

# Introduction to Bibliometric Data Sources

8/9/25 - ESSS (Granada)  
Elvira González-Salmón

# What will we learn?

1. Main bibliometric databases
2. In depth
  - a. WoS-CC
  - b. Scopus
  - c. OpenAlex
3. How to critically decide which database is more suitable

# 1. Main bibliometric databases

# Main bibliometric databases

Clarivate™



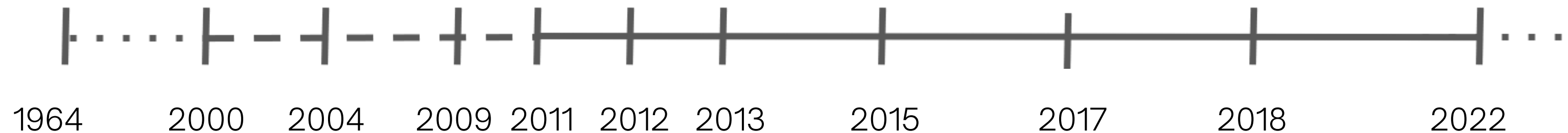
Google Scholar

Scopus®

ORCID



OpenAlex



And there are many more, such as PubMed Central (1996), Lens.org (2000), CrossRef (2000)...

Thanks to the availability of these databases, bibliometric analysis is now **accessible** to a wider audience, enabling individuals to carry out such work from home and to select the database that best aligns with their **specific needs**.



However, there is also a degree of danger in the use of certain metrics, particularly when they are applied without a **proper understanding** of their meaning or the implications they carry.



## 2. In depth...

# We will focus on...



The ones we will use on the exercises

- Indexing and coverage
- Metadata
- Data retrieval process
- Products (citations, journals)

09.45 – 10.30 **Accuracy and Beyond: Data Quality of Bibliometric Databases**  
Stephan Stahlschmidt, *German Centre for Higher Education Research and Science Studies (DZHW), Germany*

10.30 – 11.00 **Coffee break**

11.00 – 11.45 **Designing Effective Queries for Document Retrieval**  
Stephan Gauch, *Humboldt Universität zu Berlin, Germany*

11.45 – 12.30 **Data Cleaning & Integration**  
Cinzia Daraio, *Sapienza University of Rome, Italy*

# In a nutshell



Traditional bibliometric data source	Traditional bibliometric data source	Non-Traditional bibliometric data source
Requires subscription	Requires subscription	Free and Open Access
(Editorial) Quality control	(Editorial) Quality control	Limited quality control
Different download options but API not well-documented	Different download options	Different download options

# Access granted by ESSS

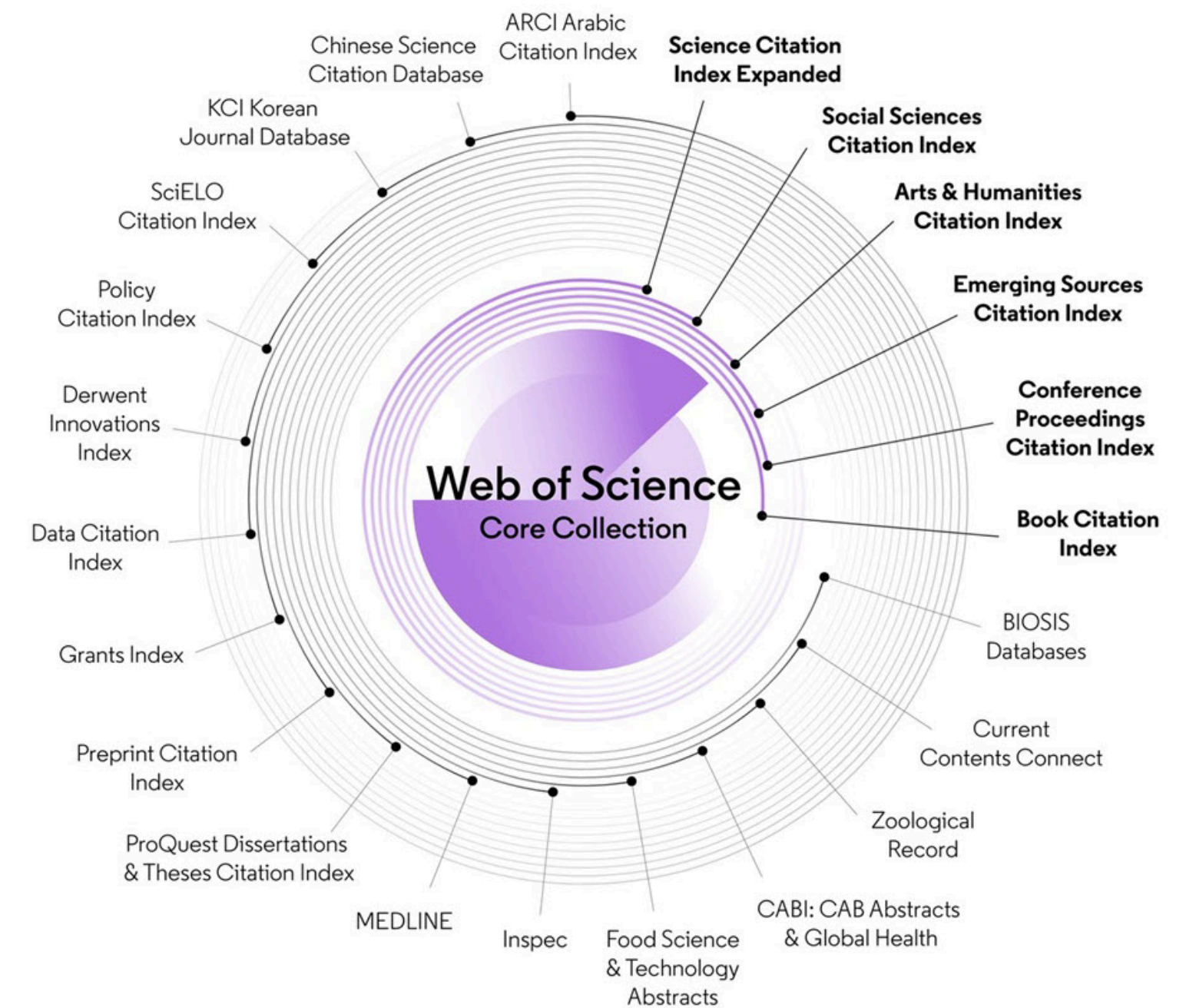
You should have received an email with access to

- Clarivate products
- Elsevier product
- Altmetrics
- No need to have one for OpenAlex

## 2.a WoS-CC

# Indexing and coverage

- Created in 1964 by Eugene Garfield
- Multidisciplinary
- English-language bias
  - +97M records
  - +2.45B cited references
  - +22K scientific journals
- Updated daily
- Databases included:
  - Core Collection\*: curated, citation-linked set of journals, books, and proceedings at the heart of WoS, used for global indicators.
  - Others: additional subject-specific or regional databases hosted on the WoS platform



\*The one we will use in the exercises

# Metadata

PSYCHOLOGY & SEXUALITY  
2021, VOL. 12, NO. 4, 332–344  
<https://doi.org/10.1080/19419899.2020.1729844>



OPEN ACCESS

## What is gender, anyway: a review of the options for operationalising gender

Anna Lindqvist<sup>a,b</sup>, Marie Gustafsson Sendén<sup>b</sup> and Emma A. Renström<sup>c</sup>

<sup>a</sup>Department of Psychology, Lund University, Lund, Sweden; <sup>b</sup>Department of Psychology, Stockholm University, Stockholm, Sweden; <sup>c</sup>Department of Psychology, University of Gothenburg, Gothenburg, Sweden

### ABSTRACT

In the social sciences, many quantitative research findings as well as presentations of demographics are related to participants' gender. Most often, gender is represented by a dichotomous variable with the possible responses of woman/man or female/male, although gender is not a binary variable. It is, however, rarely defined what is meant by gender. In this article, we deconstruct the concept 'gender' as consisting of several facets, and argue that the researcher needs to identify relevant aspects of gender in relation to their research question. We make a thorough exposition of considerations that the researcher should bear in mind when formulating questions about each facet, in order to exemplify how complex this construct is. We also remind the researcher that gender is not a binary category and discuss challenges in the balance between taking existing gender diversity into account and yet sorting participants into gender categorisations that function in statistical analyzes. To aid in this process, we provide an empirical example on how gender identity may be categorised when using a free-text response. Lastly, we suggest that other measurements than participants' gender might be better predictors of the outcome variable.

### ARTICLE HISTORY

Received 1 March 2019  
Accepted 10 February 2020

### KEYWORDS

Gender; gender identity;  
transgender; research  
methods; cisnormativity

Many quantitative research findings and demographics in the social sciences are related to gender, but what is gender and how can it be operationalised? Regarding most variables, researchers strive to generate valid instruments with low measurement errors. Despite this, a traditional use of binary gender measurements prevails in most fields of social science (Westbrook & Saperstein, 2015), even though gender is not a binary category (Ansara & Hegarty, 2014; Hyde, Bigler, Joel, Tate, & van

## What is gender, anyway: a review of the options for operationalising gender

Highly Cited Paper

By Lindqvist, A (Lindqvist, Anna)<sup>[1]</sup>,<sup>[2]</sup>; Sendén, MG (Senden, Marie Gustafsson)<sup>[2]</sup>; Renström, EA (Renstrom, Emma A.)<sup>[3]</sup>

[View Web of Science ResearcherID and ORCID](#) (provided by Clarivate)

### Source

PSYCHOLOGY & SEXUALITY  
Volume: 12 Issue: 4 Page: 332-344  
DOI: 10.1080/19419899.2020.1729844

### Published

OCT 2 2021

### Early Access

FEB 2020

### Indexed

2020-02-20

### Document Type

Review

### Abstract

In the social sciences, many quantitative research findings as well as presentations of demographics are related to participants' gender. Most often, gender is represented by a dichotomous variable with the possible responses of woman/man or female/male, although gender is not a binary variable. It is, however, rarely defined what is meant by gender. In this article, we deconstruct the concept 'gender' as consisting of several facets, and argue that the researcher needs to identify relevant aspects of gender in relation to their research question. We make a thorough exposition of considerations that the researcher should bear in mind when formulating questions about each facet, in order to exemplify how complex this construct is. We also remind the researcher that gender is not a binary category and discuss challenges in the balance between taking existing gender diversity into account and yet sorting participants into gender categorisations that function in statistical analyzes. To aid in this process, we provide an empirical example on how gender identity may be categorised when using a free-text response. Lastly, we suggest that other measurements than participants' gender might be better predictors of the outcome variable.

### Keywords

**Author Keywords:** Gender; gender identity; transgender; research methods; cisnormativity  
**Keywords Plus:** SEX; ENOUGH; CATEGORIES; NORMS

# Metadata

1,051,381 results from Web of Science Core Collection for:

gender

+ Add Keywords Quick add keywords: < + gender + gender equality + gender identity + gender stereotypes + gender dysphoria + gender differences + >

1,051,381 Documents 100 Researchers

Analyze Results Citation Report Create Alert

Refine results Export Refine

Search within results...

Quick Filters

- ☐ Highly Cited Papers 2,574
- ☐ Hot Papers 53
- ☐ Review Article 39,615
- ☐ Early Access 15,353
- ☐ Open Access 387,282
- ☐ Enriched Cited References 159,867
- ☐ Open publisher-invited reviews 1,192

Publication Years

Show Final Publication Year

- ☐ 2026 14
- ☐ 2025 43,806
- ☐ 2024 70,469
- ☐ 2023 66,674
- ☐ 2022 71,487

See all >

0/1,051,381 Add To Marked List Export Preferred Search Results Combined Sem... Sort by Relevance < 1 of 2,000 >

1 DOING GENDER

Semantic search result

West, C and Zimmerman, DH

Jun 1 1987 | GENDER & SOCIETY 1 (2) , pp.125-151

The purpose of this article is to advance a new understanding of gender as a routine accomplishment embedded in everyday interaction. To do so entails a critical assessment of existing perspectives on sex and gender and the introduction of important distinctions among sex, sex category, and gender. We argue that recognition of the analytical indep... Show more

Full Text at Publisher ...

6,978 Citations

73 References

Co-citation map

Related records

2 Cross-Cultural Patterns of Gender Differences in STEM: Gender Stratification, Gender Equality and Gender-Equality Paradoxes

Guo, JS; Marsh, HW; (...); Hu, X

Jun 2024 | EDUCATIONAL PSYCHOLOGY REVIEW 36 (2)

Our study is among the first to provide a comprehensive review of cross-national patterns of gender differences in various STEM-related constructs-achievement, beliefs, attitudes, aspirations, and participation, concerning country-level gender equality. We complement our review with empirical analyses utilizing rigorous methodologies and richer dataset ... Show more

View full text ...

7 Citations

95 References

Co-citation map

Related records

3 Troubling genderS and consumer well-being: Going across, between and beyond the

Bibliographic fields are key to:

- Improve data retrieval
- Analyze and describe the dataset
- Build variables and indicators

# Data retrieval process

To download the complete bibliographic record, max. no. of records is 500/1000 (depending on format)

API available: programmatic, large-scale, structured, suited for bibliometric analysis and automation.

The screenshot displays the Clarivate Web of Science interface. At the top, it shows '1,051,381 Documents' and '100 Researchers'. Below this, there are buttons for 'Analyze Results', 'Citation Report', and 'Create Alert'. On the left, there's a 'Refine results' section with a search bar and various filters like 'Quick Filters' and 'Publication Years'. The main area shows a list of search results. A blue arrow points to the 'Export' button in the top right of the results list. A dropdown menu is open, showing various export formats: EndNote online, EndNote desktop, Add to my researcher profile, Plain text file, RefWorks, RIS (other reference software), BibTeX, Excel, Tab delimited file, Printable HTML file, InCites, FECYT CVN, Email, and Fast 5000. The first result is 'DOING GENDER' by West, C and Zimmerman, DH, published in 'GENDER & SOCIETY' in 1987. The second result is 'Cross-Cultural Patterns of Gender Equality and Gender-Equality' by Guo, JS; Marsh, HW; (...); Hu, X, published in 'EDUCATIONAL PSYCHOLOGY' in 2024.

# Download all journals

English ▾

Products

Web of Science

Master Journal List

InCites Benchmarking & Analytics

Journal Citation Reports™

Research Horizon Navigator

Essential Science Indicators

Reference Manager

EndNote

EndNote Click

Master Journal List

Search Journals

Match Manuscript

Downloads

Help Center

Welcome, Elvira González-Salmón

Settings

Log Out

Collection List Downloads

Web of Science Core Collection

Monthly Changes Archive

Additional Web of Science Indexes

Web of Science Core Collection

Last Updated: August 18, 2025

Lists for each index are available for download to Web of Science subscribers.

Web of Science Core Collection

Last Updated: August 18, 2025

The Web of Science Core Collection™ includes the Science Citation Index Expanded™ (SCIE), Social Sciences Citation Index™ (SSCI), Arts & Humanities Citation Index™ (AHCI), and Emerging Sources Citation Index™ (ESCI). Web of Science Core Collection includes only journals that demonstrate high levels of editorial rigor and best practice. The Journal Citation Reports™ includes journals from the SCIE and SSCI.

Each collection list download includes the journal title, ISSN/eISSN, publisher name and address, language, and category.

Science Citation Index Expanded (SCIE)

Social Sciences Citation Index (SSCI)

Arts & Humanities Citation Index (AHCI)

Emerging Sources Citation Index (ESCI)

JCR 2025

# Journal Citations Reports

English ▾ Products

- Web of Science
- Master Journal List
- InCites Benchmarking & Analytics
- Journal Citation Reports™**
- Research Horizon Navigator
- Essential Science Indicators
- Reference Manager
- EndNote
- EndNote Click

**Journal Citation Reports™** Journals Categories Publishers Countries/Regions Compare My favorites Elvira González-Sal

22,249 journals

Journal name/abbreviation, ISSN/eISSN, category, publisher, country/region

Indicators: Default ▾

Copy query link Export

Customize

Filter

Journal name ▾	ISSN	eISSN	Category	Edition	Total Citations ▾	2024 JIF ▾	JIF Quartile	2024 JCI ▾	% of Citable OA ▾
<input type="checkbox"/> CA-A CANCER JOURNAL FOR CLINICIANS	0007-9235	1542-4863	ONCOLOGY	SCIE	71,799	232.4	Q1	112.16	98.51 %
<input type="checkbox"/> NATURE REVIEWS MICROBIOLOGY	1740-1526	1740-1534	MICROBIOLOGY	SCIE	61,109	103.3	Q1	9.86	3.42 %
<input type="checkbox"/> NATURE REVIEWS DRUG DISCOVERY	1474-1776	1474-1784	Multiple ▾		52,833	101.8	Q1	13.95	8.33 %
<input type="checkbox"/> NATURE REVIEWS MOLECULAR CELL BIOLOGY	1471-0073	1471-0080	CELL BIOLOGY	SCIE	73,135	100.2	Q1	7.04	0.70 %

# Journal Citations Reports

Filter

Journals (22,249)

ISSN/eISSN

Categories (254)

Publishers (8,825)

Country / region (111)

Citation Indexes

JCR Year

Open Access

JIF Quartile

JIF Range

JCI Range

JIF Percentile

Reset

Apply

Impact Metrics

Metrics focused on the citation impact of the journals.

☒ Total Citations

☒ 2024 JIF

☐ JIF Rank

☐ 5 Year JIF

☐ 5 Year JIF Quartile

☐ JIF Without Self Cites

☐ Immediacy Index

Normalized Metrics

Metrics that have been adjusted mathematically to a particular context.

☒ 2024 JCI

☐ JCI Rank

☐ JCI Quartile

☐ JCI Percentile

☐ Eigenfactor

☐ Normalized Eigenfactor

☐ Article Influence Score

☐ JIF Percentile

☒ JIF Quartile

☐ AIS Quartile

☐ AIS Rank

Source Metrics

Metrics based on the content of the journals.

☐ Citable Items

☐ % of Articles in Citable items

☐ Cited Half-Life

☐ Citing Half-Life

☐ Total Articles

☒ % of Citable OA

Name your indicators

Save

Apply

Compare

My favorites

Elvira González-Salmón

Copy query link

Export

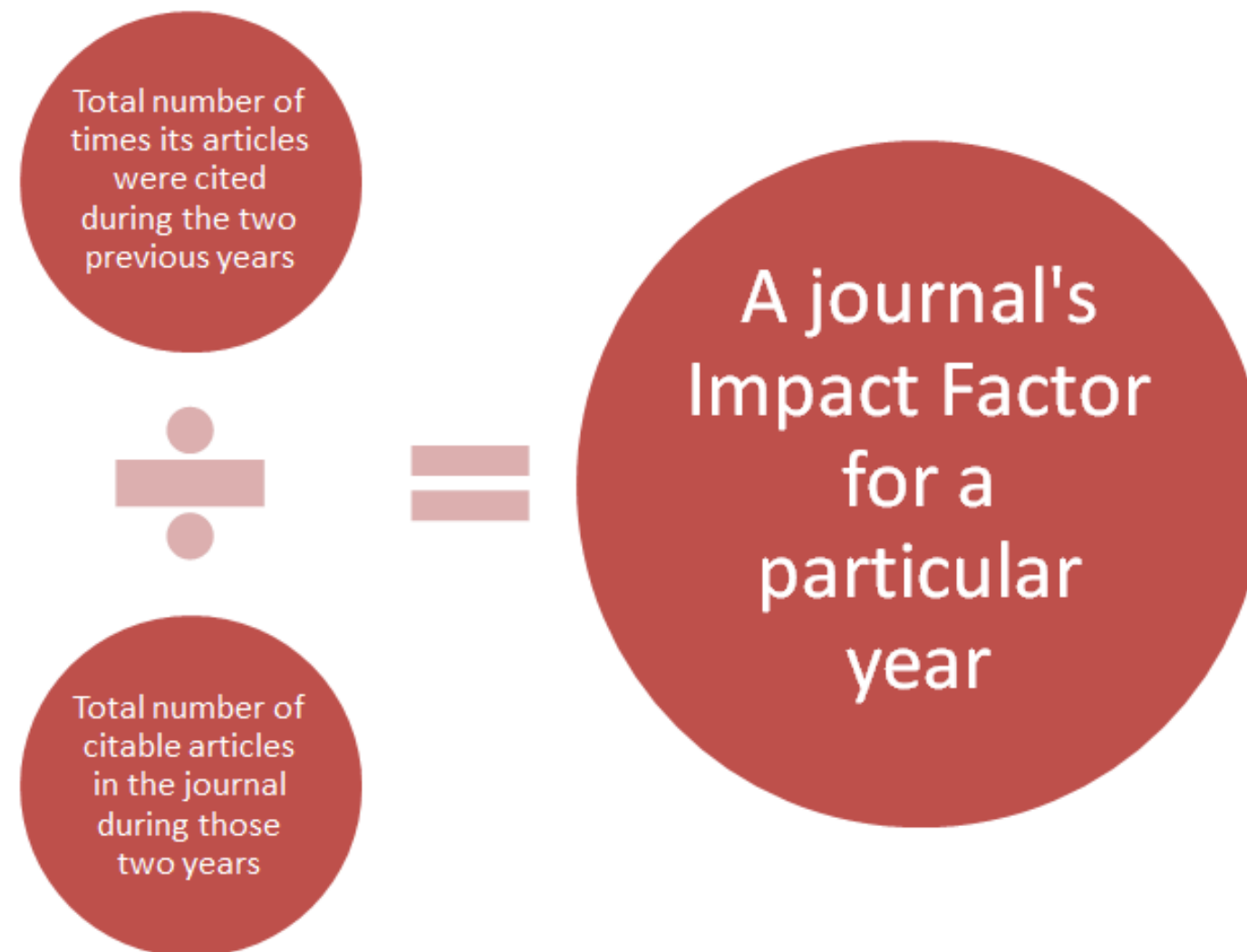
Customize

JIF Quartile	2024 JCI	% of Citable OA
Q1	112.16	98.51 %
Q1	9.86	3.42 %
Q1	13.95	8.33 %
Q1	7.94	0.70 %
Q1	5.88	0.00 %
Q1	23.28	18.13 %

16

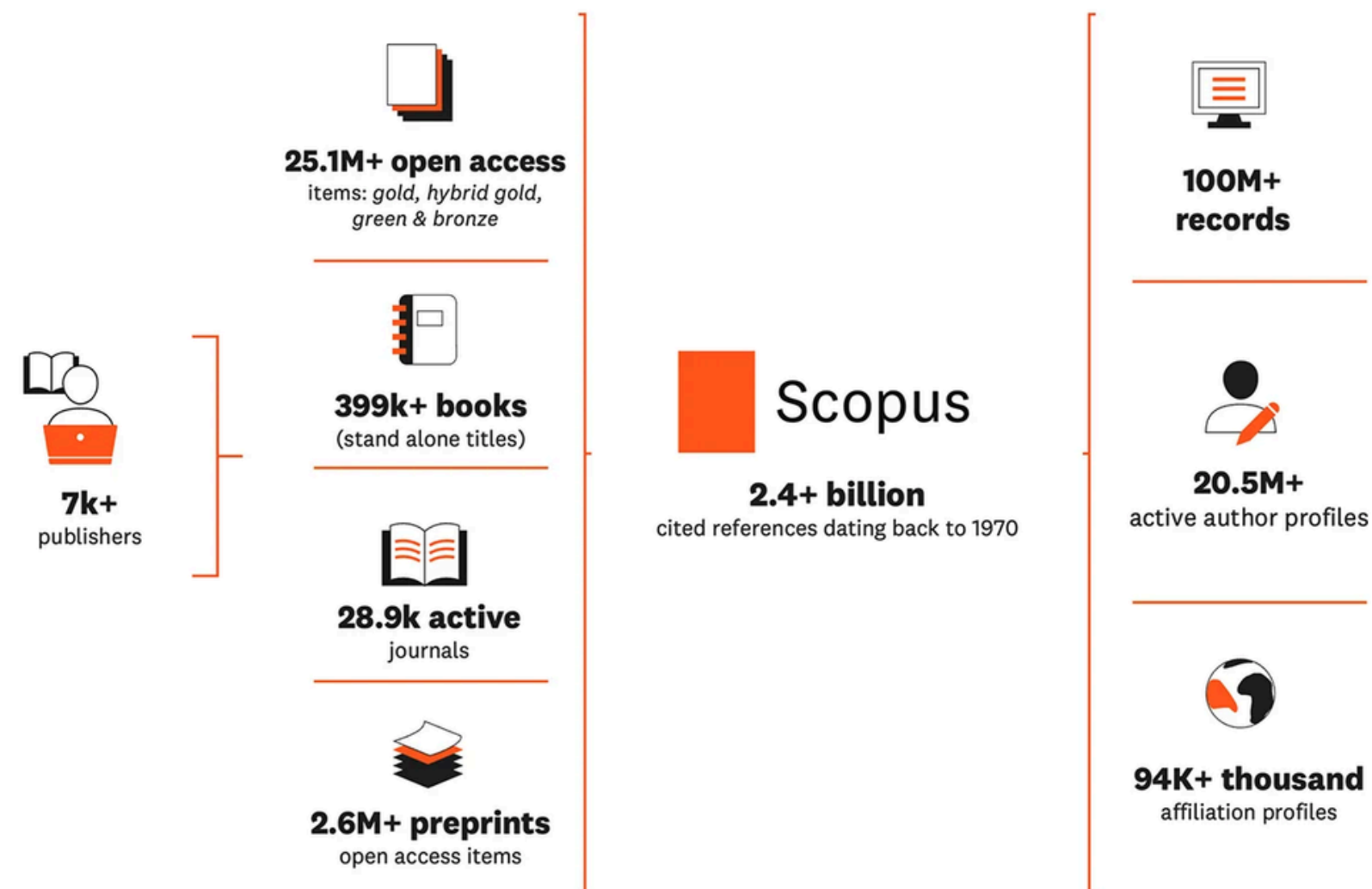
# Journal Citations Reports

- JIF (Journal Impact Factor)



## 2.b Scopus

# Indexing and coverage



- Owned by Elsevier and launched in 2004
- Main competitor of WoS
- Includes its own set of journal-level indicators
- Single citation index without subcollections

As of February 2025

Scopus®

# Metadata

## Investigating the impact of artificial intelligence development on water pollution in China

Gondwana Research • Article • 2024 • DOI: 10.1016/j.gr.2024.04.011

Ma, Xinzheng<sup>a, b</sup>; Weng, Shimei<sup>c</sup>; Zhao, Xin<sup>d</sup>; Li, Jing<sup>e</sup>; Haider, Sajjad<sup>f</sup>

<sup>a</sup>College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, 310058, China

[Show all information](#)

[View PDF](#) [Full text](#) [Export](#) [Save to list](#)

[Document](#) [Impact](#) [Cited by \(4\)](#) [References \(44\)](#) [Similar documents](#)

### Abstract

Analysing the impacts of AI development on water pollution is of great significance for improving water environment management measures and promoting global sustainable development. In this study, macro and micro panel data of China from 2008 to 2019 were used to measure the development level of AI at the industry and provincial levels in China, and a panel data regression model was constructed to analyse the effects of AI development on water pollution and regional heterogeneity. The measurement results show that the development level of AI is increasing. Among the six industries, the intelligence level of the manufacturing industry has developed the fastest, by approximately 48.74 times compared with the baseline. At the provincial level, Guangdong Province made the greatest progress in intelligence level, increasing by approximately 23.6 times. The regression results show that AI technology is helpful in reducing the level of water pollution, and this conclusion remains valid after the regression of instrumental variables and robustness testing. In terms of regional heterogeneity, AI technology can significantly reduce the degree of water pollution in eastern and western China. However, the development of AI technology in the central region has no significant impact on water pollution, and a risk of increasing the degree of water pollution exists. The reasons may be that the eastern region has high levels of economic vitality and technological innovation and a strong ability to transform AI achievements, which can significantly

4 82th percentile  
Citations

1.64  
FWCI

### Abstract

[Author keywords](#)

[Indexed keywords](#)

[Funding details](#)

[Corresponding authors](#)

- Bibliographic records show similar structure to WoS
- Consider there are differences in terms of labeling, document types, classification scheme, etc.
- Combining data from Scopus and WoS would require further data processing and homogenization.

Scopus®

# Metadata



## Detailed information ×

### Bibliographic information

Document type	Article
DOI	10.1016/j.j.gr.2024.04.011
EID	2-s2.0-85192886941
Original language	English
Publication date	August 2024
PubMed ID	
Source type	Journal
ISSN	1342937X
Publisher	Elsevier Inc.
Publication year	2024
Source title	Gondwana Research
Volume	132
Pages	182 - 192

### Authors (5)

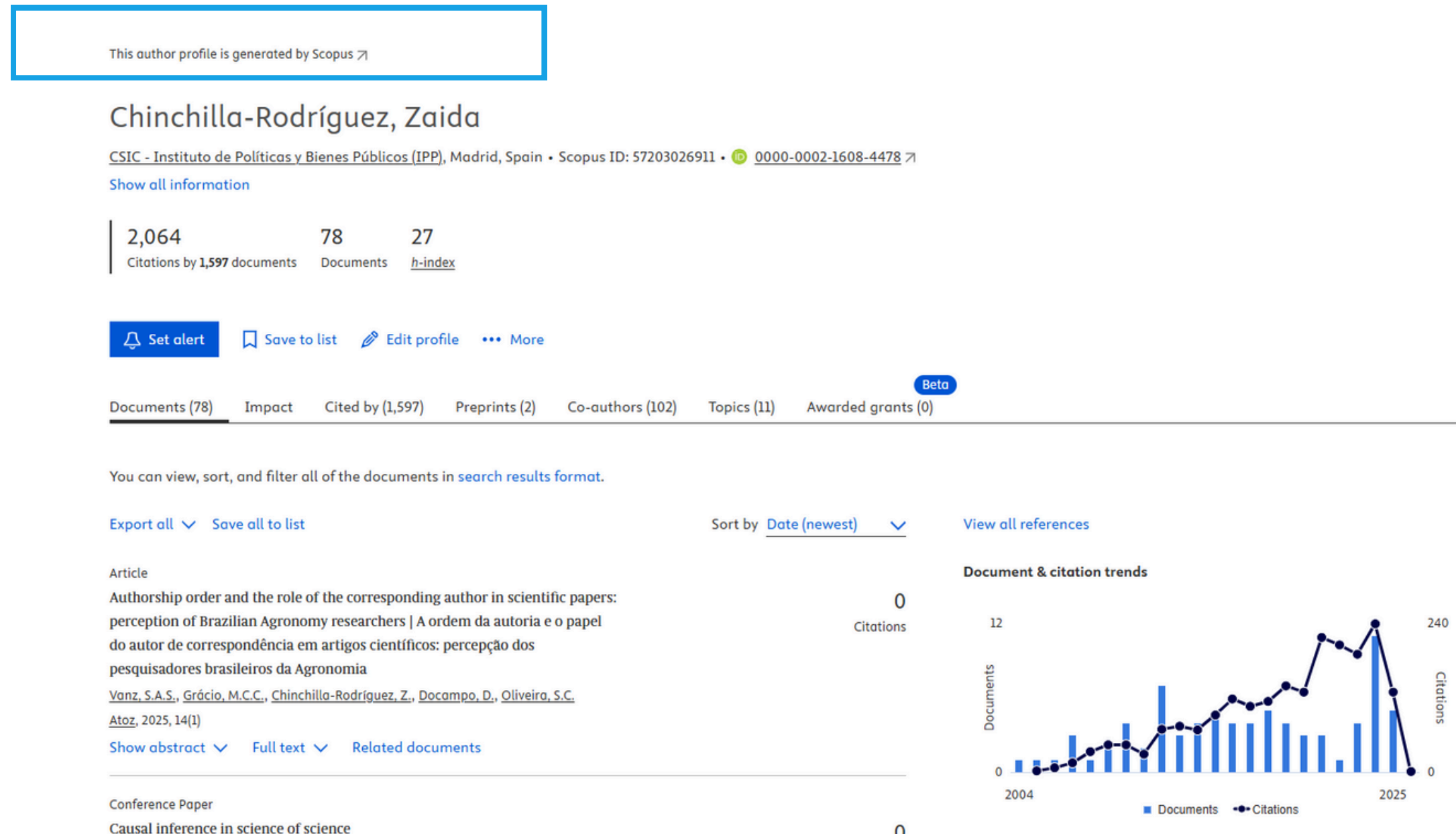
- Ma, Xinzhen<sup>a, b</sup>
- Weng, Shimei<sup>c</sup>
- Zhao, Xin<sup>d</sup>
- Li, Jing<sup>e</sup>
- Haider, Sajjad<sup>f</sup>

### Author affiliations (6)

- <sup>a</sup> College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, 310058, China
- <sup>b</sup> Polytechnic Institute, Zhejiang University, Hangzhou, 310015, China
- <sup>c</sup> School of Mathematics and Statistics, Fujian Normal University, Fuzhou, 350117, China
- <sup>d</sup> School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, 233030, China
- <sup>e</sup> School of Economics, Lanzhou University, Lanzhou, 730000, China
- <sup>f</sup> Chemical Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh, 11421, Saudi Arabia



# Metadata - author profiles



- A nice feature of Scopus is their algorithmically generated Author Profiles
- It offers a list of author-level metrics (e.g., H-Index, total publications).
- It also includes other author-level metadata (e.g., affiliation changes).

Scopus®

# Data Retrieval process

Advanced query ☐

Search within  
Article title, Abstract, Keywords

Search documents \*  
gender science

Save search

Set search alert

+ Add search field

Reset Search

Documents Preprints **Beta** Secondary documents

33,779 documents found

Analyze results

Refine search

Search within results

Filters

Year

Range Individual

from to

Author name

Subject area

☐ All ☐ Export ☐ Download ☐ Citation overview ☐ More

Show all abstracts Sort by Date (newest)

	Authors	Source	Year	Citations
<input type="checkbox"/> 1	Yang, J., Hao, P., Yanping, S., ... Yanqi, L., Qigang, C.	Chinese Journal of Tissue Engineering Research	2026	0
<input type="checkbox"/> 2	Outão, J.C., Costa, L.A., Constantinou, E., dos Santos, R.P., Serebrenik, A.	Journal of Systems and Software	2026	0
<input type="checkbox"/> 3	Mohammadi, E., Thelwall, M., Cai, Y.,	Information Processing and Management	2026	0

File types

- CSV
- RIS
- BibTeX
- Plain text

Reference managers

- Mendeley
- Refworks (RIS)
- Zotero (RIS)
- EndNote (RIS)

Platforms

- SciVal

Scopus®

# Data Retrieval process

Export 10 documents to CSV ? x

You can export up to 20,000 documents in CSV format. Some fields might not be available for export at the moment, please check back again later.

① Export Processing Time

☒ All documents on this page

☐ Documents 1 - To

What information do you want to export?

<input checked="" type="checkbox"/> Citation information	<input type="checkbox"/> Bibliographical information	<input type="checkbox"/> Abstract & keywords	<input type="checkbox"/> Funding details	<input type="checkbox"/> Other information
<input checked="" type="checkbox"/> Author(s)	<input type="checkbox"/> Affiliations	<input type="checkbox"/> Abstract	<input type="checkbox"/> Funding text	<input type="checkbox"/> Accession numbers & chemicals
<input checked="" type="checkbox"/> Document title	<input type="checkbox"/> Serial identifiers (e.g. ISSN)	<input type="checkbox"/> Author keywords		<input type="checkbox"/> Conference information
<input checked="" type="checkbox"/> Year	<input type="checkbox"/> PubMed ID	<input type="checkbox"/> Indexed keywords		<input type="checkbox"/> Include references
<input checked="" type="checkbox"/> EID	<input type="checkbox"/> Publisher			
<input checked="" type="checkbox"/> Source title	<input type="checkbox"/> Editor(s)			
<input checked="" type="checkbox"/> Volume, issues, pages	<input type="checkbox"/> Language of original document			
<input checked="" type="checkbox"/> Citation count				
<input checked="" type="checkbox"/> Source & document type				
<input checked="" type="checkbox"/> Publication stage				
<input checked="" type="checkbox"/> DOI				
<input checked="" type="checkbox"/> Open access				

Select all information ☒ Truncate to optimize for Excel ① ☐ Save as preference **Export**

Up to 20,000  
documents

API available

Scopus®

Sources

Subject area

▼

Enter subject area

CiteScore 2024 has been released.
[View CiteScore methodology >](#)

X

Filter refine list

Apply

Clear filters

Display options

☐ Display only Open Access journals

Counts for 4-year timeframe

☒ No minimum selected

☐ Minimum citations

☐ Minimum documents

Citescore highest quartile

☐ Show only titles in top 10 percent

☐ 1st quartile

☐ 2nd quartile

☐ 3rd quartile

☐ 4th quartile

Source type

☐ Journals

☐ Book Series

☐ Conference Proceedings

☐ Trade Publications

Apply

Clear filters

48,954 results

[Download Scopus Source List](#)
[Learn more about Scopus Source List](#)

☐ All
 [Export to Excel](#)
[Save to source list](#)

View metrics for year: 2024

	Source title	CiteScore	Highest percentile	Citations 2021-24	Documents 2021-24	% Cited
<input type="checkbox"/> 1	Ca-A Cancer Journal for Clinicians	1154.2	99% 1/415 Oncology	121.190	105	94
<input type="checkbox"/> 2	Foundations and Trends in Machine Learning	202.9	99% 1/490 Software	3450	17	88
<input type="checkbox"/> 3	Nature Reviews Drug Discovery	181.8	99% 1/321 Pharmacology	32.182	177	92
<input type="checkbox"/> 4	Nature Reviews Molecular Cell Biology	150.9	99% 1/410 Molecular Biology	33.659	223	87
<input type="checkbox"/> 5	MMWR Recommendations and Reports	129.9	99% 1/376 Health (social science)	2857	22	100
<input type="checkbox"/> 6	Nature Reviews Clinical Oncology	114.5	99% 2/415 Oncology	28.505	249	93
<input type="checkbox"/> 7	Nature Reviews Disease Primers	113.3	99% 1/668 General Medicine	19.833	175	83

## 2.c OpenAlex

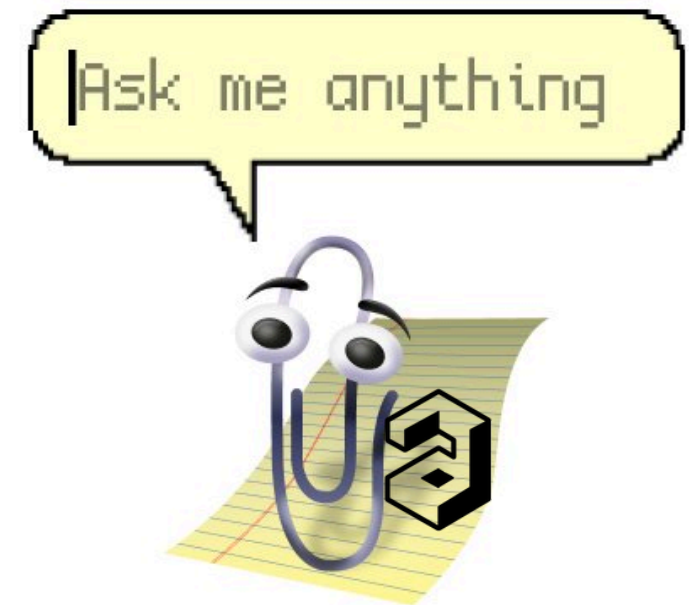
# Indexing and coverage

- Microsoft Academic Graph offered a fully open knowledge graph to over 225M records where all entities (works, authors, journals, institutions, topics, citations) were linked explicitly.
- It became a promising data source for bibliometricians
- But in 2021, Microsoft decided to discontinue the project...
- As a response to that, the Arcadia Foundation funded OpenAlex, a project led by non-profit organization Our Research, to continue with the project.
- In 2022, OpenAlex was released.



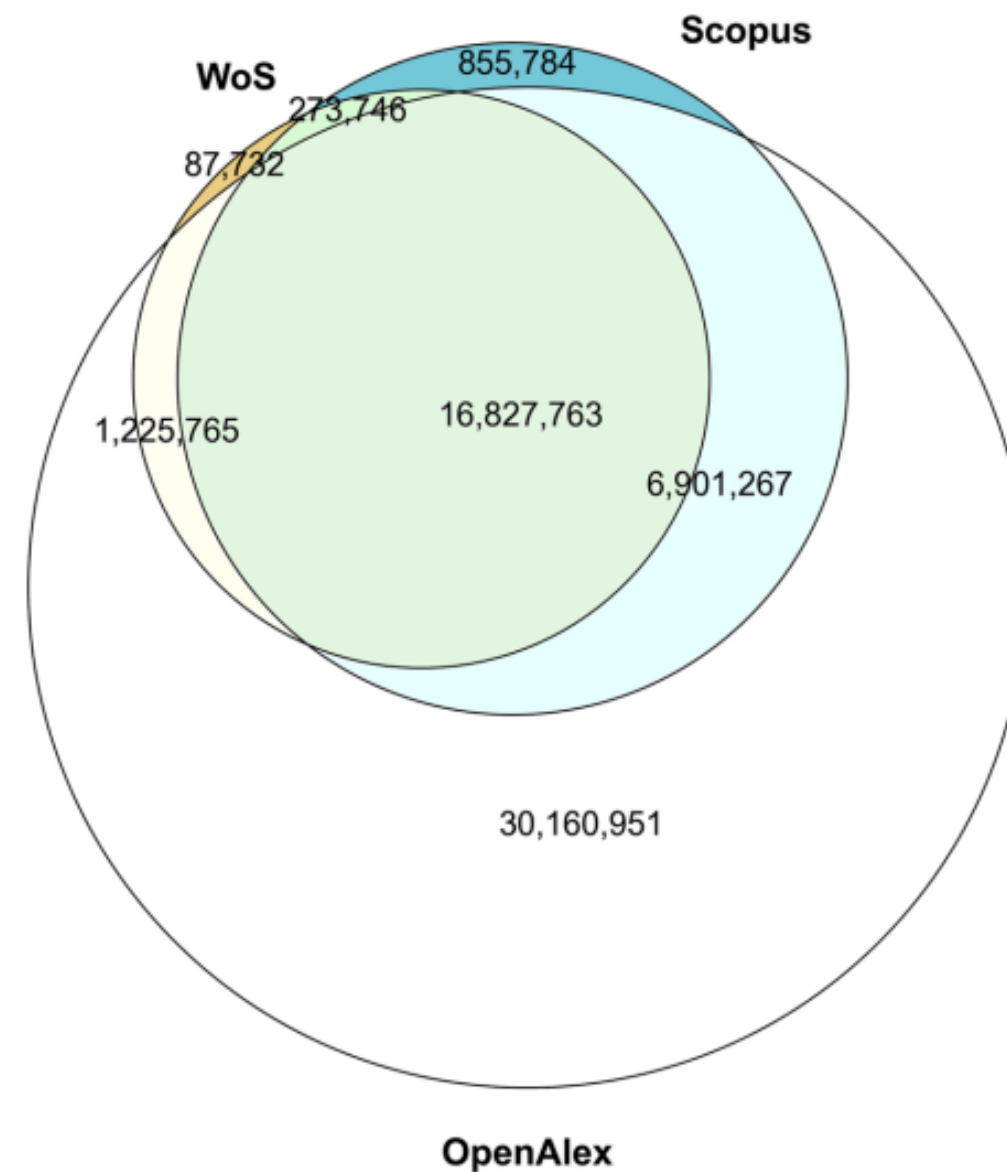
# Indexing and coverage

- It does not only feed from the MAG project but combines information from a variety of data sources.
  - Mainly MAG and Crossref
  - ORCID
  - DOAJ (Directory of Open Access Journals)
  - Unpaywall
  - PubMed
  - Internet Archive
  - Web crawls
  - and more!



# Indexing and coverage

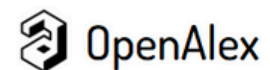
**Fig. 1** Venn diagram of the intersection sizes of unique DOIs based in each database on exact DOI match (without deduplication, i.e. cases of DOIs that have been assigned to multiple papers are now kept in the sets), for records published between 2015 and 2022



Culbert, J. H., 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(4), 2475-2492.

# Metadata

- OpenAlex has greatly improved the quality of its metadata. But of course, there is room for improvement
- The project is still in its very early stages
- Since spring 2024 it includes an interface which is in constant transformation.



[Log In](#) [Sign Up](#)


Search and analyze the world's research.

Try: [Claudia Goldin](#) [coriander OR cilantro](#) [Institution](#)



# Metadata

Revisiting the Southern Grotesque: Mikhail Bakhtin and the Case of Carson McCullers

 Work

 HTML   API  

**Year:** 2001

**Type:** article

**Source:** [The Southern literary journal](#)

**Author:** [Sarah Gleeson-White](#)

**Cites:** [35](#)

**Cited by:** [32](#)

**Related to:** [10](#)

**FWCI:**

**Citation percentile (by year/subfield):** 97,75

**Topic:** [Contemporary Literature and Criticism](#)

**Subfield:** [Literature and Literary Theory](#)

**Field:** [Arts and Humanities](#)

**Domain:** [Social Sciences](#)

**Sustainable Development Goal:** [Gender equality](#)

**Open Access status:** closed

Expect some lower-quality metadata and incomplete bibliographic records



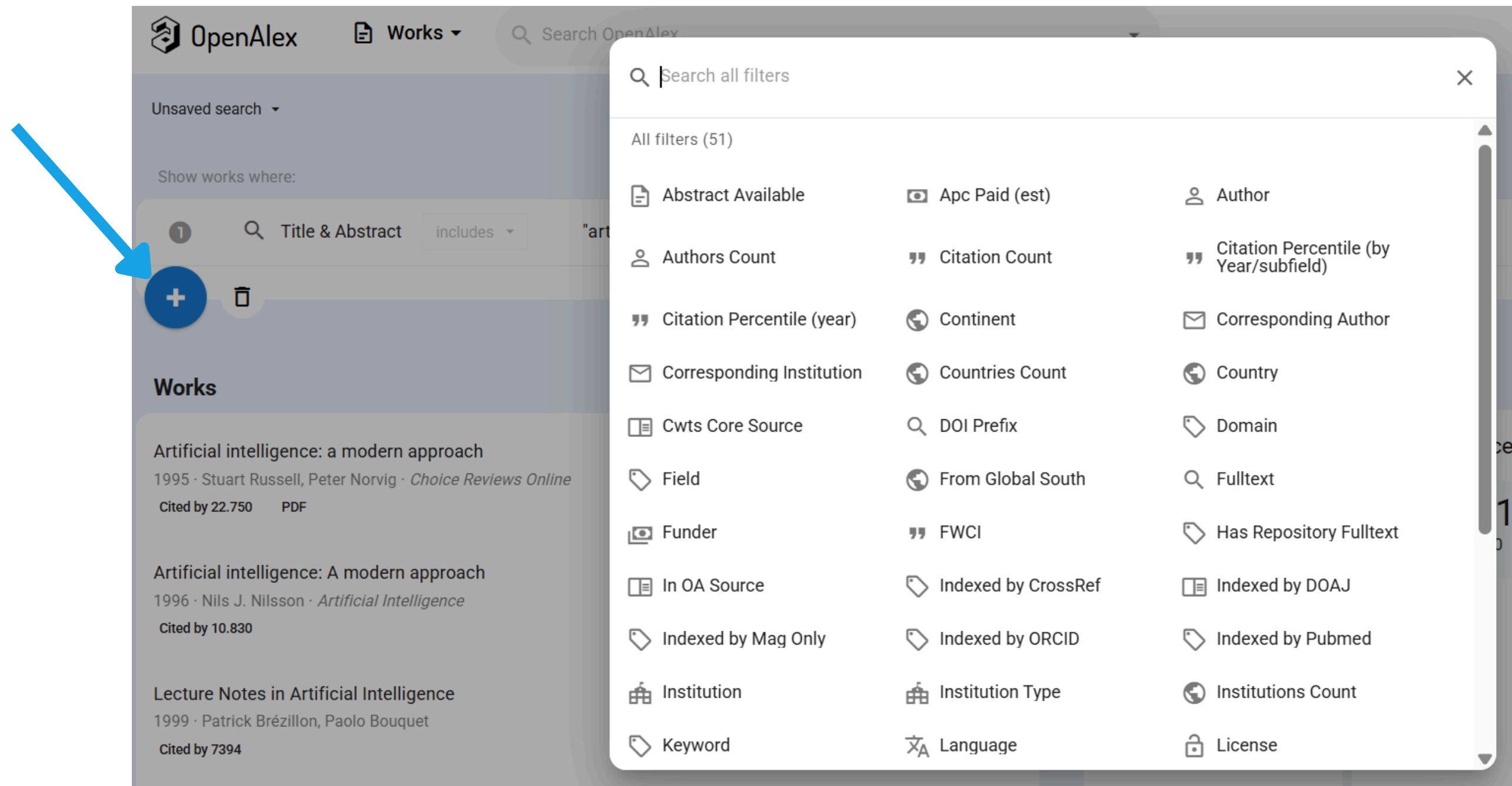
# Data Retrieval process

Record type

The screenshot shows the OpenAlex website interface. At the top, there is a search bar labeled 'Search OpenAlex' and a 'Works' dropdown menu. A blue arrow points to the 'Works' dropdown, which is open, showing options: 'Works' (Scholarly papers, books, datasets, etc.), 'Authors' (Creators of scholarly works), 'Sources' (Journals, conferences, and repositories), and 'Institutions' (Universities and research centers). Below the dropdown, the 'Works' section displays a list of search results. The first result is 'Artificial intelligence: a modern approach' by Stuart Russell and Peter Norvig, published in 1995, cited by 22,750, and available as a PDF. The second result is 'Artificial intelligence: A modern approach' by Nils J. Nilsson, published in 1996, cited by 10,830. The third result is 'Lecture Notes in Artificial Intelligence' by Patrick Brézillon and Paolo Bouquet, published in 1999. To the right of the results, there is a 'Stats' section showing '480.400 results'. Below this, there is a 'Year' histogram and a 'Topic' list. The 'Open Access' section shows a pie chart indicating that 45.1% (216,900) of the results are open access. The 'Topic' list includes: 'Artificial Intelligence in Healthcare and Education' (22,220), 'Ethics and Social Impacts of AI' (10,220), 'Big Data and Business Intelligence' (9,339), 'Online Learning and Analytics' (6,675), 'Radiomics and Machine Learning in' (6,166), and 'Machine Learning in' (6,166).

# Data Retrieval process

## Filtering options



The screenshot shows the OpenAlex website interface. A blue arrow points to a circular button with a plus sign, which is used to open the filtering options menu. The menu is titled "Search all filters" and lists 51 filters organized into three columns. The filters include various criteria such as document availability, citation metrics, author information, and institutional details.

OpenAlex Works Search OpenAlex

Unsaved search

Show works where:

1 Title & Abstract includes "art

+

Works

Artificial intelligence: a modern approach  
1995 · Stuart Russell, Peter Norvig · *Choice Reviews Online*  
Cited by 22.750 PDF

Artificial intelligence: A modern approach  
1996 · Nils J. Nilsson · *Artificial Intelligence*  
Cited by 10.830

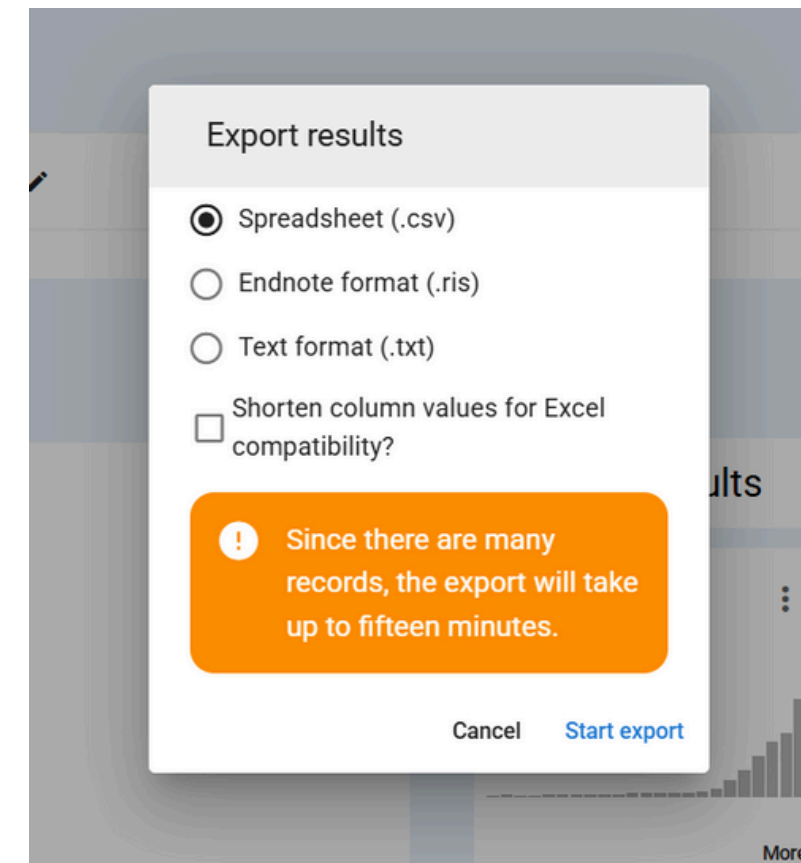
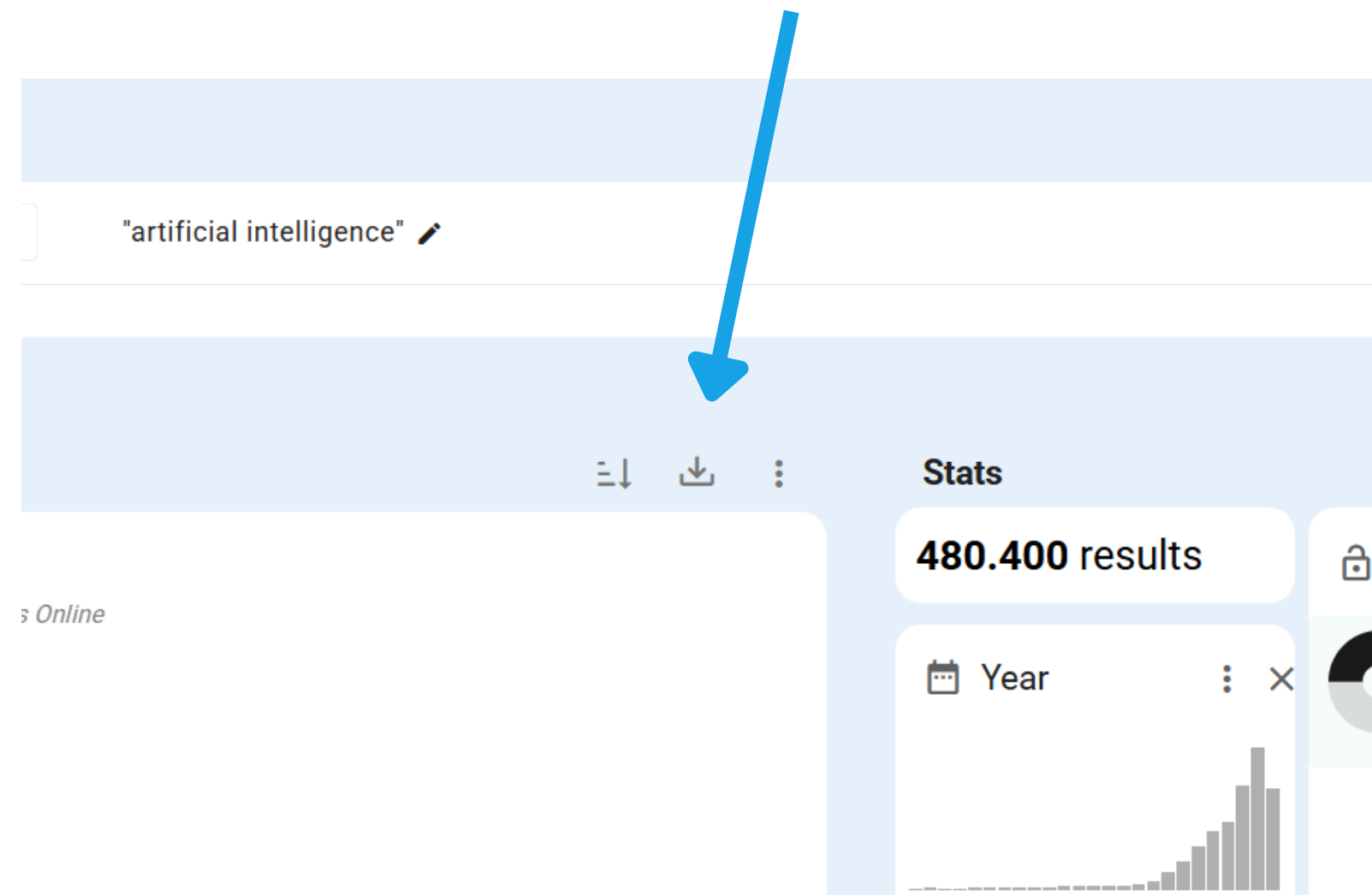
Lecture Notes in Artificial Intelligence  
1999 · Patrick Brézillon, Paolo Bouquet  
Cited by 7394

Search all filters

All filters (51)

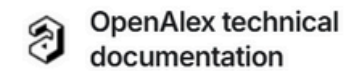
Abstract Available	Apc Paid (est)	Author
Authors Count	Citation Count	Citation Percentile (by Year/subfield)
Citation Percentile (year)	Continent	Corresponding Author
Corresponding Institution	Countries Count	Country
Cwts Core Source	DOI Prefix	Domain
Field	From Global South	Fulltext
Funder	FWCI	Has Repository Fulltext
In OA Source	Indexed by CrossRef	Indexed by DOAJ
Indexed by Mag Only	Indexed by ORCID	Indexed by Pubmed
Institution	Institution Type	Institutions Count
Keyword	Language	License

# Data Retrieval process



Export options

# Data Retrieval process



API ENTITIES  
Aboutness endpoint (/text)

HOW TO USE THE API

[API Overview](#)

Get single entities >

Get lists of entities >

Get groups of entities

Rate limits and authentication

DOWNLOAD ALL DATA

OpenAlex snapshot

Snapshot data format

Download to your machine

Upload to your database >

[HOW TO USE THE API](#)

Copy

## API Overview

The API is the primary way to get OpenAlex data. It's free and requires no authentication. The daily limit for API calls is 100,000 requests per user per day. For best performance, [add your email](#) to all API requests, like `mailto=example@domain.com`.

## Learn more about the API

- [Get single entities](#)
- [Get lists of entities](#) — Learn how to use [paging](#), [filtering](#), and [sorting](#)
- [Get groups of entities](#) — Group and count entities in different ways
- [Rate limits and authentication](#) — Learn about joining the [polite pool](#)
- [Tutorials](#) — Hands-on examples with code



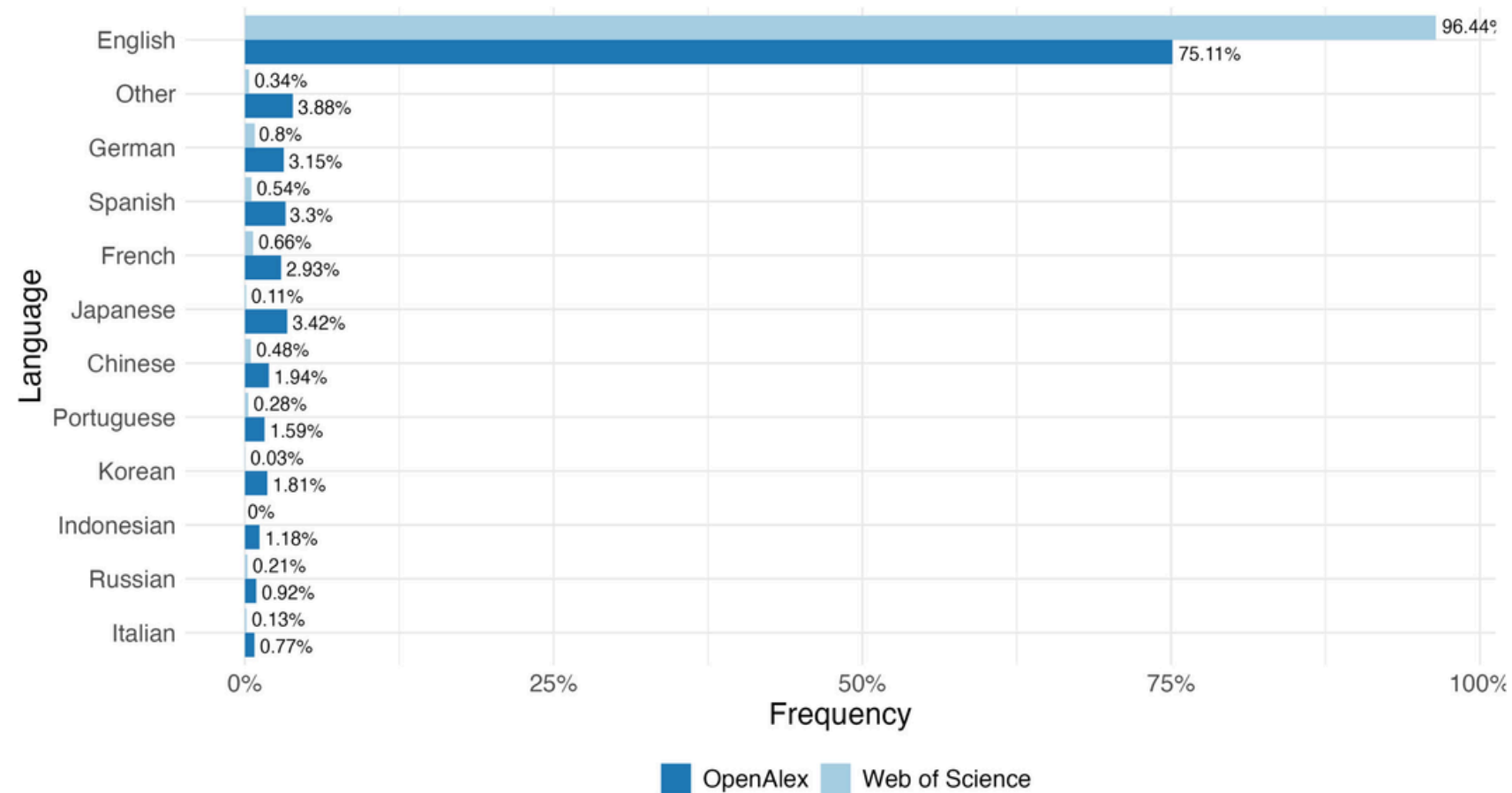


3. How to critically decide which database is more suitable

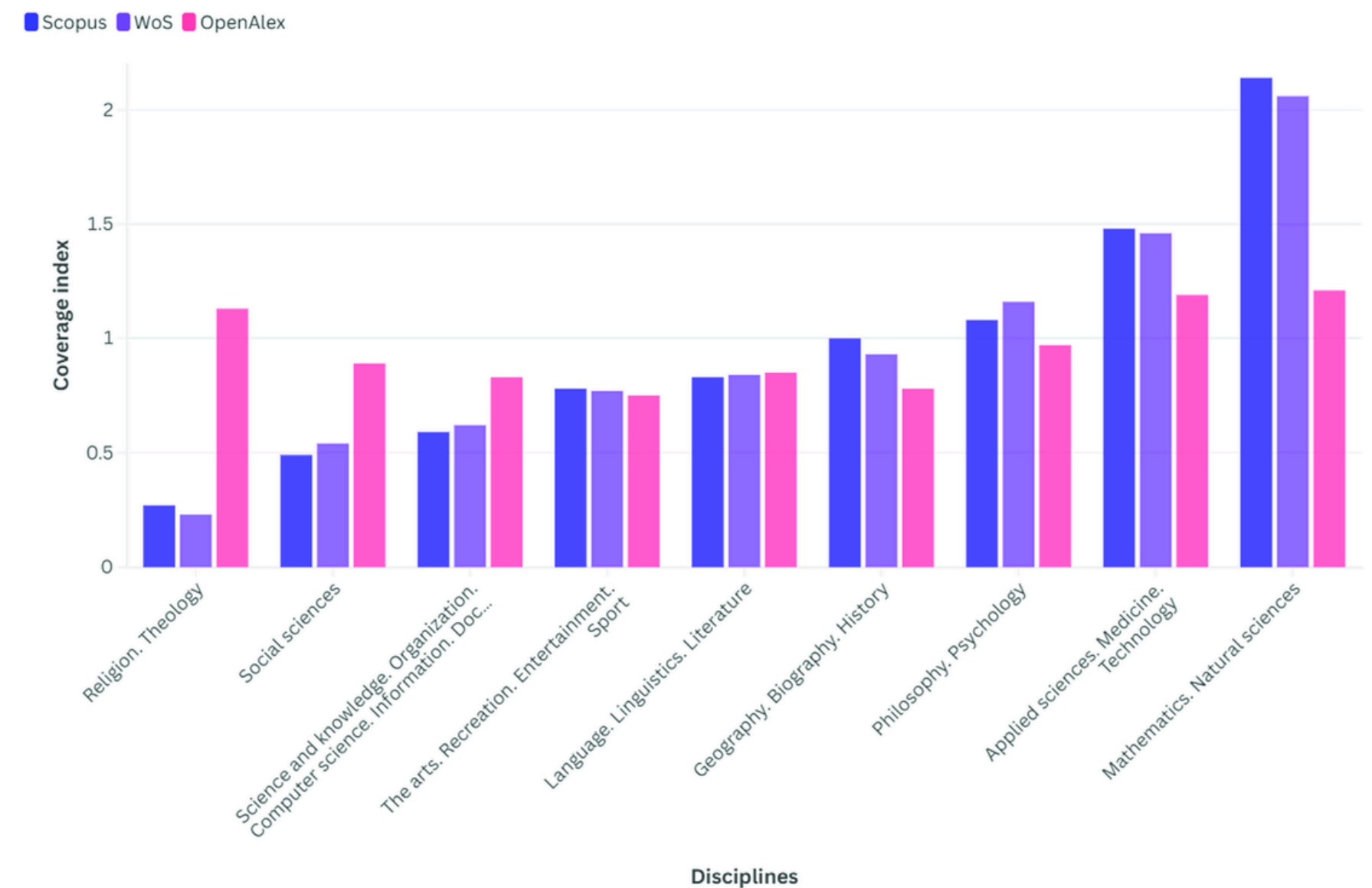
# Which database should I choose?

Things to consider:

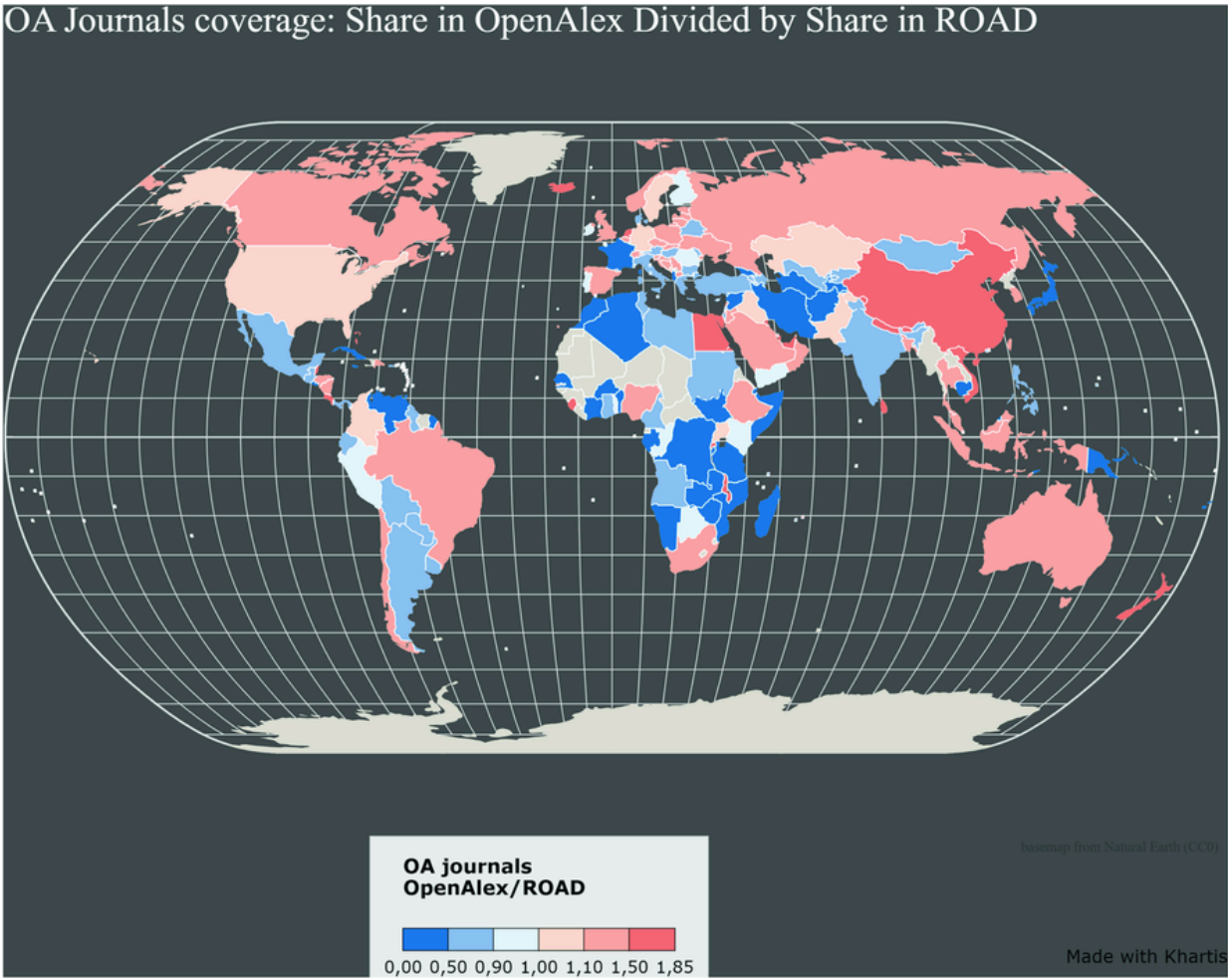
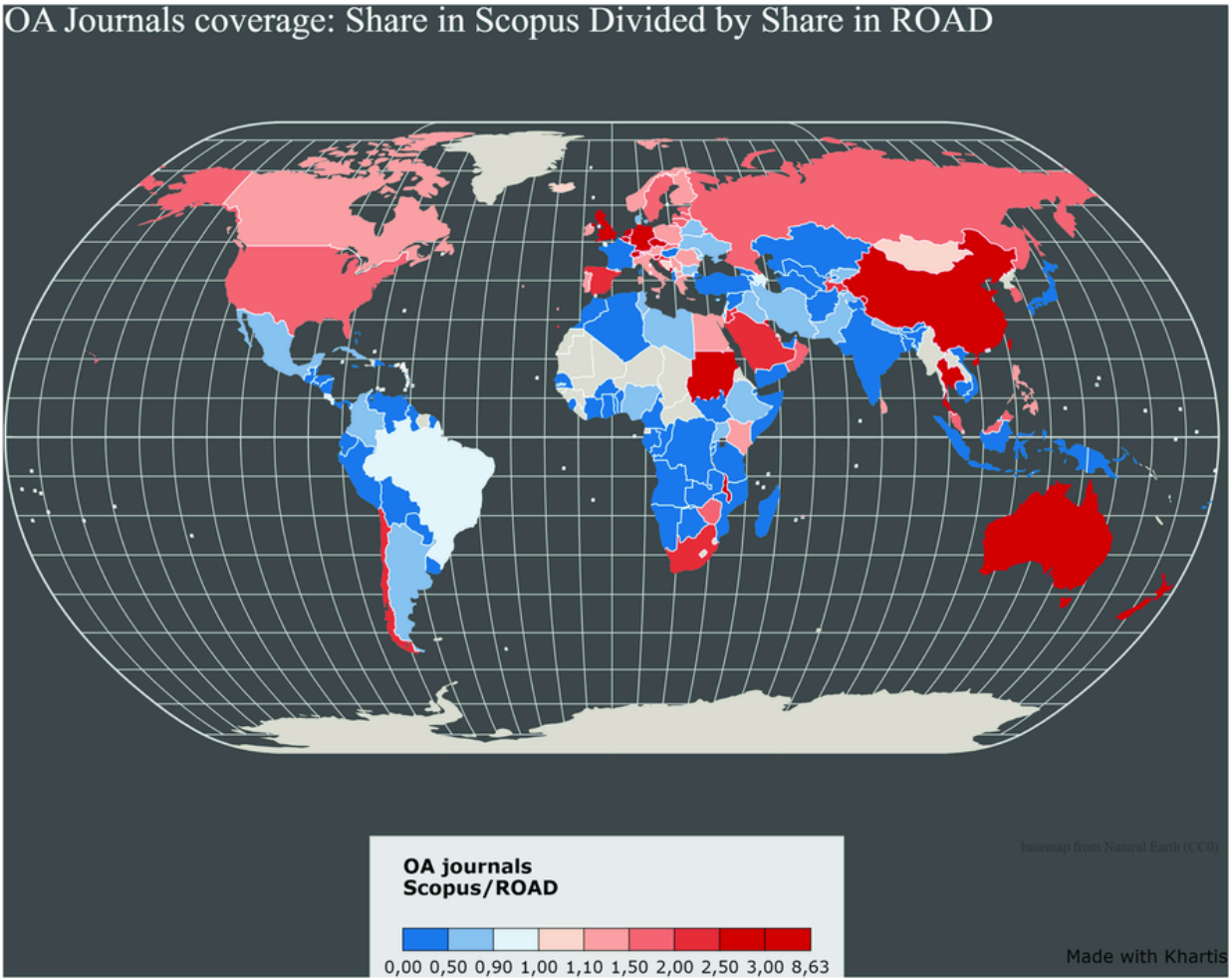
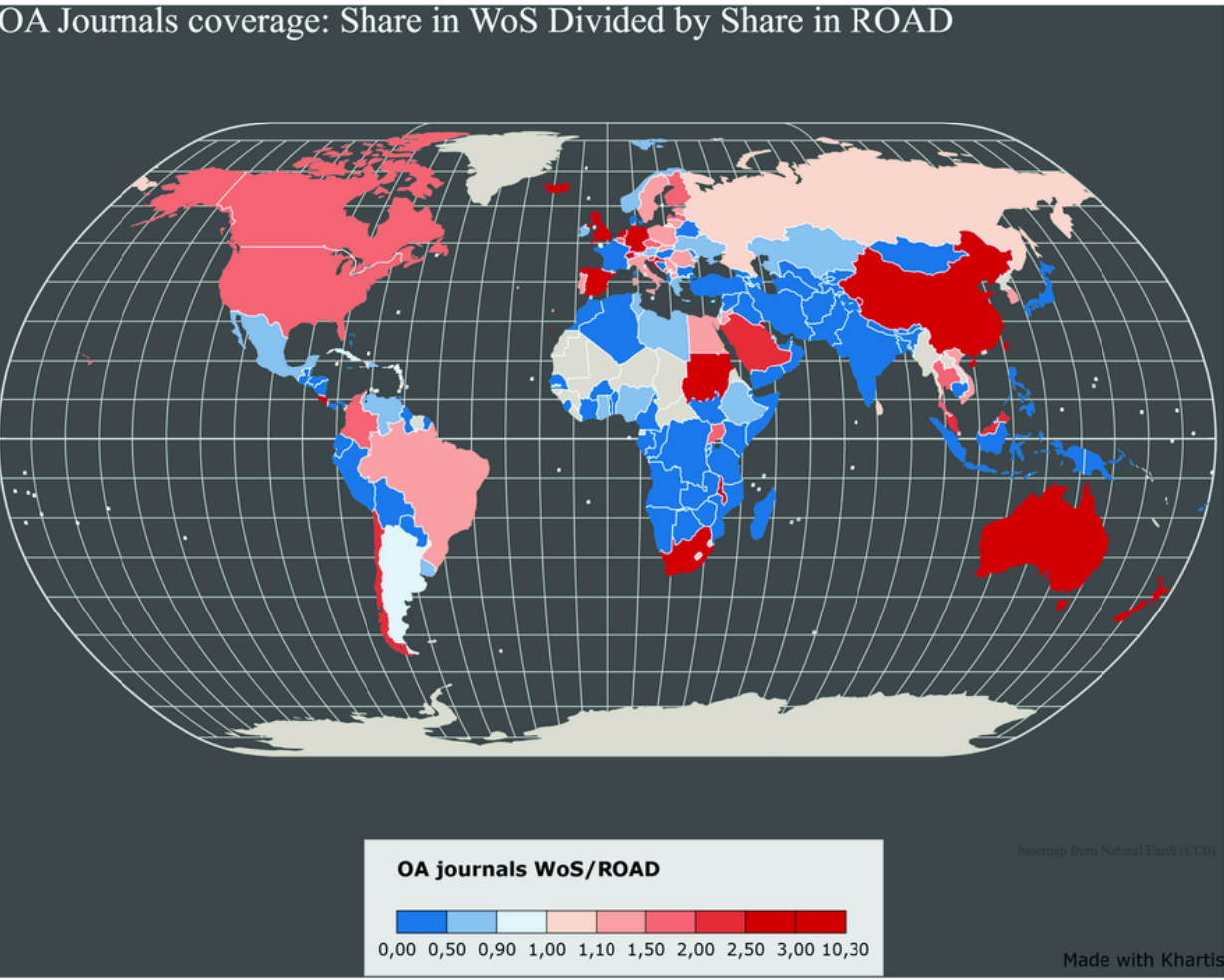
- Coverage
- No ground truth, but different angles of it
  - Global or regional?
  - All disciplines



Céspedes, L., Kozlowski, 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>



Maddi A, Maisonobe M, Boukacem-Zeghmouri C (2025) Geographical and disciplinary coverage of open access journals: OpenAlex, Scopus, and WoS. *PLOS ONE* 20(4): e0320347. <https://doi.org/10.1371/journal.pone.0320347>

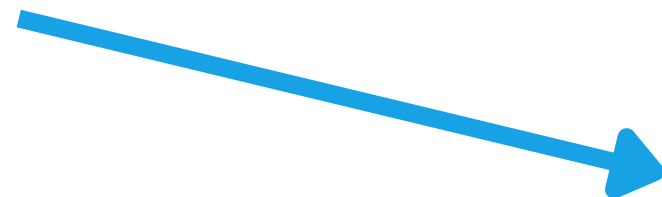


Maddi A, Maisonobe M, Boukacem-Zeghmouri C (2025)  
Geographical and disciplinary coverage of open access journals:  
OpenAlex, Scopus, and WoS. *PLOS ONE* 20(4): e0320347.  
<https://doi.org/10.1371/journal.pone.0320347>

# Which database should I choose?

Things to consider:

- Coverage
- No ground truth, but different angles of it
  - Global or regional?
  - All disciplines
- Data quality



09.45 –  
10.30

## **Accuracy and Beyond: Data Quality of Bibliometric Databases**

Stephan Stahl Schmidt, *German Centre for Higher Education Research and Science Studies (DZHW), Germany*

# Which database should I choose?

Things to consider:

- Coverage
  - Global or regional?
  - All disciplines
- Data quality
- Purpose
- Transparency and interpretability



Thank you!

Any questions?