

Open Science Factsheet No. 9

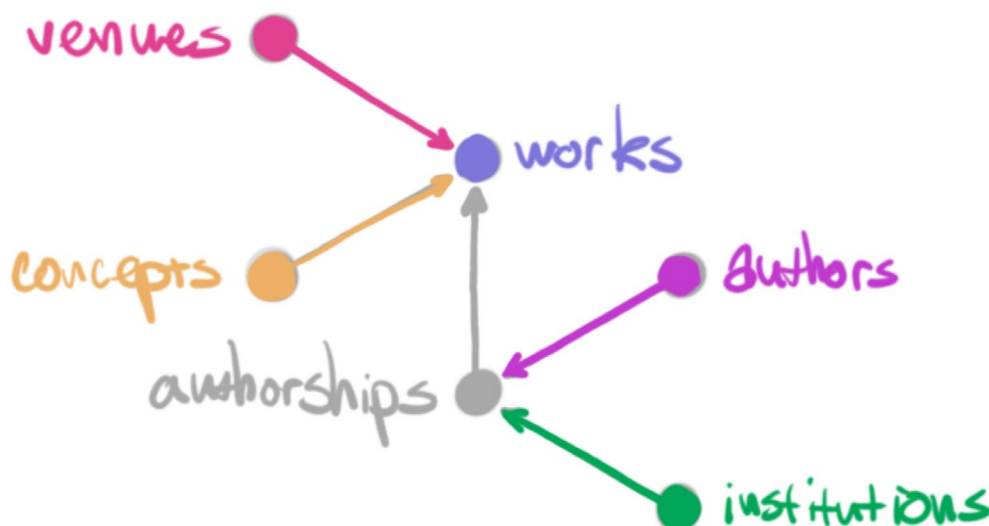
based on the 64th online seminar: OpenAlex

Background

- The new open source index, [OpenAlex](#), launched in January 2022, provides a free, open, and global catalog 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 catalog](#).
- The metadatabase OpenAlex offers a free alternative to subscription-based platforms such as Scopus, Dimensions, or Web of Science.
- OpenAlex is developed and maintained by the non-profit service company [OurResearch](#).

Data Types

The OpenAlex dataset describes scholarly entities and how those entities are connected to each other. There are five types of entities:



OpenAlex, <https://docs.openalex.org/about-the-data>
(Last accessed on June 17, 2022)

1. Works are papers, books, datasets, etc; they cite other works
2. Authors are people who create works
3. Venues are journals and repositories that host works
4. Institutions are universities and other orgs that are affiliated with works (via authors)
5. Concepts tag Works with a topic (via AI technology)

Useful tip: License information can be queried directly via the API.

Data Sources

The two most important data sources for OpenAlex are Crossref and the discontinued Microsoft Academic Graph (MAG). Other sources include:

- ORCID
- ROR
- DOAJ
- Unpaywall
- Pubmed
- Pubmed Central
- The ISSN International Centre
- Web crawls
- Subject-area and institutional repositories, including arXiv and Zenodo
- Wikidata, mainly for concept annotations
- Integration of DataCite DOIs is planned, inter alia to enhance the coverage of datasets

Data Reuse and Access

- OpenAlex data can be used to build scholarly search engines, recommender services, or knowledge graphs. It can help manage research by tracking citation impact, spotting emerging research areas, and identifying and promoting work from underrepresented groups. It can also be used for bibliometric research, science and technology studies, or for other kinds of research on research itself.
- The catalog is freely accessible via [REST API](#).
- OpenAlex is based on a fully Open Source [codebase](#).
- The release of a web-based graphical user interface and the introduction of an alert feature are planned for the summer of 2022.
- Helpful tips and tricks for using OpenAlex are offered via Twitter: [OpenAlex Tips of the Day](#)

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.
- Currently, 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.

Our speaker

Dr. Heather Piwowar



[Mail](#) | [ORCID iD](#) | [Web](#)

Twitter: [@researchremix](#)