Offene Bibliometriedaten des Kompetenznetzwerks Bibliometrie

Najko Jahn (SUB Göttingen) & Dr. Bernhard Mittermaier (FZ Jülich)

https://www.open-bibliometrics.de/

KOMPETENZ NETZWERK BIBLIOMETRIE

Bundesministerium für Forschung, Technologie und Raumfahrt

About the KB

Data infrastructure for bibliometric research and application in Germany based on the proprietary solutions **Web of Science** and **Scopus** since 2008.

The network includes 25 partners.

The current funding period (2022-25) includes a pilot to move from closed to open data practices by integrating **OpenAlex** (KB OPENBIB).

More info about KB: https://bibliometrie.info/

Aims

✓ Provision of OpenAlex within the KB data infrastructure (FIZ Karlsruhe)

✓ Coverage analyses: Comparing OpenAlex with Web of Science and Scopus

➔ Items and References (DZHW, SUB Göttingen)
➔ Authors (GESIS)
➔ Affiliations from German institutions (Bielefeld University)
➔ Document types (SUB Göttingen)
➔ Funding (DZHW)
➔ Journals and publishers (FZ Jülich)
➔ Open Access and full-texts (SUB Göttingen, GESIS)

✅ Curation and **open data release** with focus on German research landscape

Coverage analyses

Overlap of OpenAlex, Scopus and Web of Science (2015-22)

Shared publications in OpenAlex, Scopus and Web of Science
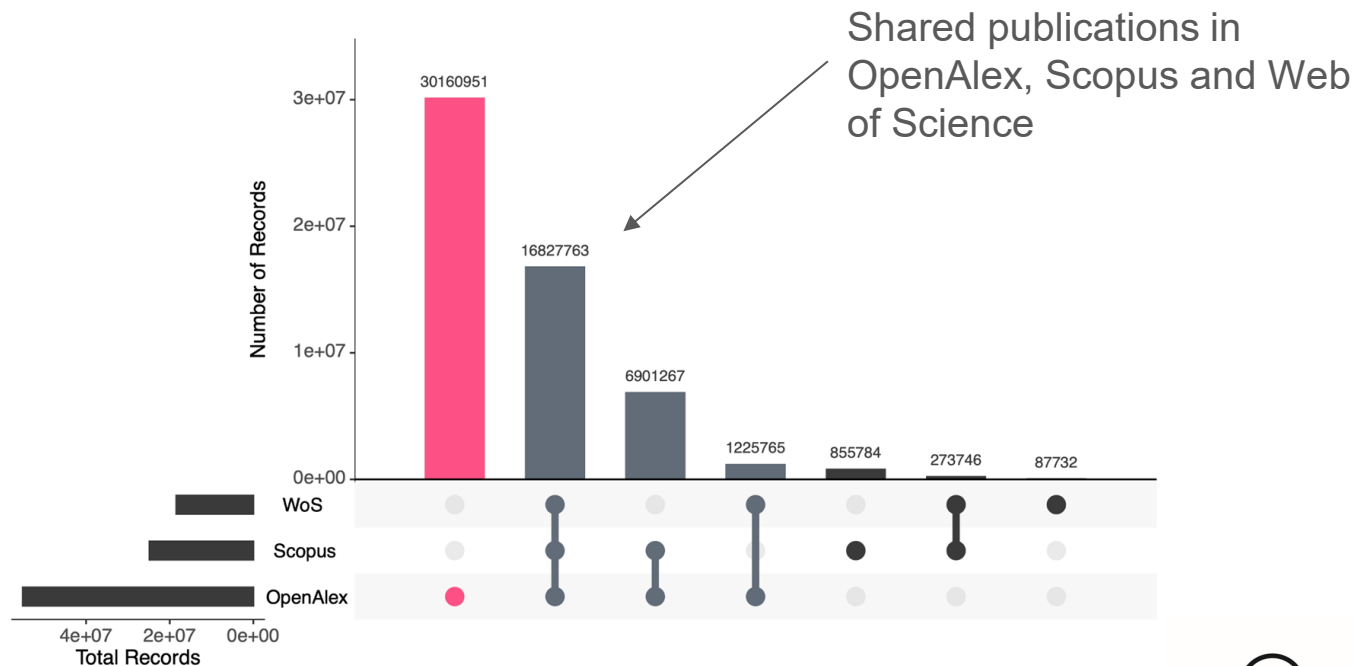
Data from: Culbert, J.H., Hobert, A., Jahn, N. *et al.* Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics* 130, 2475–2492 (2025).
https://doi.org/10.1007/s11192-025-05293-3

https://open-bibliometrics.de/

Curation: Approach

- Use of safeguarded snapshots of OpenAlex, as provided by the FIZ Karlsruhe

- Rectangular data linked by OpenAlex IDs and PIDs (DOI, ROR)

- CC0 by default → <u>No derivation from proprietary data sources possible</u>

- Automatic approaches with quality assurance that can be reproduced

- Different ways to access the data

- Provide computational use cases how to work with the data

Announcing OPENBIB Initial Data Release

DATA

On this page

Announcing OPENBIB Initial
Data Release

Initial data release

How to get it?

What's next?

AUTHORS	AFFILIATIONS
Nick Haupka	State and University Library Göttingen
Jack Culbert	GESIS – Leibniz Institute for the Social Sciences
Paul Donner	German Centrer for Higher Education Research and Science Studies (DZHW)
Najko Jahn	State and University Library Göttingen
Christopher Lenke	Universität Bielefeld
Philipp Mayr	GESIS – Leibniz Institute for the Social Sciences
Andreas Meier	Forschungszentrum Jülich
Bernhard Mittermaier	Forschungszentrum Jülich
Barbara Scheidt	Forschungszentrum Jülich
Stephan Stahlschmidt	German Centrer for Higher Education Research and Science Studies (DZHW)
Niels Taubert	Universität Bielefeld

PUBLISHED
May 6, 2025

Bundesministerium
für Forschung, Technologie
und Raumfahrt

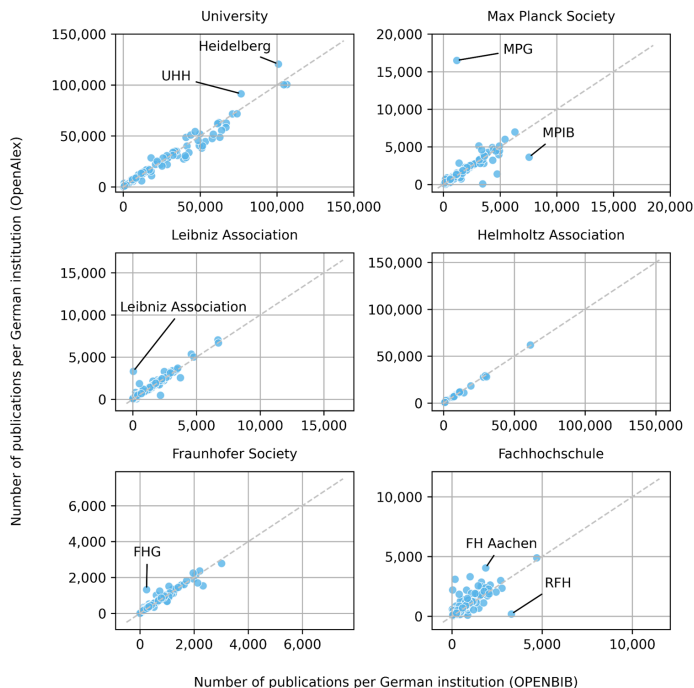# Initial data release: Entities Overview

Address information 🏛 (Uni Bielefeld)

Publishers 📚 (FZ Jülich)

Funding information 📄 (FZ Jülich, DZHW)

Document types 🗂 (SUB Göttingen)

Transformative Agreements 📝 (SUB Göttingen)

KB KOMPETENZ NETZWERK BIBLIOMETRIE

Bundesministerium
für Forschung, Technologie
und Raumfahrt

Initial data release: Address information

Based on the Bielefeld Institutionkodierung.

First time OpenAlex affiliation data are mapped to sectors of Germany's higher education and research landscape.

Starting in 2021, it enhances and quality assures address data of German research institutions.

The ratio of distinct publications to affiliations is nearly the same. For earlier publications, the affiliations in WoS and Scopus are more complete than in OpenAlex.
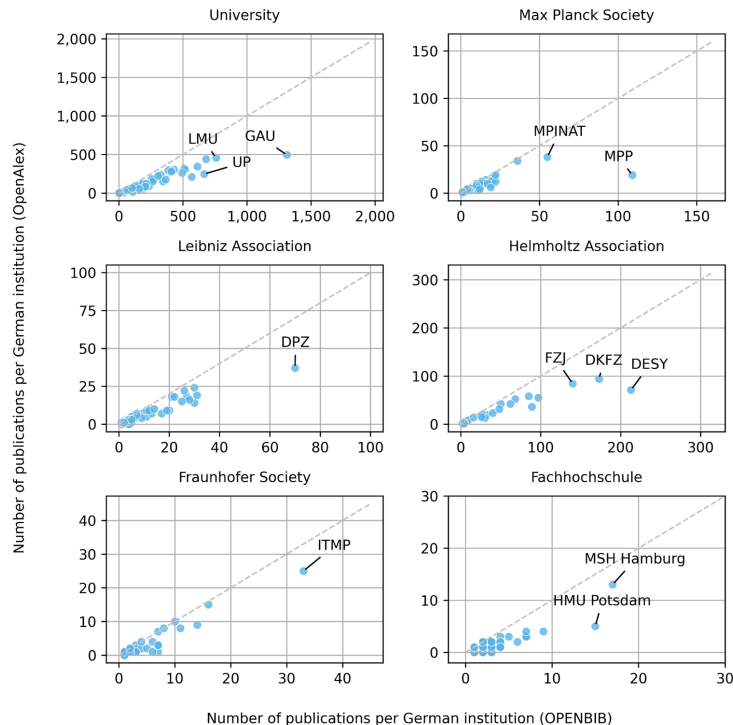
Initial data release: Publishers

publisher_id	publisher_id_orig	publisher_name	standard_name	unit_pk	wikidata	ror	url
13711	https://openalex.org/P4310320330	Adis, Springer Healthcare	Springer Healthcare	249	https://wikidata.org/wiki/Q52636673		https://springerhealthcare.com/
37020	https://openalex.org/P4310320090	Springer Medizin	Springer Medizin	80	https://wikidata.org/wiki/Q29458659		https://www.springermedizin.de/
1099		Springer Nature	Springer Nature	220	https://wikidata.org/wiki/Q21096327	https://ror.org/0117jxy09	https://www.springernature.com/
25525	https://openalex.org/P4310319965	Springer Nature	Springer Nature	220	https://wikidata.org/wiki/Q21096327	https://ror.org/0117jxy09	https://www.springernature.com/
32820	https://openalex.org/P4310320108	Springer Nature (Netherlands)	Springer Nature (Netherlands)	263	https://wikidata.org/wiki/Q30338262	https://ror.org/01ebjwm20	https://group.springernature.com/
35162	https://openalex.org/P4310319972	Springer International Publishing	Springer Nature (Switzerland)	250	https://wikidata.org/wiki/Q21820634	https://ror.org/03v60jr94	https://www.springer.com/de
20506		Springer Science and Business Media LLC	Springer Science+Business Media	79	https://wikidata.org/wiki/Q176916		https://www.springer.com/
36249	https://openalex.org/P4310319900	Springer Science+Business Media	Springer Science+Business Media	79	https://wikidata.org/wiki/Q176916		https://www.springer.com/
10756	https://openalex.org/P4310319986	Springer VS	Springer VS	275	https://wikidata.org/wiki/Q2313945		https://www.springer.com/de/springervs

Initial situation in OpenAlex: Highly standardized names for publishers available, along with data records containing a publisher ID, but issues persists.

Aim: bring together what belongs together in order to simplify evaluations at publisher level (eg parent-child relations, name variants, deceased imprints, …)

Necessary: Identify publishers and their spelling variants and merge them under a standard name

Bundesministerium
für Forschung, Technologie
und Raumfahrt

KKB KOMPETENZ NETZWERK BIBLIOMETRIE

Initial data release: Funding data

Using data about financing open access publications from the DFG "Open-Access-Publikationskosten" program as a proxy to identify DFG funding

Rather small, but research information of high quality provided by the Open Access Monitor!

Show-case for the integration of more funding data from German research funders!

https://open-bibliometrics.de/

Initial data release: Document types

Aim: Improve the identification of original research contributions like research articles

Method: Machine Learning Classifier trained on data from PubMed and Crossref

Data: The dataset includes all articles and reviews from journals in OpenAlex (Snapshot: 08/2024) with publication years 2014 to 2024
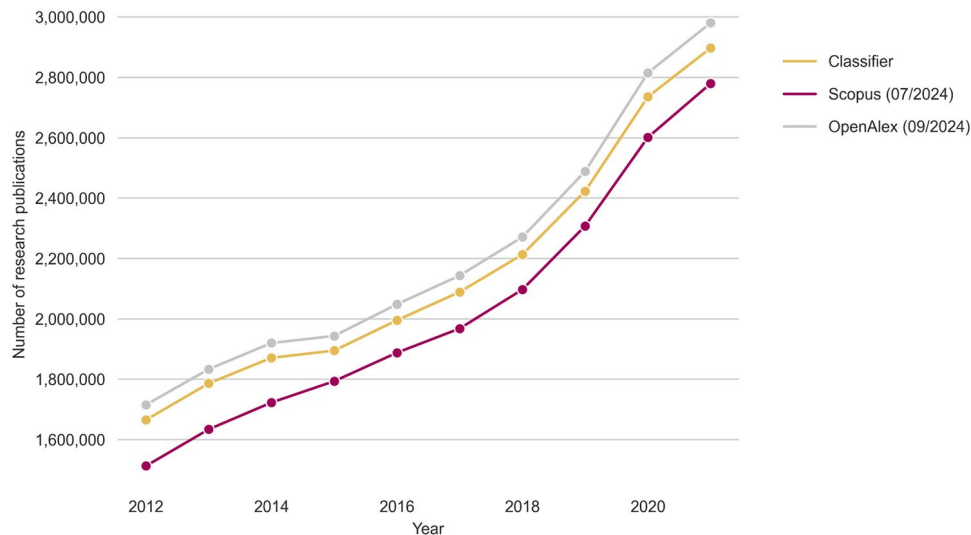
Blog Post:
https://subugoe.github.io/scholcomm_analytics/posts/oal_document_types_classifier/

Initial data release: Transformative Agreements

Growth of Transformative Agreements
Annual first-author articles from participating institutions

Legend: Other, Sage, Taylor & Francis, Wiley, Springer Nature, Elsevier
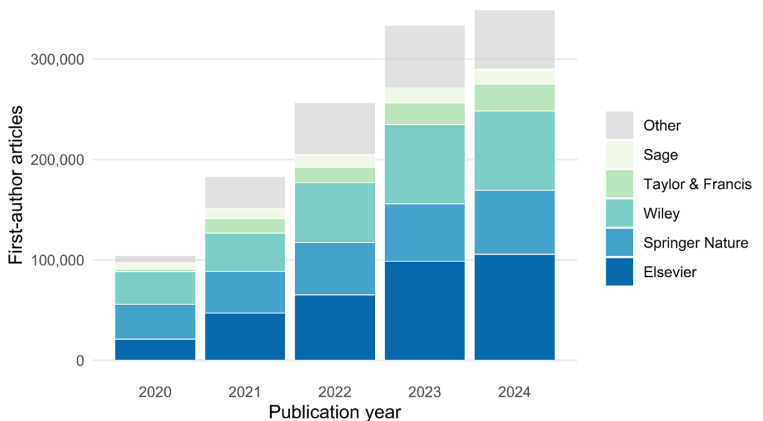
KBOPENBIB | April 2025 | CC0

Aim: Safeguard and enrich data about open access agreements

Data Sources: ESAC Registry, cOAlition S Transformative Agreement Dump, ISSN Agency and OpenAlex

The dataset comprises:

- 1,043 transformative agreements covering 17,199 eligible journals and 5,355 participating institutions
- 1,362,221 eligible articles published by first authors affiliated with participating institutions

Coverage analysis:

Jahn, N. (2025). *Estimating Transformative Agreement Impact on Hybrid Open Access: A Comparative Large-Scale* Study Using Scopus, Web of Science and Open Metadata.
https://arxiv.org/abs/2504.15038

https://open-bibliometrics.de/

Initial Data Release: Access

Website: https://open-bibliometrics.de

Announcement: https://open-bibliometrics.de/posts/20250507-OpenDataRelease/

Data provision:

– KB data infrastructure FIZ Karlsruhe `kb_project_openbib`

– Zenodo: https://zenodo.org/records/15308680

– Google Cloud BigQuery:
https://console.cloud.google.com/bigquery?ws=!1m4!1m3!3m2!1ssubugoe-collaborative!2sopenbib

Notebooks: https://github.com/kbopenbib/kbopenbib_data

Outlook

Until end of 2025

🔧 Improve automatic data curation routines

🔧 Quality assurance

Starting from 2026

🔧 Continuous curation of selected entities

🥳 OpenAlex (and friends) as a core data pillar to opening up the

KB!

https://open-bibliometrics.de/

KB KOMPETENZ NETZWERK BIBLIOMETRIE

Bundesministerium
für Forschung, Technologie
und Raumfahrt

Thank you and follow-ups!

KB OPENBIB partners