OPEN METADATA SOURCES

COMPARING OPENALEX TO CROSSREF

DATE: 18 MARCH 2022

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.

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 recent
- OpenAlex Native Format utrecht-university.OpenAlex_native.Work

Contents

There is actually a way, I think of pulling in a table of contents, but I haven't done that previously. Or it can be done manually obviously.

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 (inlcuding 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 recent
- OpenAlex Native Format: utrecht-university.OpenAlex_native.Work

Crossref Metadata

OpenAlex

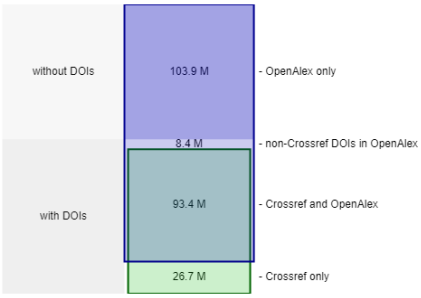
Goals of this report

Limitations

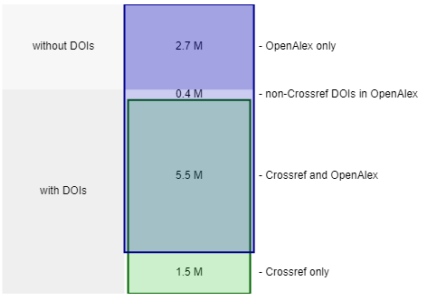
Coverage of OpenAlex vs Crossref

Comparing coverage

OpenAlex coverage all time: proportion with and without DOIs, overlap with Crossref.
OpenAlex coverage of 2020: smaller proportion publications without DOI, same coverage of Crossref

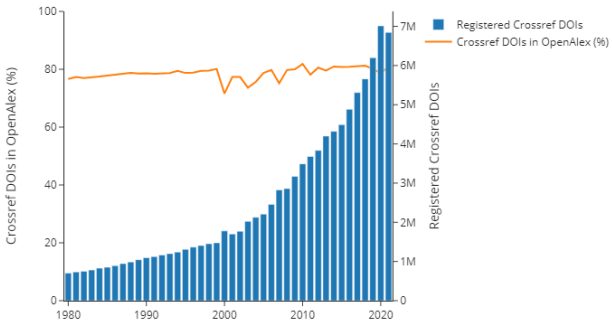


overall comparison - all time

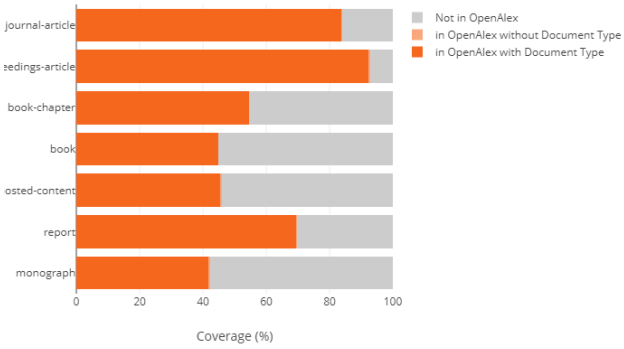


overall comparison - 2020

The proportion of Crossref that is covered in OpenAlex is stable over time, around 75-80%.
Coverage in OpenAlex of publication types in Crossref [describe]



coverage by publication date - all time

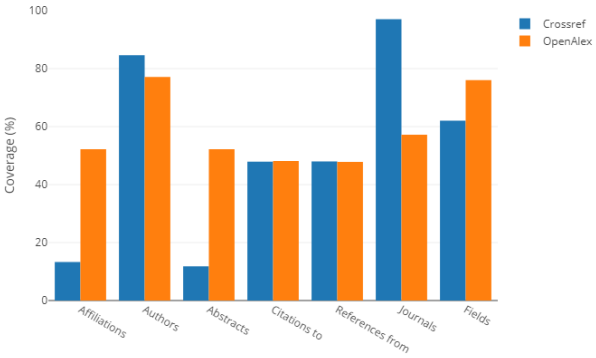


coverage by publication type - all time

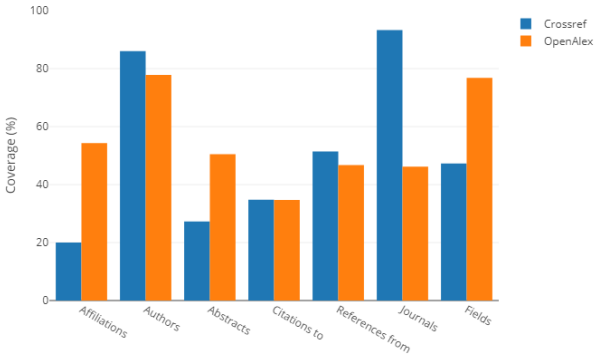
Value Add of OpenAlex to Crossref

Overview

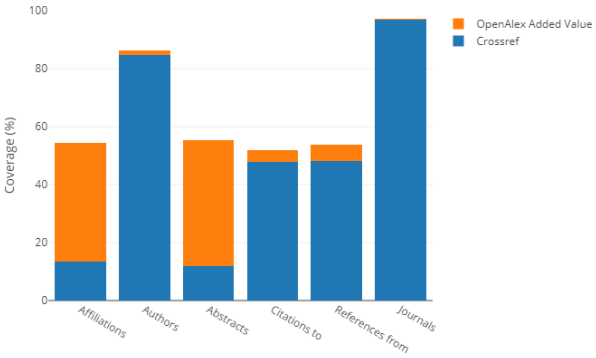
Comparing coverage of metadata types in Crossref and OpenAlex (all time and 2020) -> describe differences Added value of OpenAlex for different metadata types over all publications (all time and 2020) -> describe differences



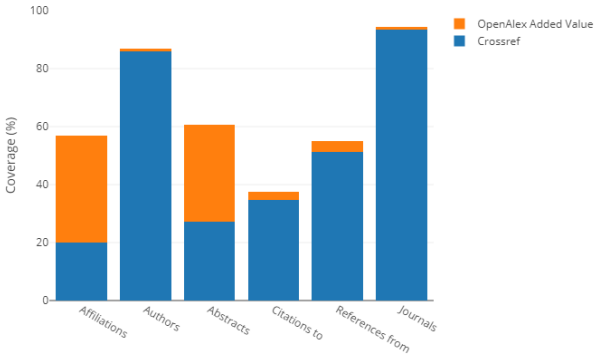
coverage comparison - all time



coverage comparison - 2020



coverage added value - all time

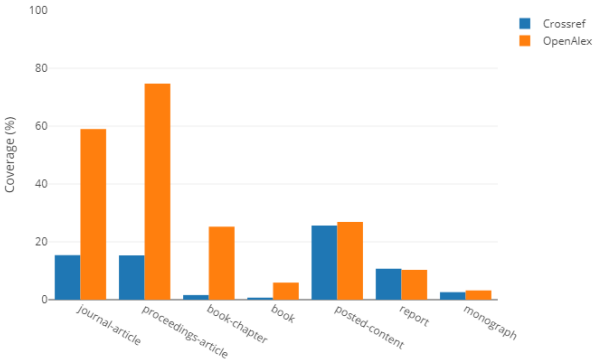


coverage added value - 2020

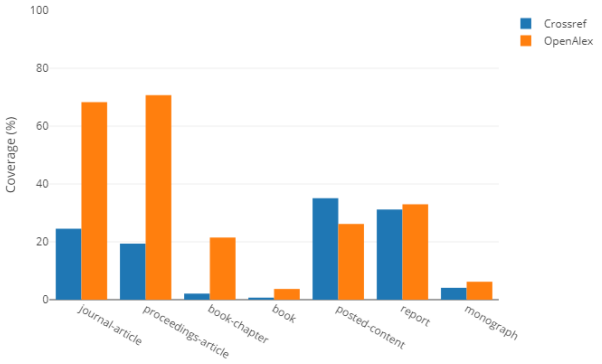
Details

We can do loops eg over the data elements. But this might be better for a supplementary data section as we will presumably want to actually comment on the graphs themselves?

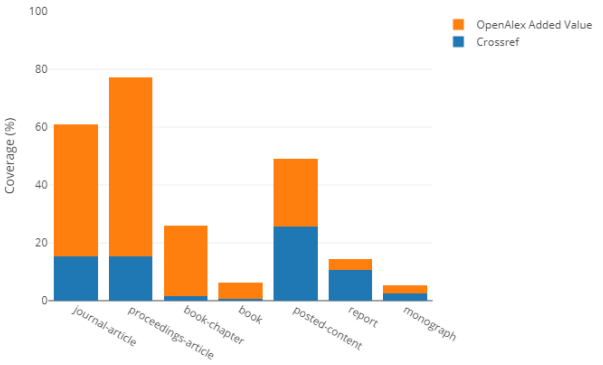
Affiliations



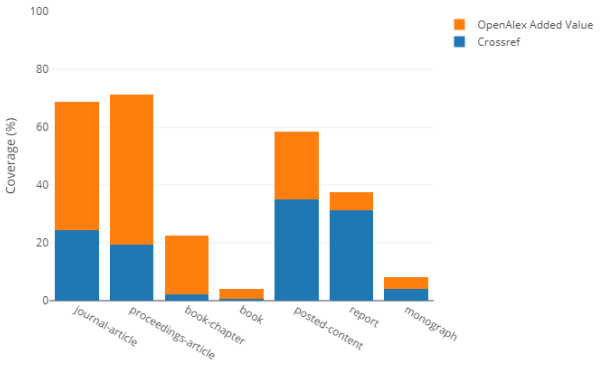
coverage comparison - all time



coverage comparison - 2020

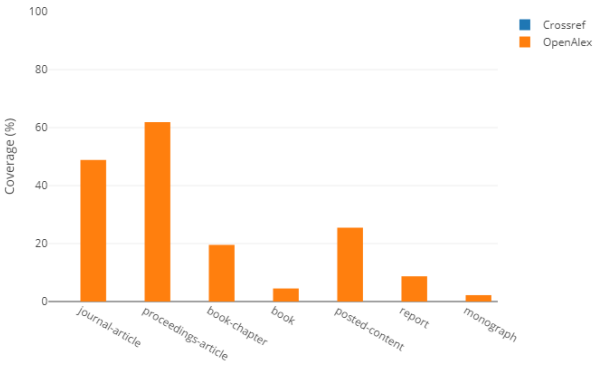


coverage added value - all time

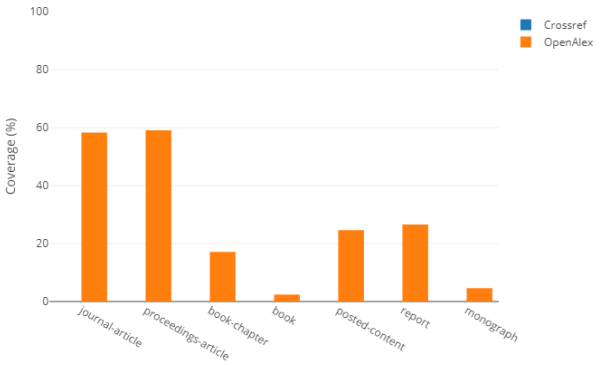


coverage added value - 2020

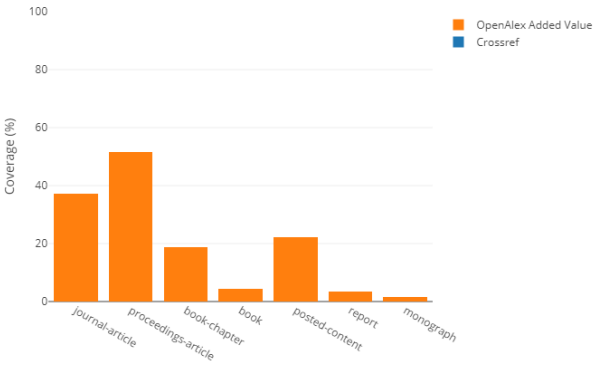
Affiliations ROR



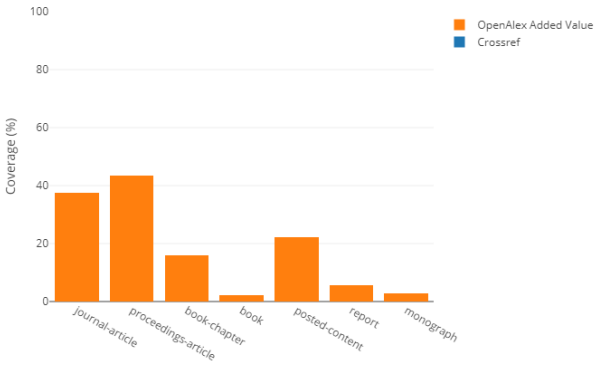
coverage comparison - all time



coverage comparison - 2020

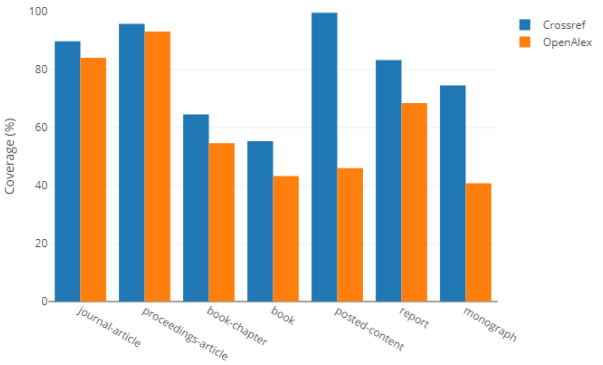


coverage added value - all time

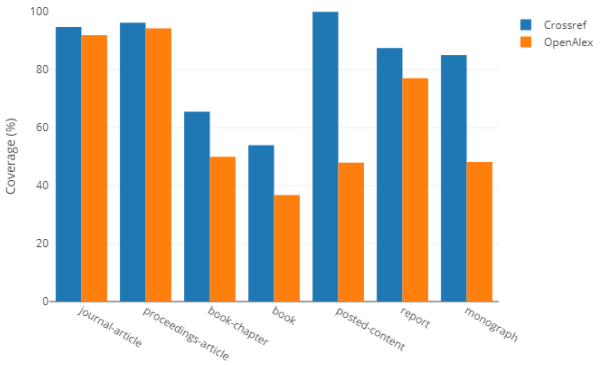


coverage added value - 2020

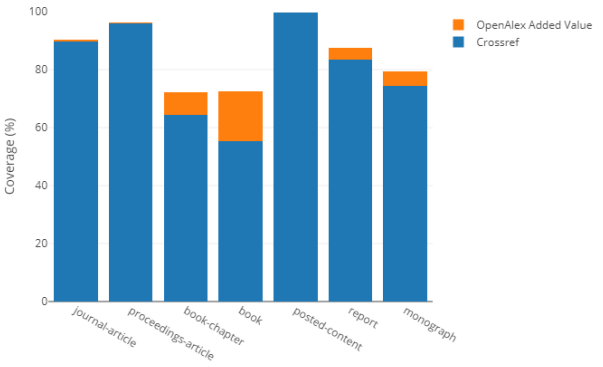
Authors



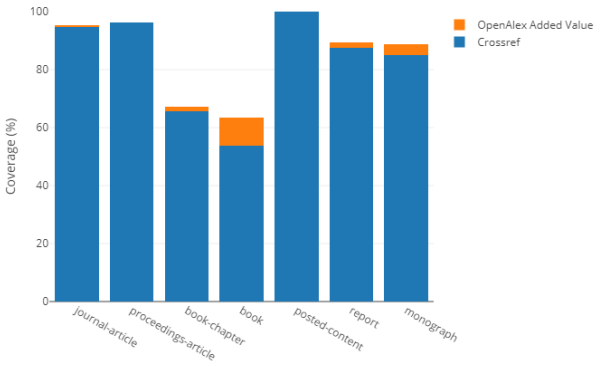
coverage comparison - all time



coverage comparison - 2020

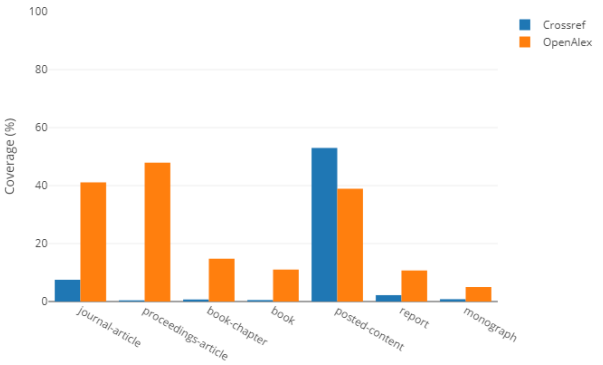


coverage added value - all time

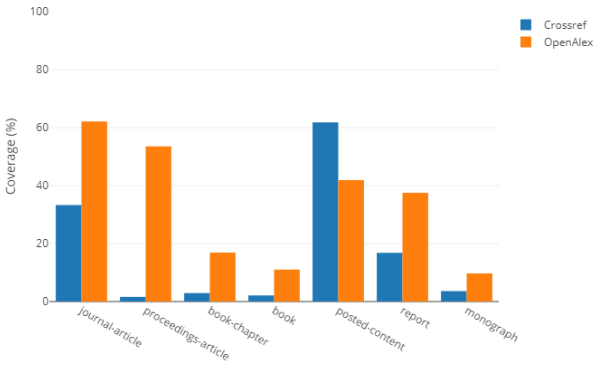


coverage added value - 2020

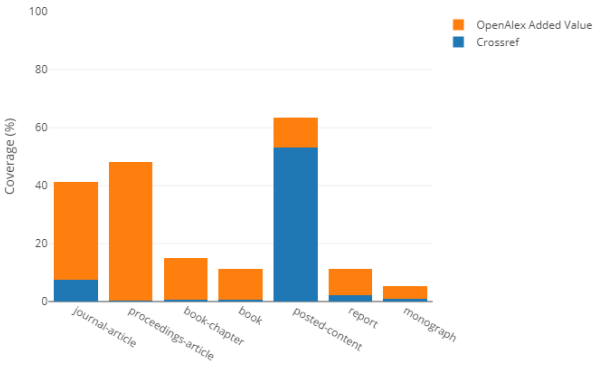
Authors ORCIDs



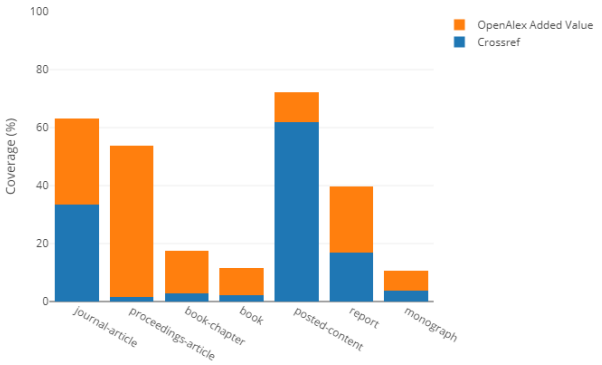
coverage comparison - all time



coverage comparison - 2020

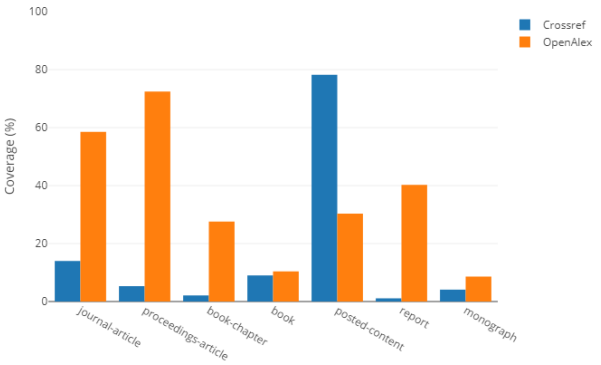


coverage added value - all time

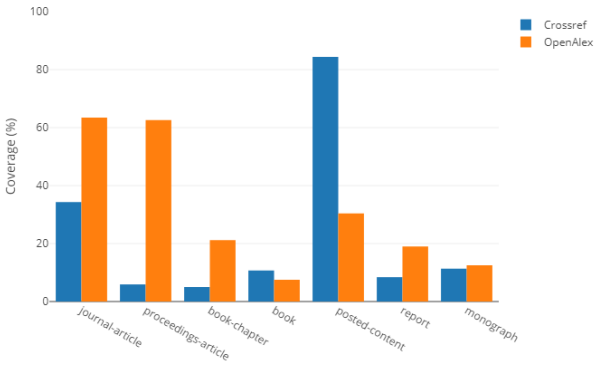


coverage added value - 2020

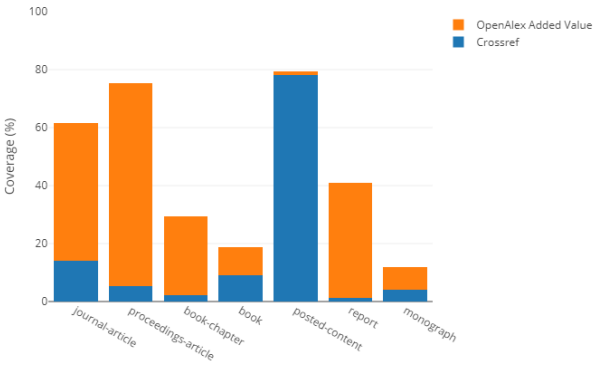
Abstracts



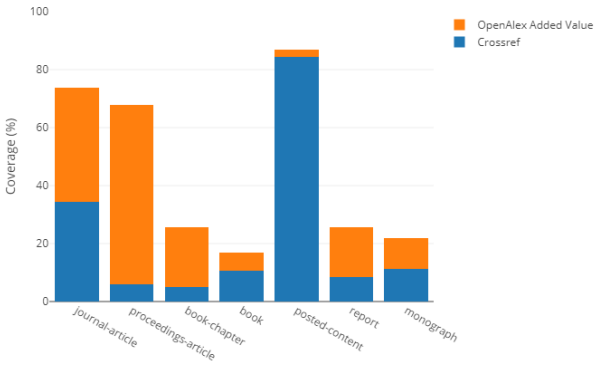
coverage comparison - all time



coverage comparison - 2020

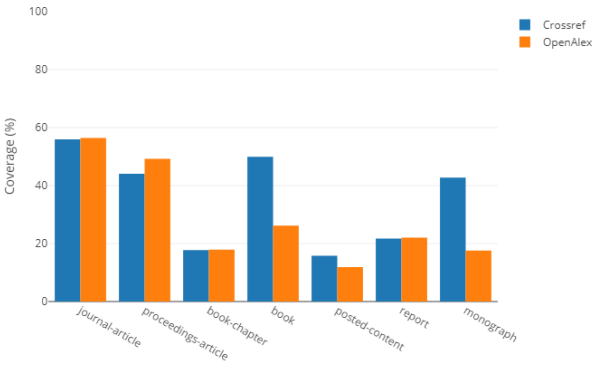


coverage added value - all time

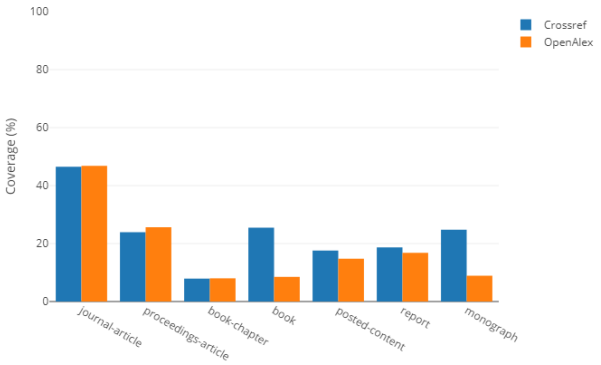


coverage added value - 2020

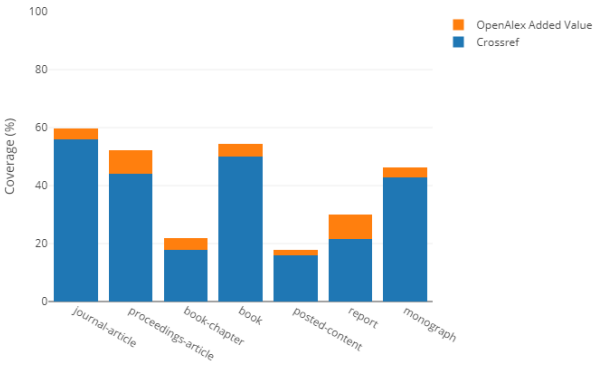
Citations to



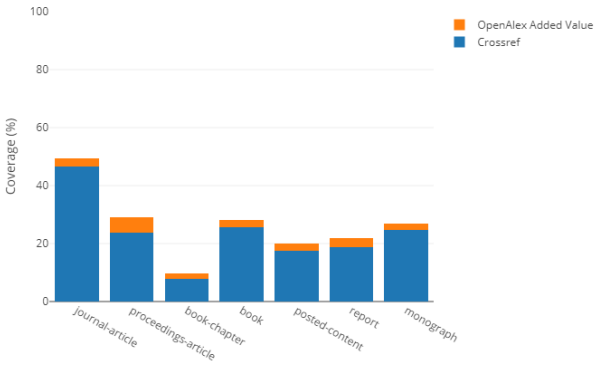
coverage comparison - all time



coverage comparison - 2020

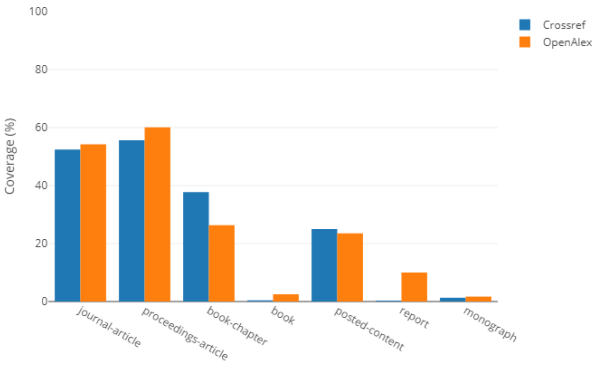


coverage added value - all time

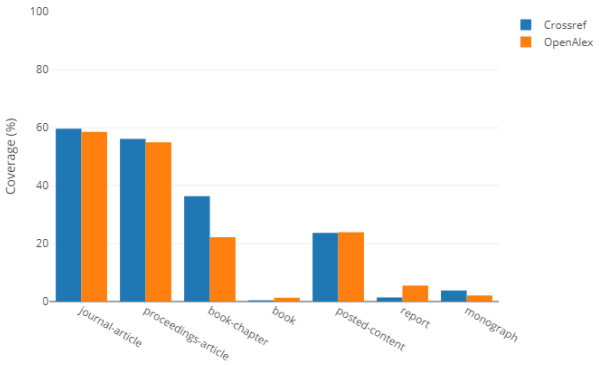


coverage added value - 2020

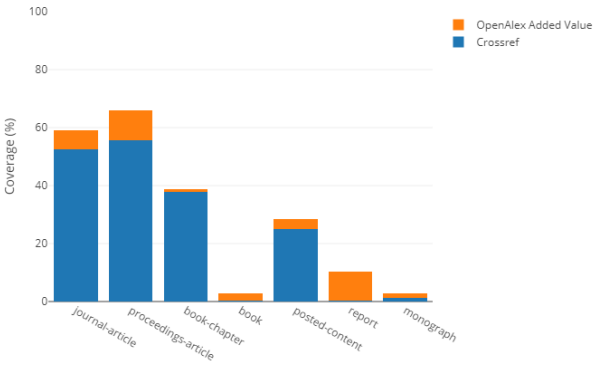
References from



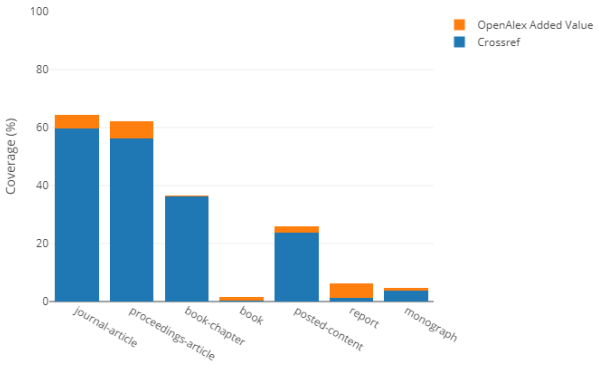
coverage comparison - all time



coverage comparison - 2020

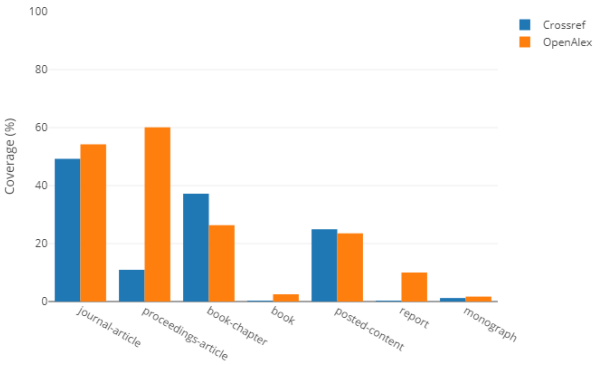


coverage added value - all time

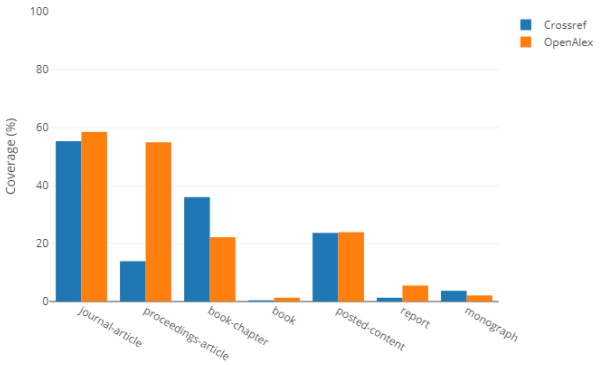


coverage added value - 2020

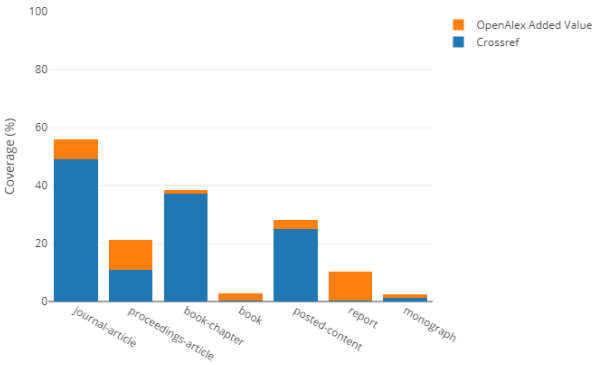
Open References from



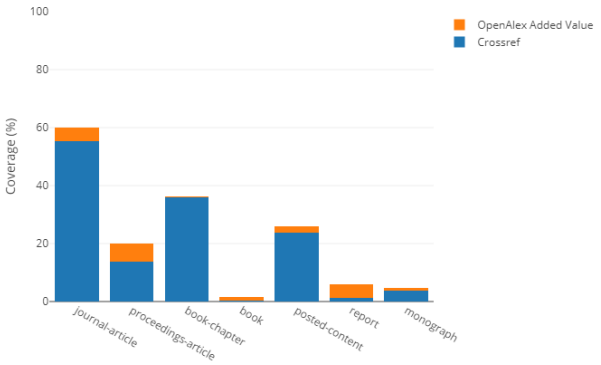
coverage comparison - all time



coverage comparison - 2020

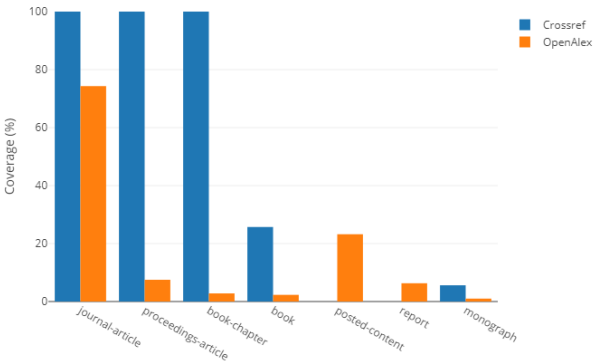


coverage added value - all time

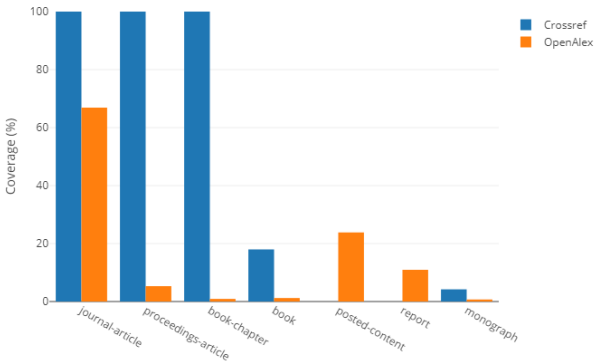


coverage added value - 2020

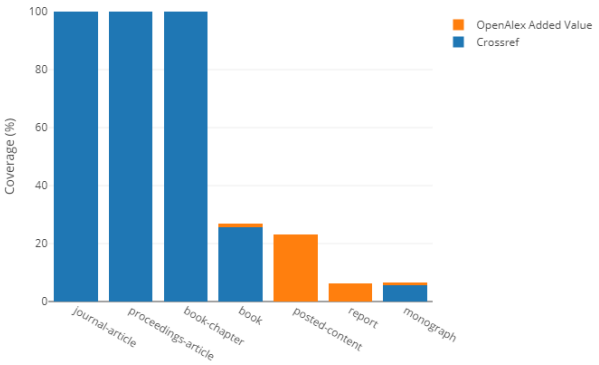
Journals



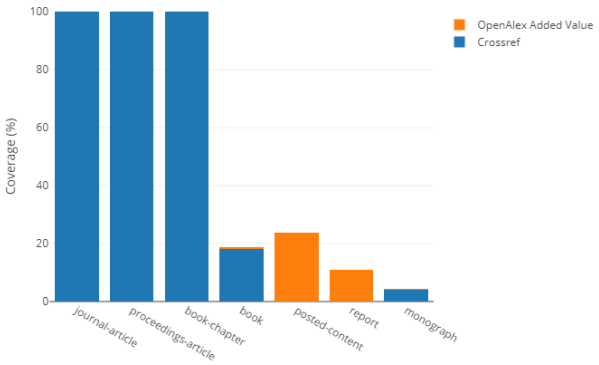
coverage comparison - all time



coverage comparison - 2020

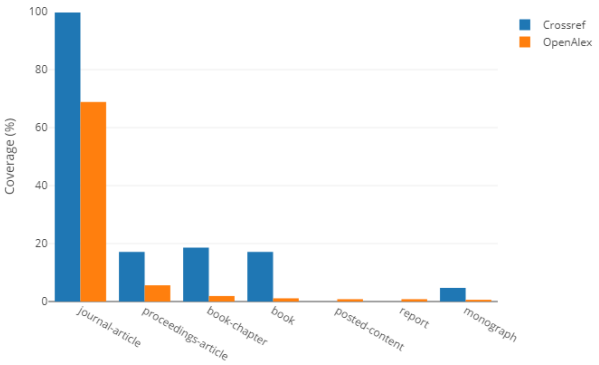


coverage added value - all time

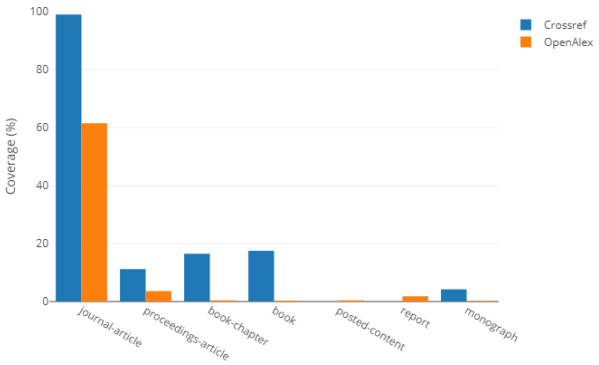


coverage added value - 2020

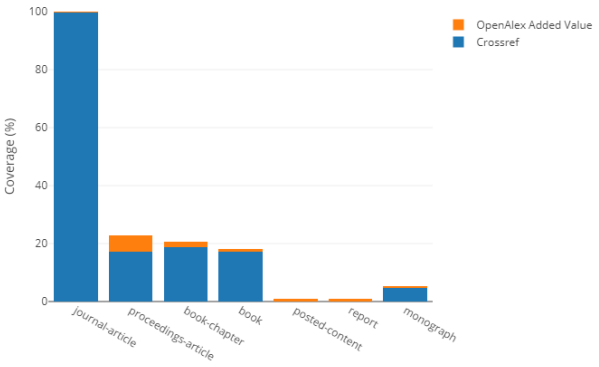
Journals ISSN



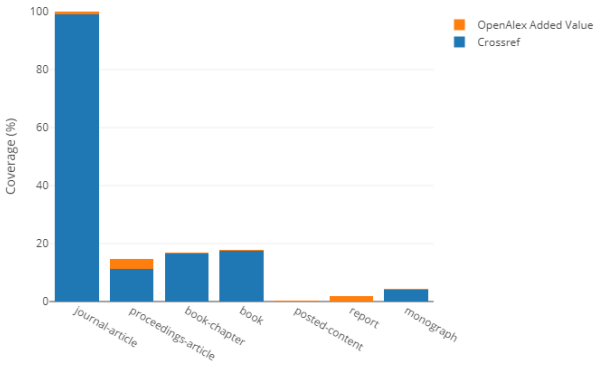
coverage comparison - all time



coverage comparison - 2020

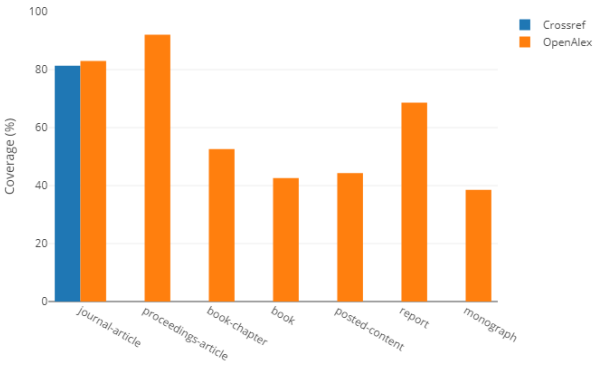


coverage added value - all time

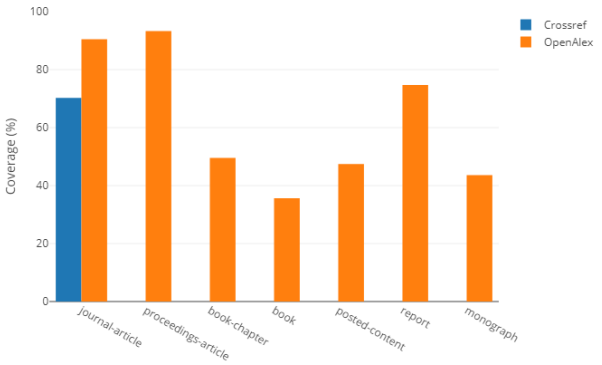


coverage added value - 2020

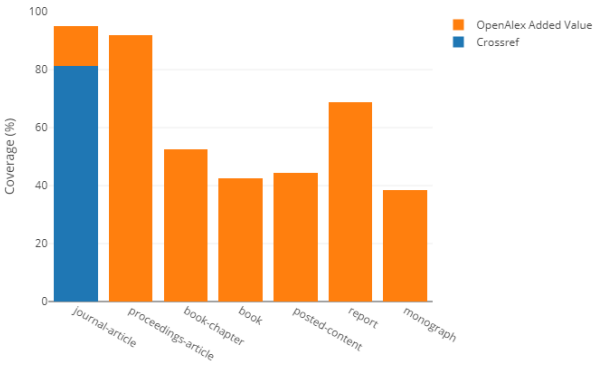
Fields



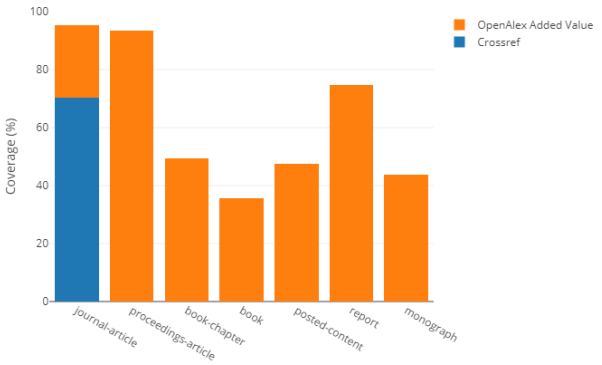
coverage comparison - all time



coverage comparison - 2020



coverage added value - all time



coverage added value - 2020

OpenAlex Coverage Beyond Crossref

Publication Types

Metadata Coverage

Overall

By publication type

By field

Methodology

Appendices

Appendix A - Complete Tables

OpenAlex Coverage

Table 1. OpenAlex Metadata Coverage of Crossref DOIs

Time Frame	Crossref DOIs	OpenAlex Coverage of DOIs
All Time	120141465	93393648
Crossref Current	20058172	16016838
Focus Year	7012560	5514414

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	20058172	17496450	5173497	3936403	5518678	9249685	9862613	18832104	14626037
Focus Year	7012560	6031297	1764980	1399557	1914610	3229854	3315446	6542494	4951558

Appendix B - Historical MAG Analysis??

1. OpenAlex - non-Crossref coverage 4a. Publication types - with and without (Crossref?) DOIs 4b. Coverage of 6 main parameters - with /without (Crossref?) DOIs 4c. Coverage of 6 main parameters per publication type
2. with/without (Crossref?) DOIs

3. Methodology
4. Appendix - tables with AllTheThingsTM