OpenAlex

An open & complete index of the global research ecosystem

Today's Agenda

Introducing OpenAlex

Why you should use it

- major features
- how it compares

How you can start using it

- access methods
- form & function of the database

Q&A

Scientific Knowledge Graphs (SKGs) are essential infrastructure.

- research discovery
- scientometrics
- research intelligence and assessment

What are SKGs?

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

OpenAlex is an open SKG.
And that's a major feature.

Limitations of Pay-to-view SKGs

Their subscriptions are costly
Their (your) results cannot be shared
You can't build on them
You inherit their exclusiveness

Their subscriptions are costly

- **Pressure on budgets** is intensifying at Universities (esp. library)
- Paywalls **systematically exclude** less wealthy regions
- After paying for subscription, your **access is limited**

OpenAlex is free, enabling equitable access across the globe

Their (your) results cannot be shared

- **transparency** in decision-making
- **reproducibility** of meta-research

Because **OpenAlex is completely open**, anyone can examine and replicate analyses and scenario-play factors influencing decisions

You can’t build on them

Possibilities limited by closed databases & their licenses:

- access to full datasets
- commercial uses of data
- integrations with internal or external dashboards
- development of derivative tools

OpenAlex data and codes are under CC0, anyone can examine and use however they wish without lawyers

You inherit their exclusiveness

Exclusiveness criteria that create biases:
- must have an English abstract
- publication status
- theses/dissertations excluded
- type of peer review
- journal has relatively few citations
- geographical diversity of authors

OpenAlex does not apply indexing criteria so **you can pick which data to include** for your purposes

OpenAlex has broader coverage than any other SKG.

SKG content coverage

	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

OpenAlex indexes works & associated metadata

Sources [232k]

Concepts [65k]

Publishers [10k]

Institutions [102k]

Works [243M]

Funders [32k]

Authors [91M]

https://api.openalex.org/counts

OpenAlex

Some of the metadata available (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]

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]

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

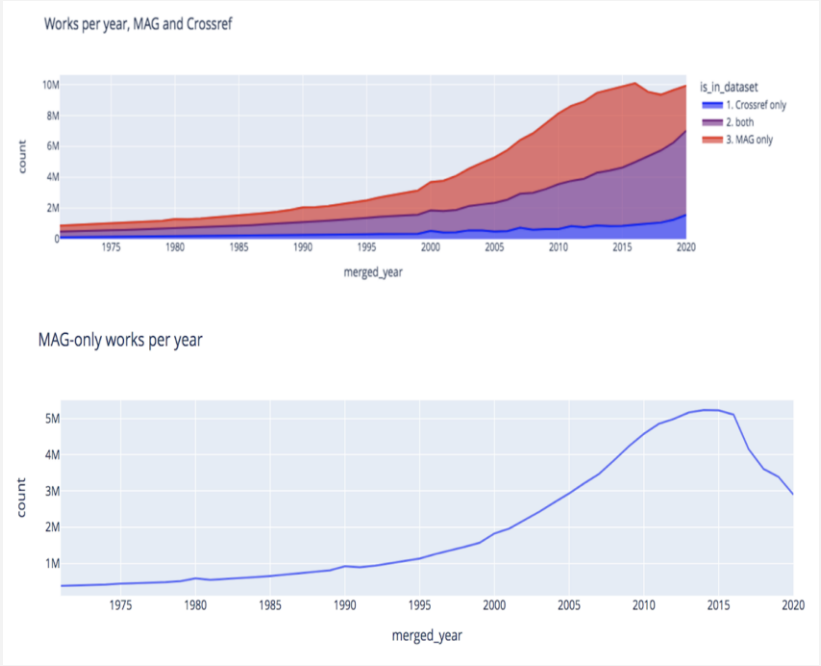
Concepts
Name
ID, Wikidata,
Concept level, Ancestor concepts
Description (multiple languages), Related concepts
Works count, Cited by count

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

- Works: DOI
- Citations: Crossref
- Authors: ORCID
- Journals: ISSN
- Institutions: ROR
- Concepts: Wikidata



Surprised it's free and better?
See what others have to say.

Testimonials

“I am the main developer of VOSviewer, one of the most popular software tools worldwide for visualizing scientific literature based on bibliographic data. VOSviewer supports a large number of bibliographic databases. However, most of these databases have important limitations:

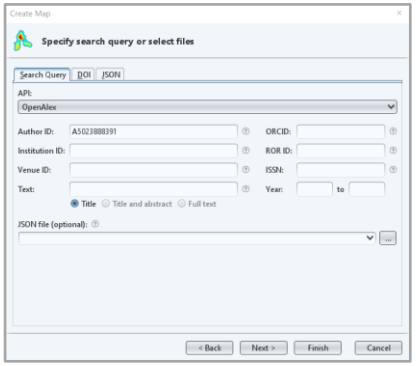
They require an expensive subscription (Web of Science, Scopus),

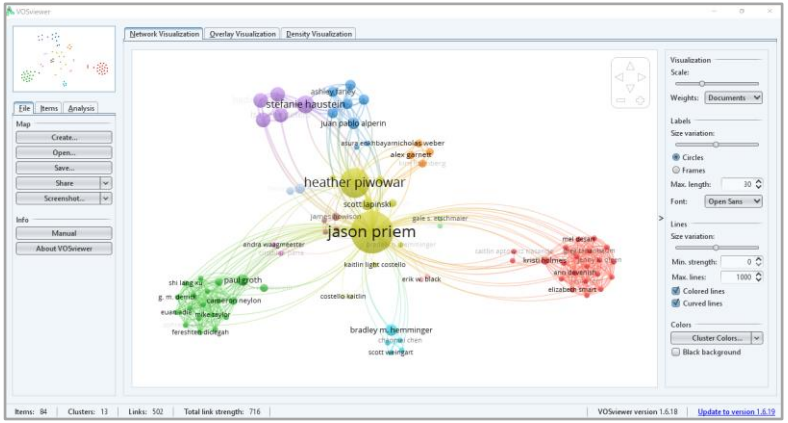
- their coverage of the scientific literature is limited (Web of Science, Scopus),
- they allow only small amounts of data to be exported (Web of Science, Scopus, Dimensions),
- they are restricted to specific disciplines (PubMed, Europe PMC),
- there are major gaps in their data (PubMed, Europe PMC, Crossref, OpenCitations), and/or
- downloading data is very slow (Crossref, OpenCitations).

OpenAlex is of crucial importance for VOSviewer users because it offers a better performance than other databases on all the above-mentioned criteria.

VOSviewer users regularly run into problems because of the limitations of bibliographic databases. When they ask me for help, I often refer them to OpenAlex as an alternative database that is likely to offer a solution to their problem. I consider OpenAlex to be a fundamental building block for an ecosystem of open infrastructures for high-quality research analytics.”

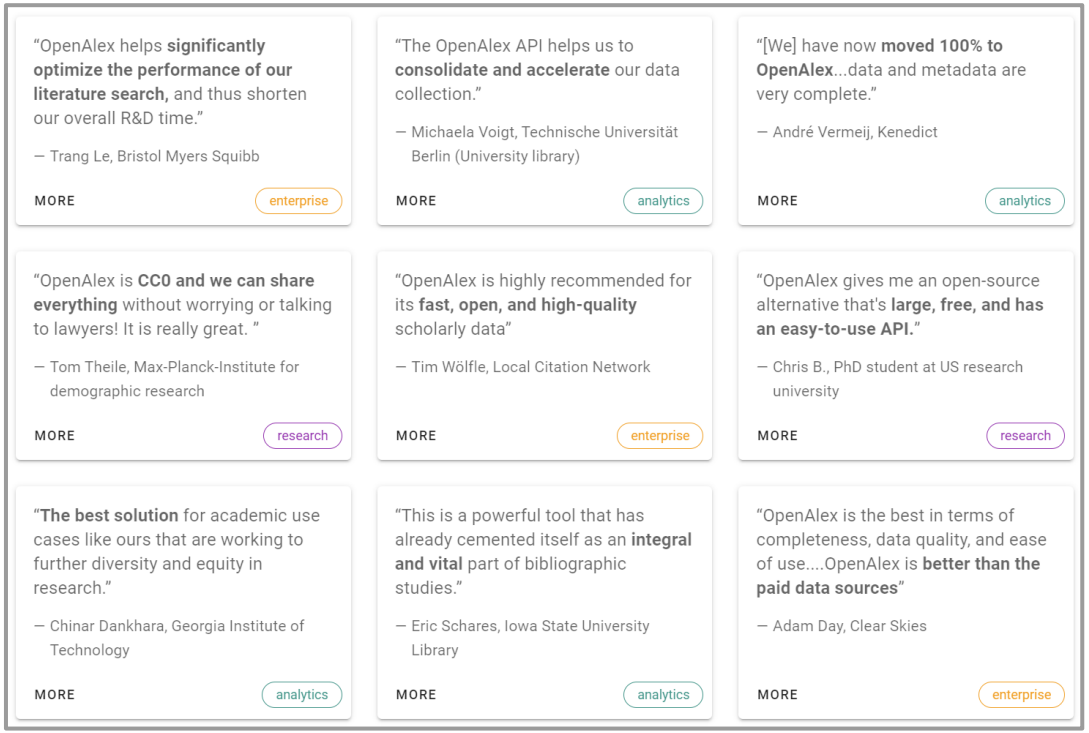
— Nees Jan van Eck, Centre for Science and Technology Studies (CWTS), Leiden University





OpenAlex

Testimonials



"OpenAlex helps **significantly optimize the performance of our literature search,** and thus shorten our overall R&D time."

— Trang Le, Bristol Myers Squibb

MORE enterprise

"The OpenAlex API helps us to **consolidate and accelerate** our data collection."

— Michaela Voigt, Technische Universität Berlin (University library)

MORE analytics

"[We] have now **moved 100% to OpenAlex**...data and metadata are very complete."

— André Vermeij, Kenedict

MORE analytics

"OpenAlex is **CC0 and we can share everything** without worrying or talking to lawyers! It is really great. "

— Tom Theile, Max-Planck-Institute for demographic research

MORE research

"OpenAlex is highly recommended for its **fast, open, and high-quality** scholarly data"

— Tim Wölfle, Local Citation Network

MORE enterprise

"OpenAlex gives me an open-source alternative that's **large, free, and has an easy-to-use API.**"

— Chris B., PhD student at US research university

MORE research

"**The best solution** for academic use cases like ours that are working to further diversity and equity in research."

— Chinar Dankhara, Georgia Institute of Technology

MORE analytics

"This is a powerful tool that has already cemented itself as an **integral and vital** part of bibliographic studies."

— Eric Schares, Iowa State University Library

MORE analytics

"OpenAlex is the best in terms of completeness, data quality, and ease of use....OpenAlex is **better than the paid data sources**"

— Adam Day, Clear Skies

MORE enterprise

More online at:

openalex.org/testimonials

OpenAlex

OpenAlex does have limitations.

OpenAlex limitations

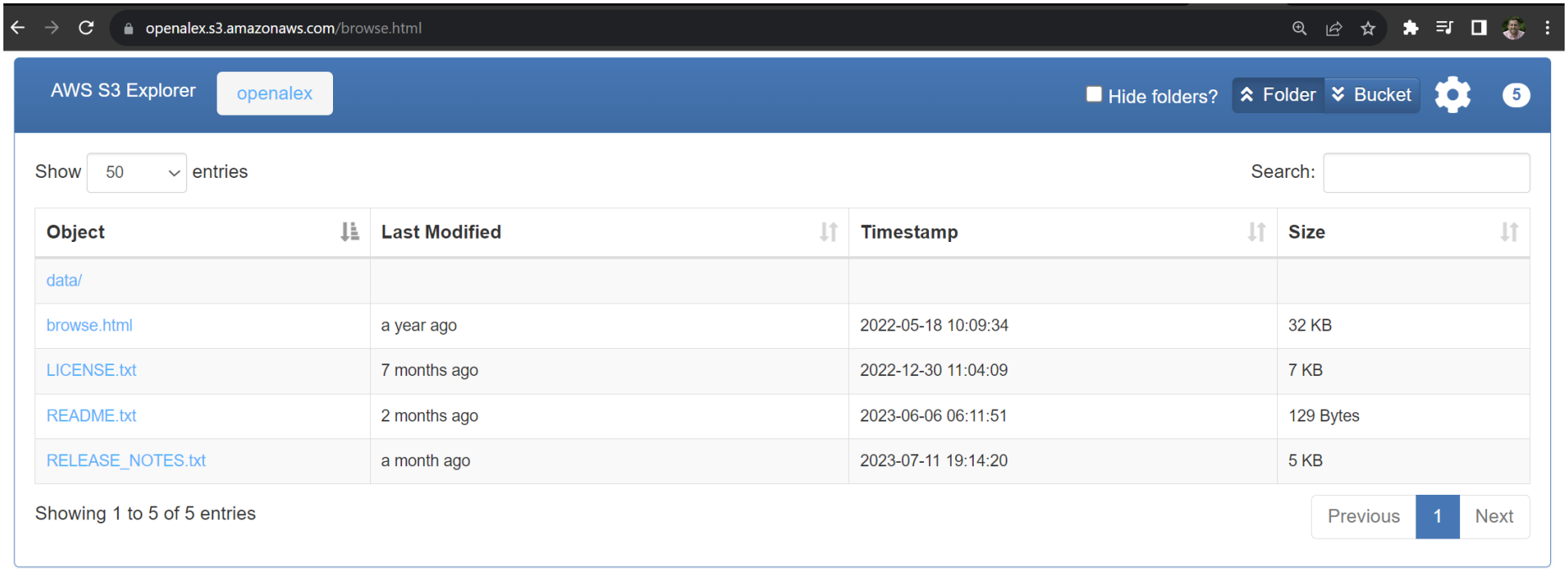
- **Bias**: Inherits from sources, esp Crossref and MAG
- **Research base**: Not much, but rapidly growing
- **Coverage**: No patents. Limited software, datasets
- **Stability**: Data changing (improving) monthly

OpenAlex History

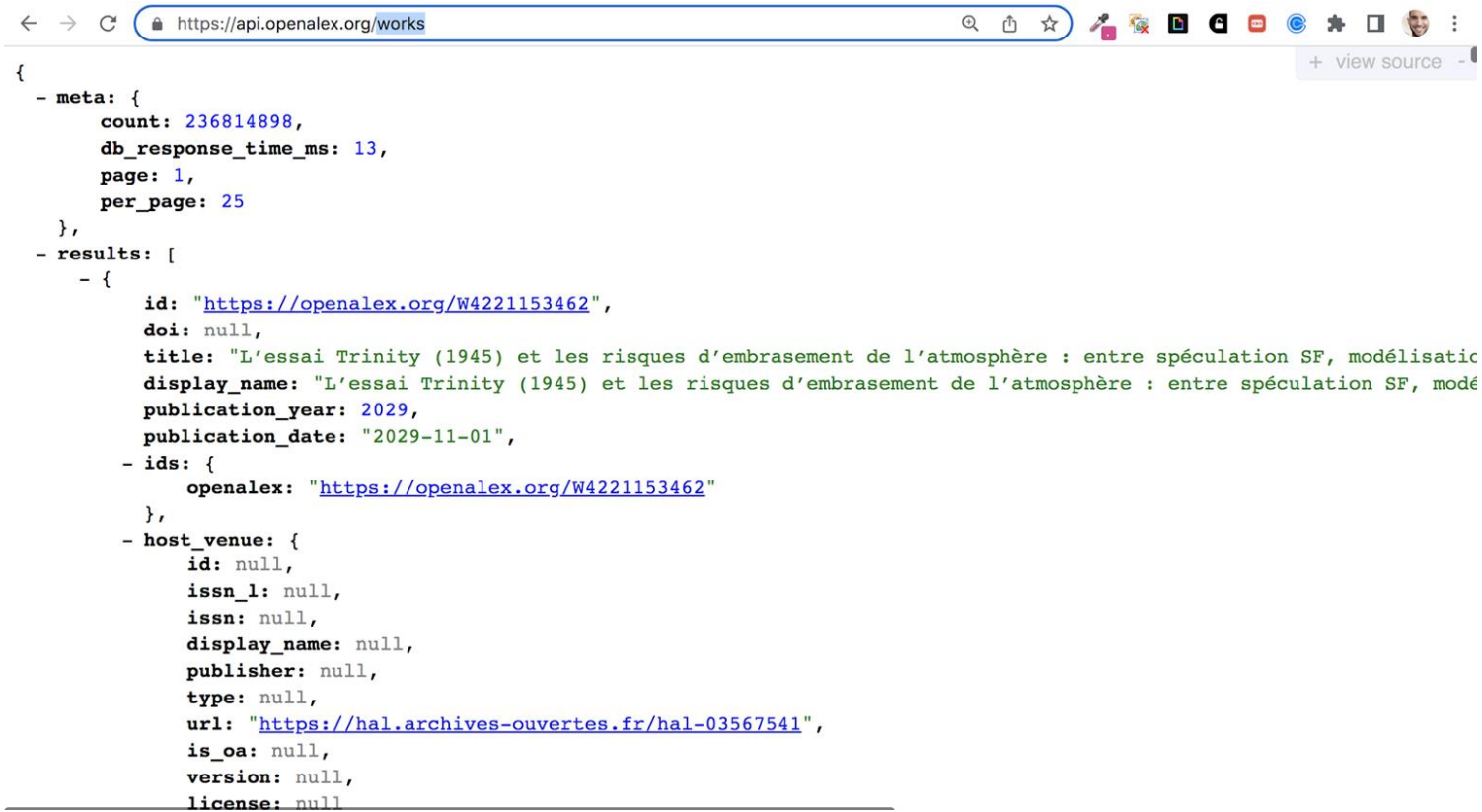
May 2021- Microsoft announced MAG sunsetting
Dec 2021- MAG discontinued
Jan 2022- OpenAlex beta launched
May 2022- User Group launched
August 2022- Full text search
December 2022- Customer support ticket system
March 2023- Premium offering launched
July 2023- Improved author disambiguation launched

How you can access OpenAlex.

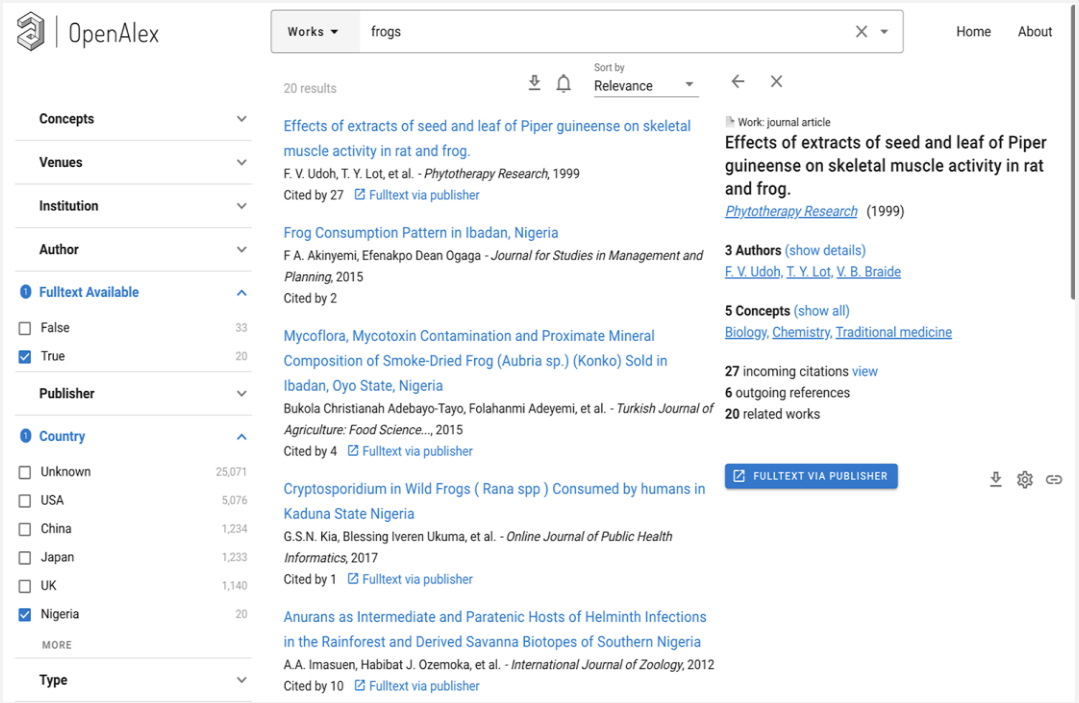
OpenAlex snapshot



OpenAlex API



OpenAlex web interface- in Beta!

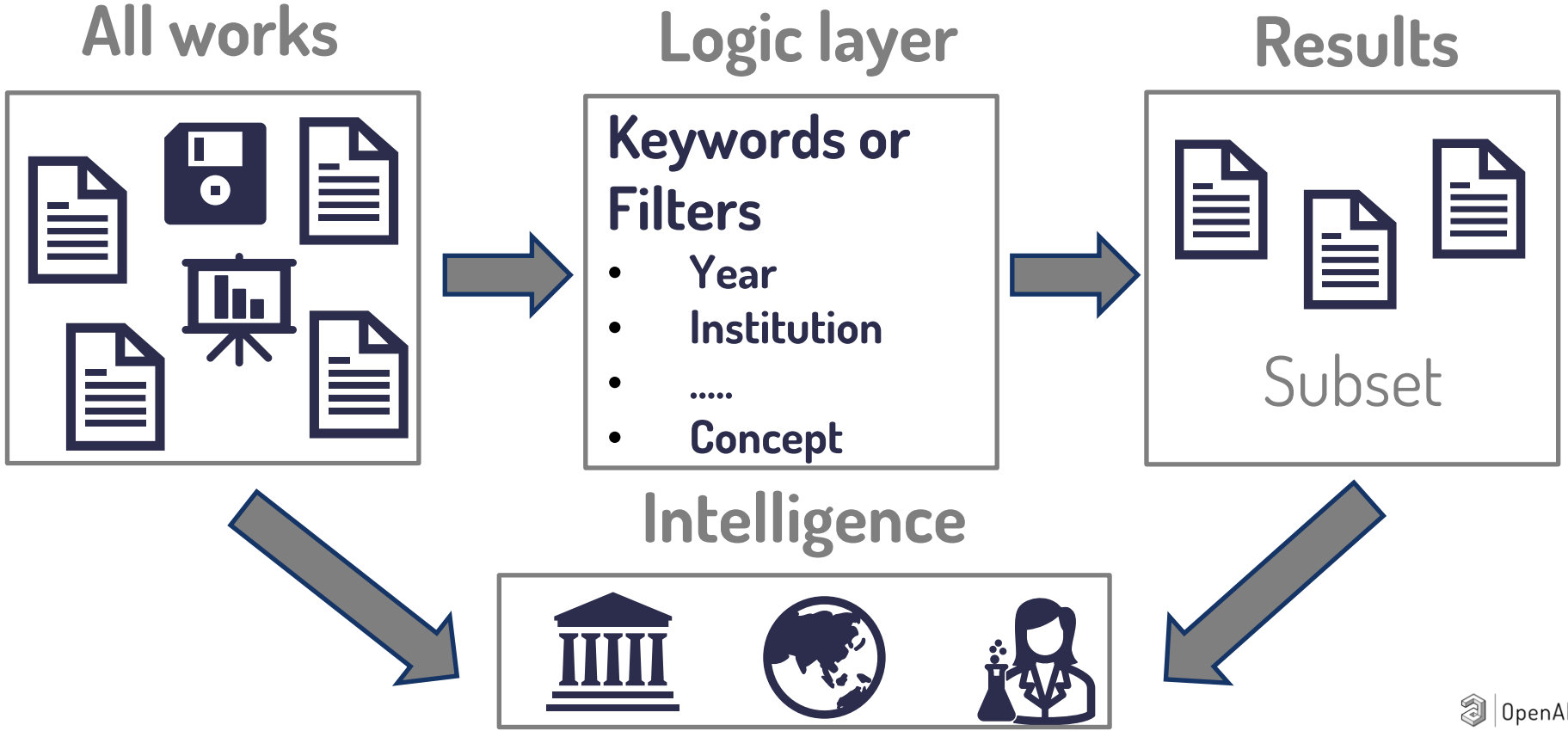


GUI for accessing specific subsets of data, building API calls, and exporting data

Walk-through coming soon!

Functions enabled by the content and metadata structure.

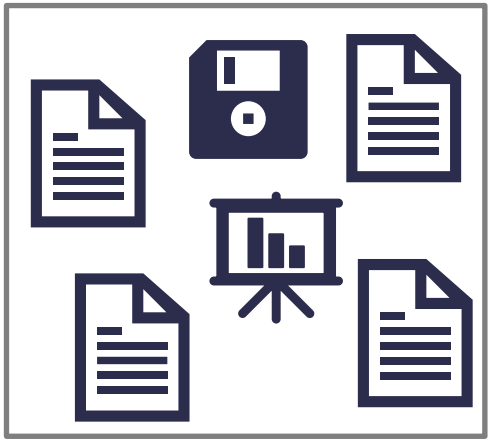
Subsetting & Analysing the Data



Find specific works

e.g., Open access journal publications mentioning kelp with an author from Canada that was funded by NSERC

All works

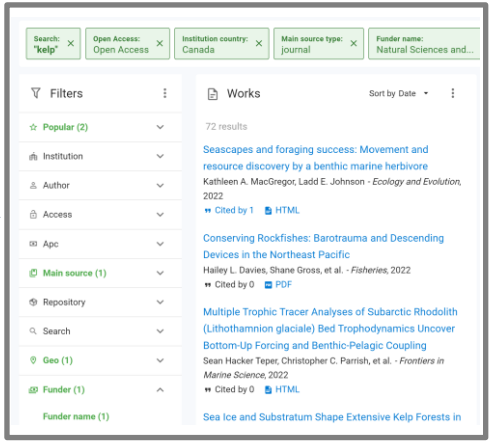


Logic layer

Keywords or Filters

- "kelp"
- OA
- Canada
- journal
- Funded by NSERC

Results

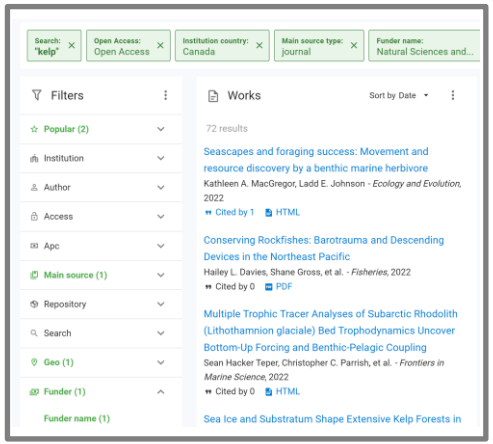


OpenAlex

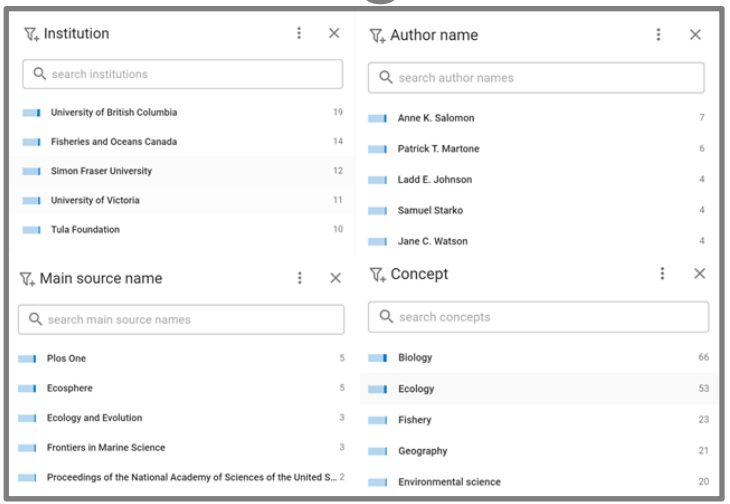
Common intelligence use cases

e.g., Open access journal publications mentioning kelp with an author from Canada that was funded by NSERC

Results

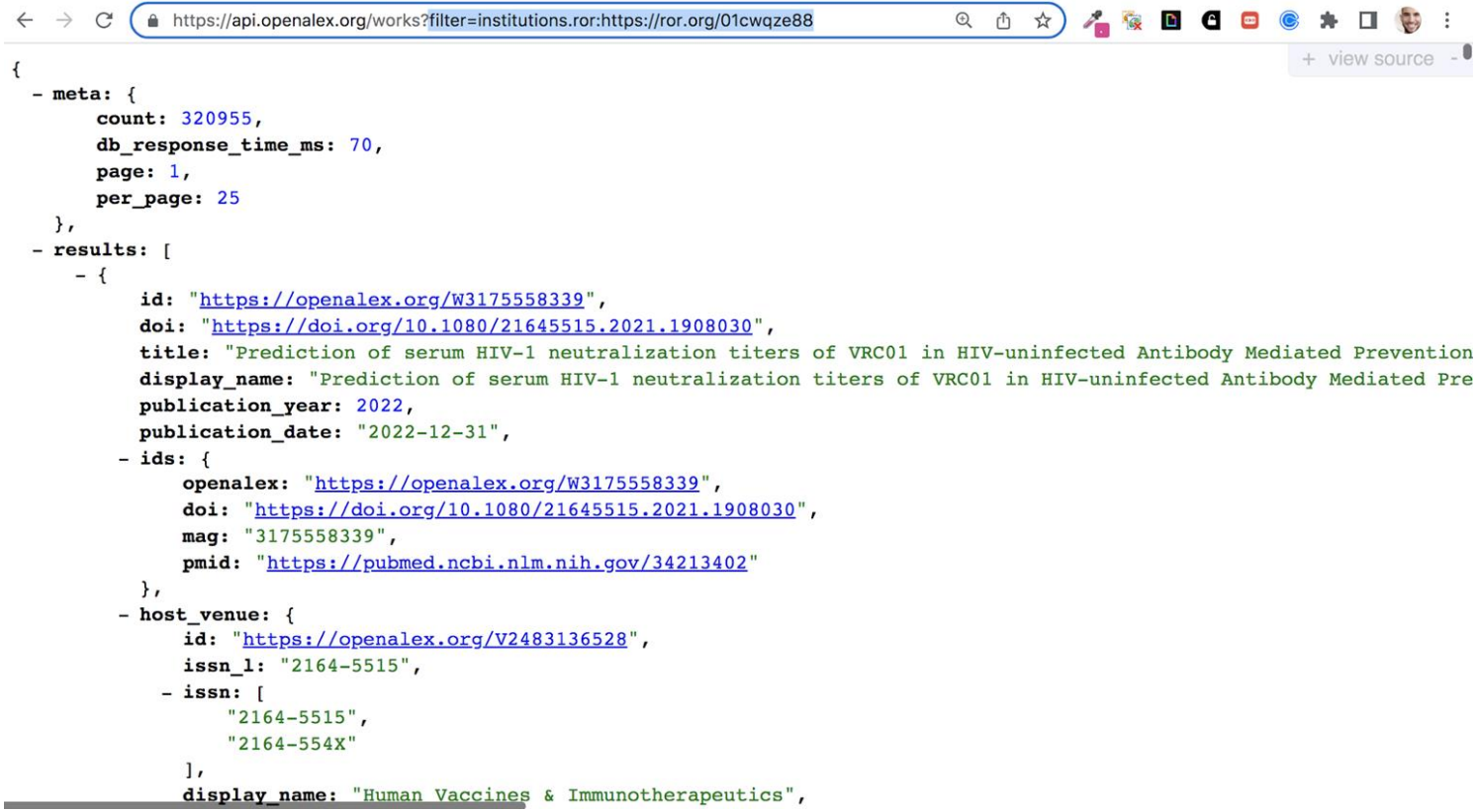


Intelligence

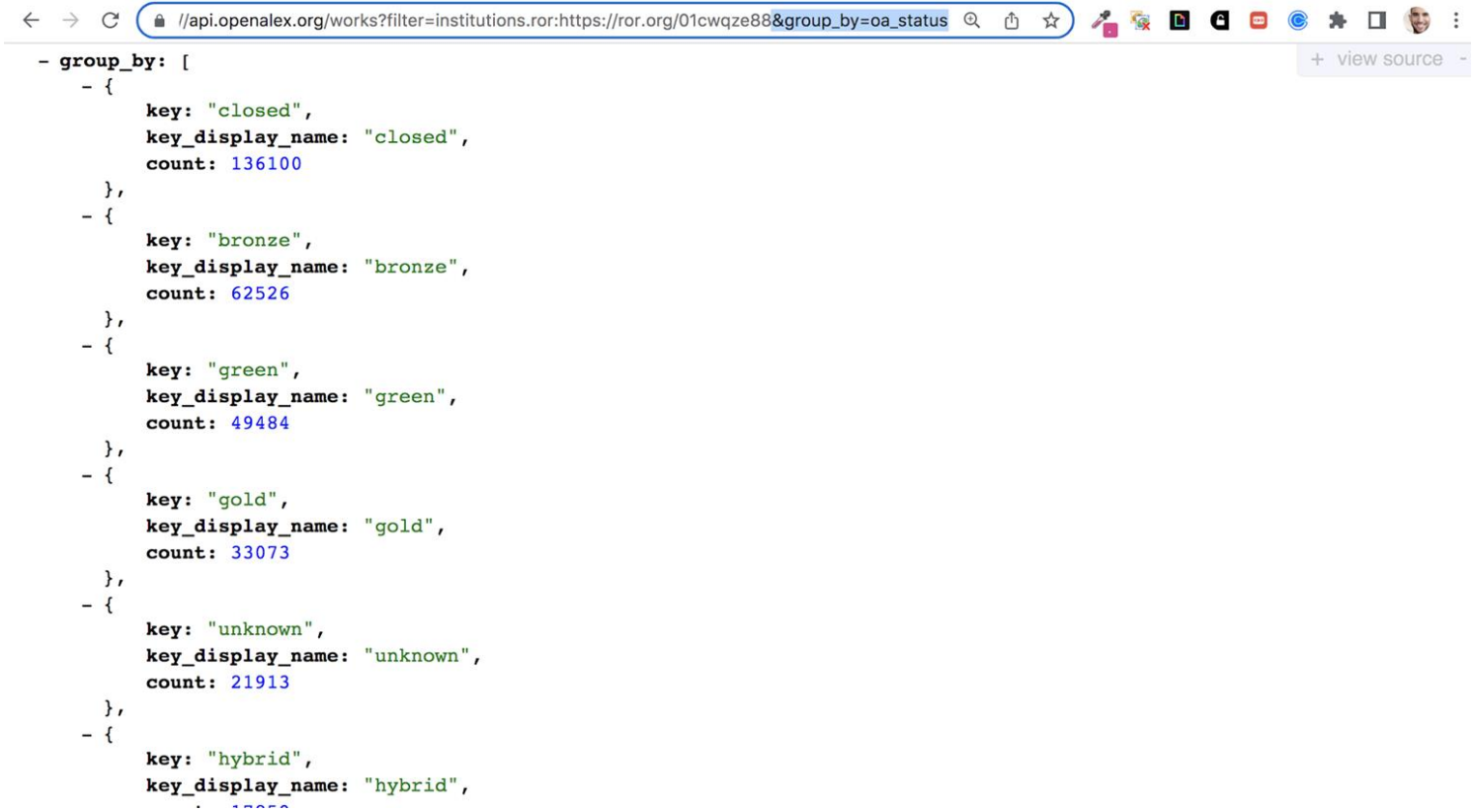


OpenAlex

OpenAlex API: filter



OpenAlex API: group



Additional Resources

OpenAlex Resources
documentation: https://docs.openalex.org/
tutorial: https://docs.openalex.org/quickstart-tutorial
google user group: https://groups.google.com/g/openalex-users
help tickets: https://openalex.org/help
upcoming webinars: https://openalex.org/webinars

Other Resources
R Library to interface with OpenAlex APIs: https://docs.ropensci.org/openalexR/

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

https://openalex.org/feedback

SKG content 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
def institutions_in_global_south():
    total = 0
    for country in GLOBAL_SOUTH_COUNTRIES:
        country_code = countries.get(country).alpha2
        r = requests.get(
            f"https://api.openalex.org/institutions?filter=country_code:{country_code}&mailto=jason@ourresearch.org"
        )
        institution_count = r.json()["meta"]["count"]
        print(
            f"{institution_count} institutions in {countries.get(country).name} ({country_code})"
        )
        total = total + institution_count
    return total


def works_from_global_south():
    total = 0
    for country in GLOBAL_SOUTH_COUNTRIES:
        country_code = countries.get(country).alpha2
        r = requests.get(
            f"https://api.openalex.org/works?filter=authorships.institutions.country_code:{country_code}&mailto=jason@ourresear
        )
        works_count = r.json()["meta"]["count"]
        print(
            f"{works_count} works from {countries.get(country).name} ({country_code})"
        )
        total = total + works_count
    return total


def authors_from_global_south():
    total = 0
    for country in GLOBAL_SOUTH_COUNTRIES:
        country_code = countries.get(country).alpha2
        r = requests.get(
            f"https://api.openalex.org/authors?filter=last_known_institution.country_code:{country_code}&mailto=jason@ourresear
        )
        authors_count = r.json()["meta"]["count"]
        print(
            f"{authors_count} from {countries.get(country).name} ({country_code})"
        )
        total = total + authors_count
    return total


if __name__ == "__main__":
    institutions_count = institutions_in_global_south()
    print(f"{institutions_count:,} institutions in the Global South")
    # 22,073 institutions in the Global South

    works_count = works_from_global_south()
    print(f"{works_count:,} works from the Global South")
    # 22,978,980 works from the Global South

    authors_count = authors_from_global_south()
    print(f"{authors_count:,} authors with last known affiliation in Global South")
```