

OPEN METADATA SOURCES

COMPARING OPENALEX (MAG
FORMAT) TO CROSSREF

DATE: 26 MARCH 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 (MAG format) compared to Crossref, and of DOIs vs non-DOIs in OpenAlex (MAG format), 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 at:

<https://github.com/Curtin-Open-Knowledge-Initiative/open-metadata-report>

with all images and data belonging to this report located in [/reports/run_20220326_1](#)

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_metadata20220107
- Crossref Member Data: utrecht-university.crossref.member_data with date 20220311
- OpenAlex (MAG format):

Complete data and code are available at:

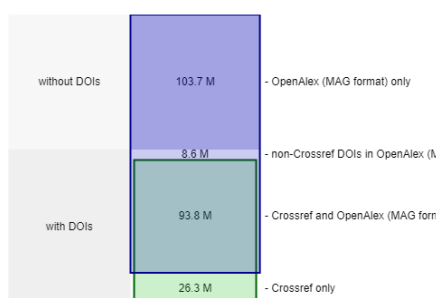
<https://github.com/Curtin-Open-Knowledge-Initiative/open-metadata-report>

with all images and data belonging to this report located in [/reports/run_20220326_1](#)

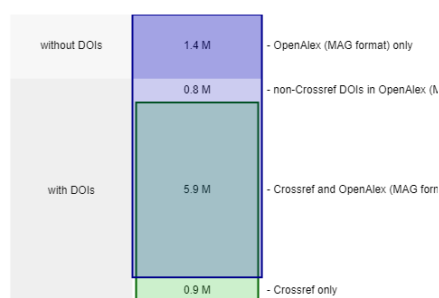
Coverage of OpenAlex (MAG format) vs Crossref

Comparing coverage

Overview

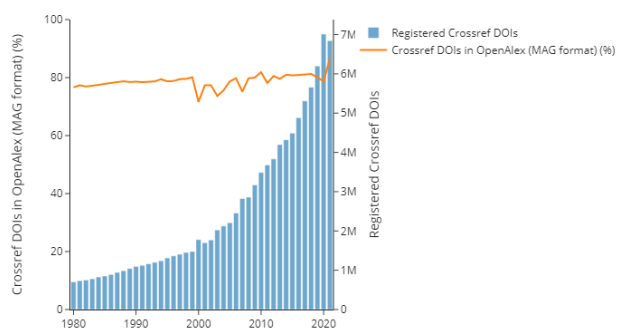


overall comparison - all time

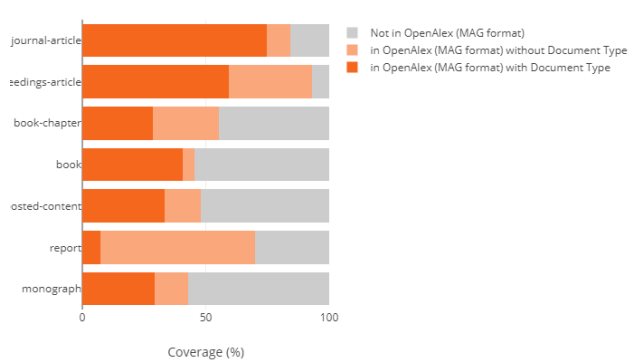


overall comparison - 2021

By year and publication type



coverage by publication date - all time

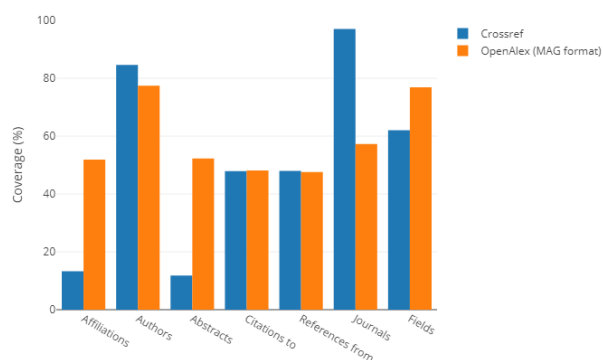


coverage by publication type - all time

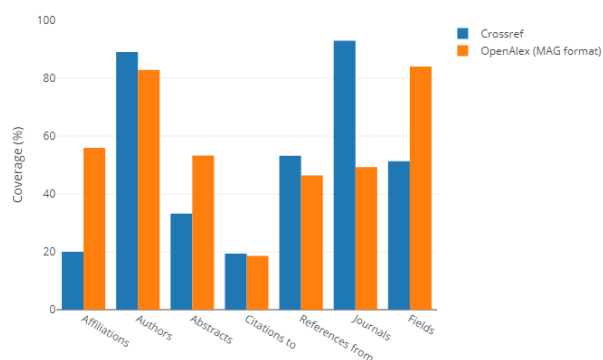
Value Add of OpenAlex (MAG format) to Crossref

Overview

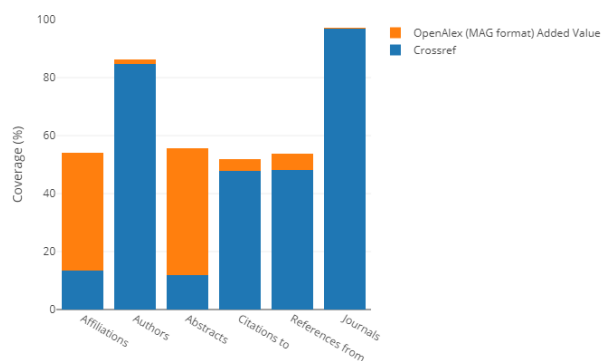
Comparing coverage of metadata types in Crossref and OpenAlex (MAG format)



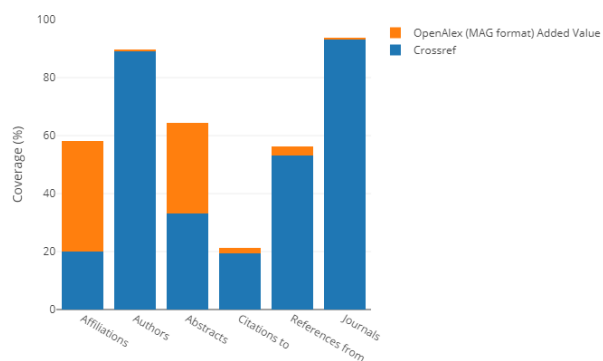
coverage comparison - all time



coverage comparison - 2021



coverage added value - all time

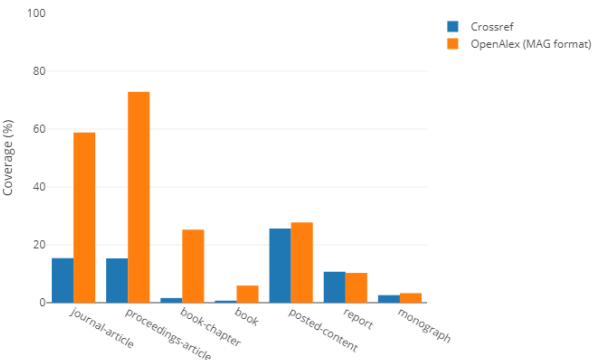


coverage added value - 2021

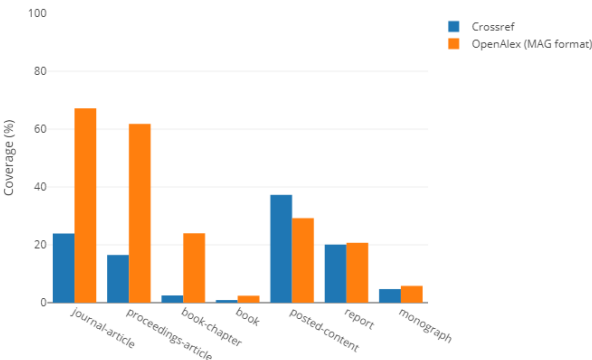
Details

Metadata coverage in OpenAlex (MAG format) and Crossref by publication type

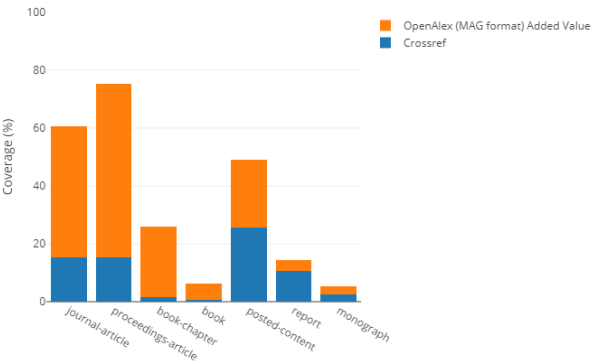
Affiliations



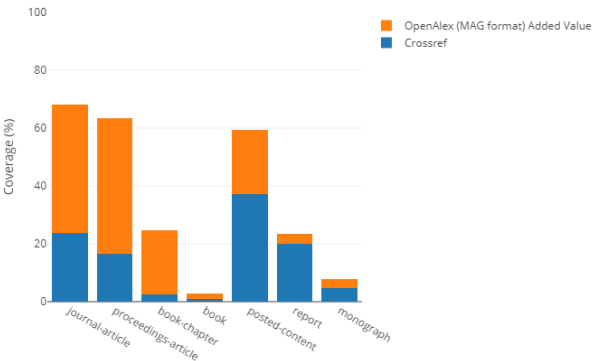
coverage comparison - all time



coverage comparison - 2021

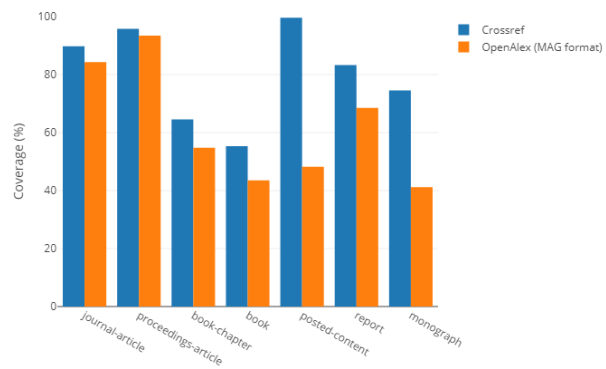


coverage added value - all time

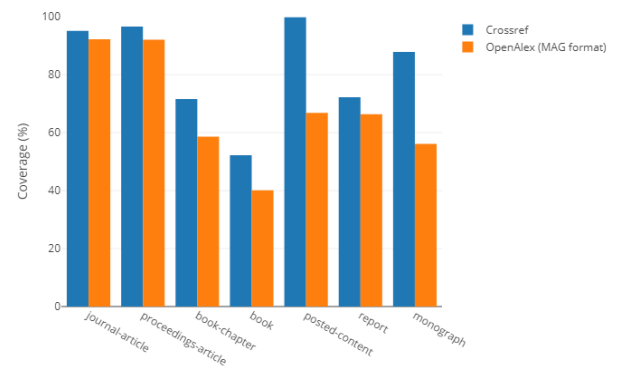


coverage added value - 2021

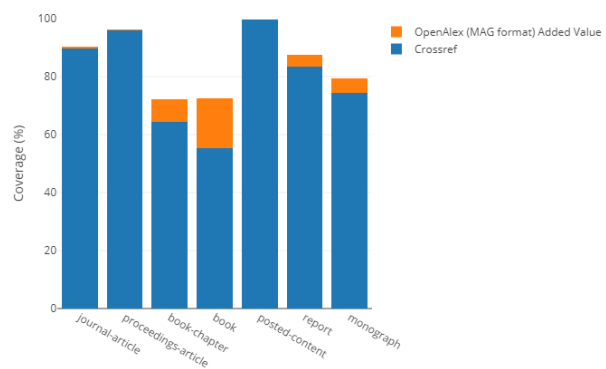
Authors



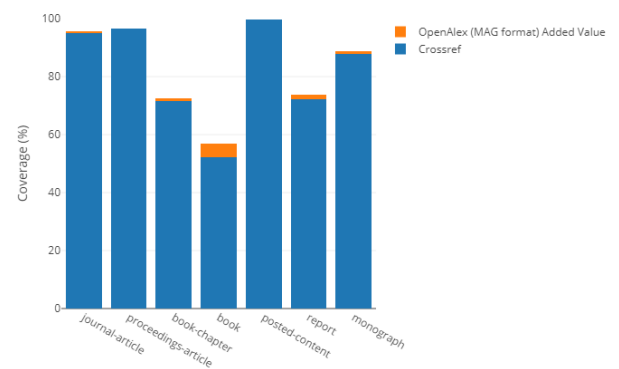
coverage comparison - all time



coverage comparison - 2021

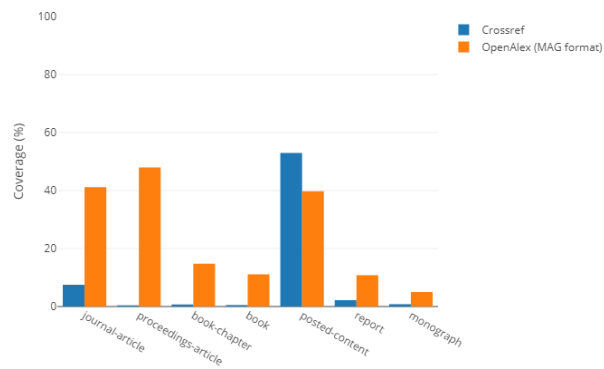


coverage added value - all time

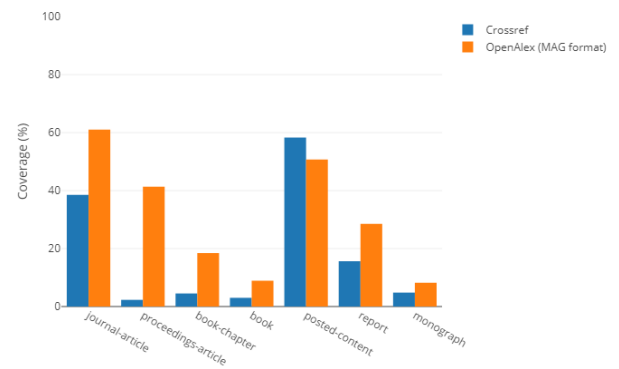


coverage added value - 2021

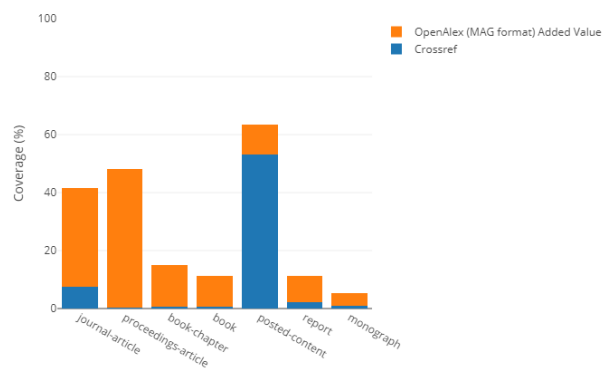
Authors ORCIDiDs



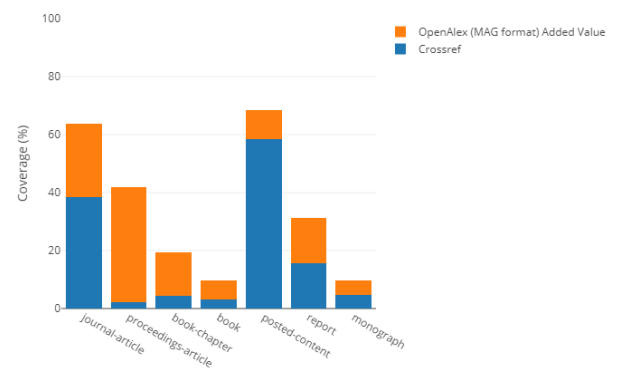
coverage comparison - all time



coverage comparison - 2021

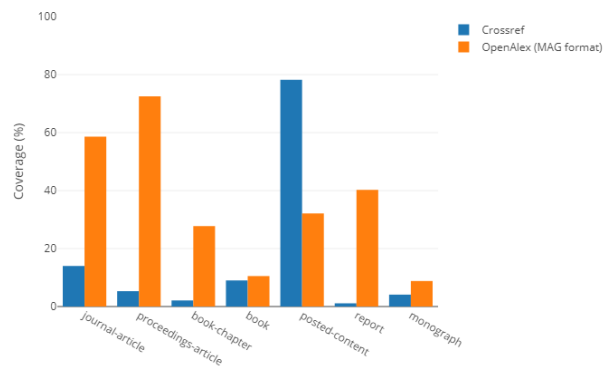


coverage added value - all time

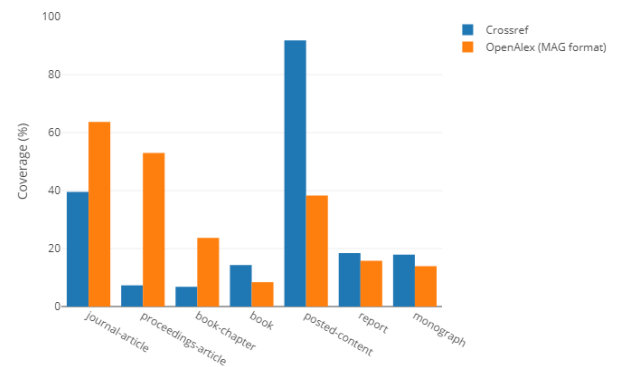


coverage added value - 2021

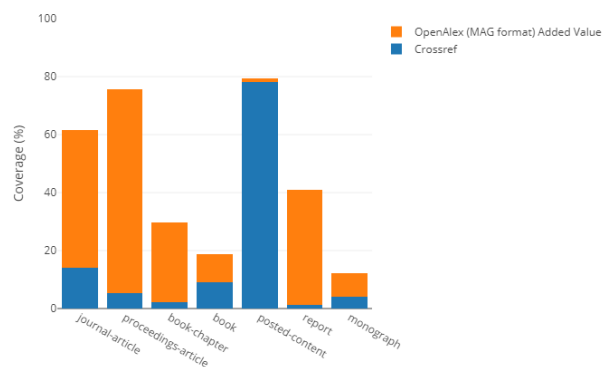
Abstracts



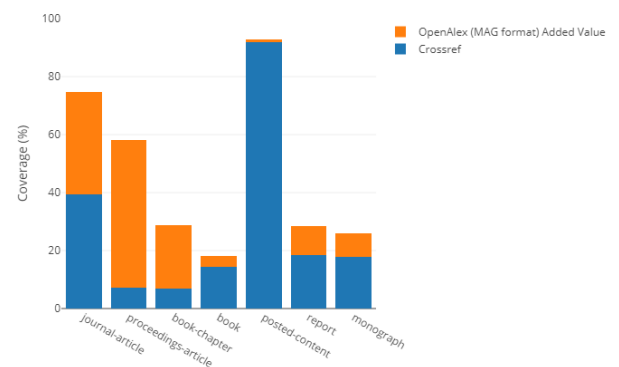
coverage comparison - all time



coverage comparison - 2021

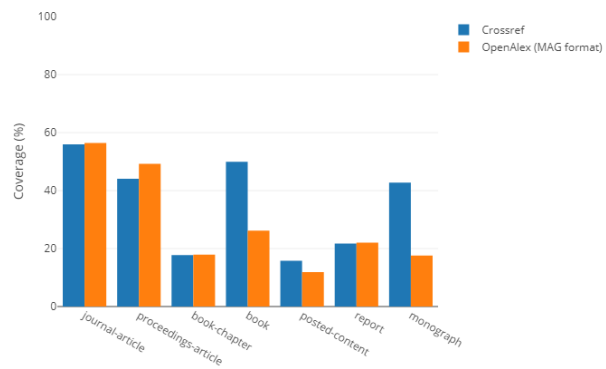


coverage added value - all time

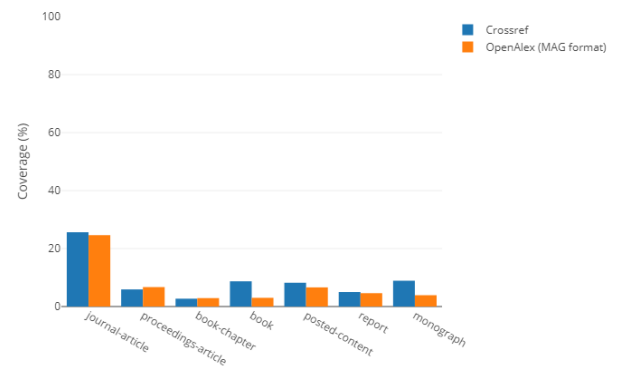


coverage added value - 2021

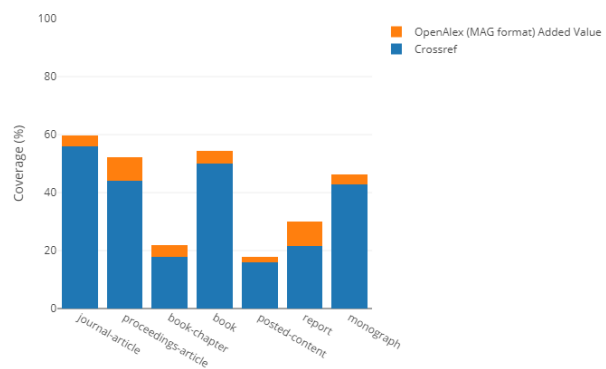
Citations to



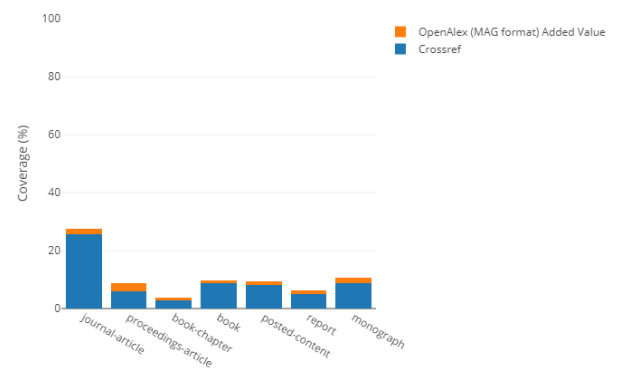
coverage comparison - all time



coverage comparison - 2021

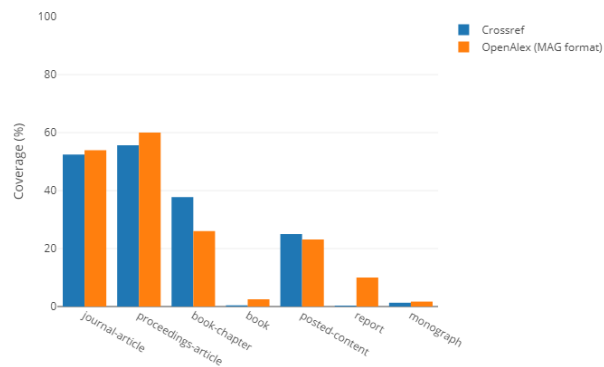


coverage added value - all time

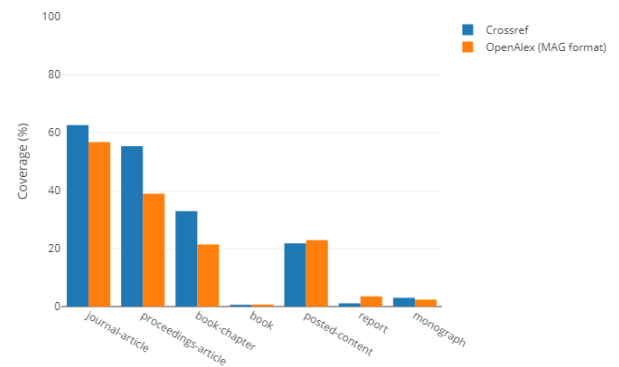


coverage added value - 2021

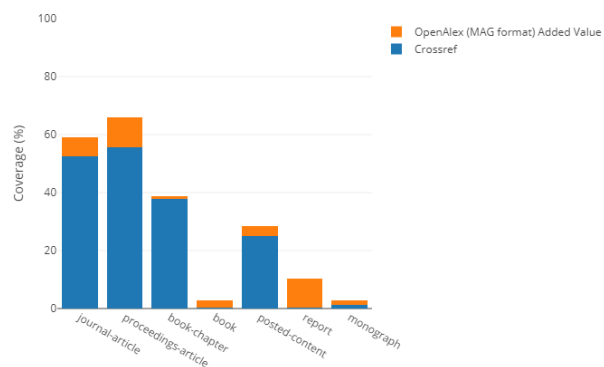
References from



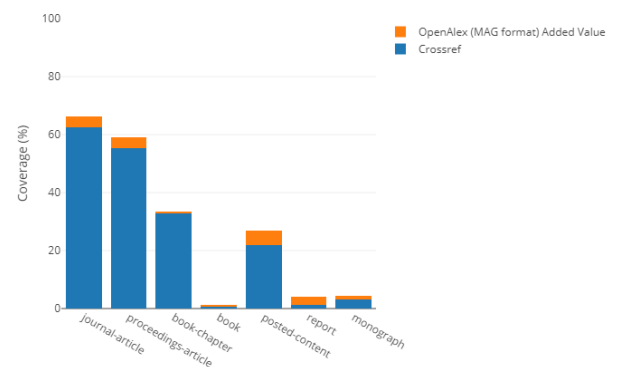
coverage comparison - all time



coverage comparison - 2021

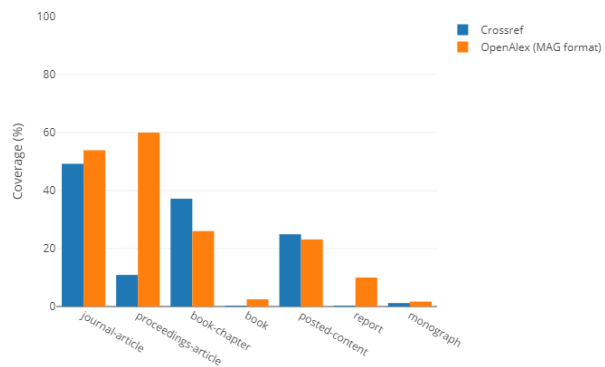


coverage added value - all time

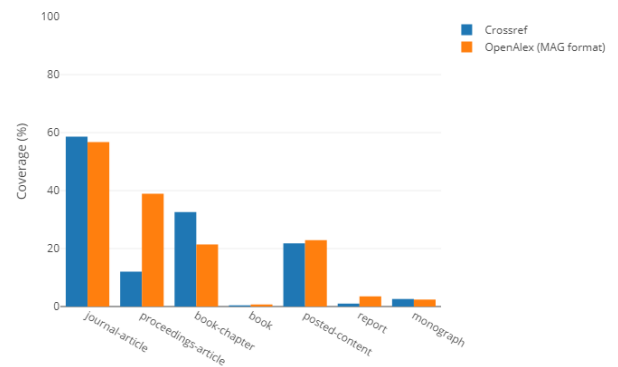


coverage added value - 2021

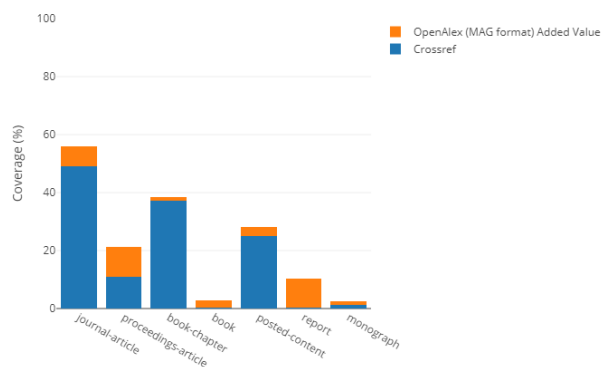
Open References from



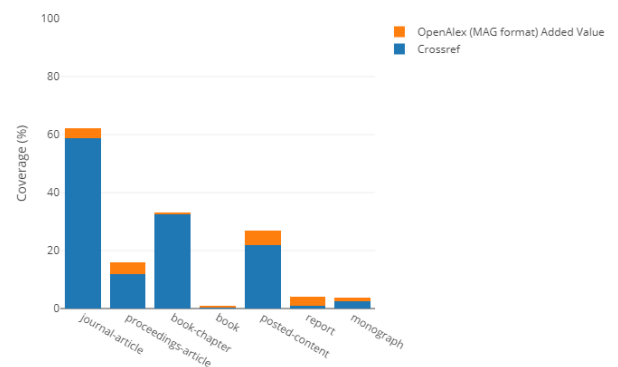
coverage comparison - all time



coverage comparison - 2021

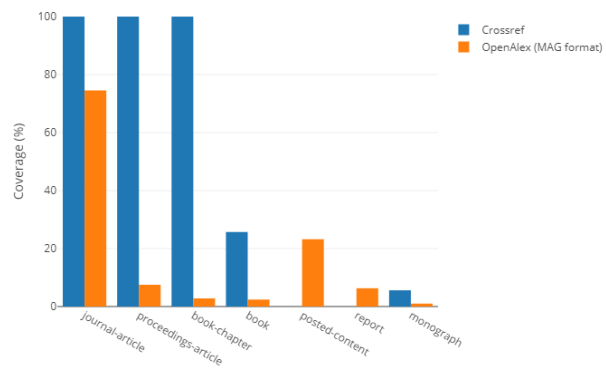


coverage added value - all time

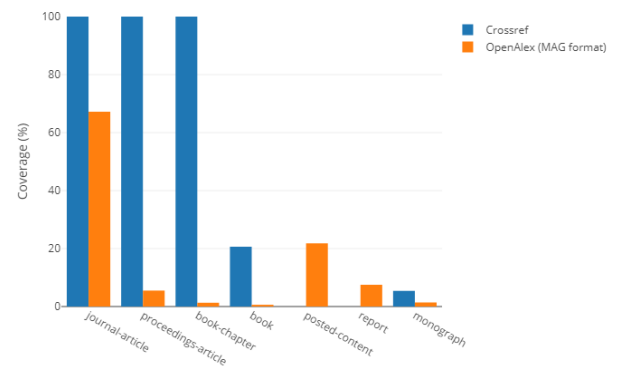


coverage added value - 2021

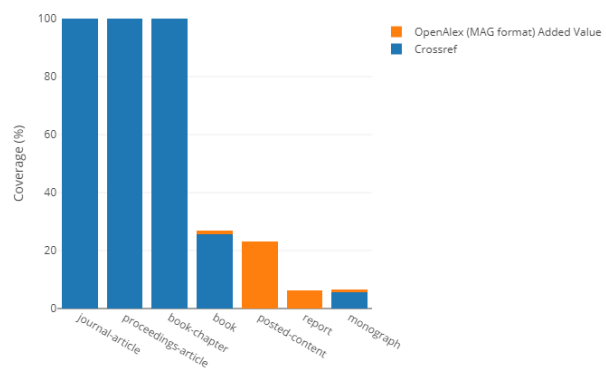
Journals



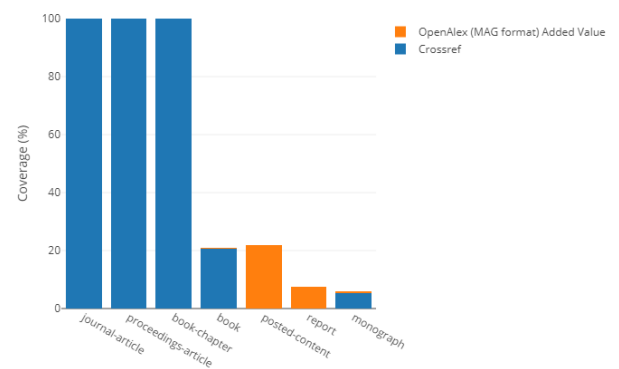
coverage comparison - all time



coverage comparison - 2021

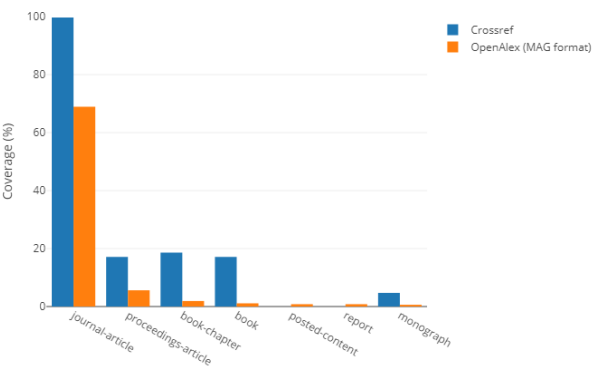


coverage added value - all time

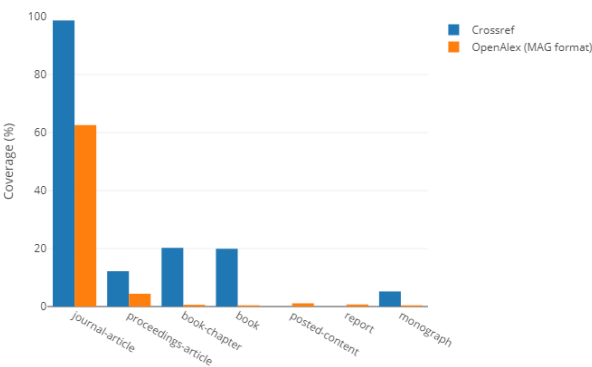


coverage added value - 2021

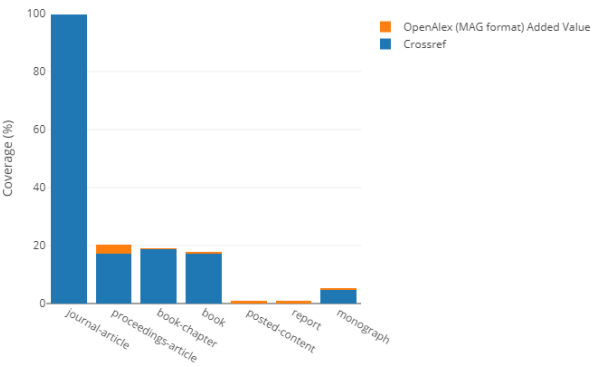
Journals ISSN



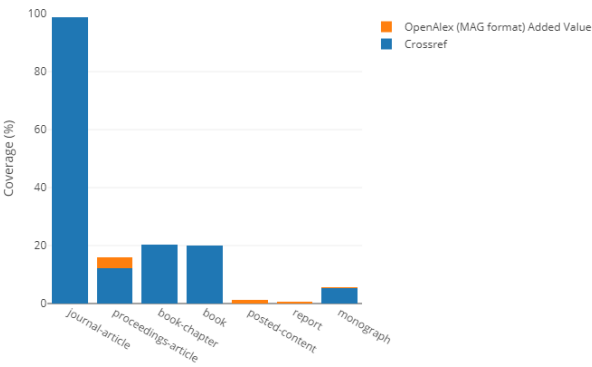
coverage comparison - all time



coverage comparison - 2021



coverage added value - all time

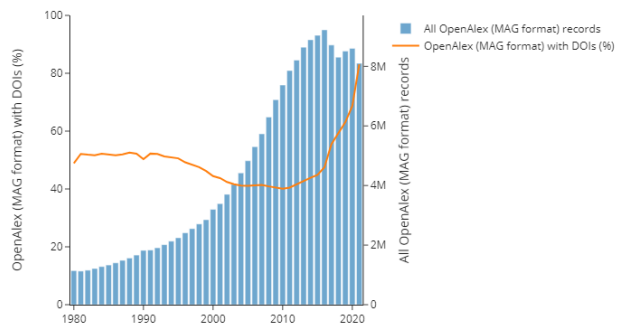


coverage added value - 2021

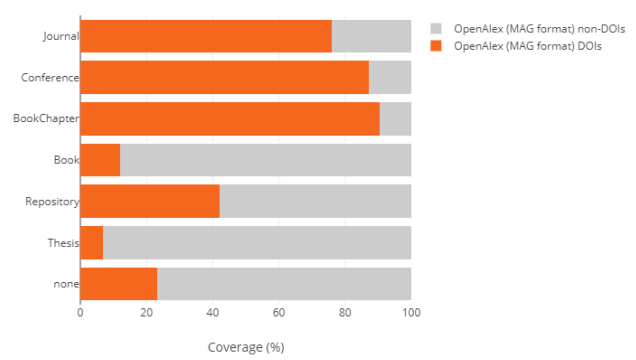
OpenAlex (MAG format) Coverage Beyond Crossref

DOIs vs non-DOIs

By year and publication type



coverage by publication date - all time

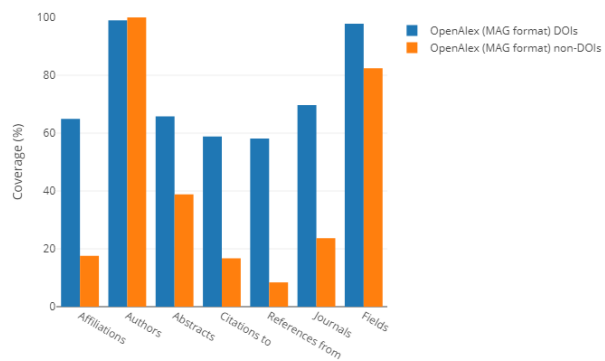


coverage by publication type - all time

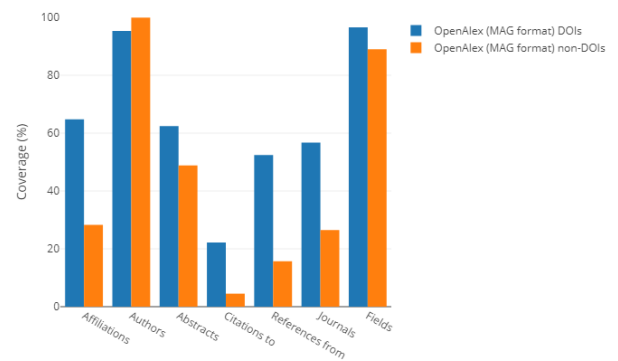
Metadata Coverage

Overview

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



coverage comparison - all time



coverage comparison - 2021

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 = 2021

OpenAlex (MAG format) Coverage

Table 1. OpenAlex (MAG format) Metadata Coverage of Crossref DOIs

Time Frame	Crossref DOIs	OpenAlex Coverage of DOIs
All Time	120141465	93837117
Crossref Current	14480074	11942877
Focus Year	6845252	5926237

Crossref Coverage

Table 2. Crossref Metadata Coverage of Crossref DOIs

Time Frame	Crossref DOIs	Author Strings	Author ORCIDs	Affiliation Strings	Abstracts	Open Abstracts	Field Classification	Venue Names	ISSNs
All Time	120141465	101589631	7654447	15929784	14187606	51699495	74432176	116534739	95732822
Crossref Current	14480074	12694350	4072967	2878598	4348439	6887490	7244910	13499622	10590883
Focus Year	6845252	6101054	2084965	1369435	2271715	3270231	3510576	6366601	5131324