrastruktur für die bibliometrische Forschung geschaffen werden.

2 Ablauf des Vorhabens

Das Teilprojekt konzentrierte sich auf zwei Schwerpunkte: den Vergleich der Datenqualität von OpenAlex mit WoS und Scopus sowie die Optimierung von Adress- und Institutionsdaten deutscher Forschungseinrichtungen mittels des Verfahrens der Institutionen-Kodierung.

Die Analysen zeigten, dass die Institutionen-Kodierung erfolgreich auf die Adressdaten von OpenAlex angewendet werden kann. Die erzielten Zuordnungsergebnisse waren teilweise sogar besser als in den proprietären Vergleichsdatenbanken. Zudem verfügt OpenAlex über eine deutlich höhere Anzahl unterschiedlicher Adress-Dokument-Kombinationen und enthält Institutionen, die in WoS und Scopus bislang nicht erfasst sind. Gleichzeitig wurde ein verbleibendes Optimierungspotenzial sichtbar: Rund 12 % der Adress-Dokument-Kombinationen konnten noch nicht eindeutig zugeordnet werden. Im Bereich der Datenkuratierung wurden die ursprünglich geplanten zusätzlichen Datenquellen ORCID und DEAL geprüft, erwiesen sich jedoch aufgrund unzureichender Datenqualität beziehungsweise fehlender Verfügbarkeit als ungeeignet. Stattdessen wurden bisher nicht erfasste Institutionen identifiziert, in die Basistabellen aufgenommen und die Zuordnungsregeln gezielt an die Besonderheiten der OpenAlex-Daten angepasst. Insgesamt wurden Textmuster für 491 Institutionen überarbeitet und zahlreiche neue Einrichtungen integriert.

3 Wesentliche Ergebnisse

Ein wesentliches Projektergebnis war die Überführung der Institutionen-Kodierung für OpenAlex in den Regelbetrieb. Seit 2025 werden die kodierten OpenAlex-Adressdaten regelmäßig für die Partner des Kompetenznetzwerks bereitgestellt. Darüber hinaus wurden die kuratierten Metadaten unter einer offenen Lizenz veröffentlicht. Mit dem Datensatz OPENBIB: Selected curated open metadata based on OpenAlex steht der Forschungscommunity erstmals eine qualitätsgesicherte offene bibliometrische Datenbasis zur Verfügung. Zur Förderung der Nachnutzung wurde außerdem das gesamte Verfahren der Institutionen-Kodierung einschließlich Quellcode, Basistabellen, Zuordnungsregeln und Dokumentation in deutscher und englischer Sprache veröffentlicht. Während der Projektlaufzeit erschienen zwei öffentlich zugängliche Versionen des Verfahrens. Insgesamt zeigt das Projekt, dass offene Bibliometriedatenbanken inzwischen eine leistungsfähige und transparente Alternative zu proprietären Systemen darstellen. Durch die entwickelte Infrastruktur und die etablierten Kuratierungsverfahren wurden wichtige Voraussetzungen für eine nachhaltige, offene bibliometrische Informationsversorgung der deutschen Forschungscommunity geschaffen.

Teil 2

Ausführlicher Bericht

1 Ausgangslage

Im Zuge der fortschreitenden Digitalisierung der Wissenschaft ist eine Reihe frei zugänglicher Datenquellen entstanden, die auch für die bibliometrische Forschung interessant sind. Zu nennen sind hier insbesondere Crossref, OpenAlex, OpenCitations und der OpenAIRE Research Graph. Daneben finden sich in einigen Ländern bereits Projekte, die offene Bibliometriedatenbanken zu verschiedenen Zwecken nutzen. Die fünf antragstellenden Einrichtungen strebten im Rahmen des Kompetenznetzwerks Bibliometrie (KnB) in der Förderphase 2022–2024 an, sich mit offenen Bibliometriedaten zu beschäftigen. Ziel des Vorhabens war es, eine offene Bibliometriedatenbank aufzubauen und ihre Leistungsfähigkeit gegenüber den etablierten Produkten Web of Science (WoS) und Scopus (SCP) zu vergleichen. Zudem sollten Kuratierungsverfahren für ausgewählte Datentypen entwickelt bzw. optimiert und in den Standardbetrieb überführt werden.

Eine solche Beschäftigung war aus vier Gründen geboten: Erstens haben die Erfahrungen mit WoS und Scopus gezeigt, dass die Nutzung der Datenbanken teilweise erheblichen Restriktionen unterliegt. Durch den Aufbau einer offenen Bibliometriedatenbank sollten diese Restriktionen proprietärer Datenbanken vermieden und ein hohes Maß an Datenhoheit realisiert werden. Zweitens war und ist die Nutzung proprietärer Datenbanken im Rahmen des KnB mit erheblichen Lizenzkosten verbunden, deren Höhe von der Anzahl der beteiligten Institutionen und Personen abhängt. Durch den Aufbau und die Kuration einer offenen Bibliometriedatenbank sollten Kosten von der Anzahl der Nutzer:innen der Datenbank entkoppelt werden. Drittens führt die Nutzung proprietärer Datenbanken zu einem Black-Boxing der Daten. Durch die Nutzung einer offenen Bibliometriedatenbank sollte dies vermieden und Transparenz und Nachvollziehbarkeit erhöht werden. Viertens bildet die Kommodifizierung von Daten über die Wissenschaft für kommerzielle Anbieter die Grundlage, um Einfluss auf die Wissenschaft zu gewinnen. Dem sollte mit dem Aufbau einer offenen Bibliometriedatenbank begegnet werden.

2 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Ziel des Vorhabens war der Aufbau einer offenen Bibliometriedatenbank innerhalb des KnB. Die am Verbundvorhaben beteiligten Partner waren im Zuge einer Kartierung dieses Feldes zur gemeinsamen Überzeugung gelangt, dass die Datenbank OpenAlex der Non-Profit-Organisation OurResearch das größte Potenzial besitzt und insbesondere der eigenständigen Zusammenführung unterschiedlicher offener Datenquellen klar vorzuziehen ist. Wesentliche Quellen der Bibliometriedaten sind Crossref und der Microsoft Academic Graph (MAG). OpenAlex zeichnet sich durch ein hohes Maß an Standardisierung aus, indem es einschlägige Identifier umfasst. Die Daten sind in insgesamt fünf Kerntabellen strukturiert und enthalten Informationen zu wissenschaftlichen Veröffentlichungen – mit Schwerpunkt auf Zeitschriftenartikeln – sowie Zitationsdaten, normierte Angaben zu Autor:innen und Einrichtungen und eine Zeitschriftendisambiguierung.

Ein Alleinstellungsmerkmal ist die geplante Weiterentwicklung der MAG-Fachzuordnung auf Artekelebene. Mittelfristig angekündigt hat OpenAlex zudem die Aufnahme von Förderinformationen. Zu beachten ist, dass sich die Datenbank in einem Beta-Status befindet, sodass Änderungen des Datenschemas im Projektverlauf erwartbar waren. Aufgrund der Eigenschaften dieser Datenbank setzte das Vorhaben auf OpenAlex und nutzte die von OurResearch bereits geleistete Arbeit. Der Fokus richtete sich dabei auf die Herstellung der Nutzbarkeit dieser Datenbank innerhalb der deutschen Forschungscommunity sowie die iterative Analyse und Kuratierung der Metadaten von Publikationen, die unter Beteiligung von Autor:innen an deutschen Forschungseinrichtungen entstanden sind.

Um auf der Grundlage von OpenAlex eine offene Bibliometriedatenbank innerhalb des KnB kooperativ zu etablieren, realisierte das Verbundvorhaben drei aufeinander abgestimmte Teilziele:

- **Datenbankbereitstellung:** Bereitstellung einer freien und maschinenlesbaren Entwicklerinstanz der Bibliometriedatenbank OpenAlex als Grundlage für die Kuratierung deutscher Publikationsdaten unter Nutzung einer offenen Lizenz.
- **Datenbankvergleich:** Vergleichende Analyse der Abdeckung (Coverage) und Qualität der offenen Bibliometriedatenbank OpenAlex gegenüber den proprietären Datenbanken WoS und Scopus (Ist-Analyse).
- **Datenkuratierung:** Entwicklung und Anwendung technischer Verfahren zur Kuratierung der Metadaten von Publikationen, die unter Beteiligung von Autor:innen deutscher Forschungseinrichtungen entstanden sind.

Im Zuge der Verlängerung des Projekts im Förderzeitraum 2025 sollten die Voraussetzungen für eine routinemäßige Nutzung der Institutionen-Kodierung für deutsche Adressdaten geschaffen und die Verfahren der Institutionen-Kodierung zur Nachnutzung unter einer offenen Lizenz publiziert werden.

3 Planung und Ablauf des Vorhabens

Während die Datenbankbereitstellung von anderen Partnern des Verbundprojekts geleistet wurde, fokussierte sich die Arbeit dieses Teilprojekts auf den Datenbankvergleich und die Datenkuratierung, insbesondere im Bereich der Adressinformationen. Im Einzelnen ging es darum, die Zuordnung von Adressinformationen deutscher Forschungseinrichtungen in OpenAlex mithilfe des Verfahrens der Institutionen-Kodierung zu erproben, die Zuordnungsergebnisse mit denen proprietärer Datenbanken zu vergleichen und das Verfahren für die Anwendung auf OpenAlex zu optimieren. Zudem sollte das Verfahren in den Regelbetrieb überführt und die Institutionen-Kodierung zur Nachnutzung veröffentlicht werden.

3.1 Arbeitsbereich 1: Datenbankvergleich

Im Fall von Adressinformationen sollte die Institutionen-Kodierung (Rimmert et al., 2017; Donner et al., 2020) auf die in den drei Datenbanken vorliegenden Adressen aus Deutschland angewendet und ein Identifier aus DOI und Institutions-Identifier `kb_inst_id` gebildet werden. Auf dieser Grundlage

sollte die Abdeckung der Adressinformationen in Web of Science, Scopus und OpenAlex analysiert werden.

3.2 Arbeitsbereich 2: Datenkuratierung

Im Fall der Adressdaten und Affiliationen existiert in OpenAlex bereits eine Zuordnung von Publikationen (Works) zu Institutionen. Zu diesem Zweck wertet OpenAlex Adressstrings aus strukturierten (z. B. PubMed) und unstrukturierten (Webseiten) Quellen aus und ordnet sie einer ROR-ID zu. Mithilfe der Institutionen-Kodierung sollte die Zuordnung für die Adressinformationen deutscher Forschungseinrichtungen optimiert werden. Hierzu wurde das Verfahren zunächst an die Spezifika der in OpenAlex vorliegenden Adressinformationen angepasst und angewendet.

Durch die Anpassung oder Ergänzung bestehender Textmuster sollte die Zuordnungsleistung der Institutionen-Kodierung weiter erhöht werden. Zudem sollte geprüft werden, ob durch die Hinzunahme weiterer Datenquellen die Adressinformationen von Autor:innen deutscher Forschungseinrichtungen vervollständigt werden können.

3.3 Überführung der Institutionen-Kodierung in den Regelbetrieb

Zur Überführung der Institutionen-Kodierung in den Regelbetrieb sollte der datenvorbereitende Transformationsalgorithmus an die Besonderheiten des OpenAlex-Adressstrings angepasst und bestehende Transformationsregeln optimiert bzw. neu definiert werden. Daneben sollten Institutionen, die ausschließlich in OpenAlex vorkommen, in die Basistabellen aufgenommen und weitere Textmuster zur Zuordnung ergänzt werden.

Neben den Arbeiten mit qualitätsverbesserndem Charakter sollten im Jahr 2025 analog zu Web of Science und Scopus vierteljährlich Kodierungen der Adressinformationen von OpenAlex erstellt und den KnB-Partnern im Rahmen der aufzubauenden OpenAlex-Bibliometriedatenbank zur Nutzung zur Verfügung gestellt werden.

3.4 Veröffentlichung der Verfahren zur Adressbereinigung

Zur Nachnutzung der Institutionen-Kodierung auch in andere Bereiche, in denen es auf eine Zuordnung von Adress-Informationen zu Forschungseinrichtungen ankommt, war es weiteres Ziel des Projekts, die Verfahren der Institutionen-Kodierung zu veröffentlichen. Dies umfasste die Publikation des gesamten Verfahrens mit sämtlichen Skripten, die dafür notwendigen Basistabellen mit Organisations-bezogenen Informationen sowie die für die Zuordnung notwendigen Textmuster

4 Inhaltliche Ergebnisse

4.1 Datenbankvergleich

Sämtliche Arbeiten im Arbeitsbereich Datenbankvergleich konnten wie geplant innerhalb der Projektlaufzeit abgeschlossen werden. Für die weiteren Projektarbeiten, insbesondere im Bereich der Datenkuratierung, wurden die folgenden Ergebnisse erzielt:

- Erstens konnte im Berichtsjahr demonstriert werden, dass das Verfahren grundsätzlich auf die Adressdaten von OpenAlex anwendbar ist. Die Institutionen-Kodierung stellt somit einen vielversprechenden Ausgangspunkt für die Verbesserung der Qualität deutscher Adressdaten in OpenAlex dar.
- Zweitens zeigte sich, dass das Verfahren der Institutionen-Kodierung bei OpenAlex zu guten Zuordnungsergebnissen führte, die zum Teil besser waren als bei proprietären Datenbanken.
- Drittens weist OpenAlex im Vergleich zu den Datenbanken Web of Science (WoS) und Scopus (SCP) deutlich mehr distinkte Adress-Dokument-Kombinationen auf.
- Viertens wurde eine Reihe von Institutionen identifiziert, die in den Affiliationsdaten von WoS und Scopus bislang nicht enthalten waren.
- Fünftens existiert ein Anteil von ungefähr 12% nicht zugeordneten distinkten Adress-Dokument-Kombinationen, der auf Optimierungspotenzial innerhalb der Institutionen-Kodierung hinweist.

Die Ergebnisse des Datenbankvergleichs wurden in einem Blogpost dokumentiert:

Lenke, C., & Taubert, N. (2025, September 23). Address Information in OpenAlex, Web of Science, and Scopus: First Insights. <https://open-bibliometrics.de/posts/20250923-AddressInsights/>

4.2 Datenkuratierung

Anreicherung von Adressdaten: Der Vergleich von OpenAlex mit Scopus ergab Hinweise darauf, dass Affiliationsdaten in OpenAlex unvollständiger abgebildet werden als in den beiden Vergleichsdatenbanken. Im Projektantrag wurden hierfür zwei potenzielle Datenquellen in Erwägung gezogen: Die im Projekt DEAL abgebildeten Adressdaten (corresponding authors) sowie Adressdaten aus ORCID. Leider erwiesen sich beide Quellen als nicht zielführend.

- Im Fall von ORCID bestätigten sich die bereits geäußerten Zweifel an der Qualität selbstgemeldeter Adressdaten (Baglioni et al. 2021). Diese Bedenken verfestigten sich im Projektverlauf, sodass ORCID zumindest derzeit als wenig geeignete Quelle für Affiliationsdaten erscheint.
- Im Hinblick auf die von DEAL Operations gesammelten Affiliationsdaten wurde uns auf Anfrage mitgeteilt, dass keine Datensätze bereitgestellt werden könnten, die Affiliationen von Autor:innen mit Publikations- und Kostendaten verknüpfen.

Aus diesen Gründen fokussierten sich die Arbeiten auf die folgenden beiden Schwerpunkte:

- **Aufnahme neuer Institutionen:** Zur Verbesserung der Zuordnungsquote wurden bislang nicht bekannte Institutionen identifiziert, in den Basistabellen der Institutionen-Kodierung ergänzt und entsprechende Textmuster zur Erkennung in das Verfahren aufgenommen. Eine Liste der für OpenAlex aufgenommene Institutionen findet sich im Anhang A.
- **Optimierung der Textmuster auf Besonderheiten der OpenAlex-Daten:** Im zweiten Schritt wurden die Textmuster gezielt an die Besonderheiten der Adressdaten in OpenAlex angepasst. Hierzu wurden nicht zugeordnete Adressvarianten entsprechend ihrer Häufigkeit

durchgesehen, und bestehende Textmuster wurden so erweitert, dass sie auch OpenAlex-typische Schreibweisen korrekt zuordnen können. Im Berichtszeitraum wurden Textmuster für 491 Institutionen überarbeitet.

4.3 Überführung der Institutionen-Kodierung in den Regelbetrieb

Im Verlauf des Projekts konnten die Voraussetzungen geschaffen werden, um die Institutionen-Kodierung in den Regelbetrieb zu überführen. In 2025 wurden dazu die folgenden Institutionen-Kodierungen erstellt und den Partnern des KnB über die gemeinsame Datenbank-Infrastruktur am FIZ Karlsruhe zur Verfügung gestellt.

Benennung der Kodierung	Bereitstellung am
oa_b_20250103	22.05.2025
oa_b_20250580	20.08.2025
oa_b_20250711	13.10.2025
oa_b_20260116	08.05.2026

Neben der Bereitstellung der Kodierungsergebnisse im KnB wurden die Projektergebnisse auch unter einer offenen Datenbanklizenz interessierten Nachnutzer:innen jenseits des KnB zur Verfügung gestellt.

Hauptka, 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>

Während der Laufzeit des Projekts wurden zwei Beta-Version bereitgestellt, die Veröffentlichung der ersten regulären Version erfolgte nach Abschluss des Projekts im Februar 2026.

4.4 Veröffentlichung der Institutionen-Kodierung

Wie geplant, wurde die Institutionen-Kodierung mit sämtlichen für die Nachnutzung notwendigen Komponenten für die Veröffentlichung vorbereitet. Neben der Durchsicht des Codes und die Sichtung notwendiger Komponenten wurde eine Dokumentation in deutscher und englischer Sprache verfasst, die Nachnutzer:innen einen eigenständigen Umgang mit dem Verfahren ermöglichen sollen. Diese wurden auf die Gitlab-Instanz der Universität Bielefeld verbracht und dort veröffentlicht. Da das Verfahren der Institutionen-Kodierung fortlaufend verbessert wird, ist eine regelmäßiger Veröffentlichungsrhythmus vorgesehen. Während der Projektlaufzeit wurden zwei Versionen der Institutionen-Kodierung veröffentlicht.

Lenke, C., & Taubert, N. (2024). Institutionen-Kodierung für Adressdaten wissenschaftlicher Publikationen [Dataset]. <https://doi.org/10.4119/unibi/2999367>

Lenke, C., & Taubert, N. (2025). Institutionen-Kodierung für Adressdaten wissenschaftlicher Publikationen, Version 2.0 [Dataset]. <https://doi.org/10.4119/unibi/3006783>

5 Wichtigste Positionen des zahlenmäßigen Nachweises

Die wichtigsten Positionen des zahlenmäßigen Nachweises finden sich im beiliegenden Abschlussnachweis.

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.

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

Im Projekt konnte OpenAlex als zusätzliche Datenquelle im KB etabliert werden, womit eine Alternative zu kostenpflichtigen Angeboten wie Scopus und Web of Science angeboten werden kann. Die im Projekt geleisteten Abdeckungsanalysen legen dabei Stärken und Schwächen des neuen Service offen und können Nutzenden helfen zwischen OpenAlex und proprietären Datenquellen abzuwägen. Gleichzeitig helfen die im Projekt erstellten Datenkurationen dabei, das Angebot von OpenAlex gezielt zu verbessern.

Gleichzeitig erlauben die Veröffentlichung und die Bereitstellung der Verfahren, Textmuster und Stammdaten der Institutionen-Kodierung unter einer offenen Lizenz die Nachnutzung der Institutionen-Kodierung über den engeren Anwendungskontext der Bibliometrie hinaus. Im Förderzeitraum wurden zwei Versionen der Institutionen-Kodierung veröffentlicht (Lenke & Taubert 2024, 2025).

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

Während der Projektlaufzeit gab es umfangreiche Änderungen am Datenbestand seitens OpenAlex. Im Zusammenhang mit der Institutionen-Kodierung zu berichten ist hier insbesondere ein Abfall der Abdeckung von Adressdaten im Zuge des so genannten OpenAlex ‚Walden‘ release (Jahn, 2025). Die Verringerung ist auf die Einführung von Anti-AI-Tools auf der Seite von Wissenschaftsverlagen

zurückzuführen. Dabei waren vom Absinken eine Reihe größerer Verlage betroffen. Die das Scraping von Adress-Informationen durch OpenAlex unterbinden. 40% der Adress-Informationen von Open Alex stammen vom Webseitenscraping von Verlagen (Porter, 2026). Für die zukünftige Beschäftigung mit Adress-Informationen von OpenAlex bedeutet dies, die Entwicklung der Abdeckung von Adressinformationen weiterhin intensive zu beobachten.

9 Erfolgte oder geplante Veröffentlichungen der Projektergebnisse

9.1 Blogposts

Haupka, N., Culbert, J., Donner, P., Jahn, N., Lenke, C., Mayr, P., Meier, A., Mittermaier, B., Scheidt, B., Stahlschmidt, S., & Taubert, N. (2026, February 11). New OPENBIB Data Release.

<https://open-bibliometrics.de/posts/20260211-OpenDataRelease/>

Lenke, C., & Taubert, N. (2025, September 23). Address Information in OpenAlex, Web of Science, and Scopus: First Insights. <https://open-bibliometrics.de/posts/20250923-AddressInsights/>

Haupka, N., Culbert, J., Donner, P., Jahn, N., Lenke, C., Mayr, P., Meier, A., Mittermaier, B., Scheidt, B., Stahlschmidt, S., & Taubert, N. (2025, May 6). Announcing OPENBIB Initial Data Release.

<https://open-bibliometrics.de/posts/20250507-OpenDataRelease/>

9.2 Datenpublikation

Lenke, C., & Taubert, N. (2024). Institutionen-Kodierung für Adressdaten wissenschaftlicher Publikationen [Dataset]. <https://doi.org/10.4119/unibi/2999367>

Lenke, C., & Taubert, N. (2025). Institutionen-Kodierung für Adressdaten wissenschaftlicher Publikationen, Version 2.0 [Dataset]. <https://doi.org/10.4119/unibi/3006783>

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

Donner, P., Rimmert, C., & van Eck, N. J. (2020). Comparing institutional-level bibliometric research performance indicator values based on different affiliation disambiguation systems.

Quantitative Science Studies, 1(1), 150–170. https://doi.org/10.1162/qss_a_00013

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

Porter, S. (2026, February 11). No shortcuts to research information citizenship. *Digital Science*.

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

Rimmert, C., Schwechheimer, H., & Winterhager, M. (2017). Disambiguation of Author Addresses in Bibliometric Databases - Technical Report. Report. <https://pub.uni-bielefeld.de/record/2914944>.

Anhang A: Zur Optimierung von OpenAlex neu aufgenommene Institutionen

- 3M Deutschland GmbH
- 50Hertz Transmission GmbH
- AbsInt Angewandte Informatik GmbH
- Active Fiber Systems GmbH
- Ada Health GmbH
- ADAMA Deutschland GmbH
- ADIDAS AG
- AEG-Telefunken AG
- AESKU.GROUP GmbH
- AGT International
- Altair Engineering GmbH
- Amazon Deutschland
- Amprion GmbH
- Andrologicum München
- Apogenix AG
- Asklepios Medical School GmbH
- attocube systems AG
- Averbis GmbH
- AVL Deutschland GmbH
- Baker Hughes (Deutschland) GmbH
- BasCat - UniCat BASF JointLab
- Bavarian Nordic GmbH
- Bayerische Zentrum für Krebsforschung (BZKF)
- Berghof Foundation Operations gGmbH
- BerlinAnalytix GmbH
- Berlin Cures Holding AG
- Bertelsmann Stiftung
- Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege
- Berufsgenossenschaft Holz und Metall
- Berufsgenossenschaft Nahrungsmittel und Gastgewerbe
- Berufsgenossenschaft Rohstoffe und Chemische Industrie
- Berufskolleg am Wasserturm
- Berufsverband der Frauenärzte e.V. (BVF)
- biobyte solutions GmbH
- BioMed X GmbH
- BioTeSys GmbH
- BRAIN Biotech AG
- Brainlab AG
- Brockmann Consult GmbH
- Bundesarchiv
- Bundessortenamt
- Cadence Design Systems GmbH
- CADFEM Germany GmbH
- Center for Population and Health e.V. - Marburg
- Center for Scalable Data Analytics and Artificial Intelligence
- Charles River Deutschland
- Cherenkov Telescope Array Observatory gGmbH
- CHT Germany GmbH
- Clara Angela Foundation gGmbH
- COATEMA Coating Machinery GmbH
- Continental AG
- CUTANEON - Skin & Hair Innovations GmbH
- Cytel Inc.
- Daiichi Sankyo Europe GmbH
- Daimler Truck AG
- DAK-Gesundheit
- Danfoss Deutschland
- Datev e.G
- DENSO AUTOMOTIVE Deutschland GmbH
- Desitin Arzneimittel GmbH
- Detecon International GmbH
- Deutsche Bank AG
- Deutsche Diabetes Stiftung
- Deutsche Forschungsgemeinschaft e.V.
- Deutsche Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V. (DGAUM)
- Deutsche Gesellschaft für Ernährung e.V.
- Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie
- Deutsche Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e. V. (DGK)
- Deutsche Physikalische Gesellschaft e. V.
- Deutsches Meeresmuseum - Museum für Meereskunde und Fischerei · Aquarium
- Deutsche Welle
- Diamond Materials GmbH & Co. KG
- DNV SE
- Drägerwerk AG & Co. KGaA
- DRK Schmerz-Zentrum Mainz
- Dürr Aktiengesellschaft
- EFCNI - European Foundation for the Care of Newborn Infants
- ELSEVIER Deutschland
- Epilepsiezentrum Kleinwachau gemeinnützige GmbH
- Europäisches Radiochirurgie Centrum München
- European Centre for Medium-Range Weather Forecasts
- European Severe Storms Laboratory (ESSL)
- Experimental Pharmacology & Oncology Berlin-Buch GmbH
- Exzellenzcluster Cellular Stress Responses in Aging-Associated Diseases (CECAD)
- FEAAM GmbH
- Fertility Center Berlin

- Festo Vertrieb GmbH & Co. KG
- Fusion for Energy Germany
- General Electric Deutschland Holding GmbH
- geneXplain GmbH
- Geoverbund ABC/J
- gerinnungszentrum rhein-ruhr
- GKN Powder Metallurgy GMBH
- Global Labor Organization (GLO) gGmbH
- glyXera GmbH
- Google Deutschland
- Hamburgisches Krebsregister
- Heidelberg Engineering GmbH
- Heidelberg Materials AG
- Heliatek GmbH
- Herrenknecht AG
- Hüttenwerke Krupp Mannesmann GmbH
- Hydroisotop GmbH
- IDT Biologika GmbH
- IEEE Germany Section
- IFN Schönow e.V.
- IKF Pneumologie GmbH & Co. KG - Institut für klinische Forschung Pneumologie
- Immatix Biotechnologies GmbH
- itemis AG
- JENOPTIK AG
- Johnson & Johnson GmbH
- kbo-Kinderzentrum München gemeinnützige GmbH
- KHD Humboldt Wedag International AG
- Kinderhospital Osnabrück am Schölerberg
- Klinikum Mutterhaus der Borromäerinnen gGmbH
- Kodak Holding GmbH
- Kompetenznetz Vorhofflimmern e.V.
- Krebsregister Saarland
- Krebsregister Schleswig-Holstein
- KUKA Aktiengesellschaft
- Labor Dr. Wisplinghoff
- Landeskrebsregister NRW gGmbH
- Landschaftsverband Westfalen-Lippe (LWL)
- Lehr- und Versuchsanstalt für Viehhaltung, Hofgut Neumühle
- Leica Microsystems GmbH
- Lenze SE
- LOI Thermprocess GmbH
- LSE Space GmbH
- MAHLE GmbH
- Maschinenfabrik Reinhausen GmbH
- Math2Market GmbH
- Max-Planck-Institut Luxemburg für internationales, europäisches und regulatorisches Verfahrensrecht
- Medical Park SE
- Medcover GmbH
- Medipan GmbH
- Medtronic GmbH
- Messer SE & Co. KGaA
- METOP GmbH
- Microsoft Germany
- Mitsubishi Electric Europe B.V.
- Mitsubishi Heavy Industries EMEA, Ltd.
- Muehlhan Holding GmbH
- Müller-BBM Industry Solutions GmbH
- Museum für Naturkunde Chemnitz
- MVZ genetikum® GmbH
- MVZ HPH Institut für Pathologie und Hämatopathologie GmbH
- MVZ Medizinische Genetik Mainz
- NanoTemper Technologies GmbH
- Nephrologisches Zentrum Villingen-Schwenningen
- Netze BW GmbH
- numares AG
- Nuvisan GmbH
- NVIDIA Deutschland
- NVision Imaging Technologies GmbH
- Ober Scharrer Gruppe GmbH
- PAION
- PharmImage Biomarker Solutions GmbH
- PharmBioTec gGmbH
- PharMetrX
- Phenex Pharmaceuticals AG
- Phoenix Contact GmbH & Co. KG
- Privatpraxis Prof. Jonas und Dr. Panda-Jonas
- Proteomics GmbH
- PTV Planung Transport Verkehr GmbH
- PTW Freiburg GmbH
- Radiologische Allianz eG&R
- Raith GmbH
- R-Biopharm AG
- Rigaku Europe SE
- RUHRGAS AG
- Sacher Lasertechnik GmbH
- Salzgitter AG
- ScaDS.AI
- Schaeffler Technologies AG & Co. KG
- Schneider Electric GmbH
- Science-Consulting in Diabetes GmbH
- Sensitec GmbH
- SGS Holding Deutschland B.V. & Co. KG
- SGS INSTITUT FRESENIUS GmbH
- Shell Deutschland GmbH
- SICK AG
- Signatope GmbH
- SMA Solar Technology AG

- Sozialstiftung Bamberg
- SpaceTech GmbH
- spectrumK GmbH
- SPORTHOPAEDICUM STRAUBING – BERLIN
– REGensburg – MÜNCHEN MVZ-
TRÄGERGESELLSCHAFT GBR
- Stallergenes GmbH
- StatConsult Gesellschaft für klinische und
Versorgungsforschung mbH
- STEAG GmbH
- Steinbeis-Stiftung für
Wirtschaftsförderung (StW)
- Stiftung Deutsche Krebshilfe
- Stiftung Neanderthal Museum
- St. Josef-Stift Sendenhorst
- Sunfire GmbH
- Technische Hochschule Leuna-Merseburg
- Telespazio Germany GmbH
- TenneT TSO GmbH
- TEVA GmbH
- Texas Instruments
- THE CROP TRUST
- Thüringer Institut für Nachhaltigkeit und
Klimaschutz (ThINK)
- Thüringisches Landesamt für
Denkmalpflege und Archäologie
- TOMTEC Imaging Systems GmbH
- TSG ResearchLab gGmbH
- TÜV NORD AG
- TÜV RHEINLAND AG
- TÜV SÜD AG
- UltraFast Innovations GmbH
- Valeo
- Vector Informatik GmbH
- Verein Deutscher Ingenieure e. V. (VDI)
- VivaCell Biotechnology GmbH
- VODAFONE Deutschland
- VON ARDENNE GmbH
- VPIphotonics GmbH
- WIG2 GmbH
- Wiley-VCH GmbH
- Wintershall Dea AG
- Wissenschaftliches Institut der AOK
(WidO)
- Würth Elektronik eiSos GmbH & Co. KG
- X-Fab Germany
- Zalando SE
- Zoologische Gesellschaft Frankfurt von
1858 e.V.