

# OPENALEX: a public infrastructure (SKG) for analysis and monitoring of research activities

Community over commercialization

*Direzione Performance, Assicurazione Qualità, Valutazione e Politiche di Open Science, University of Milan*

## Contents

- The problem of independent authoritative research metadata repository
- What is OpenAlex: is OpenAlex a community tool?
- Use Cases

## The traditional SKGs (e.g. WOS, Scopus)

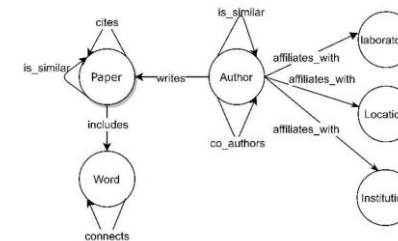
- Contain errors
- Hardly index non-English literature (< 8%)
- Do not index output other than articles
- For many regions of the world are unaffordable (very expensive)
- Data used for science analysis cannot be shared

# Open Data and SKGs

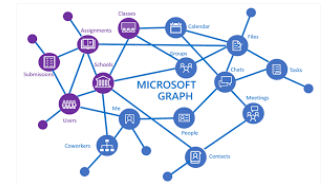
□ The problem

□ What is Open Alex

□ Use Cases



WEB OF SCIENCE™



Scopus®

- Open data - Open Data is a concept that is receiving increased attention and support in academic environments: shared data may be reused in further research!
- Traditional Scientific Knowledge Graphs (SKGs) are important for understanding research, but *closed* (available through subscription) and *broken* (excluding less wealthy regions from understanding and managing research).

Cost examples:

2023 -> Web of science and Journal Citation Reports: about 60 K€

2023 -> SCOPUS (Base + API + VAT) about 46K€

*"When commercial interests are prioritized over those of the communities that research seeks to serve, many concerning issues arise. This year's theme provides an opportunity to join together, take action, and raise awareness around the importance of community control of knowledge sharing systems."*

# Needs (possible) Solutions

We need a repository which is

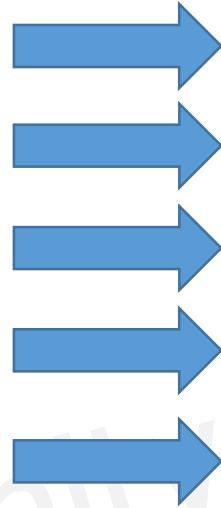
Authoritative

Independent

Complete

Open

Free



OpenAlex is an index of hundreds of millions of interconnected entities across the global research system (<https://openalex.org/about>)

OpenAlex will be used by Leiden University for its next ranking (2024)

OpenAlex is the official source for University of Milan research reports (RAR: <https://www.unimi.it/en/research>)



# Coverage - end of 2022

(source Webinar: [Introducing OpenAlex](#))

The problem

What is Open Alex

Use Cases

	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



**3/10/2023 => Works:245M**

Source: <https://api.openalex.org/counts>

## An open & complete index of the global research system

### Open by default

- Open source code: MIT
- Open data: CC0
- Useably open: API, web UI, snapshot
- Sustainably open: non-profit

Launched in January 2022, provides a free, open, and global catalogue of scholarly papers, researchers, journals, and institutions and their interrelations.

The name “**OpenAlex**” was inspired by the ancient library of Alexandria in Egypt, which maintained the world’s first library catalogue.

OpenAlex is developed and maintained by the non-profit service company OurResearch.



## Main features



- Transparency** – all is well documented
- Free of charge** – no fee is requested
- Reproducibility** – data can be shared
- Incredibly fast** – very efficient

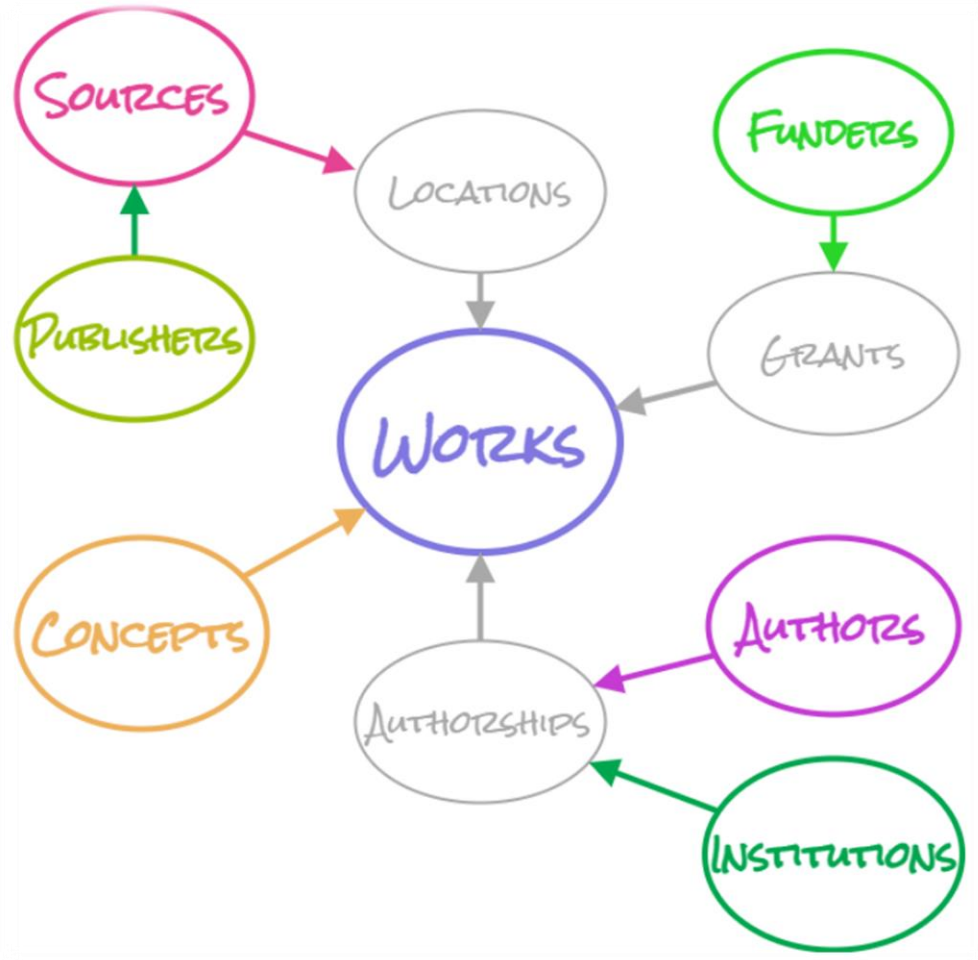
- Integrity** –

- Sources: mainly MAG (Microsoft academic graph) and Crossref, plus ORCID, ROR, DOAJ, Unpaywall, Pubmed, Pubmed Central, The ISSN International Centre, Subject-area and institutional repositories from arXiv to Zenodo and everywhere in between.
- Automatic normalization/deduplication process

- Coverage** – OpenAlex has broader coverage than any other SKGs.

<https://openalex.org/sources>  
Get *all* sources in OpenAlex

# The data model



**Authors** People who create works

**Works** Journal articles, books, datasets, and theses

**Funders** Organizations that fund research

**Publishers** Companies and organizations that distribute works

**Concepts** Topics assigned to works

**Institutions** Universities and other organizations to which authors claim affiliations

**Sources** Journals and repositories that host works

Download

## Some of the metadata available ([docs.openalex.org/](https://docs.openalex.org/))



### Work

Bibliographical [title, publication date/year]  
Related PIDs [ID, DOI, PMID, PMCID, MAG,]  
Access [License, OA Color, OA version, OA Link],  
Authors & Institutions [#, Names, IDs, Corresponding]  
APC [List price, Paid, Paid: provenance]  
Repository locations [Name, Availability, Distinct #]  
Funder [Name, Grant ID]  
Citations [Refenced, Citing, Cited by count]  
Related works, Concepts, SDGs, MeSH, ngrams  
Type, Paratext, Retraction status, Language,  
Source [Name, Type, Publisher, OA, DOAJ]

### Institution

Name [Display, Acronyms, Alternatives], ID, Type  
Location [City, Country, Region, Lat/Lon]  
Repositories, Associated institutions  
Roles (institution, publisher, funder)  
Works [Counts, Citations, Concepts]

### Author

Name (display, alternatives), ID, ORCID,  
Works [Counts, Citations, H-index, i10\_index, Concepts]  
Last known institution

### Source

Name  
ID, ISSN, MAG, Wikidata,  
Hosting organizations, Societies  
OA? In DOAJ?  
APC price & currency  
Works [Counts, Citations, Concepts]

### Concepts

Name  
ID, Wikidata,  
Concept level, Ancestor concepts  
Description (multiple languages), Related concepts  
Works count, Cited by count

# Open Alex – Is it sustainable?

Supported by **OurResearch**, a nonprofit association building tools to help make scholarly research more open, connected, and reusable. OurResearch is the creator of Unpaywall, you can learn more about how Unpaywall started in this *Nature* article, ["How Unpaywall is transforming open science."](#)

**OurResearch** has a ten-year track record of building sustainable scholarly infrastructure, and a formal commitment to sustainability as part of its commitment to sustainability as part of adoption to the POSI principles (<https://openscholarlyinfrastructure.org/>).

*Source: <https://docs.openalex.org/additional-help/faq>*

## Data Quality

OpenAlex uses an automatic normalization and deduplication process, e. g. for author or publisher names. Potential errors can be reported by the users.

Unpaywall is an important source of publication data, so it helps if users make sure that the data is displayed correctly there.

More options for user contributions are planned for the future, such as customization of affiliations.

## Why is Open Alex broader than other SKGs?

Works: Crossref: DOI, citations

22M works

$\frac{1}{3}$  of works are in non-English language

Authors: ORCID

Journals: ISSN

Institutions: ROR

Concepts: Wikidata

## Access to Open Alex

There are three ways to obtain data from OpenAlex:

- Full data dump** (updated fortnightly)
- REST API** (updated daily),
- Web-based GUI** (built on the REST API).

All these are free and require no registration or permission.



## Work in progress: areas for improvement

- Bias: Bias from sources (mainly Crossref and MAG).
- Research base: growing.
- Coverage: No patents.
- Stability: still reaching a maturity model (revision of the model is ongoing)



## Open Alex:

Find number of open access publications UNIMI last 10 years:

[https://openalex.org/works?page=1&filter=open access.is oa%3Atrue,authorships.institutions.lineage%3A189158943,publication\\_year%3A2013-&group\\_by=publication\\_year](https://openalex.org/works?page=1&filter=open%20access.is%20oa%3Atrue,authorships.institutions.lineage%3A189158943,publication_year%3A2013-&group_by=publication_year)

Find number of open access publications ITA last 10 years:

[https://openalex.org/works?page=1&filter=open access.is oa%3Atrue,publication\\_year%3A2013-,institutions.country\\_code%3AIT&group\\_by=publication\\_year](https://openalex.org/works?page=1&filter=open%20access.is%20oa%3Atrue,publication_year%3A2013-,institutions.country_code%3AIT&group_by=publication_year)

## Open Alex:

Find number of open access UNIMI publications last 10 years grouped by SDG

[https://openalex.org/works?page=1&filter=open access.is oa%3Atrue,authorships.institutions.lineage%3AI189158943,publication\\_year%3A2013-&group by=sustainable development goals.id](https://openalex.org/works?page=1&filter=open%20access.is%20oa%3Atrue,authorships.institutions.lineage%3AI189158943,publication_year%3A2013-&group%20by=sustainable%20development%20goals.id)

Find number of open access UNIMI on COVID 19 last 4 years (since 2019)

[https://openalex.org/works?page=1&filter=authorships.institutions.lineage%3AI189158943,publication\\_year%3A2019-,open access.is oa%3Atrue,default.search%3Acovid%2019](https://openalex.org/works?page=1&filter=authorships.institutions.lineage%3AI189158943,publication_year%3A2019-,open_access.is%20oa%3Atrue,default.search%3Acovid%2019)

Or via API:

[https://api.openalex.org/works?page=1&filter=open access.is oa:true,authorships.institutions.lineage:I189158943,default.search:covid+19,publication\\_year:2019-](https://api.openalex.org/works?page=1&filter=open%20access.is%20oa:true,authorships.institutions.lineage:I189158943,default.search:covid+19,publication_year:2019-)

## Open Alex:

- Find all works In 2022 from institutions.ror:<https://ror.org/00wjc7c48>  
(University di Milano)

[https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication\\_year:2022&mailto=name.surname@unimi.it](https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication_year:2022&mailto=name.surname@unimi.it)

- Find all works In 2022 from institutions.ror:<https://ror.org/00wjc7c48>  
(University di Milano) with search words climate change

[https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication\\_year:2022&search=climate%20change&mailto=name.surname@unimi.it](https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication_year:2022&search=climate%20change&mailto=name.surname@unimi.it)

## Open Alex: some experiences with REST API (Green metrics)

- Find all works In 2022 from institutions.ror:https://ror.org/00wjc7c48 (University di Milano) with search words green, sustainability, environment, circular economy, renewable energy, climate change, organic, agroecology

[https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication\\_year:2022&search=green%20OR%20sustainability%20OR%20environment%20OR%20circular%20economy%20OR%20renewable%20energy%20OR%20climate%20change%20OR%20organic%20OR%20agroecology&mailto=name.surname@unimi.it](https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication_year:2022&search=green%20OR%20sustainability%20OR%20environment%20OR%20circular%20economy%20OR%20renewable%20energy%20OR%20climate%20change%20OR%20organic%20OR%20agroecology&mailto=name.surname@unimi.it)

[Link](#)

# Querying data

```
▼ meta:
  count: 5
  db_response_time_ms: 159
  page: 1
  per_page: 200
  results: []
▼ group_by:
  ▼ 0:
    key: "gold"
    key_display_name: "gold"
    count: 3459
  ▼ 1:
    key: "closed"
    key_display_name: "closed"
    count: 2814
  ▼ 2:
    key: "green"
    key_display_name: "green"
    count: 1333
  ▼ 3:
    key: "hybrid"
    key_display_name: "hybrid"
    count: 1220
  ▼ 4:
    key: "bronze"
    key_display_name: "bronze"
    count: 592
  ▼ 5:
    key: "unknown"
    key_display_name: "unknown"
```

Find UNIMI gold green bronze gold open access **2022** from **institutions.ror:https://ror.org/00wjc7c48** grouped by status

[https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication\\_year:2022&group\\_by=oa\\_status](https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication_year:2022&group_by=oa_status)

# Querying data

```
▼ meta:
  count: 5
  db_response_time_ms: 188
  page: 1
  per_page: 200
  results: []
▼ group_by:
  ▼ 0:
    key: "closed"
    key_display_name: "closed"
    count: 62452
  ▼ 1:
    key: "gold"
    key_display_name: "gold"
    count: 53527
  ▼ 2:
    key: "green"
    key_display_name: "green"
    count: 20905
  ▼ 3:
    key: "hybrid"
    key_display_name: "hybrid"
    count: 19281
  ▼ 4:
    key: "bronze"
    key_display_name: "bronze"
    count: 8748
  ▼ 5:
    key: "unknown"
    key_display_name: "unknown"
```

find ITA gold green bronze gold open access 2022 from institutions.ror:https://ror.org/00wjc7c48 grouped by status

[https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication\\_year:2022&group by=oa status](https://api.openalex.org/works?filter=institutions.ror:https://ror.org/00wjc7c48,publication_year:2022&group_by=oa_status)