OPEN METADATA SOURCES

COMPARING OPENALEX TO CROSSREF

DATE: 03 JULY 2022

[PRELIMINARY VERSION]

Executive Summary

In January 2022, OpenAlex was launched as a source of open bibliographic metadata. Intended both as a replacement of and improvement on Microsoft Academic, it provides structured data on publications, authors, institutions and publication venues.

In this project, we assess and compare the value added by OpenAlex to Crossref metadata, both in coverage of publications and other research output (with and without DOIs) as well as in coverage of metadata (including identifiers) for authors, institutions, publication venues and disciplines.

The report currently contains all the graphs comparing metadata coverage of OpenAlex compared to Crossref, and of DOIs vs non-DOIs in OpenAlex, as well as some basic tables. More explanatory text and interpretation of findings will be added in a later version.

Complete data and code are available on Github:
https://github.com/Curtin-Open-Knowledge-Initiative/open-metadata-report
All images and data belonging to this report are located in the directory reports\run_20220703_2 in this repository.

Introduction and Background

In January 2022, OpenAlex was launched as a source of open bibliographic metadata. Intended both as a replacement of and improvement on Microsoft Academic, it provides structured data on publications, authors, institutions and publication venues.

Many tools, projects and services relied on Microsoft Academic as source of largely open metadata, and might consider switching to OpenAlex. More broadly, the launch of OpenAlex has increased interest in the potential of open metadata to enable discovery, linking and integration of data on research processes and outputs.

Unlike metadata from closed sources, open metadata can be combined and enriched to provide a rich open metadata landscape. Transparency and provenance allow identifying and addressing existing gaps and biases in coverage and quality.

In this project, we assess and compare the value added by OpenAlex to Crossref metadata, both in coverage of publications and other research output (with and without DOIs) as well as in coverage of metadata (including identifiers) for authors, institutions, publication venues and disciplines.

Data sources

This report was run using the following tables as source data:

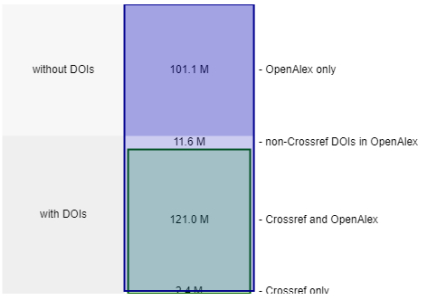
- Crossref: academic-observatory.crossref.crossref_metadata20220507
- OpenAlex: academic-observatory.openalex.Work with date 20220612

Complete data and code are available on Github:
https://github.com/Curtin-Open-Knowledge-Initiative/open-metadata-report
All images and data belonging to this report are located in the directory reports\run_20220703_2 in this repository.

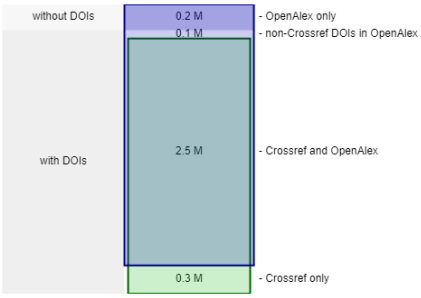
Coverage of OpenAlex vs Crossref

Comparing coverage

Overview

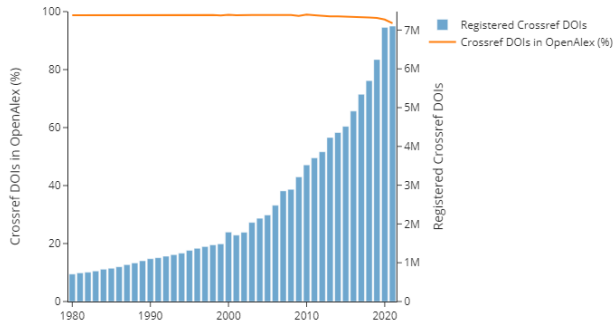


overall comparison - all time

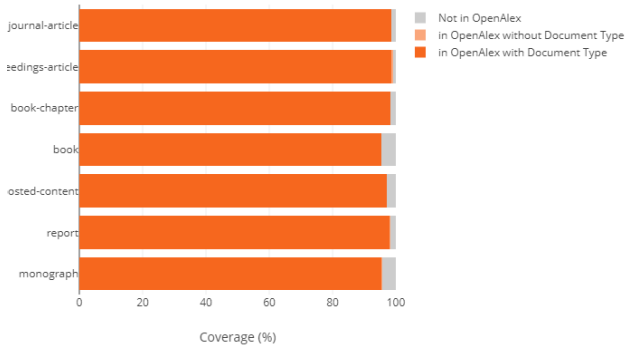


overall comparison - 2022

By year and publication type



coverage by publication date - all time

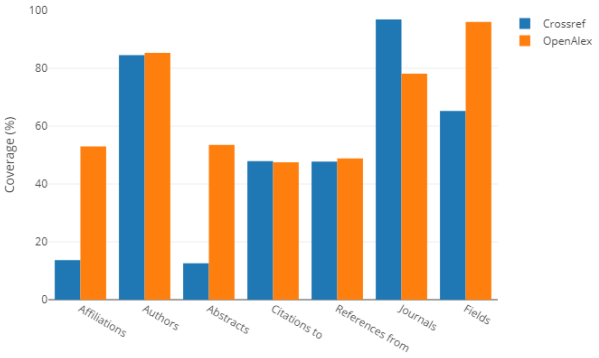


coverage by publication type - all time

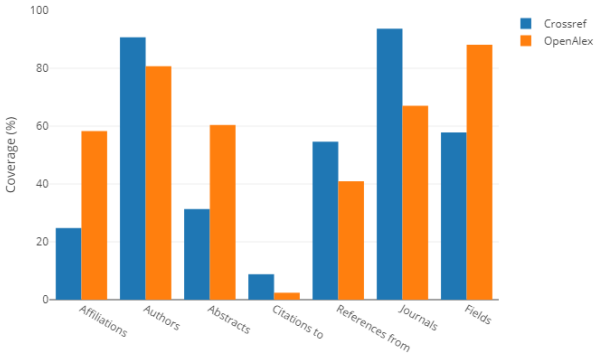
Value Add of OpenAlex to Crossref

Overview

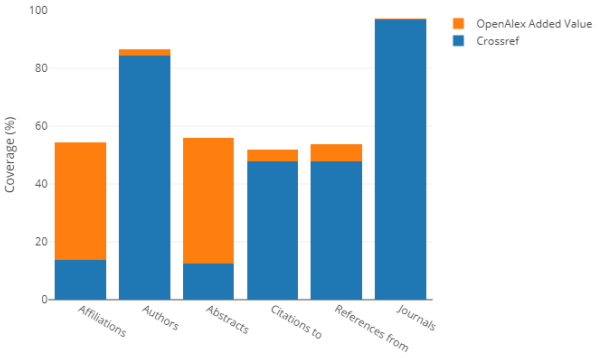
Comparing coverage of metadata types in Crossref and OpenAlex



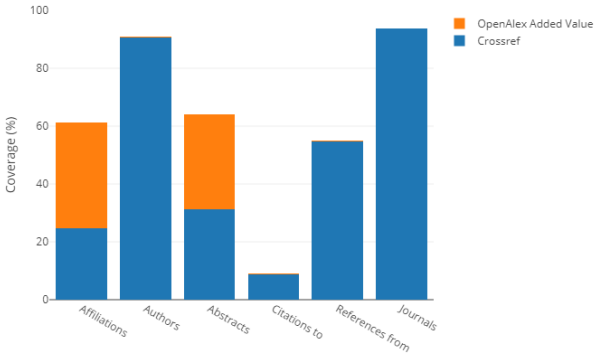
coverage comparison - all time



coverage comparison - 2022



coverage added value - all time

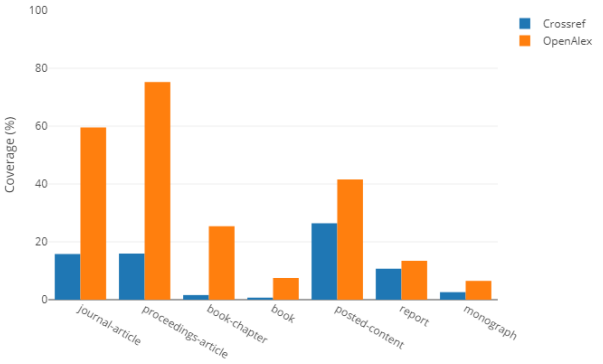


coverage added value - 2022

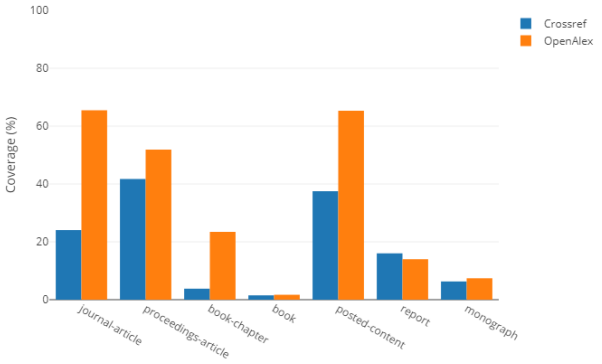
Details

Metadata coverage in OpenAlex and Crossref by publication type

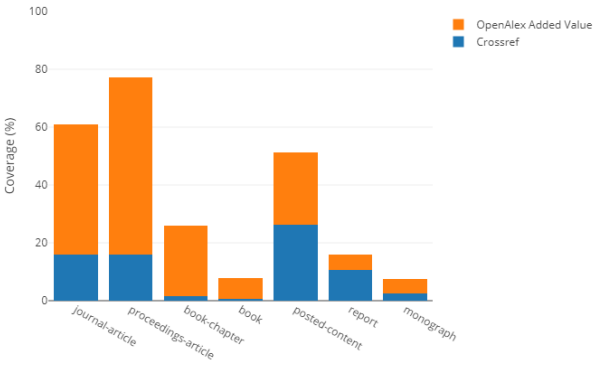
Affiliations



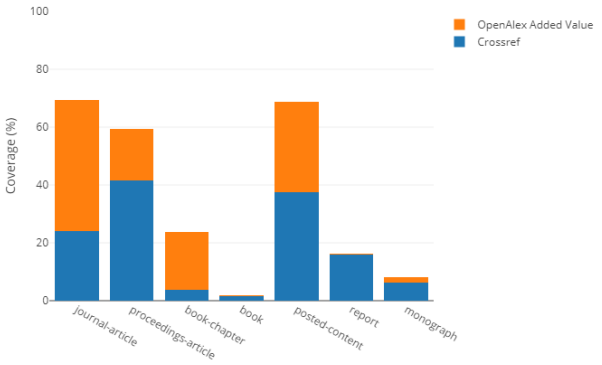
coverage comparison - all time



coverage comparison - 2022

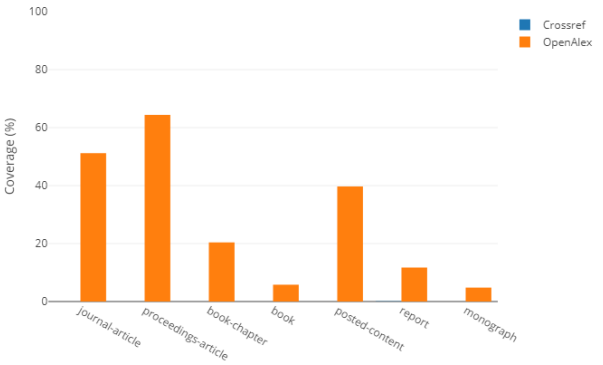


coverage added value - all time

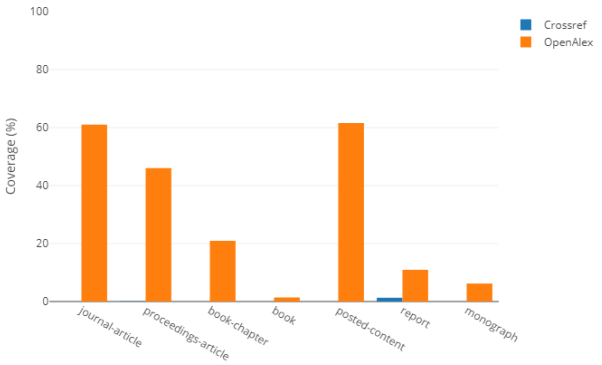


coverage added value - 2022

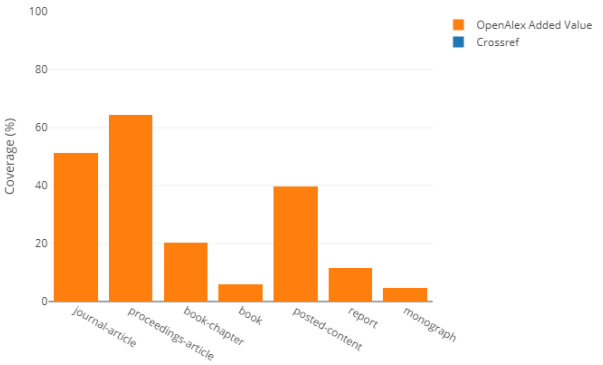
Affiliations ROR



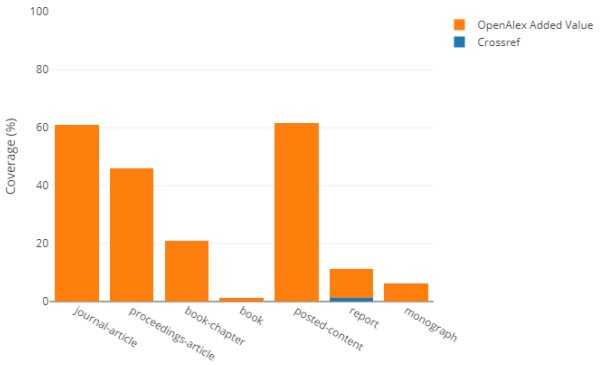
coverage comparison - all time



coverage comparison - 2022

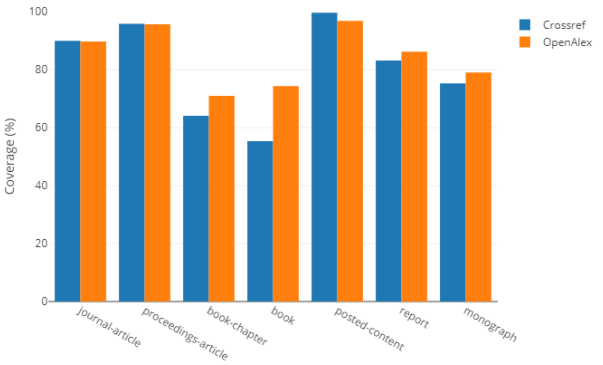


coverage added value - all time

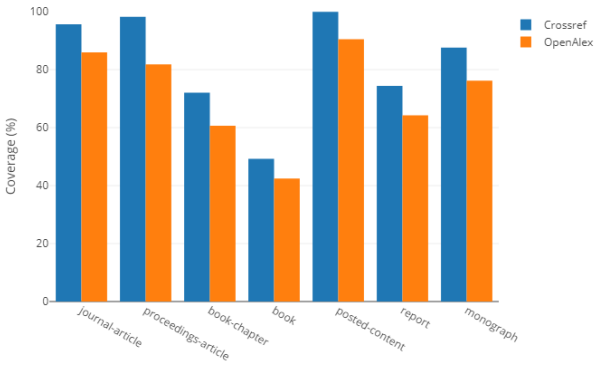


coverage added value - 2022

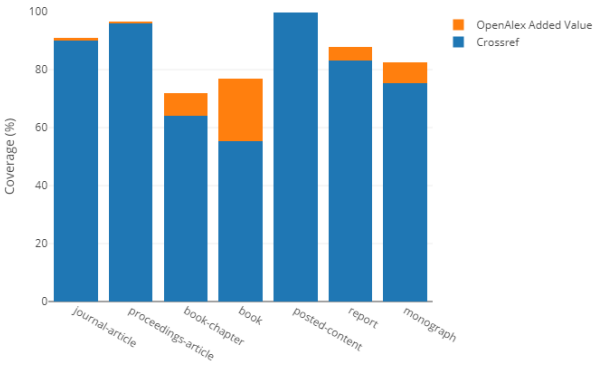
Authors



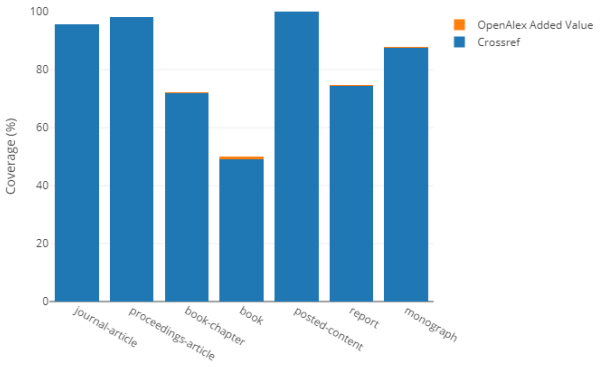
coverage comparison - all time



coverage comparison - 2022

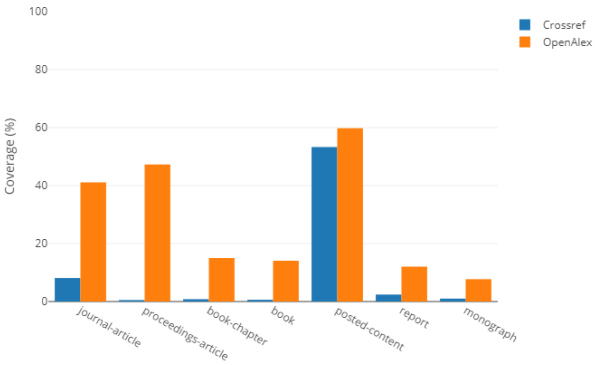


coverage added value - all time

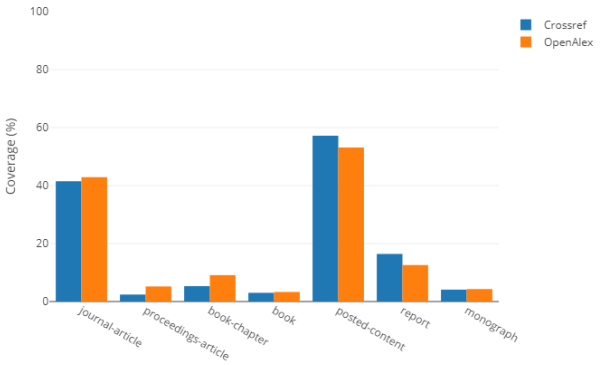


coverage added value - 2022

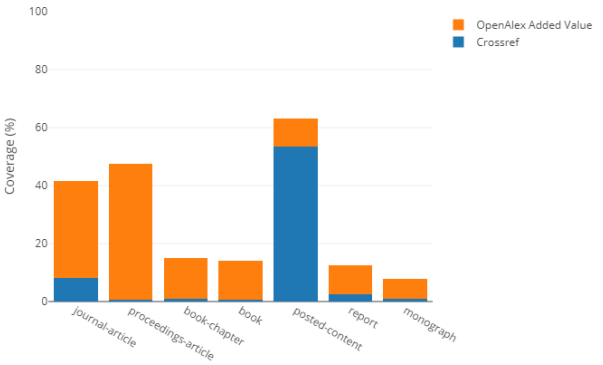
Authors ORCIDs



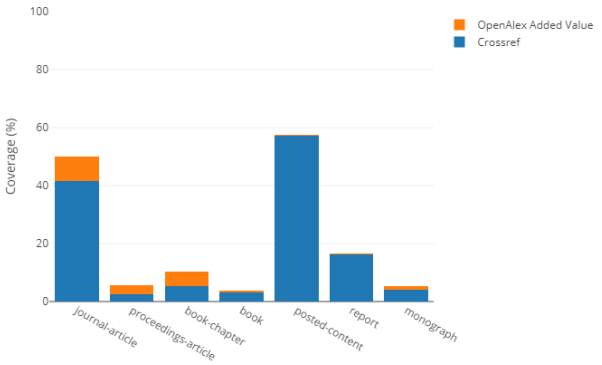
coverage comparison - all time



coverage comparison - 2022

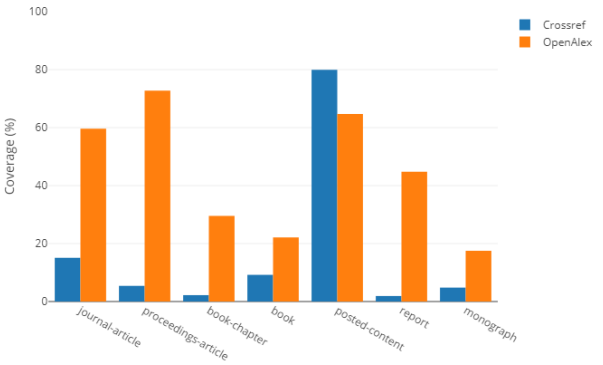


coverage added value - all time

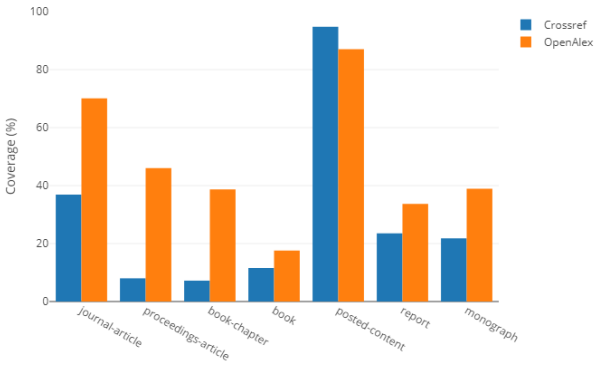


coverage added value - 2022

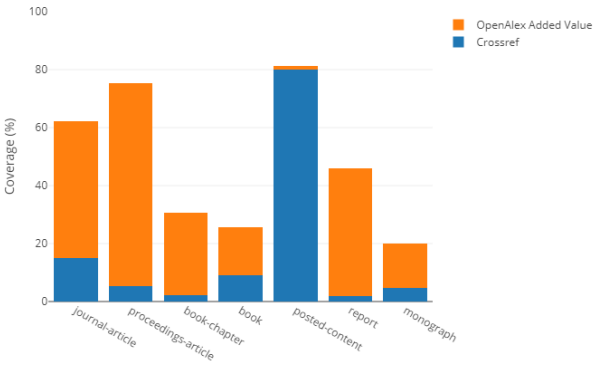
Abstracts



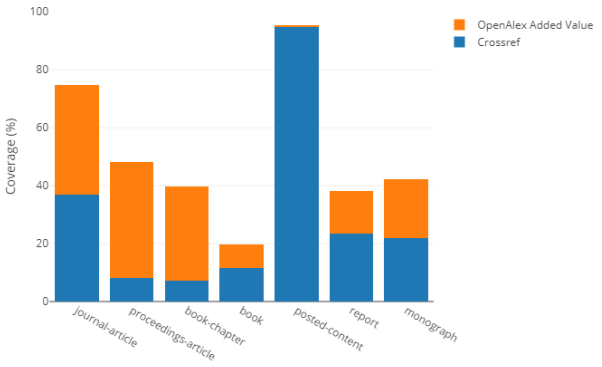
coverage comparison - all time



coverage comparison - 2022

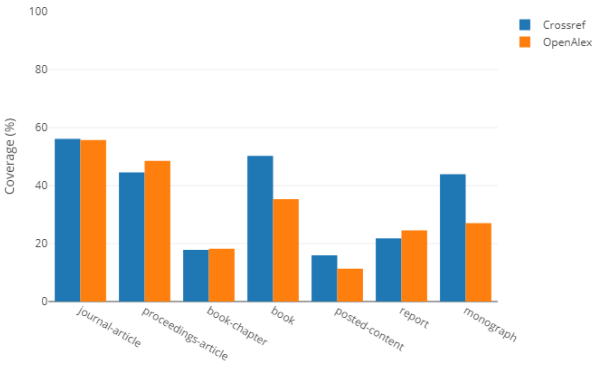


coverage added value - all time

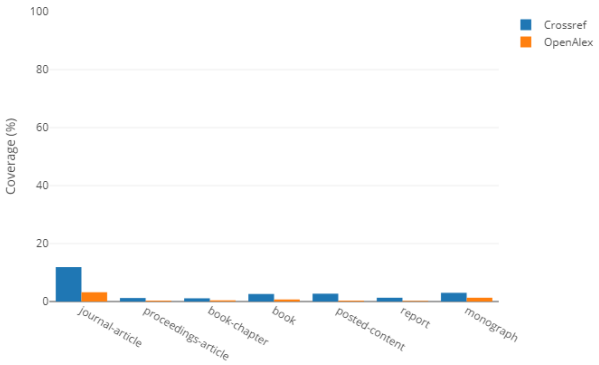


coverage added value - 2022

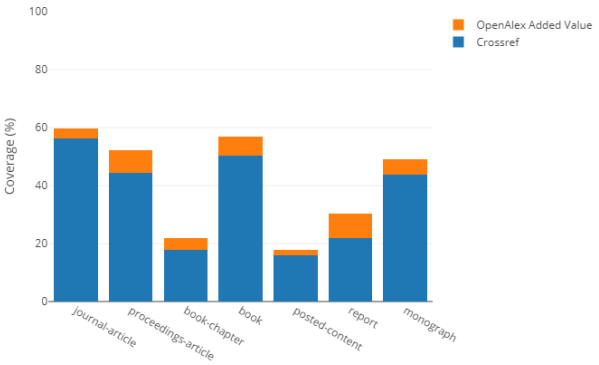
Citations to



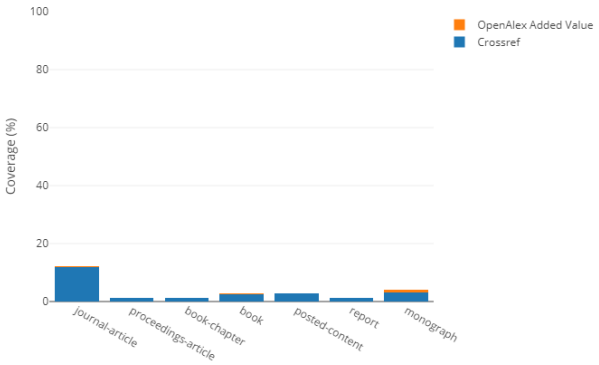
coverage comparison - all time



coverage comparison - 2022

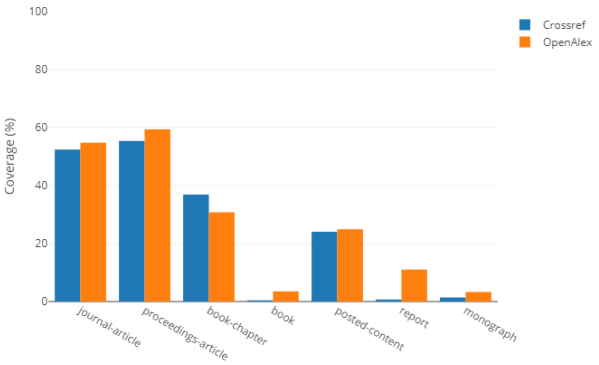


coverage added value - all time

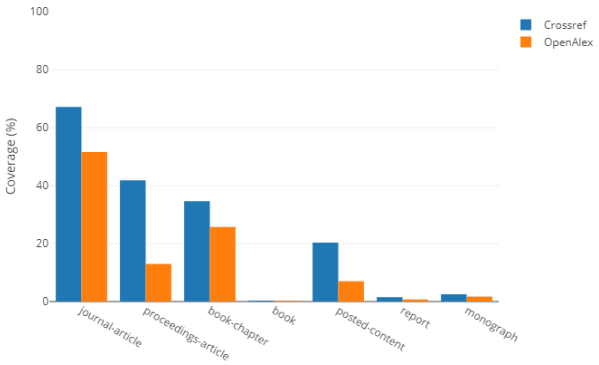


coverage added value - 2022

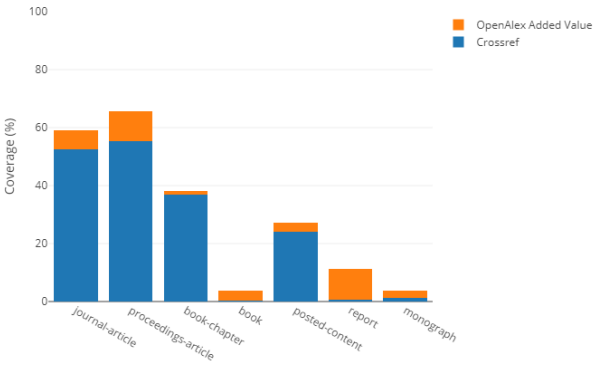
References from



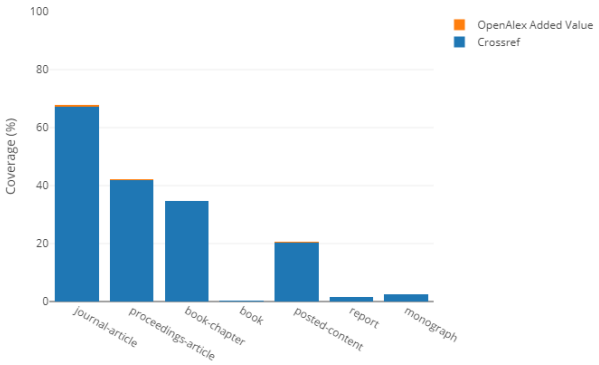
coverage comparison - all time



coverage comparison - 2022

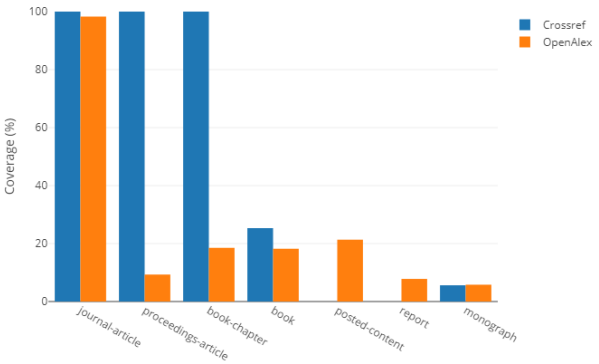


coverage added value - all time

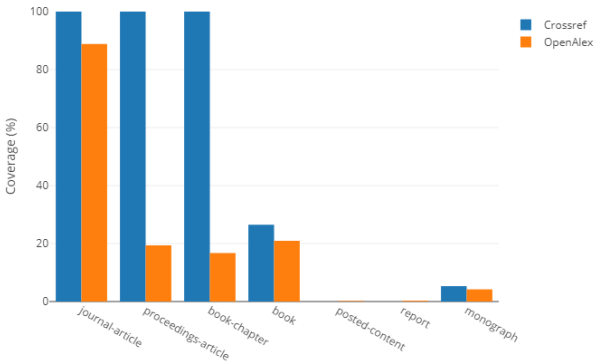


coverage added value - 2022

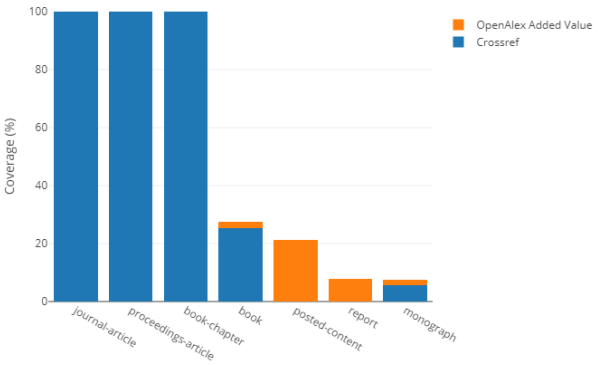
Journals



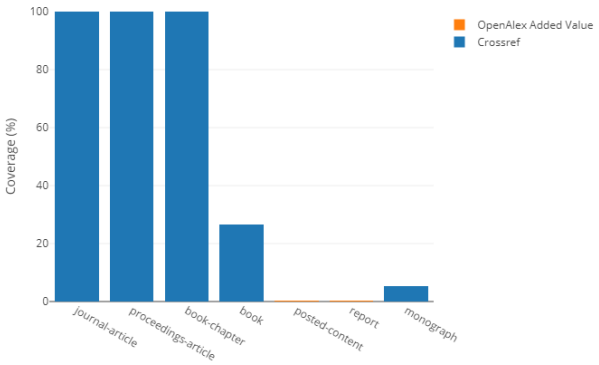
coverage comparison - all time



coverage comparison - 2022

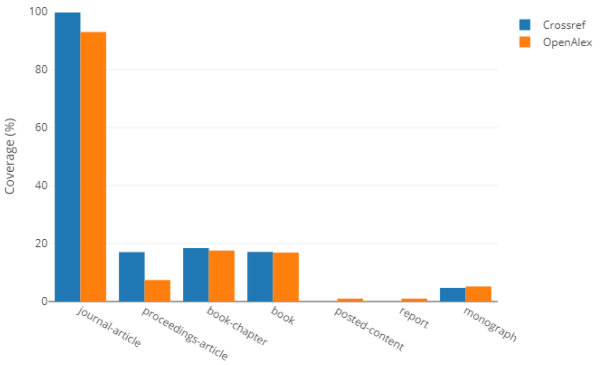


coverage added value - all time

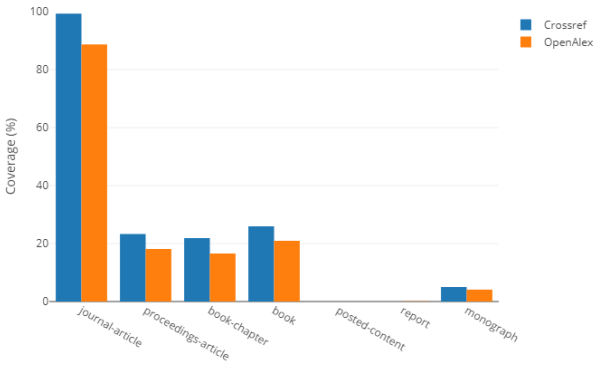


coverage added value - 2022

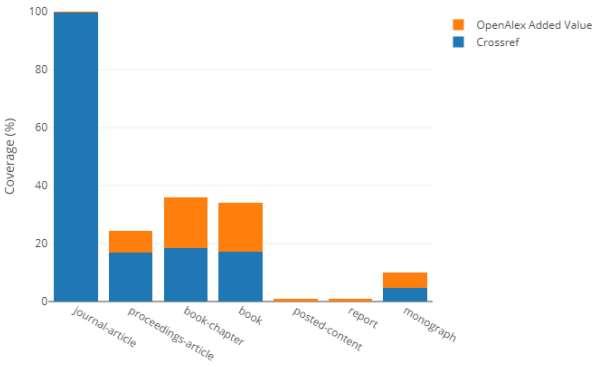
Journals ISSN



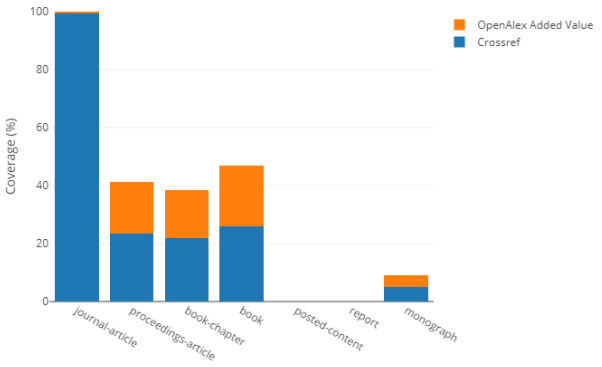
coverage comparison - all time



coverage comparison - 2022

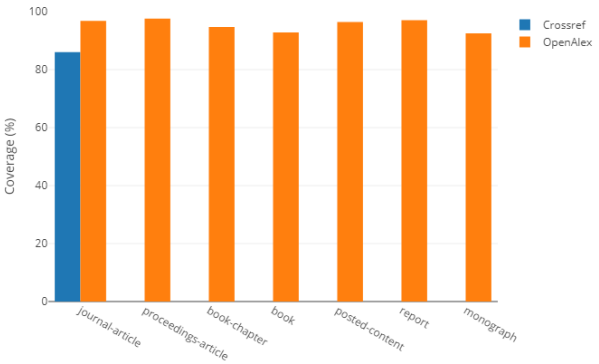


coverage added value - all time

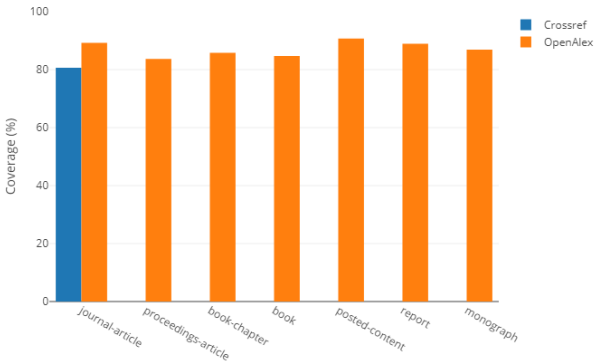


coverage added value - 2022

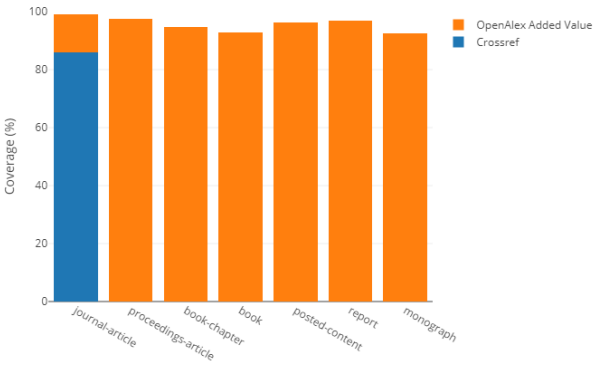
Fields



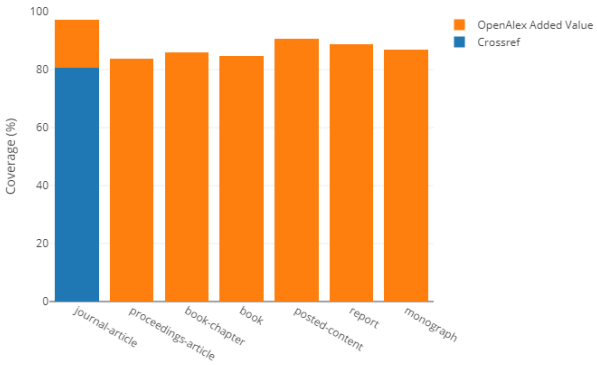
coverage comparison - all time



coverage comparison - 2022



coverage added value - all time

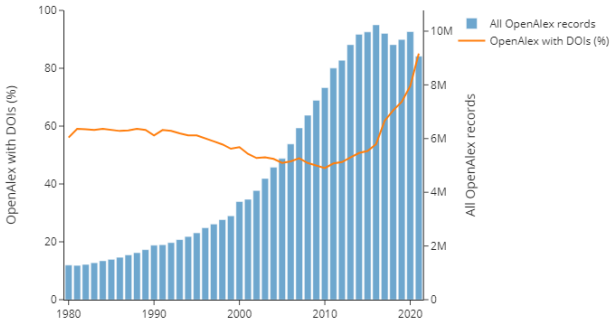


coverage added value - 2022

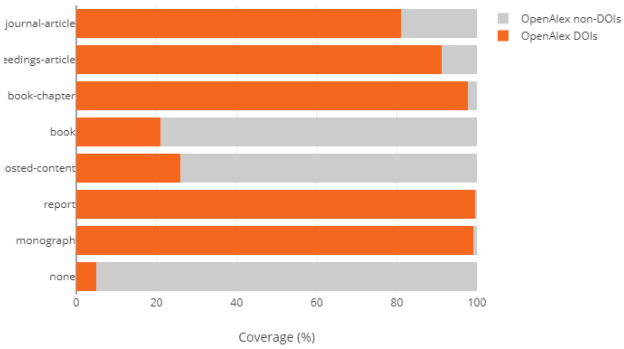
OpenAlex Coverage Beyond Crossref

DOIs vs non-DOIs

By year and publication type



coverage by publication date - all time

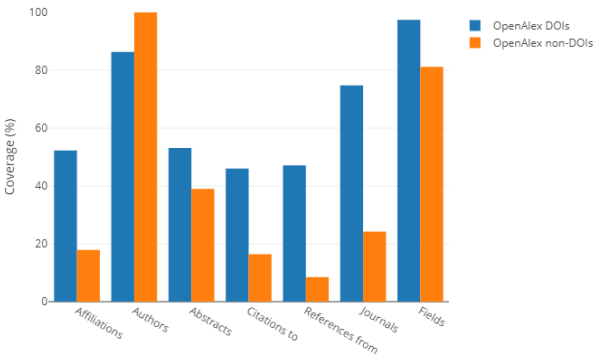


coverage by publication type - all time

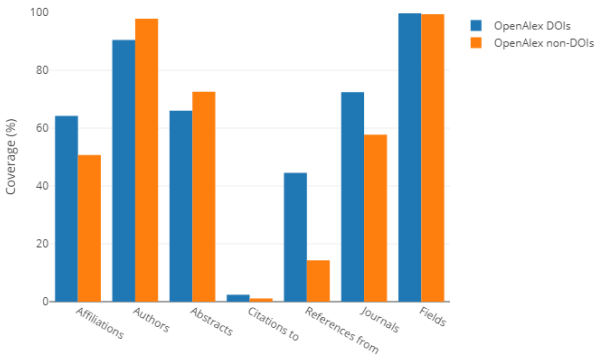
Metadata Coverage

Overview

Comparing coverage of metadata types for DOIs and non-DOIs in OpenAlex



coverage comparison - all time

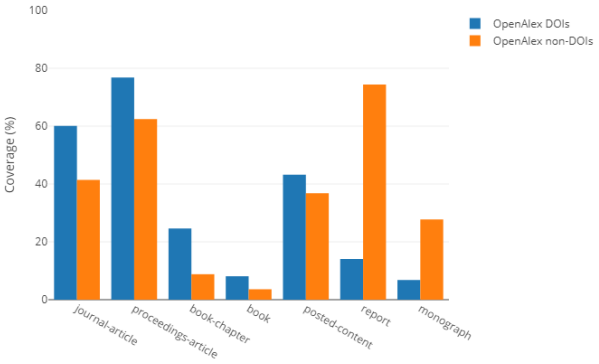


coverage comparison - 2022

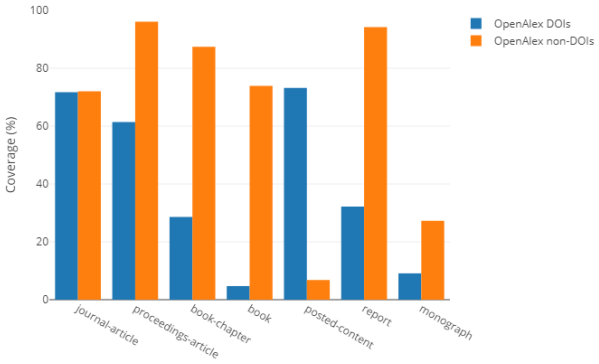
Details

Metadata coverage for DOIs and non-DOIs by publication type

Affiliations

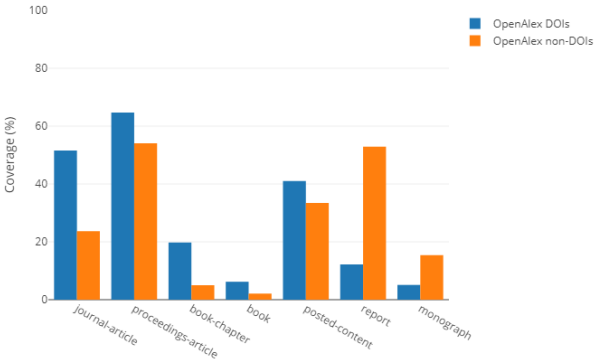


coverage comparison - all time

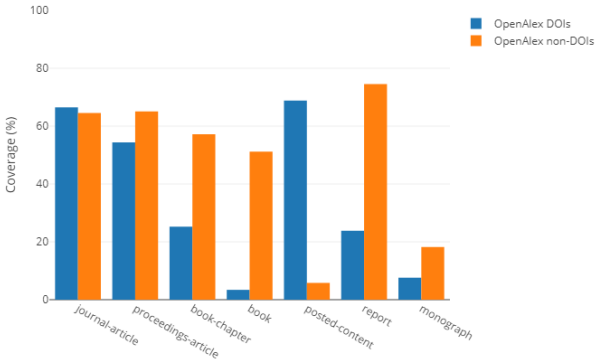


coverage comparison - 2022

Affiliations ROR

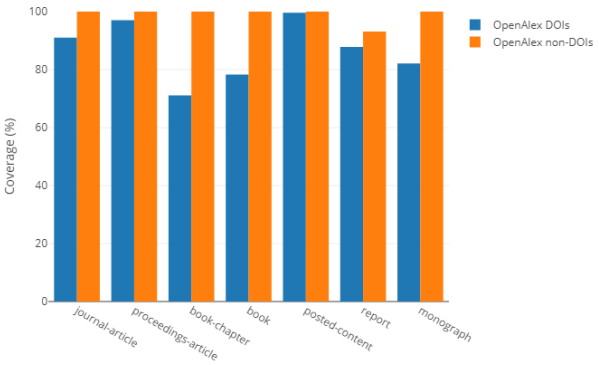


coverage comparison - all time

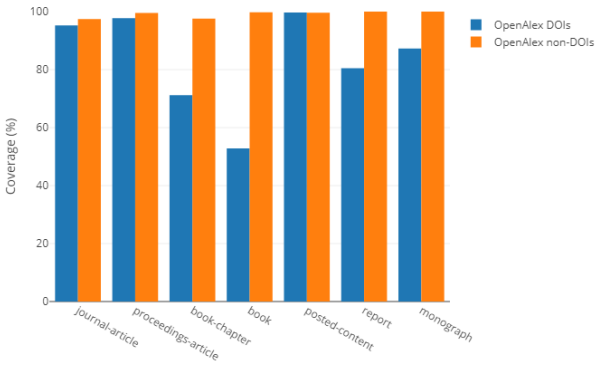


coverage comparison - 2022

Authors

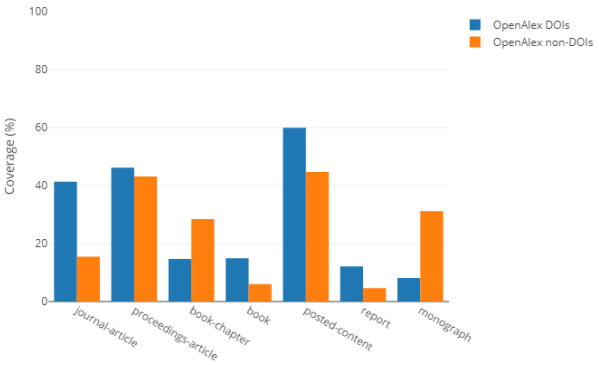


coverage comparison - all time

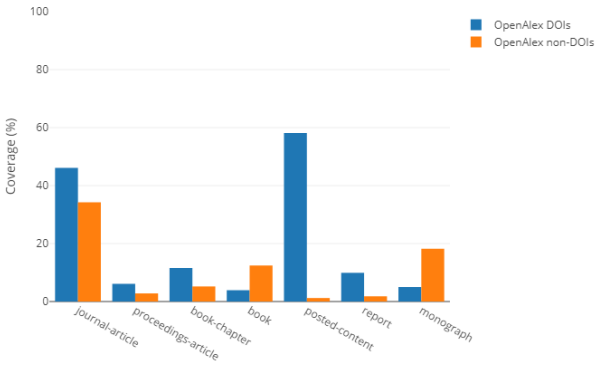


coverage comparison - 2022

Authors ORCIDs

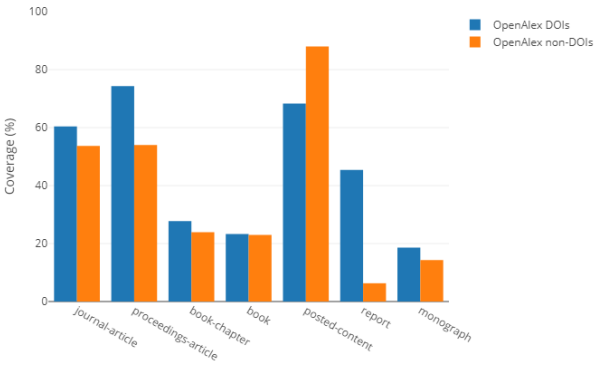


coverage comparison - all time

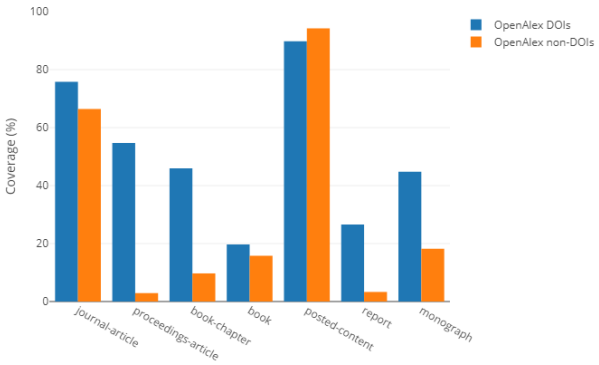


coverage comparison - 2022

Abstracts

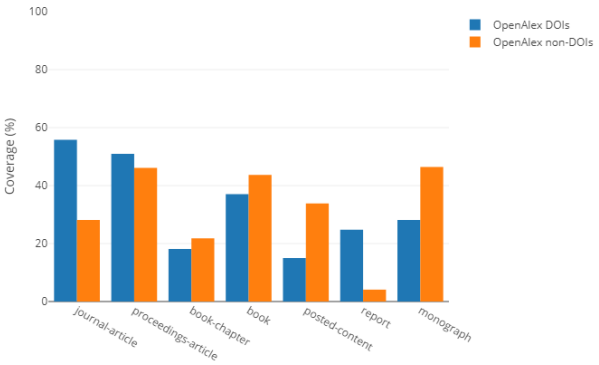


coverage comparison - all time

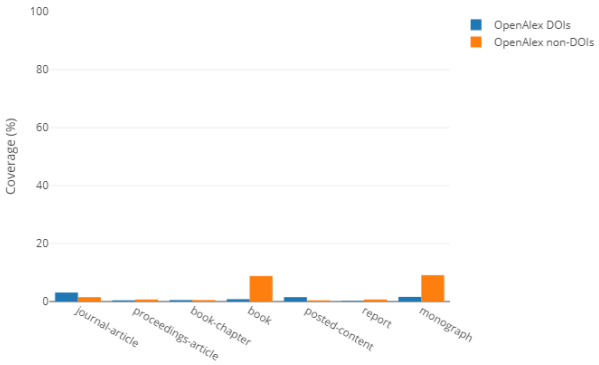


coverage comparison - 2022

Citations to

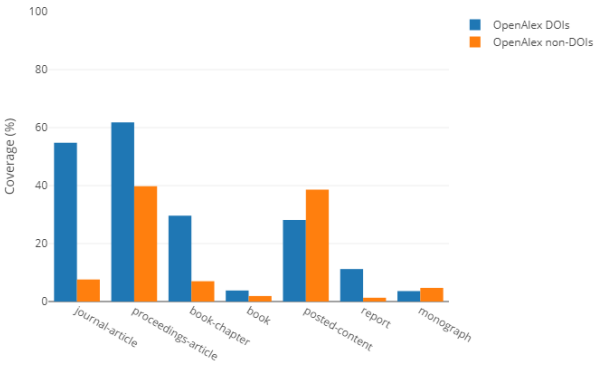


coverage comparison - all time

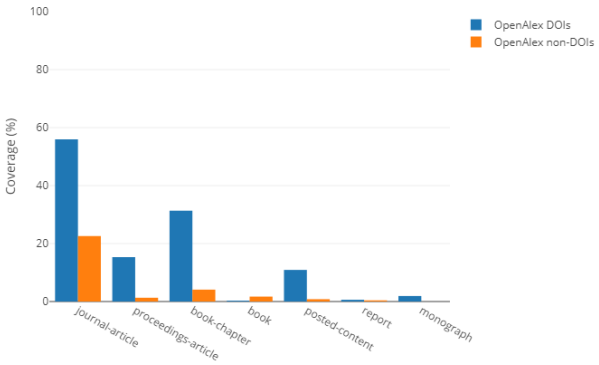


coverage comparison - 2022

References from

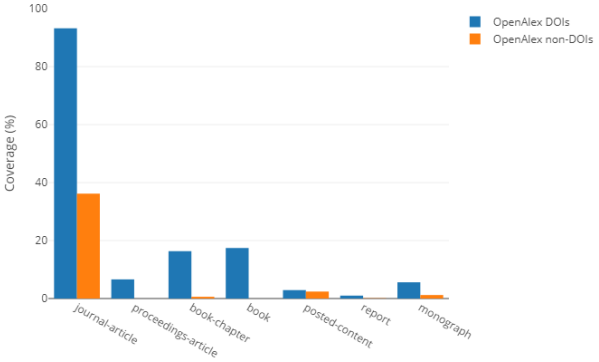


coverage comparison - all time

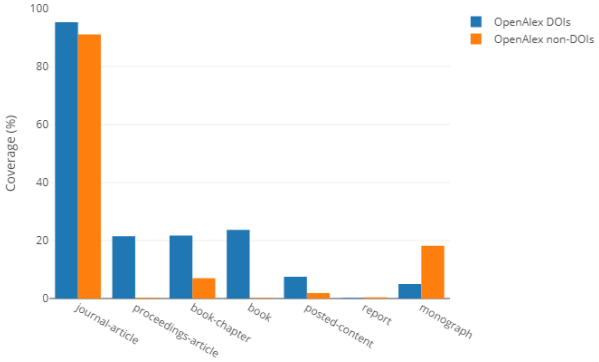


coverage comparison - 2022

Journals ISSN

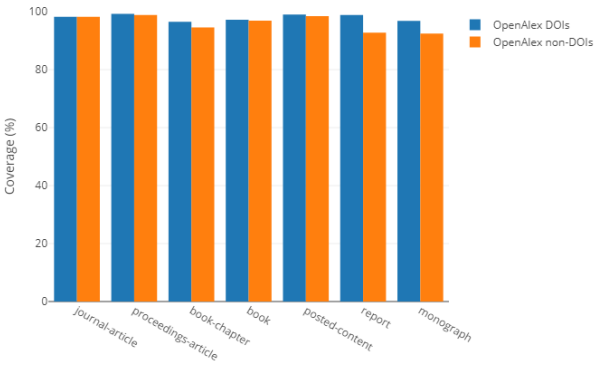


coverage comparison - all time

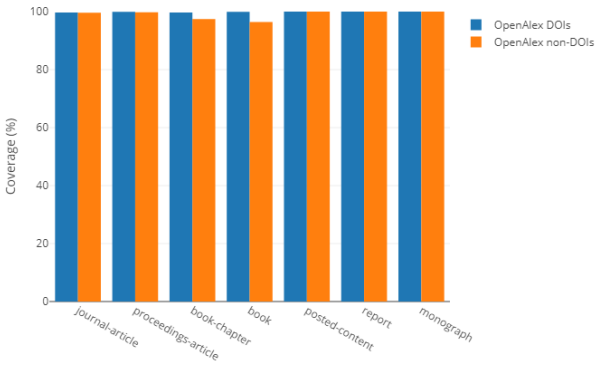


coverage comparison - 2022

Fields



coverage comparison - all time



coverage comparison - 2022

Appendix A - Tables

This section contains tables with summary counts. More tables will be added in a later version.

Crossref Current = 2020-2022
Focus Year = 2022

OpenAlex Coverage

Table 1. OpenAlex Metadata Coverage of Crossref DOIs

Time Frame	Crossref DOIs	OpenAlex Coverage of DOIs
All Time	123390682	121006622
Crossref Current	17003593	16210017
Focus Year	2829867	2517291

Crossref Coverage

Table 2. Crossref Metadata Coverage of Crossref DOIs

Time Frame	Crossref DOIs	Author Strings	Author ORCIDs	Affiliation Strings	Affiliation RORs	Abstracts	Field Classification	Venue Names	ISSNs
All Time	123390682	104214913	8375904	16858936	4061	15598258	80453587	119578458	97790273
Crossref Current	17003593	14939285	4784710	3546363	3086	5209761	9004200	15853385	12373221
Focus Year	2829867	2565941	921938	702269	1869	884343	1636961	2652599	2134995