OPEN METADATA SOURCES

COMPARING MAG 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 MAG compared to Crossref, and of DOIs vs non-DOIs in MAG, 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_metadata20211007
- Crossref Member Data: utrecht-university.crossref.member_data with date 20220311
- MAG: {'Papers': 'academic-observatory.mag.Papers20211011', 'Affiliations': 'academic-observatory.mag.Affiliations20211011', 'Authors': 'academic-observatory.mag.Authors20211011', 'Journals': 'academic-observatory.mag.Journals20211011', 'ConferenceInstances': 'academic-observatory.mag.ConferenceInstances20211011', 'ConferenceSeries': 'academic-observatory.mag.ConferenceSeries20211011', 'PaperAuthorAffiliations': 'academic-observatory.mag.PaperAuthorAffiliations20211011', 'FieldsOfStudy': 'academic-observatory.mag.FieldsOfStudy20211011', 'FieldOfStudyExtendedAttributes': 'academic-observatory.mag.FieldOfStudyExtendedAttributes20211011', 'PaperAbstractsInvertedIndex': 'academic-observatory.mag.PaperAbstractsInvertedIndex20211011', 'PaperFieldsOfStudy': 'academic-observatory.mag.PaperFieldsOfStudy20211011', 'PaperExtendedAttributes': 'academic-observatory.mag.PaperExtendedAttributes20211011', 'PaperResources': 'academic-observatory.mag.PaperResources20211011', 'PaperUrls': 'academic-observatory.mag.PaperUrls20211011', 'PaperMeSH': 'academic-observatory.mag.PaperMeSH20211011', 'doi': 'academic-observatory.mag.doi20211011', 'Author': 'academic-observatory.mag.Author20211011', 'Concept': 'academic-observatory.mag.Concept20211011', 'Institution,Venue': 'academic-observatory.mag.Institution,Venue20211011', 'Work':

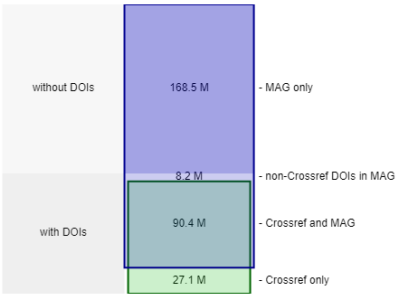
'academic-observatory.mag.Work20211011', 'additional_source_journal_fields': '', 'additional_source_org_fields': '', 'additional_truthtable_fields': '\n , CASE\n WHEN CHAR_LENGTH(journal.Issn) > 0\n THEN TRUE\n ELSE FALSE\n END\n as has_venue_issn,\n CASE\n WHEN CHAR_LENGTH(journal.Issn) > 0\n THEN 0\n ELSE 1\n END\n as count_venue_issn\n '}

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

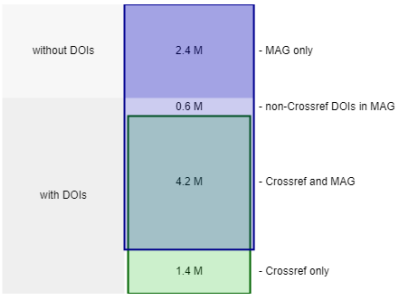
Coverage of MAG vs Crossref

Comparing coverage

Overview

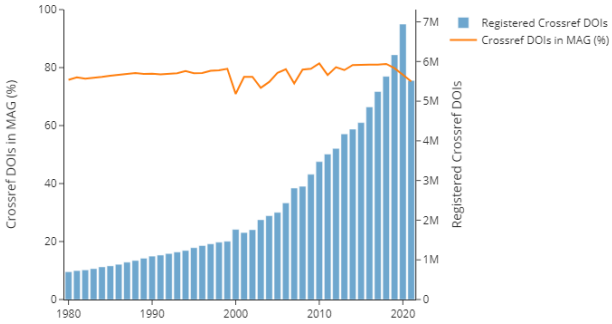


overall comparison - all time

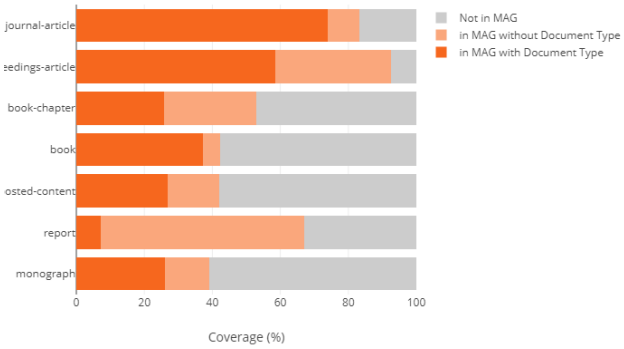


overall comparison - 2021

By year and publication type



coverage by publication date - all time

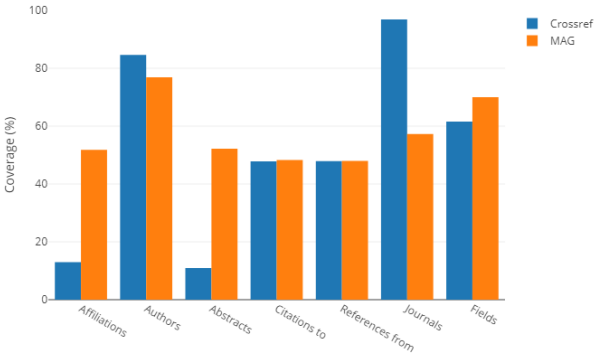


coverage by publication type - all time

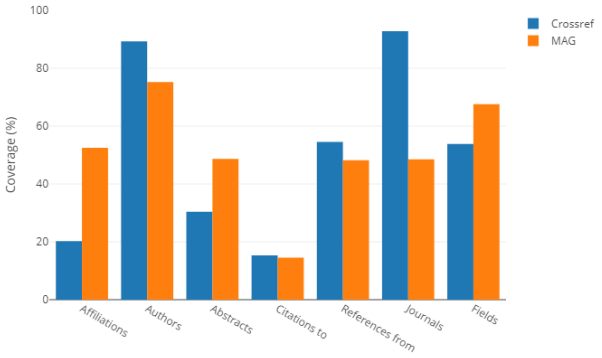
Value Add of MAG to Crossref

Overview

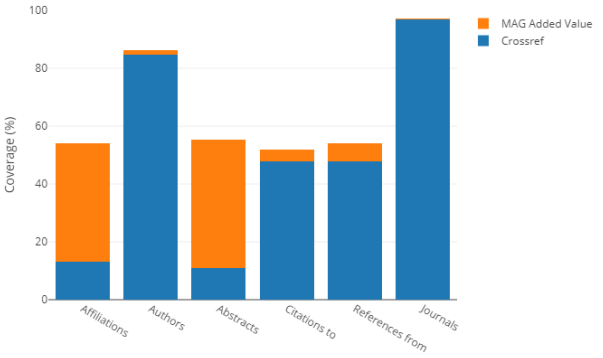
Comparing coverage of metadata types in Crossref and MAG



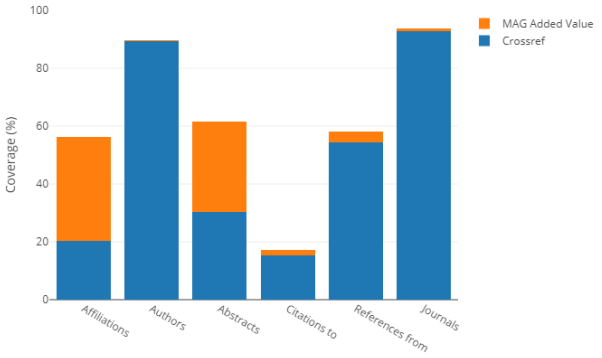
coverage comparison - all time



coverage comparison - 2021



coverage added value - all time

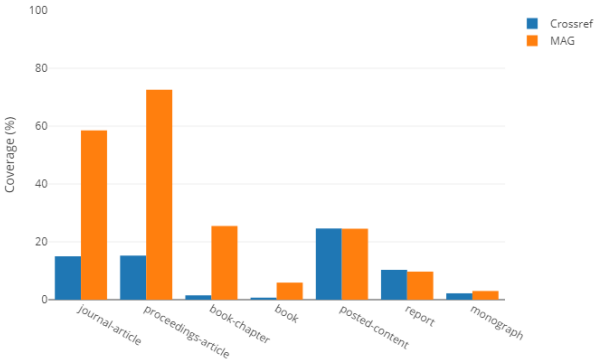


coverage added value - 2021

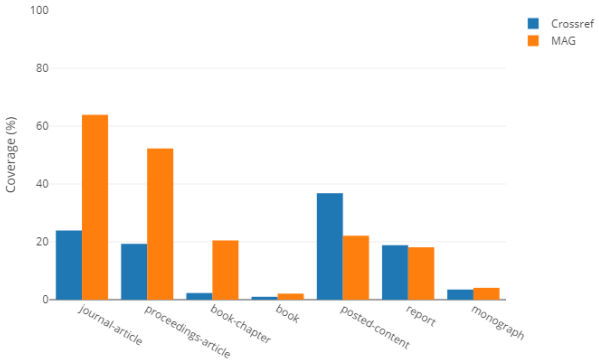
Details

Metadata coverage in MAG and Crossref by publication type

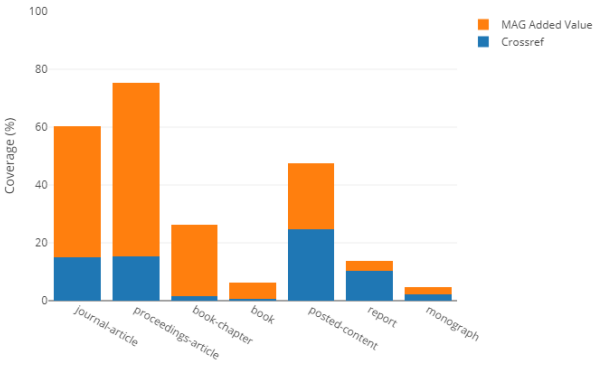
Affiliations



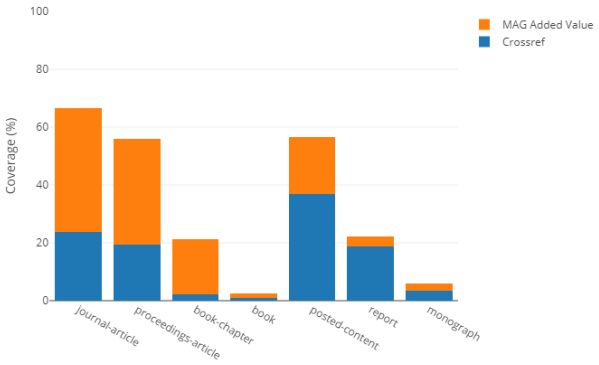
coverage comparison - all time



coverage comparison - 2021

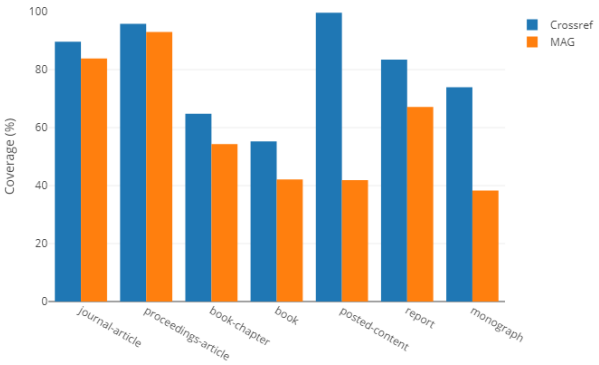


coverage added value - all time

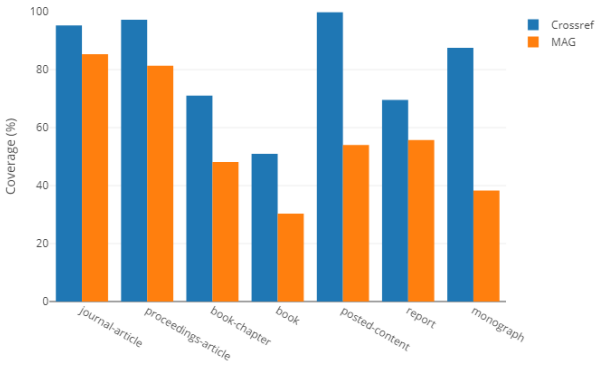


coverage added value - 2021

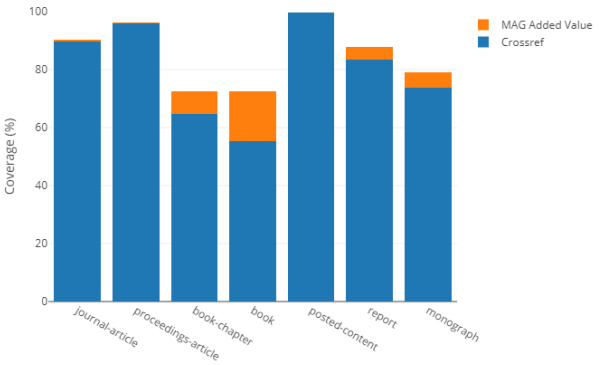
Authors



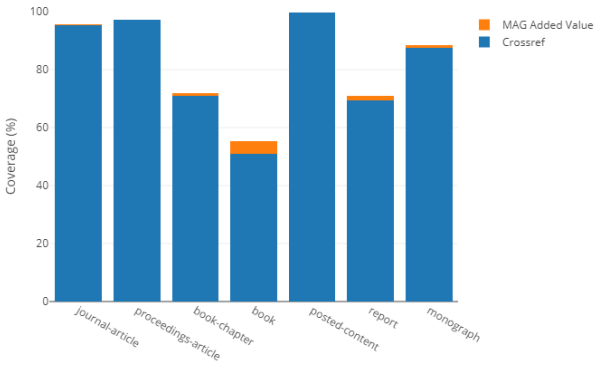
coverage comparison - all time



coverage comparison - 2021

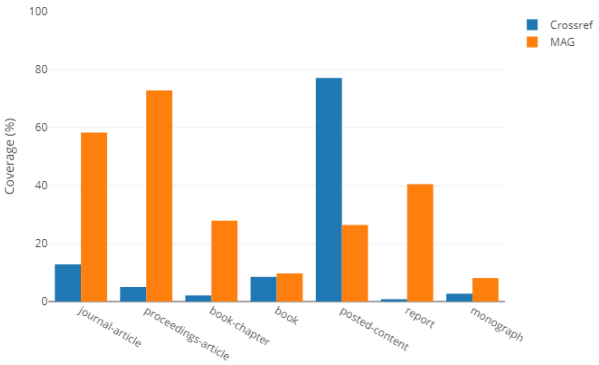


coverage added value - all time

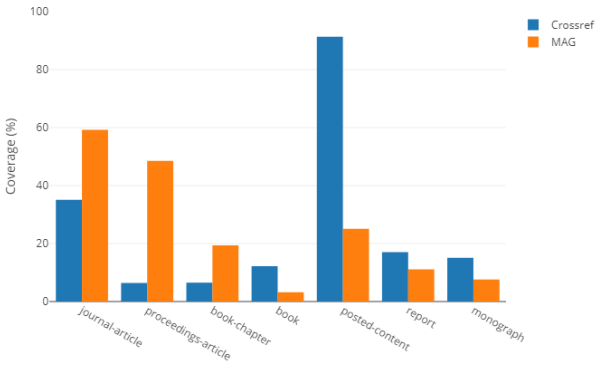


coverage added value - 2021

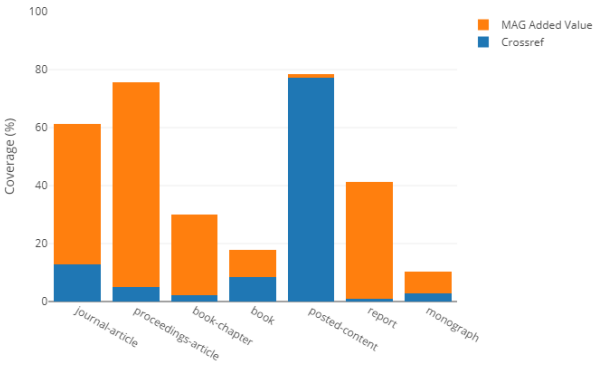
Abstracts



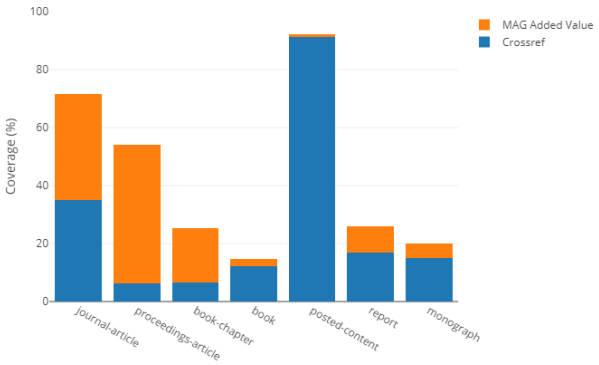
coverage comparison - all time



coverage comparison - 2021

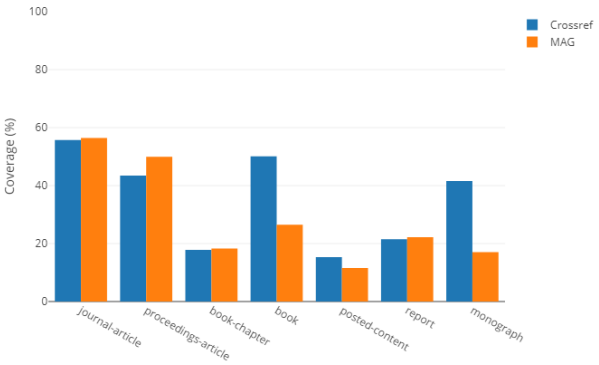


coverage added value - all time

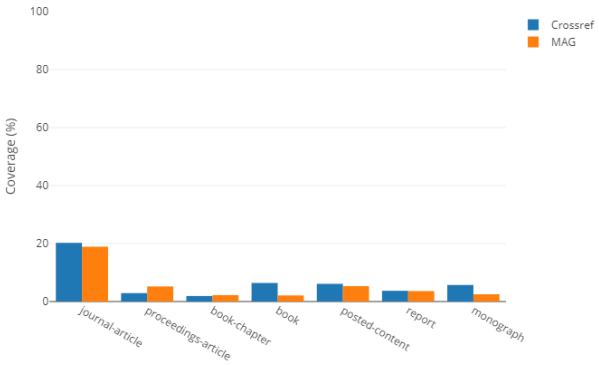


coverage added value - 2021

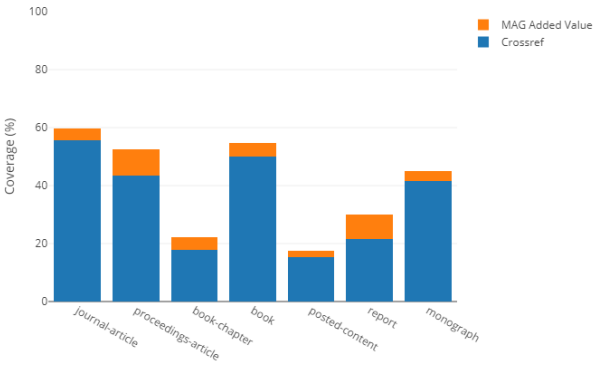
Citations to



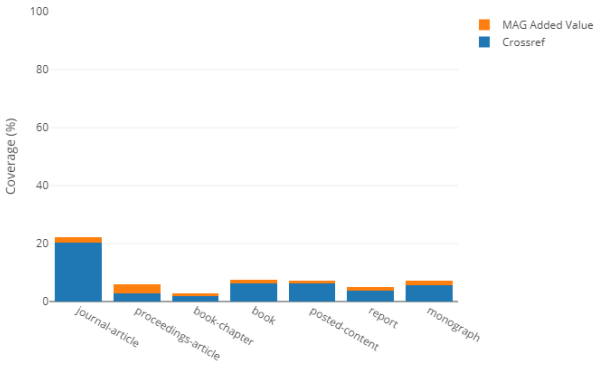
coverage comparison - all time



coverage comparison - 2021

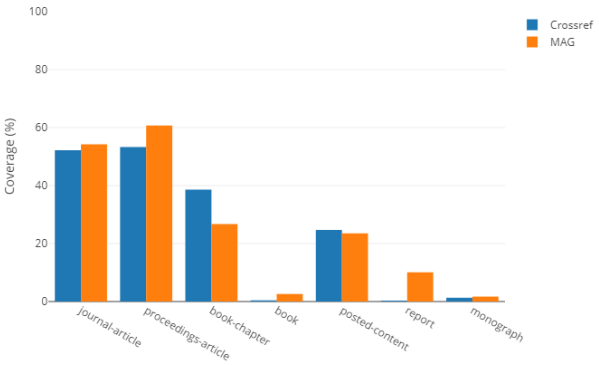


coverage added value - all time

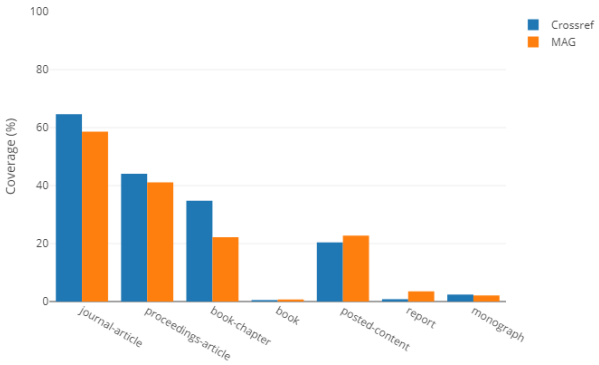


coverage added value - 2021

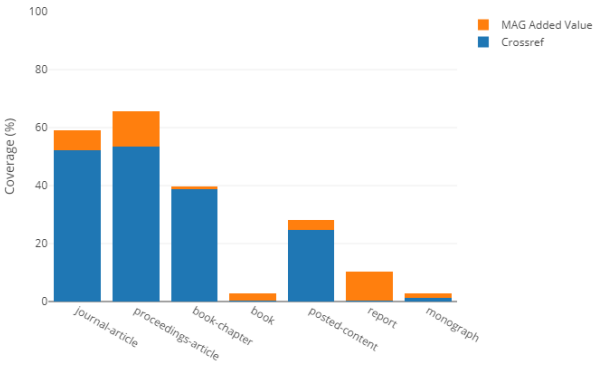
References from



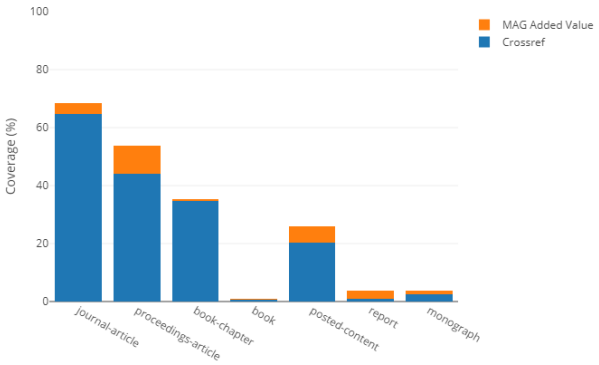
coverage comparison - all time



coverage comparison - 2021

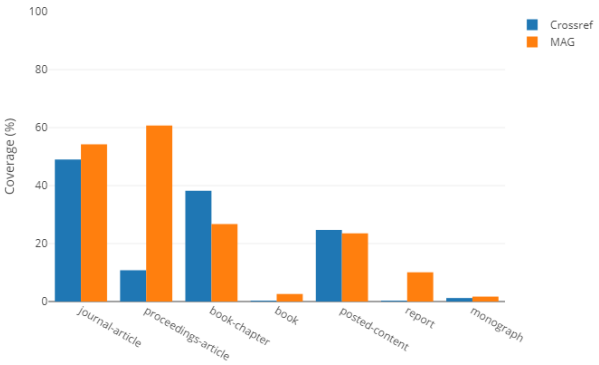


coverage added value - all time

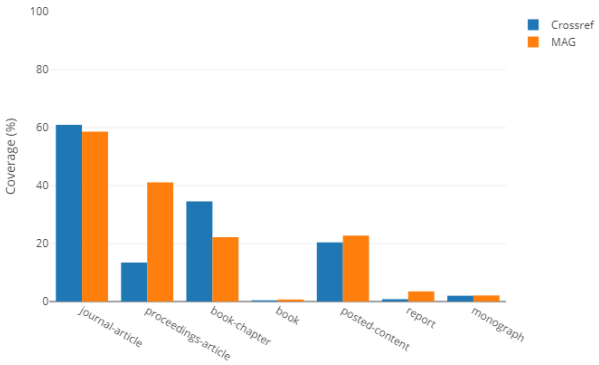


coverage added value - 2021

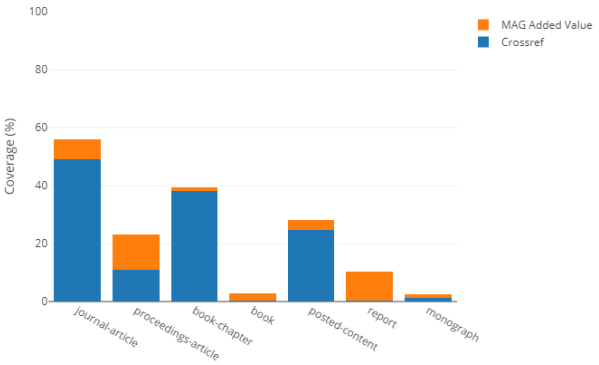
Open References from



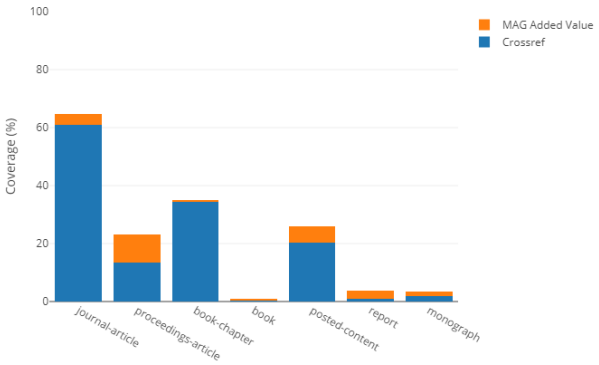
coverage comparison - all time



coverage comparison - 2021

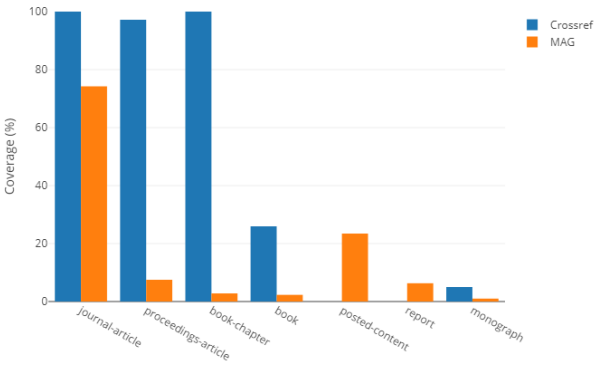


coverage added value - all time

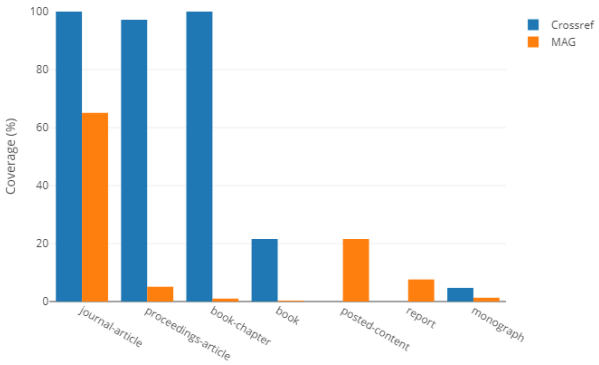


coverage added value - 2021

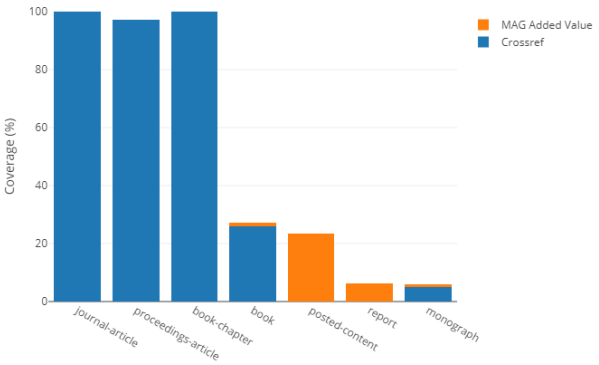
Journals



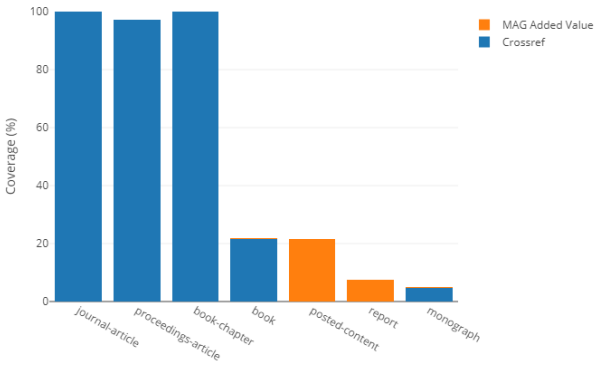
coverage comparison - all time



coverage comparison - 2021

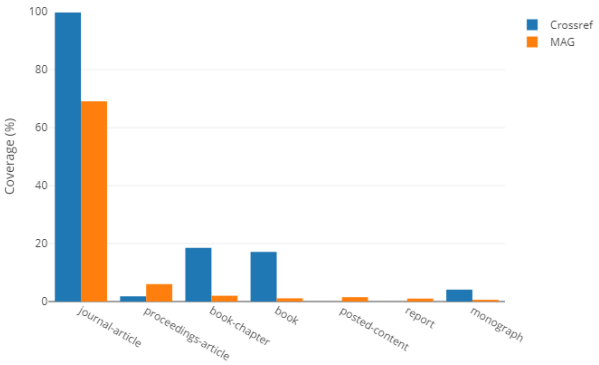


coverage added value - all time

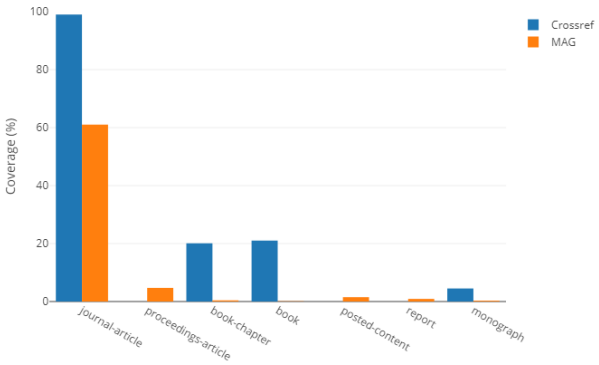


coverage added value - 2021

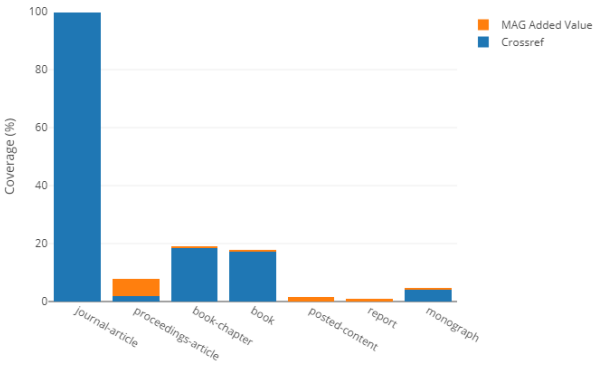
Journals ISSN



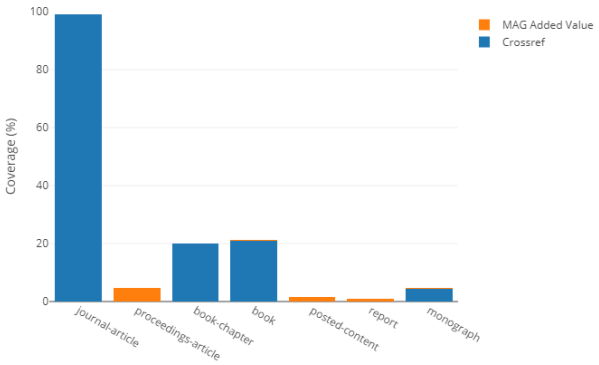
coverage comparison - all time



coverage comparison - 2021

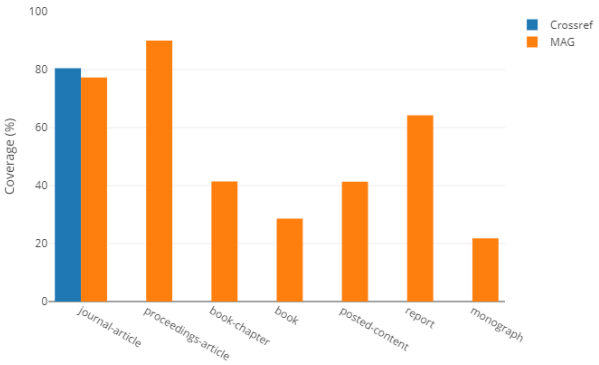


coverage added value - all time

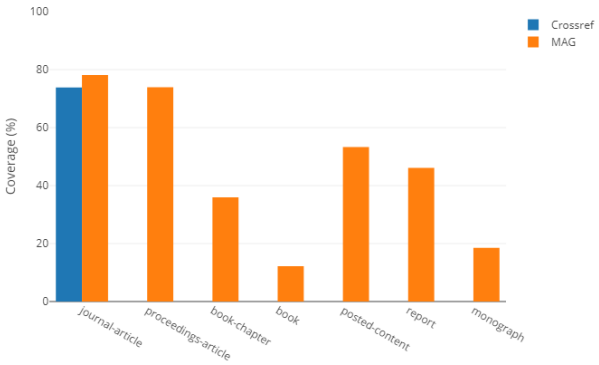


coverage added value - 2021

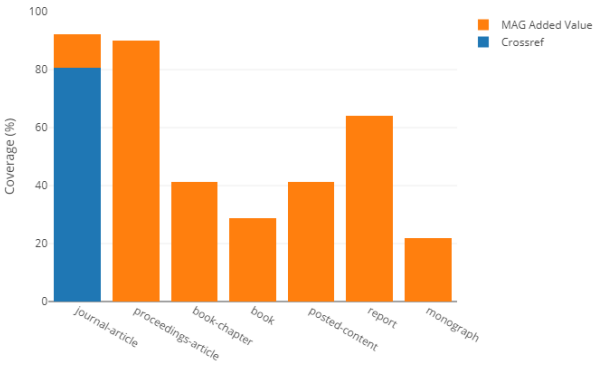
Fields



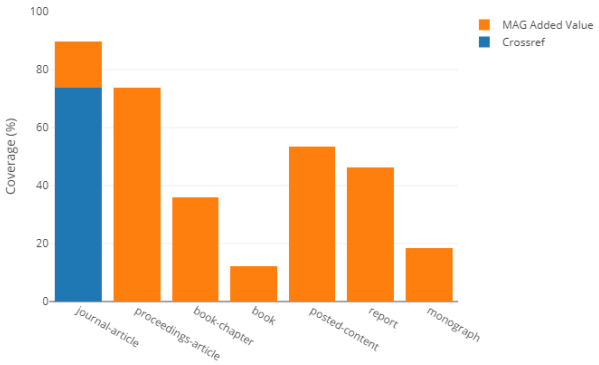
coverage comparison - all time



coverage comparison - 2021



coverage added value - all time

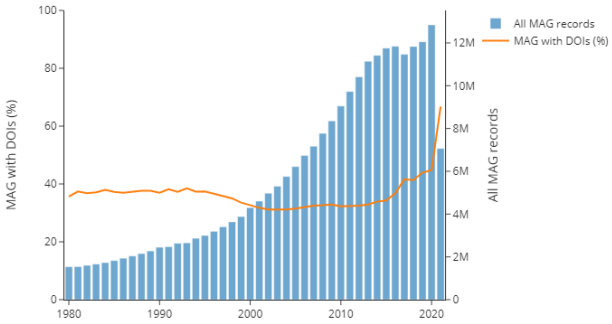


coverage added value - 2021

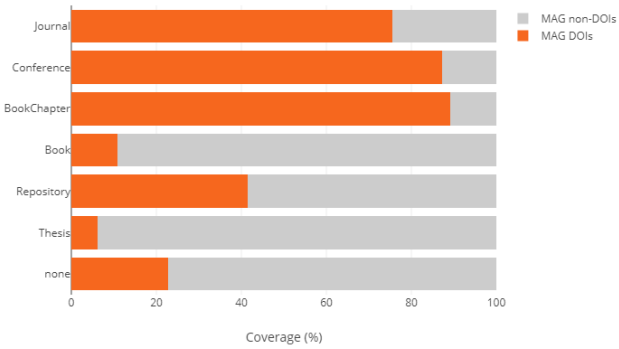
MAG Coverage Beyond Crossref

DOIs vs non-DOIs

By year and publication type



coverage by publication date - all time

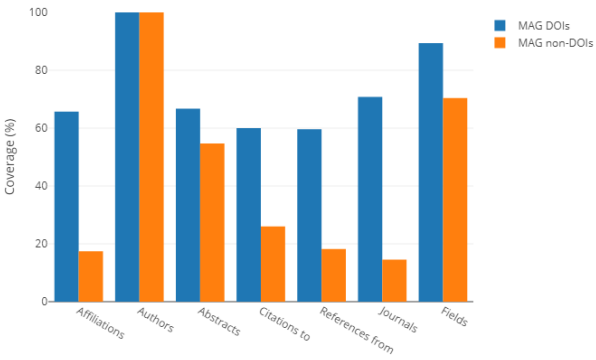


coverage by publication type - all time

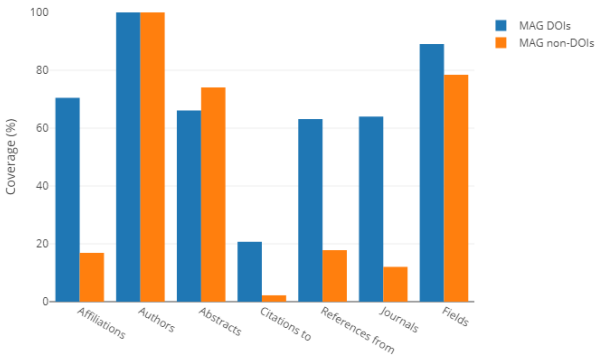
Metadata Coverage

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

MAG Coverage

Table 1. MAG Metadata Coverage of Crossref DOIs

Time Frame	Crossref DOIs	MAG Coverage of DOIs	Author Strings	Author ORCIDs	Affiliation Strings	Abstracts	References	Field Classification	Venue Names	ISSNs
All Time	117531122	90414430	90414430	nan	60833634	61318453	56388851	82244295	67402491	62354538
Crossref Current	12516988	9560727	9560727	nan	6655581	6185554	5926120	8500072	5913995	5408745
Focus Year	5522198	4155213	4155213	nan	2896950	2690078	2662976	3730973	2675796	2464455

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	117531122	99409110	7077220	15322758	12826809	50680820	72386377	113910712	93033865
Crossref Current	12516988	10956639	3515517	2508605	3473233	6029832	6205686	11639248	9074723
Focus Year	5522198	4929884	1740246	1117559	1679994	2776749	2970744	5122257	4170845