

OpenAlex im Kompetenznetzwerk Bibliometrie

Niels Taubert

Vortrag im Rahmen des Kolloquiums der Universitätsbibliothek,
Freitag, 03. Mai 2024



Inhalt

1. Rückblick: Proprietäre Bibliometriedatenbanken
2. Unzulänglichkeiten und Probleme proprietärer Bibliometriedatenbanken
3. OpenAlex
4. Was ist indexiert?
5. OpenAlex im Kompetenznetzwerk Bibliometrie
6. Stand und Erkenntnisse
7. Ausblick

1. Rückblick: Proprietäre Bibliometriedatenbanken

Ursprünge

Web of Science

- Gegründet im Jahr 1945 von Eugene Garfield
- Informationswissenschaftler und Geschäftsmann
- Entwicklung: Science Citation Index (ab 1945), Social Science Citation Index (ab 1956), Arts and Humanities Index (ab 1975), Conference Proceedings Index (ab 1990), Book Citation Index (ab 2005)
- Eigentümer: Institute for Science Information (bis 1992), Thompson/Reuters (bis 2016), Clarivate (bis dato)

Scopus

- Gegründet im Jahr 2004,
- Produkt des Verlags Elsevier
- ScienceDirect Metadaten als eine Datenbasis
- Konkurrenz zwischen Scopus und Web of Science

1. Rückblick: Proprietäre Bibliometriedatenbanken

Verwendungszweck 1: Recherche-Tool

MERTON RK		VOL	PG	YR	
JACOBS M	SOCIAL RES	36	606	69	
56 SOCIAL THEORY SOCIAL		393			
DOREIAN P	SOCIOL REV	17	47	69	
57 AM SOCIOLOGICAL REV		22	635		
DOWNNEY KJ	SOCIOL Q	10	438	69	
LEMAINE G	REV FR SOC	10	139	69	
OROMANER MJ	AM SOCIOL	4	332	69	
57 BRIT J SOCIOLOGY		110			
IRVING HH	J EDUC SOC	5	49	69	
57 SOCIAL STRUCTURE SOC		14	21	69	
MCKELVEY WW	ADM SCI QUA				
57 SOCIAL THEORY SOCIAL		53	118	69	
ACKERMAN NW	MENTAL HYG	N	30	733	69
ADLER N	Q J STUD AL		42	368	69
BACCHUS MK	SOCIOL EDUC	N	3	417	69
BELL C	SOCIOLOGY		48	213	69
BUCHER R	SOCIAL FORC		17	234	69
CAHALAN D	SOCIAL PROB		25	775	69
COHEN HA	PSYCHOL REP		10	16	69
COLOMBOT.J	J HEALTH SO		115	133	69
DAVOL SH	J GENET PSY		34	922	69
DENZIN NK	AM SOCIOL R		10	438	69
DOWNNEY KJ	SOCIOL Q		11	187	69
ELAZAR DJ	JEW J SOCIO		10	30	69
ENGEL GV	J HEALTH SO		10	314	69
FABREGA H	COMP PSYCHI	R	1969	1	69
FLECHTER C	SOCIOL TRAV		2	559	69
GAINES R	J LEARN DI		25	121	69
GOLD M	J SOC ISSUE		48	176	69
GORDON M	SOCIAL FORC		39	1	69
GROCHLA E	Z BETRIEBSW		11	243	69
HANNA JL	ANTHROPOLOG		21	513	69
HARSANYI JC	WORLD POLIT		17	189	69
JENSEN GF	SOCIAL PROB		75	389	69
KADUSHIN C	AM J SOCIOL		37	468	69
KLEIN EB	J PERSONAL		13	104	69
KLEINER RJ	AM BEHAV SC		4	236	69
KOBLEN AJF	MAN		50	136	69
KUNKEL JH	SOCIAL SCI		14	58	69
KURZMAN PA	SOCIAL WORK				

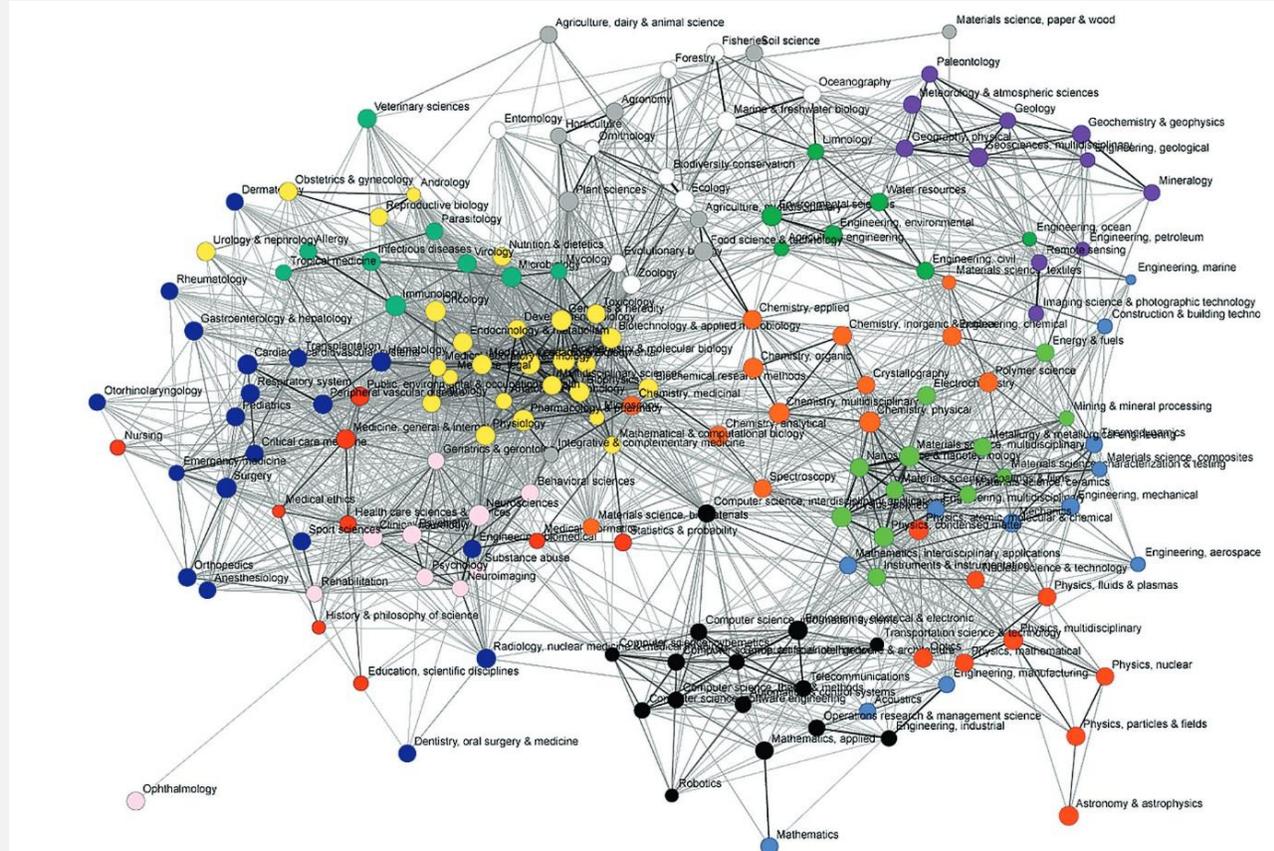
“It is best described as an **association-of-ideas index**, and it gives the reader as much leeway as he requires.” (Garfield 1955: 108)

Merton, Robert K. 1957: *Social Theory and Social Structure*. New York: The Free Press.

1. Rückblick: Proprietäre Bibliometriedatenbanken

Verwendungszweck 2: Forschungsinstrument

„Bounds of
disciplines“



Leydesdorff, Loet; Rafols, Ismael 2009: A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology*, 60(2), 348-362.

1. Rückblick: Proprietäre Bibliometriedatenbanken

Verwendungszweck 3: Unterstützung der Literaturerwerbung

INFORMATION SCIENCE & LIBRARY SCIE... ✕
 ESCI ✕
 JCR Year: 2022 ✕

Journal name ▼	ISSN	eISSN	Category	Total Citations ▼	2022 JIF ▼	JIF Quartile	2022 JCI ▼	% of OA Gold ▼
<input type="checkbox"/> Quantitative Science Studies	N/A	2641-3337	INFORMATION SCIENCE & LIBRARY SCIENCE - ESCI	968	6.4	N/A	1.60	99.04 %
<input type="checkbox"/> Bottom Line	0888-045X	2054-1724	INFORMATION SCIENCE & LIBRARY SCIENCE - ESCI	392	4.1	N/A	1.00	5.13 %
<input type="checkbox"/> Publications	N/A	2304-6775	INFORMATION SCIENCE & LIBRARY SCIENCE - ESCI	2,024	3.8	N/A	0.84	100.00 %
<input type="checkbox"/> Information and Learning Sciences	2398-5348	2398-5356	INFORMATION SCIENCE & LIBRARY SCIENCE - ESCI	593	3.4	N/A	0.77	4.38 %
<input type="checkbox"/> VINE Journal of Information and Knowledge Management Systems	2059-5891	1474-1032	INFORMATION SCIENCE & LIBRARY SCIENCE - ESCI	942	2.8	N/A	0.66	2.18 %

1. Rückblick: Proprietäre Bibliometriedatenbanken

Verwendungszwecke 4: Evaluative Bibliometrie

“In this regard, the SCI data base is being used to do such things as evaluate the research role of individual journals, scientists, organizations, and communities; refine the relationship between journals and between journals and fields of study; measure the impact of current research” (Garfield 1979: 62)

2. Unzulänglichkeiten und Probleme proprietärer Bibliometriedatenbanken

Unzulänglichkeiten

- Lizenzierung der Datenbanken ist teuer
- Datenqualität lässt oftmals zu wünschen übrig
- Datenbanken sind selektiv, es gibt Pfadabhängigkeiten, bedingt durch ihre Historie

Probleme

- Auswirkungen der Lizenzbestimmungen: Rohdaten dürfen nicht weitergegeben werden
 - Bibliometrische Forschung: Korrekte Durchführung statistischer/bibliometrischer Verfahren kann im peer review nicht überprüft werden
 - Evaluative Bibliometrie: Datensammlung und Auswertung kann durch die evaluierten Einheiten nicht geprüft werden
- Kontrolle der indexierten/nicht-indexierten Publikationsmedien durch Datenbankanbieter
- Kontrolle des Feldes für Dienstleistungen durch Datenbankanbieter
- Beschränkte Möglichkeit der Zusammenarbeit (Datenbereinigung, Verbesserung der Datenqualität)
- ‚Kultur der Geschlossenheit‘ in den Bibliometrie-

3. OpenAlex

Relevanz von OpenAlex

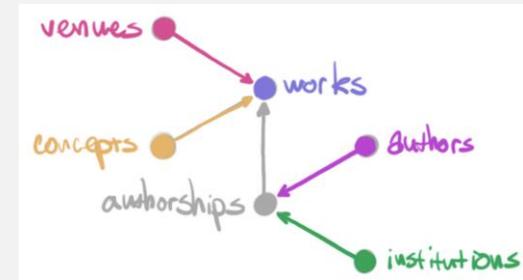
- **Coalition for Advancing Research Assessment (CoARA)** betont, dass offene Daten und Verfahren für künftige Forschungsevaluationen unerlässlich sind
- **Ministère de l'Enseignement supérieur et de la Recherche (MESR)** fördert den Aufbau einer offenen Bibliometriedatenbank auf Grundlage von OpenAlex
- In Frankreich haben zudem die **Sorbonne Université** und das **CNRS** ihre Lizenzen für proprietäre Bibliometriedatenbanken mit Verweis auf OpenAlex nicht verlängert.
- **CWTS** (Einrichtung für bibliometrische Forschung in den Niederlanden) hat OpenAlex Ende Januar 2024 in das **Leiden Ranking** integriert.

3. OpenAlex

OpenAlex ...

- ... basiert auf Microsoft Academic Graph (MAG) sowie Crossref als zentrale Datenquellen. Andere Datenquellen werden bei Bedarf hinzugezogen
- ... wird seit 2021 von OurResearch entwickelt, hat von der Arcadia Foundation eine Anschubfinanzierung von 4,5 Mio. US Dollar erhalten
- ... ist mit einer Creative Commons CC0-Lizenz versehen
- OurResearch ist ein Unternehmen, dass nach den „Principles of Open Scholarly Infrastructures“ (POSI) abreitet, darunter fallen auch die Prinzipien von Open Source, Open Data und Available Data

4. OpenAlex – Was ist indexiert?



- **Works:** Publikationen aller Art (z.B. Zeitschriftenartikel, Bücher, Forschungsdaten usw.). Zitationsbeziehungen zwischen Works werden in der DB abgebildet.
- **Authors:** Personen (derzeit 217 mio.) die an einen Werk beteiligt sind. Daten basieren auf MAG, Crossref, PubMed, ORCID, ein Algorithmus soll für die Disambiguierung sorgen.
- **Sources (ehem. Venues):** Orte, an denen Werke erscheinen (Journale und Repositorien). Derzeit sind in OpenAlex 249,000 Publikationsorte .
- **Institutions** Universitäten und außeruniversitäre Forschungseinrichtungen, aus den Affiliationsangaben der Autoren. Derzeit erfasst OpenAlex 109,000 Institutionen mit ROR ID
- **Topics:** OpenAlex ordnet Werke auf der Grundlage verfügbarer Metadaten bestimmten Topics (Gegenständen) zu. Diese Fachklassifikation besitzt vier Ebenen Topics (4.500), Subfields (252), Fields (26) und Domains (4)
- **Publisher:** Unternehmen und andere Organisationen, die Werke veröffentlichen. Derzeit finden sich in OpenAlex 10.000 Publisher (Wikidata ID).
- **Funder:** Organisationen, die Forschung finanzieren. Derzeit finden sich in OpenAlex insgesamt 32.000 Funder, Daten warden aus Crossref, Wikidata und ROR bezogen.

5. OpenAlex im Kompetenznetzwerk Bibliometrie

Kompetenznetzwerk Bibliometrie

- Existiert (incl. Vorläuferprojekt) seit 2009
- Finanziert vom BMBF
- Bereitstellung eines Zugangs zu den Bibliometriedatenbanken WoS und Scopus für 22 Mitgliedsorganisationen

KB-Projekt ‚Offene Bibliometriedaten‘

- Verbund unter Beteiligung der folgenden Partner:
 - FIZ Karlsruhe: Datenbereitstellung
 - DZHW: Abdeckungsanalyse Werke & Zitationen, Dokumenttypen, Förderinformationen
 - FZ Jülich: Zeitschriften, Verlage
 - GESIS: Autorennamen
 - SUB Göttingen: Abdeckungsanalyse Works & Zitationen, OA-Nachweise
 - Uni Bielefeld: Affiliationen

5. OpenAlex im Kompetenznetzwerk Bibliometrie

Ziel von KBOPENBIB

„Ziel des Vorhabens ist der Aufbau einer offenen Bibliometriedatenbank innerhalb des Kompetenznetzwerks Bibliometrie. Dadurch wird die Möglichkeit für die Wissenschafts- und Hochschulforschung geschaffen, innovative und offene Datenquellen als Alternative zu den proprietären Bibliometriedatenbanken zu nutzen und zudem die Abhängigkeit des Kompetenznetzwerks Bibliometrie von kommerziellen Anbietern verringert. Zugleich verspricht der Aufbau einer frei zugänglichen Bibliometriedatenbank ein eigenständiges Analysepotential bezogen auf solche Publikationsorte und -weisen, die in den proprietären Daten nicht abgedeckt sind.“

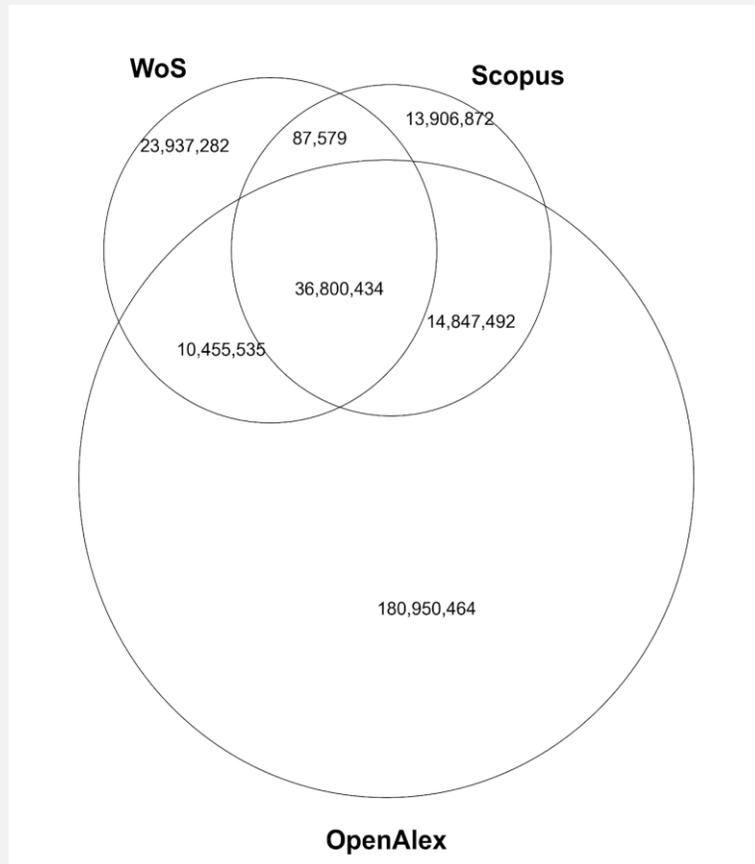
(Zielsetzung laut Projektantrag)

Zweischrittiges Vorgehen

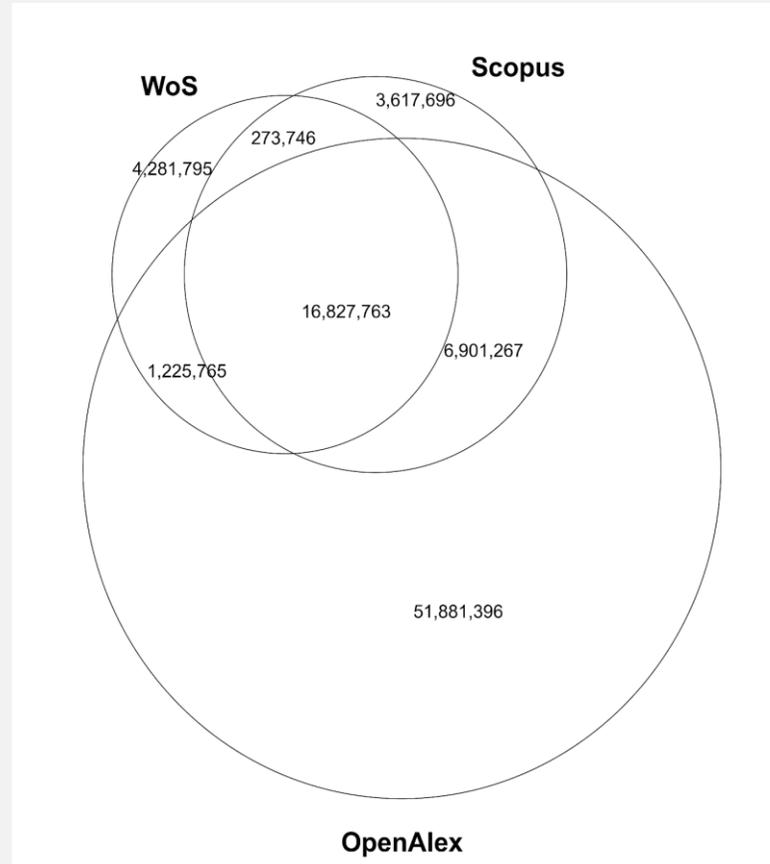
- Abdeckungsanalyse
- Entwicklung von Kuratierungsverfahren für Publikationen unter Beteiligung deutscher Institutionen

6. Stand und Erkenntnisse

Abdeckungsanalyse: Works und References



Gesamtkorpus



2015-2022

Befunde

- Die Anzahl der Werke sind im Fall von OpenAlex wesentlich höher bei den beiden proprietären Datenbanken.
- Die meisten der im WoS und Scopus indexierten Dokumente finden sich auch in OpenAlex
- Dies gilt insbesondere für rezente Jahrgänge

6. Stand und Erkenntnisse

Abdeckungsanalyse: Works und References

	WoS	Scopus	OpenAlex
<i>Whole Corpus</i>			
Number of Records	71,280,830	65,642,377	243,053,925
Number of References	1,765,281,799	2,033,522,623	1,845,379,285
<i>Whole Corpus - Articles Only</i>			
Number of Records	42,678,632	43,579,595	200,665,940
Number of References	1,400,958,343	1,422,650,789	1,636,497,394
<i>Published 2015-2022</i>			
Number of Records	22,609,069	27,620,472	76,836,191
Number of References	786,437,547	1,035,750,923	840,730,834
<i>Shared Corpus (2015-2022)</i>			
Number of Records	16,788,282	16,788,282	16,788,282
Number of References	725,008,043	727,056,725	585,616,069

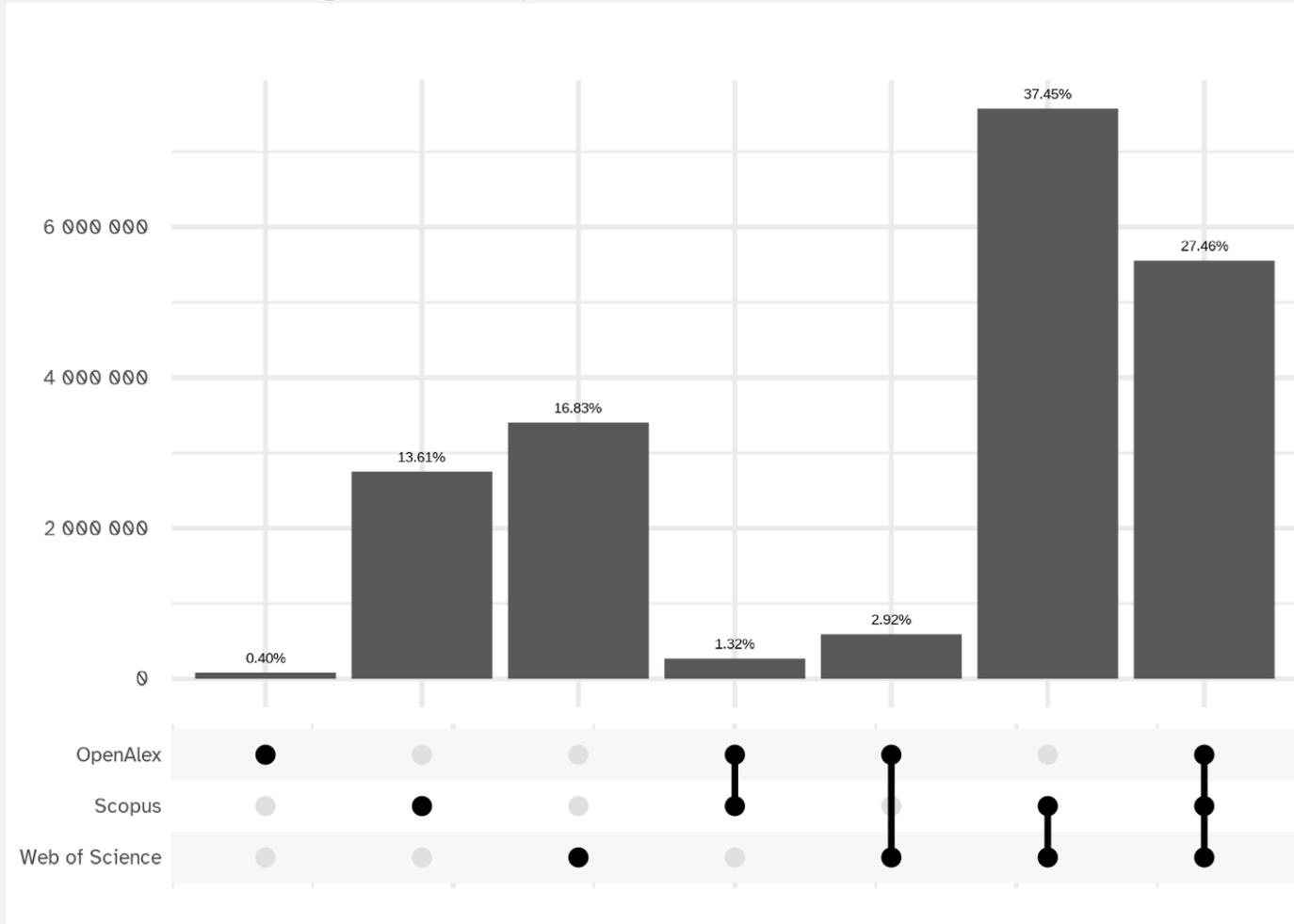
Table 1 Sizes of databases and of the Shared Corpus dataset, with the number of references in each

Befunde

- Das Verhältnis von Works zu den in ihnen vorhandenen Referenzen bewegt sich im Fall des WoS und Scopus in ähnlichen Größenordnungen, ist im Fall von OpenAlex aber deutlich geringer.
- Dies gilt auch für rezente Jahrgänge.
- Im Fall der Core-Publikationen ist das Verhältnis von Works zu Referenzen bei OpenAlex weniger ungünstig, es fehlen aber immer noch eine Vielzahl von Referenzen

5. Stand und Erkenntnisse

Abdeckungsanalyse: Förderinformationen



Befunde

OpenAlex: Für lediglich 32% der Core-Publikationen sind Förderinformationen verfügbar.

Wos: Für mehr als 84% der Core-Publikationen sind Förderinformationen verfügbar.

Scopus: für fast 80% der Core-Publikationen sind Förderinformationen verfügbar.

6. Stand und Erkenntnisse

Abdeckungsanalyse: Affiliationen (Adressdaten aus Deutschland)

	WoS	Scopus	OpenAlex
Anzahl distinkter Adressen	2.269.047	3.449.043	3.955.516
Zuordnungsquote	85,58%	84,7%	89,91%
Anzahl distinkter Adress-Dokument-Kombinationen	7.393.265	6.973.092	7.495.858
Zuordnungsquote	93,16%	88,53%	89,46%
Zuordnungsquote 2006-2015	93,98%	88,58%	89,79%

Befunde

- Verhältnis von distinkten Adressen und Adress-Dokument-Kombinationen ist im Fall von OpenAlex niedriger. Dies ist möglicherweise ein Hinweis darauf, dass nicht alle existierenden Dokument-Adress-Kombinationen abgebildet werden.
- First und Last Address sind eindeutig, die Reihenfolgen dazwischen stehender Adresse nicht (alle ,middle).
- Der Zuordnungsalgorithmus von OpenAlex führt zu Zuordnungsfehlern. Der Umfang wird derzeit ermittelt.

6. Ausblick

- OurResearch hat es in erstaunlich kurzer Zeit geschafft, eine Alternative zu proprietären Bibliometriedatenbanken zu entwickeln.
- Der Vergleich mit proprietären Datenbanken zeigt, dass sich OpenAlex zum derzeitigen noch nicht für einen professionellen Einsatz eignet.
- Angesichts der Dynamik, mit der sich die Datenbank derzeit verbessert, ist mit einer Veränderung dieser Situation zu rechnen.
- Im Projekt 'Offene Bibliometriedaten' wird damit gerechnet, dass OpenAlex ab 2026 eine der beiden proprietären Datenbanken im KB ablösen kann.
- Mit der Umstellung auf eine offene Datenbank ist die Hoffnung verbunden, dass eine Reihe der angesprochenen strukturellen Probleme abgestellt werden kann (stärkere Nachvollziehbarkeit, Überprüfbarkeit, Kultur der Offenheit und kollaborative Weiterentwicklung der Datenbank).