OpenAlex

An open & complete index of the
global research system

- Scientific Knowledge Graphs (SKGs)
  are vital, but closed and broken
- OpenAlex is an open SKG with excellent coverage
- You should use OpenAlex

OpenAlex

Scientific Knowledge Graphs (SKGs) are essential infrastructure.

What are SKGs?

- Web of Science
- Scopus
- Google Scholar?
- Microsoft Academic Graph
- Dimensions
- Crossref, COCI
- OpenAIRE

OpenAlex

Pay-to-view SKGs are badly broken.

Pay-to-view SKGs are broken

Can't afford them
Can't share them
Can't build on them.

Can't afford them
Can't share them
Can't build on them.

OpenAlex

We **can't afford** toll-access SKGs

Who can't afford them?

We can't afford toll-access SKGs

We're **systematically excluding** less wealthy regions from understanding and managing research.

OpenAlex

We can't afford toll-access SKGs

I wonder if that's why Scopus/WoS have <8% non-english articles? 🤔

OpenAlex

Pay-to-view SKGs are broken

Can't afford them

Can't share them

Can't build on them.

We can't share toll-access SKGs

- transparency
- reproducibility
- integrity

Pay-to-view SKGs are broken

Can't afford them
Can't share them
Can't build on them.

OpenAlex is an <u>open</u> SKG.

OpenAlex is an open SKG

Open source **code**: MIT
Open **data**: CC0
Usably open: API, web UI, snapshot
Sustainably open: POSI nonprofit

OpenAlex

OpenAlex API

OpenAlex API: filter

OpenAlex API: group

URL bar: `//api.openalex.org/works?filter=institutions.ror:https://ror.org/01cwqze88&group_by=oa_status`

```
- group_by: [
    - {
        key: "closed",
        key_display_name: "closed",
        count: 136100
    },
    - {
        key: "bronze",
        key_display_name: "bronze",
        count: 62526
    },
    - {
        key: "green",
        key_display_name: "green",
        count: 49484
    },
    - {
        key: "gold",
        key_display_name: "gold",
        count: 33073
    },
    - {
        key: "unknown",
        key_display_name: "unknown",
        count: 21913
    },
    - {
        key: "hybrid",
        key_display_name: "hybrid",
```

+ view source -

OpenAlex

OpenAlex snapshot

OpenAlex web interface

OpenAlex can be open because it's super cheap to make.

Why is OpenAlex so cheap?

Open MAG data

Open Source ML

Open Access literature

Open PID graphs

OpenAlex

Open PID graphs

Works: Crossref: DOI, citations
Authors: ORCID
Journals: ISSN
Institutions: ROR
Concepts: Wikidata

DOI covers more and more.

Works per year, MAG and Crossref
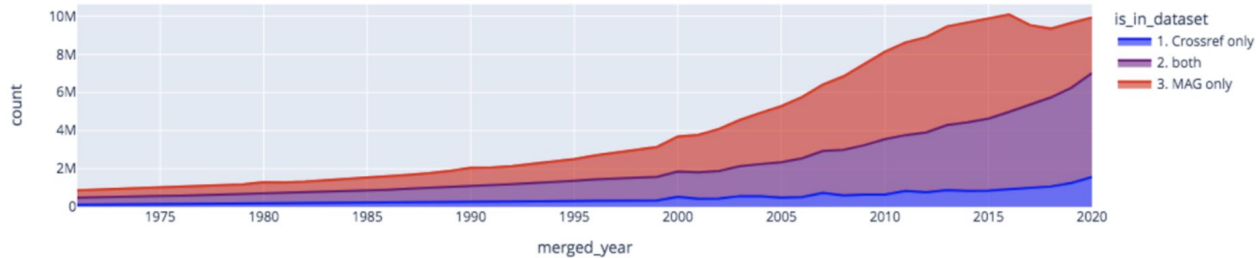
MAG-only works per year

OpenAlex has broader coverage than any other SKG.

SKG coverage (and Google Scholar)

	Works	Citations	Authors	Venues	Institutions
Web of Science (core)	82M	1.8B	17M	24k	11k
Scopus	82M	1.7B	17M	40k	80k
Google Scholar	390M	?	4M?	?	-
MAG	204M	1.7B	206B	49k	27k
Dimensions	148M	1.6B	28k	74k	100k
Crossref/COCI	135M	1.3B	4.5M	104k	1k
OpenAlex	236M	1.8B	243M	124k	108k

SKG **coverage** sources

* https://link.springer.com/article/10.1007/s11192-018-2958-5
* https://link.springer.com/article/10.1007/s11192-021-03948-5
* https://www.crossref.org/06members/53status.html
* https://api.openalex.org
* https://api.crossref.org/works
* https://app.dimensions.ai/discover/publication
* https://clarivate.libguides.com/librarianresources/coverage
* https://api.crossref.org/works?facet=ror-id:*
* queries on closed-source WoS database
* https://www.mdpi.com/2304-6775/9/1/12
* https://twitter.com/albertomartin/status/1534088434427604992
* https://twitter.com/digitalsci/status/1534182383066525696

OpenAlex coverage of the Global South

22k institutions
21M authors
22M works
⅓ of works are in
non-English language

source: our API!

```python
188   def institutions_in_global_south():
189       total = 0
190       for country in GLOBAL_SOUTH_COUNTRIES:
191           country_code = countries.get(country).alpha2
192           r = requests.get(
193               f"https://api.openalex.org/institutions?filter=country_code:{country_code}&mailto=jason@ourresearch.org"
194           )
195           institution_count = r.json()["meta"]["count"]
196           print(
197               f"{institution_count} institutions in {countries.get(country).name} ({country_code})"
198           )
199           total = total + institution_count
200       return total
201
202
203   def works_from_global_south():
204       total = 0
205       for country in GLOBAL_SOUTH_COUNTRIES:
206           country_code = countries.get(country).alpha2
207           r = requests.get(
208               f"https://api.openalex.org/works?filter=authorships.institutions.country_code:{country_code}&mailto=jason@ourresear
209           )
210           works_count = r.json()["meta"]["count"]
211           print(
212               f"{works_count} works from {countries.get(country).name} ({country_code})"
213           )
214           total = total + works_count
215       return total
216
217
218   def authors_from_global_south():
219       total = 0
220       for country in GLOBAL_SOUTH_COUNTRIES:
221           country_code = countries.get(country).alpha2
222           r = requests.get(
223               f"https://api.openalex.org/authors?filter=last_known_institution.country_code:{country_code}&mailto=jason@ourresear
224           )
225           authors_count = r.json()["meta"]["count"]
226           print(
227               f"{authors_count} from {countries.get(country).name} ({country_code})"
228           )
229           total = total + authors_count
230       return total
231
232
233   if __name__ == "__main__":
234       institutions_count = institutions_in_global_south()
235       print(f"{institutions_count:,} institutions in the Global South")
236       # 22,073 institutions in the Global South
237
238       works_count = works_from_global_south()
239       print(f"{works_count:,} works from the Global South")
240       # 22,978,980 works from the Global South
241
242       authors_count = authors_from_global_south()
243       print(f"{authors_count:,} authors with last known affiliation in Global South")
```

OpenAlex

OpenAlex connects five entity types: works, authors, venues, institutions, & concepts.

OpenAlex entities

- **Works**: papers, datasets, software, etc.
- **Authors**: people who *create* works
- **Venues**: journals and repositories that *host* works
- **Institutions**: *affiliations claimed* by authors of works
- **Concepts**: what works are *about*

OpenAlex

OpenAlex Works

236M

DOI, PMID

OpenAlex

OpenAlex Authors

243M

ORCID

OpenAlex Venues

108k
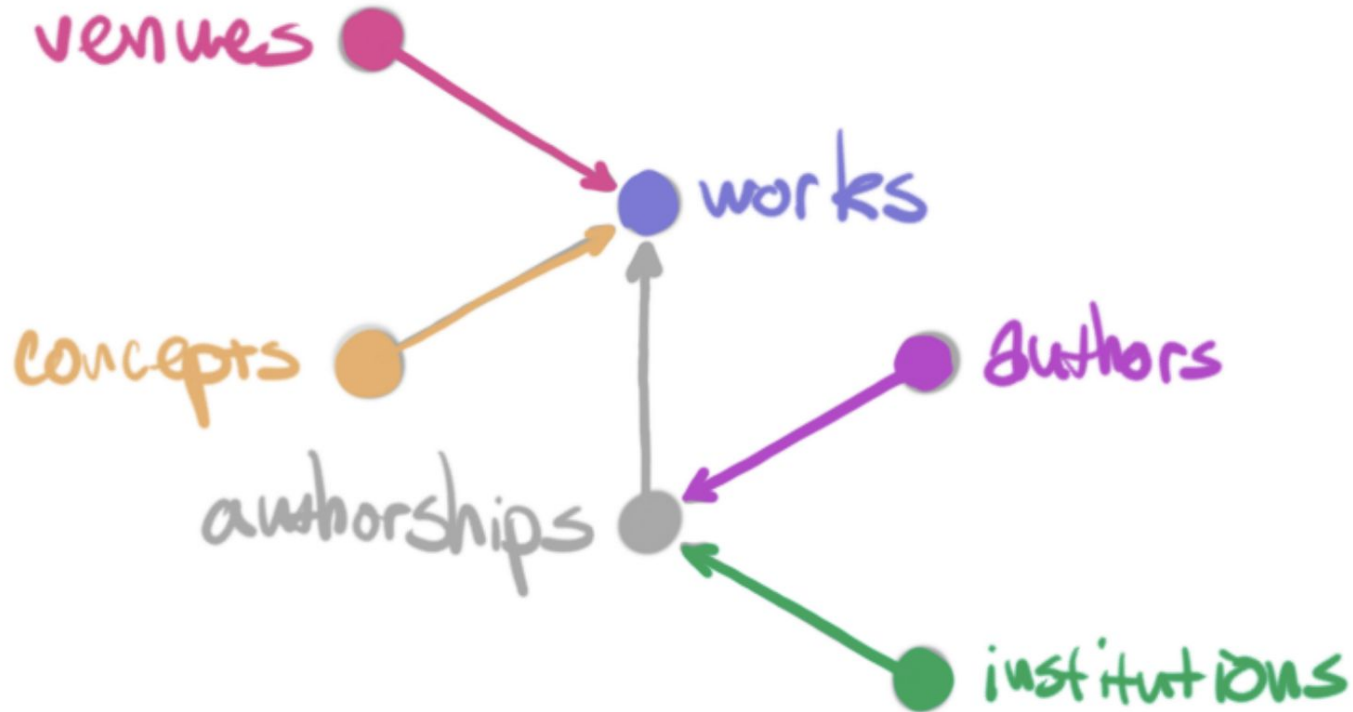
ROR

OpenAlex Institutions

124k

ORCID

60k
Wikidata

OpenAlex

Basic data model

OpenAlex has limitations.

OpenAlex limitations

- **Bias**: inherits from sources, esp Crossref and MAG
- **Research base**: not much, yet.
- **Coverage**: No funders or patents. Limited software, datasets, books.
- **Stability**: Data changing (improving) monthly

OpenAlex is open and ready for use.
Go play with it!