



OPEN DATA ON OPEN SCIENCE

**APPLYING REPRODUCIBLE AND OPEN APPROACHES TO
VALIDATE THE INTRODUCTION OF OPEN ALEX**

@MsPhelps @CameronNeylon @COKIProject



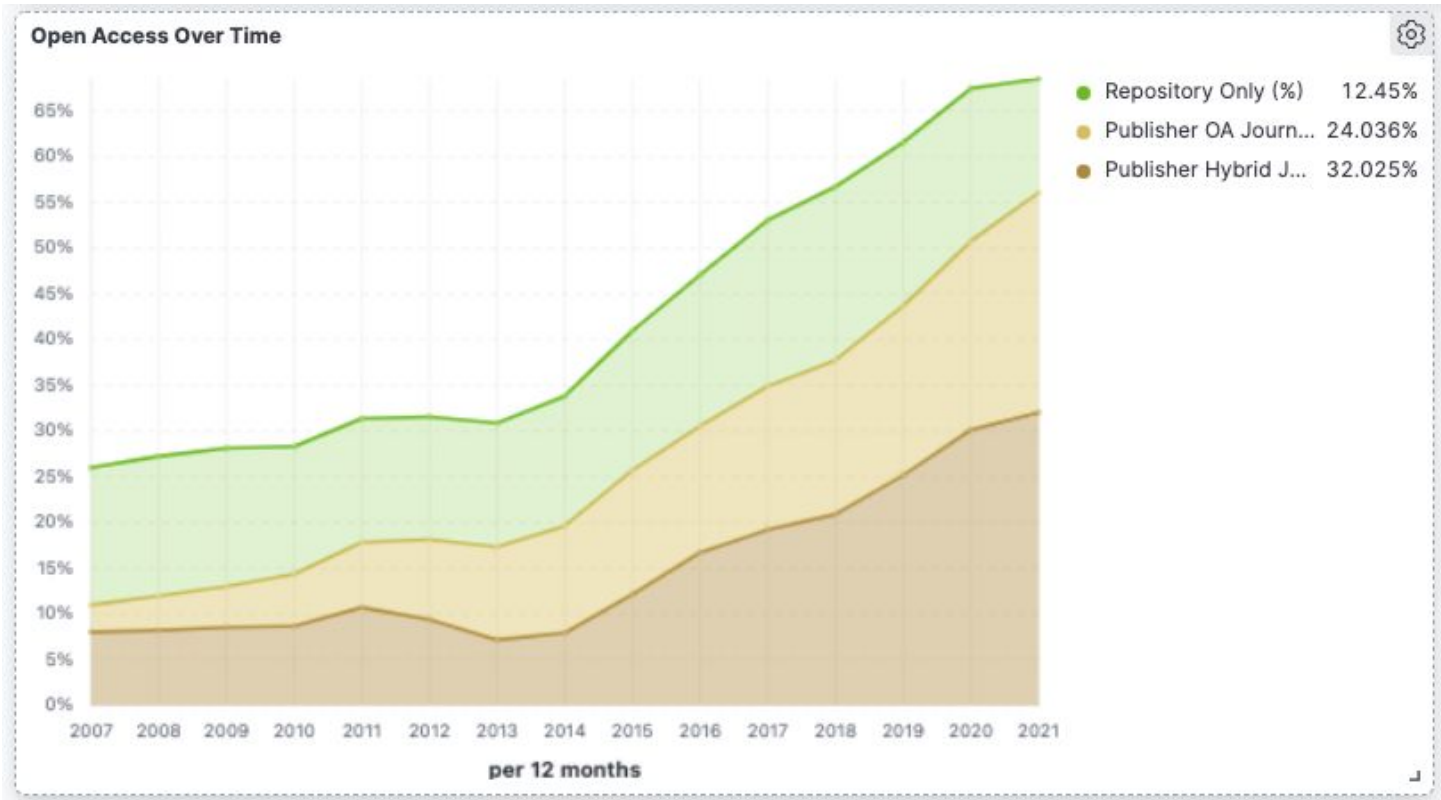
This slide deck: <https://bit.ly/oax-kzb>

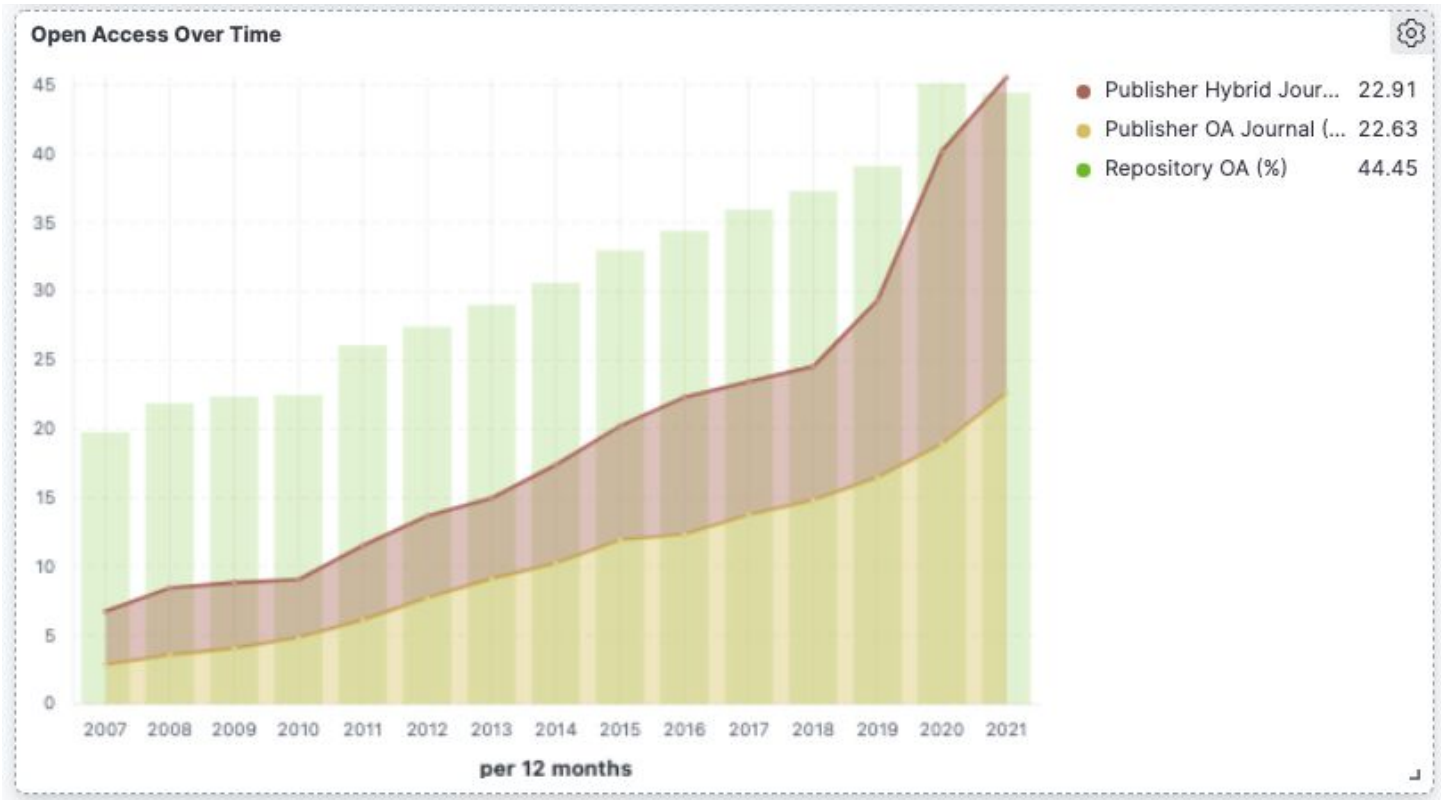


“Our goal is to change the stories that universities tell about themselves, placing open knowledge at the heart of that narrative.”



Being in control of our own story
requires that we be able to control the
data, analytics and algorithms that
makes up that story





```
academic-observatory-workflows X +
https://github.com/The-Academic-Observatory/academic-observatory-workflows/blob/develop/academic_observatory_workflows/databas: ☆

70 genre as output_type,
71 publisher,
72 journal_name,
73 name.name as normalised_journal_name,
74 journal_issn_l,
75
76 {#
77 ### Is Open Access:
78
79 We use the is_oa tag from Unpaywall directly to populate general OA status. This includes bronze.
80 #}
81 is_oa,
82 journal_is_in_doaj as is_in_doaj,
83 journal_is_oa,
84 oa_status as unpaywall_oa_status,
85
86 {#
87 ### Gold Open Access:
88
89 gold OA is defined as either the journal being in DOAJ or the best_oa_location being a publisher and a
90 license being detected. This works because Unpaywall will set the publisher as the best oa location if
91 it identifies an accessible publisher copy.
92 #}
93 CASE
94   WHEN journal_is_in_doaj OR (best_oa_location.host_type = "publisher" AND best_oa_location.license is not null AND not journal_is_in_doaj) THEN TRUE
95   ELSE FALSE
96 END
97   as gold,
98
99 {#
100 ### Gold Open Access in DOAJ Journal:
101
102 gold_just_doaj is determined directly from the Unpaywall statement that the journal is in DOAJ. No further
103 checking is done on this, so articles that Unpaywall does not capture as being accessible that are in DOAJ
104 journals will be characterised as gold_just_doaj.
105 #}
106 CASE
107   WHEN journal_is_in_doaj THEN TRUE
108   ELSE FALSE
109 END
110   as gold_just_doaj,
111 ...
```



GERMANY



Germany, officially the Federal Republic of Germany, is a country in Central Europe. It is the second most populous country in Europe after Russia, and the most populous member state of the European Union.

Derived from [Wikipedia](#) licensed [CC-BY-SA](#)

WIKIPEDIA

DOWNLOAD

EMBED

Western Europe

Europe



OPEN ACCESS
PERCENTAGE
40%

TOTAL
PUBLICATIONS
2.3m

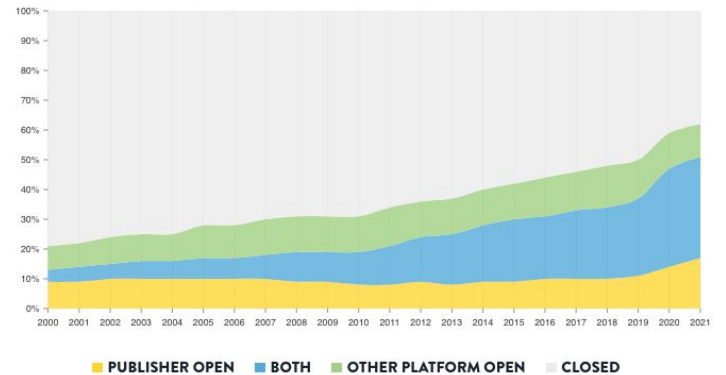
TOTAL OPEN
PUBLICATIONS
925k

TOTAL
CITATIONS
63.4m

BREAKDOWN



PERCENTAGE OF OPEN ACCESS OVER TIME



KENYA



Kenya, officially the Republic of Kenya, is a country in Eastern Africa. At 580,367 square kilometres, Kenya is the world's 48th largest country by area. With a population of more than 47.6 million in the 2019 census, Kenya is the 29th most populous country.

Derived from [Wikipedia](#) licensed [CC-BY-SA](#)

WIKIPEDIA

DOWNLOAD

EMBED

Sub-Saharan Africa

Africa



OPEN ACCESS
PERCENTAGE
60%

TOTAL
PUBLICATIONS
33k

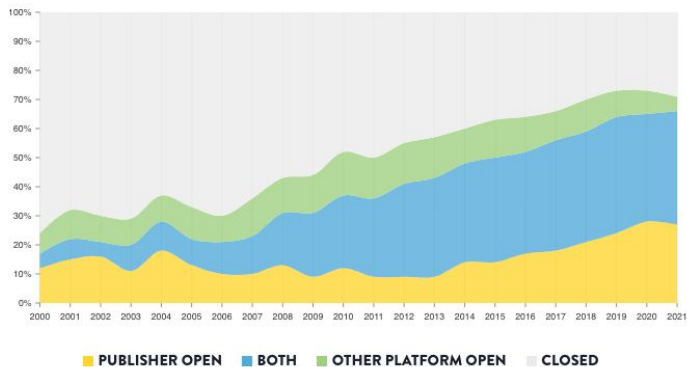
TOTAL OPEN
PUBLICATIONS
20k

TOTAL
CITATIONS
751k

BREAKDOWN



PERCENTAGE OF OPEN ACCESS OVER TIME



GERMANY



Germany, officially the Federal Republic of Germany, is a country in Central Europe. It is the second most populous country in Europe after Russia, and the most populous member state of the European Union.

Derived from [Wikipedia](#) licensed [CC-BY-SA](#)

WIKIPEDIA

DOWNLOAD

EMBED

Western Europe

Europe



OPEN ACCESS
PERCENTAGE
40%

TOTAL
PUBLICATIONS
2.3m

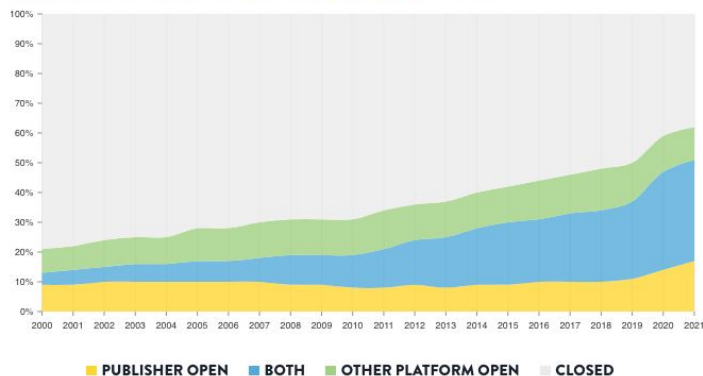
TOTAL OPEN
PUBLICATIONS
925k

TOTAL
CITATIONS
63.4m

BREAKDOWN



PERCENTAGE OF OPEN ACCESS OVER TIME



UGANDA



Uganda officially the Republic of Uganda, is a landlocked country in East Africa. The country is bordered to the east by Kenya, to the north by South Sudan, to the west by the Democratic Republic of the Congo, to the south-west by Rwanda, and to the south by Tanzania. Derived from [Wikipedia](#) licensed [CC-BY-SA](#).



OPEN ACCESS
PERCENTAGE
67%

TOTAL
PUBLICATIONS
19k

TOTAL OPEN
PUBLICATIONS
13k

TOTAL
CITATIONS
442k

WIKIPEDIA

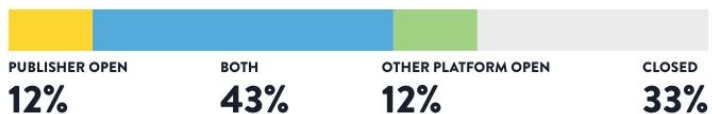
DOWNLOAD

EMBED

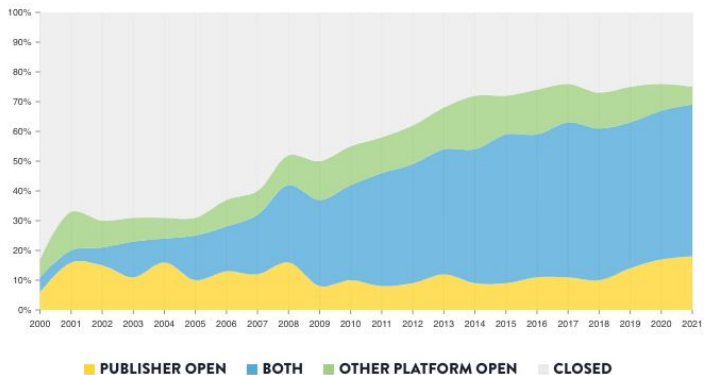
Sub-Saharan Africa

Africa

BREAKDOWN



PERCENTAGE OF OPEN ACCESS OVER TIME



GERMANY



Germany, officially the Federal Republic of Germany, is a country in Central Europe. It is the second most populous country in Europe after Russia, and the most populous member state of the European Union.

Derived from [Wikipedia](#) licensed [CC-BY-SA](#).



OPEN ACCESS
PERCENTAGE
40%

TOTAL
PUBLICATIONS
2.3m

TOTAL OPEN
PUBLICATIONS
925k

TOTAL
CITATIONS
63.4m

WIKIPEDIA

DOWNLOAD

EMBED

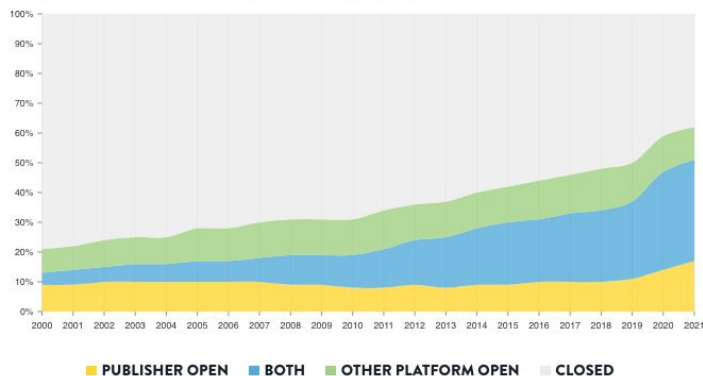
Western Europe

Europe

BREAKDOWN



PERCENTAGE OF OPEN ACCESS OVER TIME



ETHIOPIA



Ethiopia, officially the Federal Democratic Republic of Ethiopia, is a landlocked country in the Horn of Africa. It shares borders with Eritrea and Djibouti to the north, Somalia to the east and northeast, Kenya to the south, South Sudan to the west, and Sudan to the northwest. Derived from [Wikipedia](#) licensed [CC-BY-SA](#).

WIKIPEDIA

DOWNLOAD

EMBED

Sub-Saharan Africa

Africa



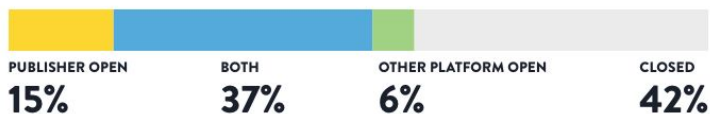
OPEN ACCESS
PERCENTAGE
58%

TOTAL
PUBLICATIONS
55k

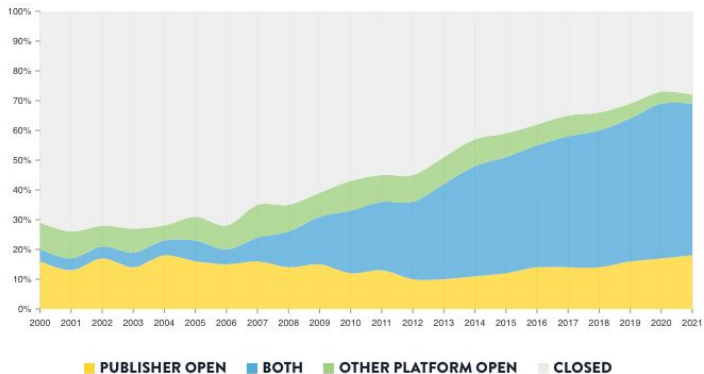
TOTAL OPEN
PUBLICATIONS
32k

TOTAL
CITATIONS
941k

BREAKDOWN



PERCENTAGE OF OPEN ACCESS OVER TIME



GERMANY



Germany, officially the Federal Republic of Germany, is a country in Central Europe. It is the second most populous country in Europe after Russia, and the most populous member state of the European Union.

Derived from [Wikipedia](#) licensed [CC-BY-SA](#).

WIKIPEDIA

DOWNLOAD

EMBED

Western Europe

Europe



OPEN ACCESS
PERCENTAGE
40%

TOTAL
PUBLICATIONS
2.3m

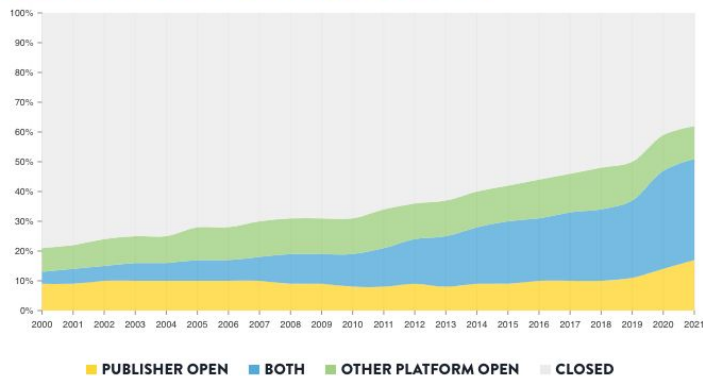
TOTAL OPEN
PUBLICATIONS
925k

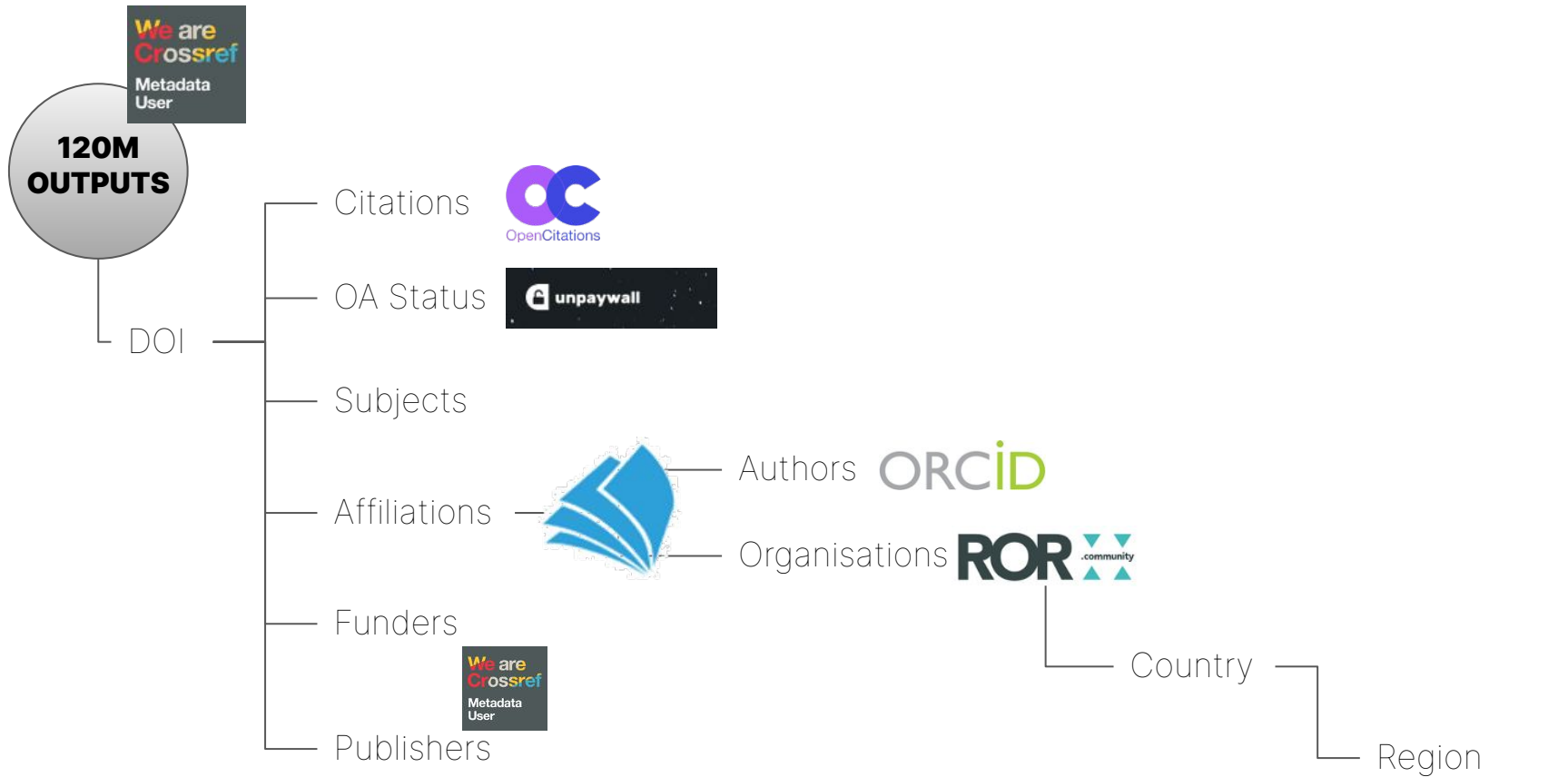
TOTAL
CITATIONS
63.4m

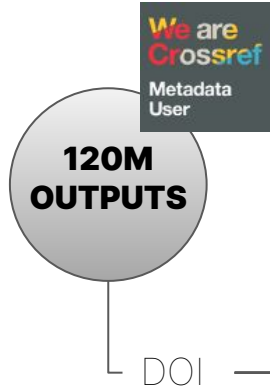
BREAKDOWN



PERCENTAGE OF OPEN ACCESS OVER TIME







We are Crossref
Metadata User

120M
OUTPUTS

DOI

Citations



OA Status



Subjects

Affiliations



Authors

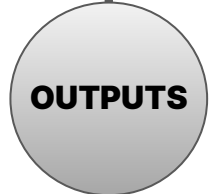
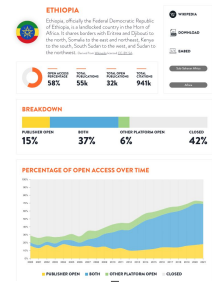
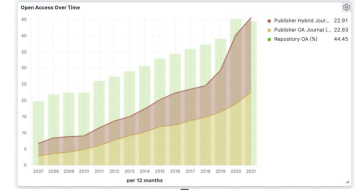
ORCID

Funders



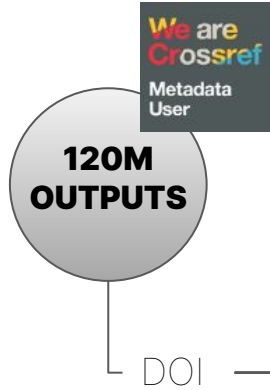
Publishers

Organisations

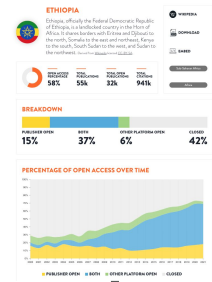
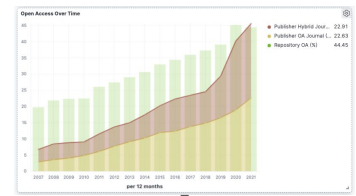
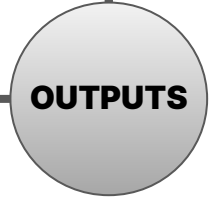
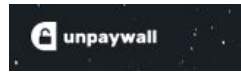


Country

Region



- Citations
- OA Status
- Subjects
- Authors
- Affiliations
- Funders
- Publishers



120M
OUTPUTS

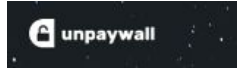


DOI

Citations



OA Status



Subjects

Affiliations



Authors



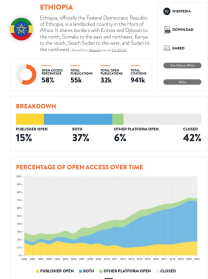
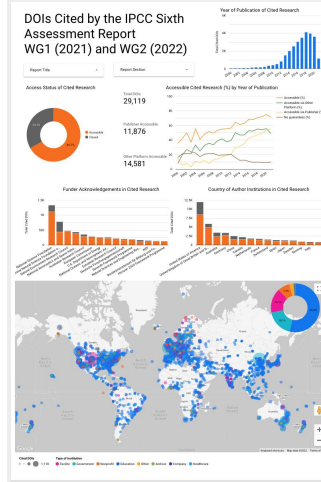
Organisations



Funders



Publishers



OUTPUTS

Country

Region

120M
OUTPUTS



DOI

Citations



OA Status



Subjects



Authors



Affiliations

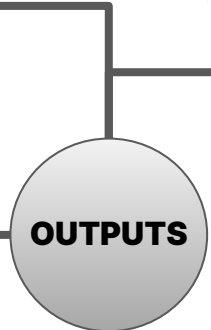
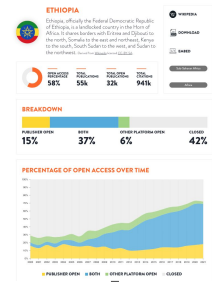
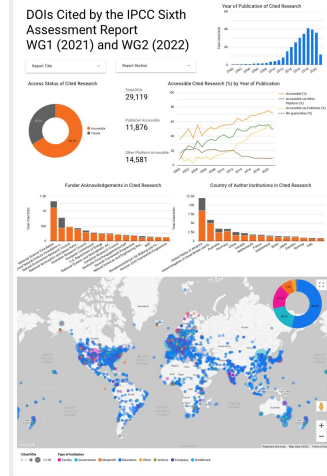
Organisations



Funders



Publishers



Country

Region



The loss of Microsoft Academic is a challenge...but also an opportunity...



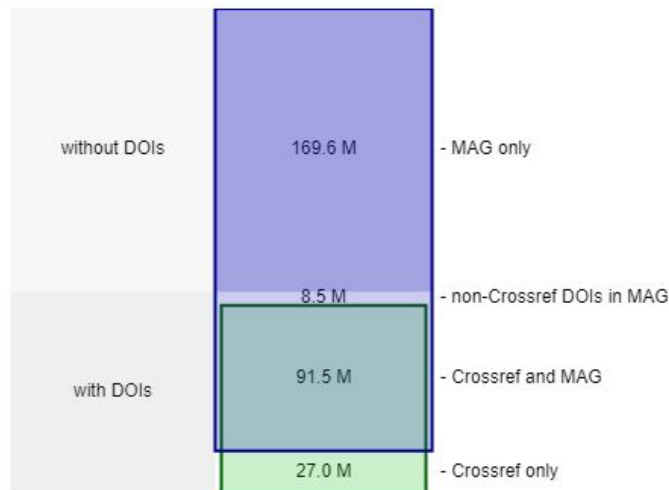
...what if we actually critically validated
a dataset *before* we started relying on
it for strategic decision making?

VALIDATING OPEN ALEX

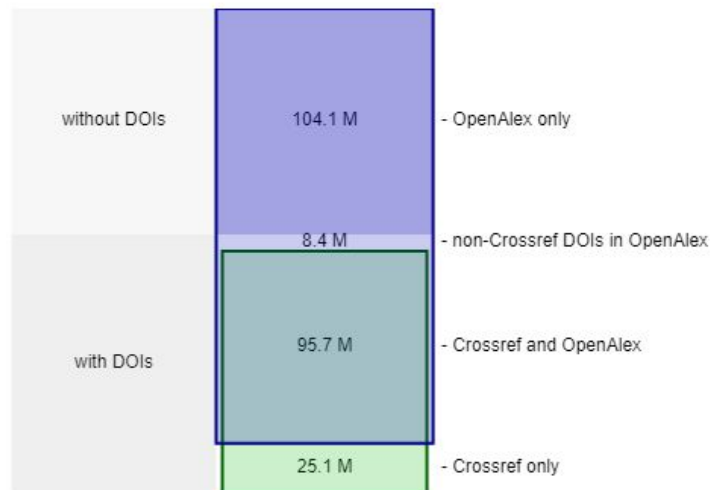
OPEN ANALYSIS OF OPEN DATA



COMPARING MAG AND OPENALEX COVERAGE

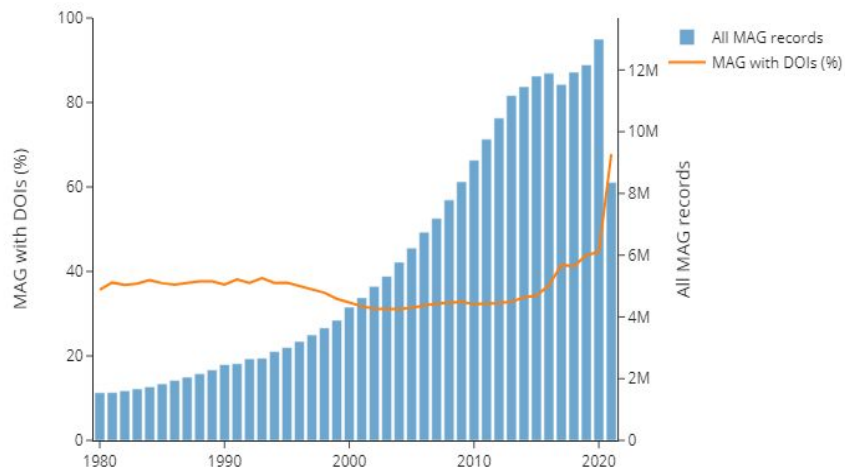


MAG Last Dump - Dec 2021
(271 M records)

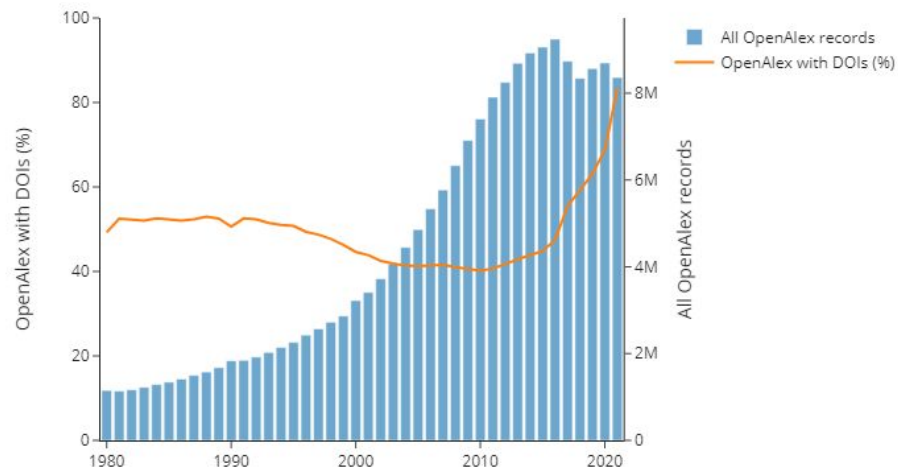


OpenAlex - March 2022
(209 M records, no patents)

COMPARING MAG AND OPENALEX COVERAGE

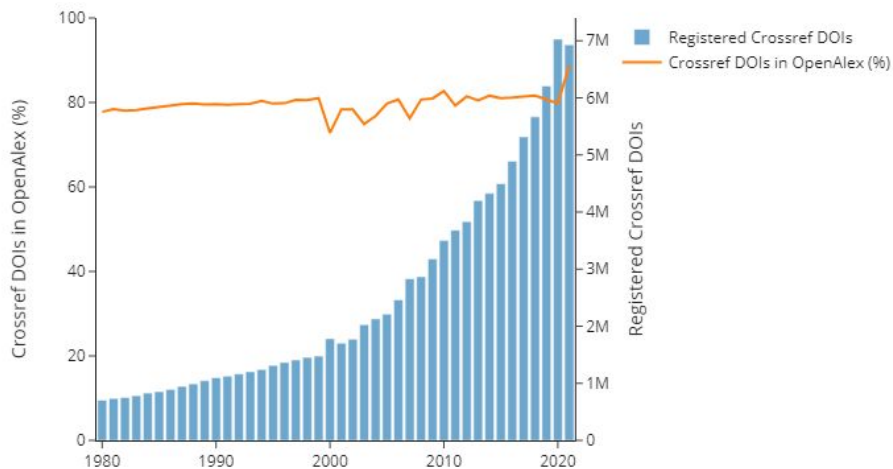


MAG Last Dump - Dec 2021

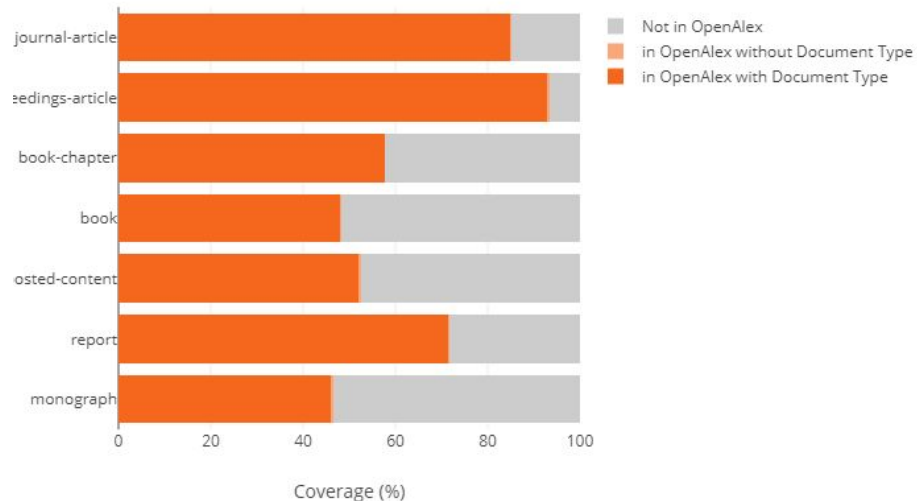


OpenAlex - March 2022

OPENALEX ADDED VALUE TO CROSSREF DATA

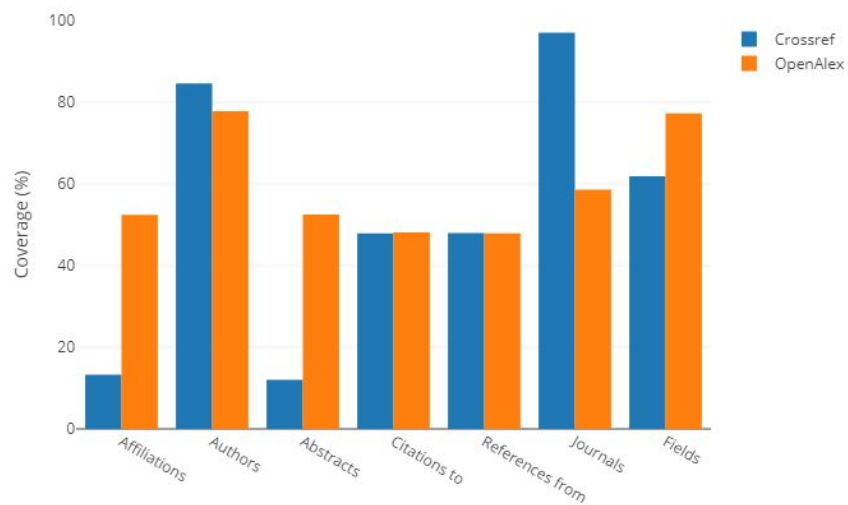


coverage of Crossref DOIs
by publication year

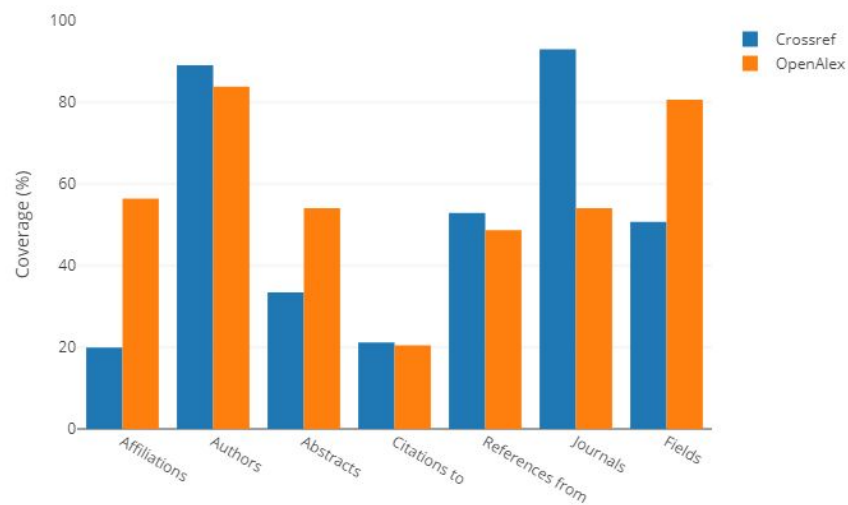


coverage of Crossref DOIs
by publication type

OPENALEX ADDED VALUE TO CROSSREF DATA

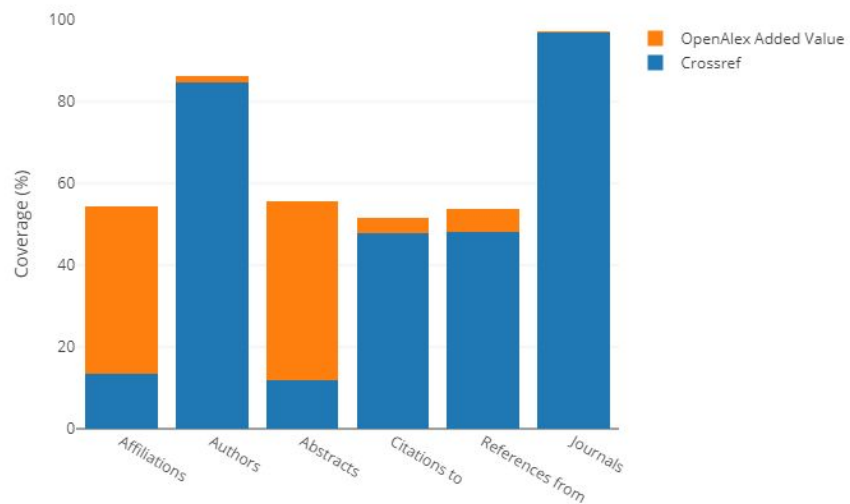


coverage comparison - all time

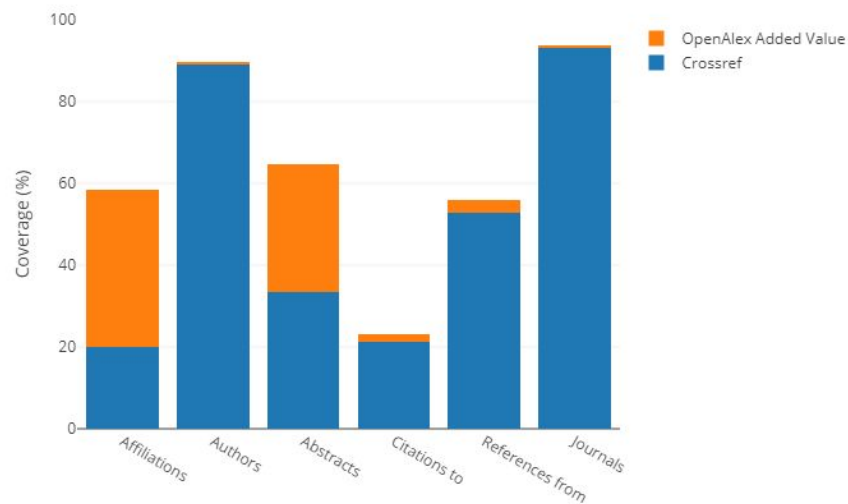


coverage comparison - 2021

OPENALEX ADDED VALUE TO CROSSREF DATA



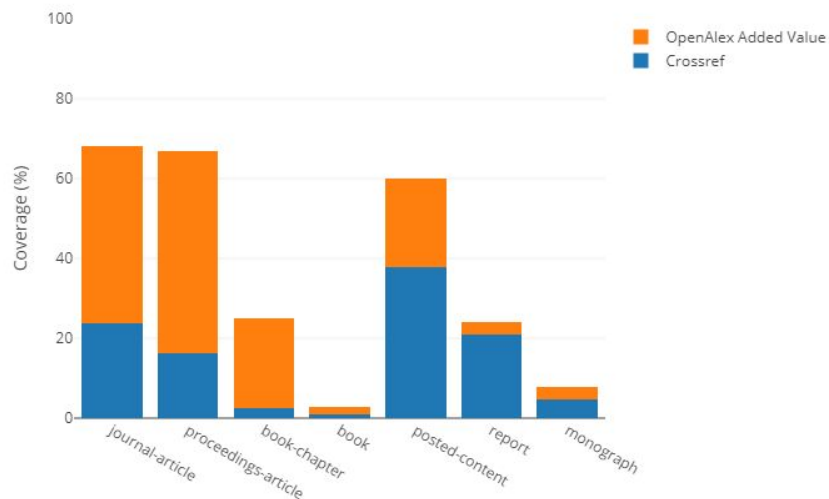
added value - all time



added value - 2021

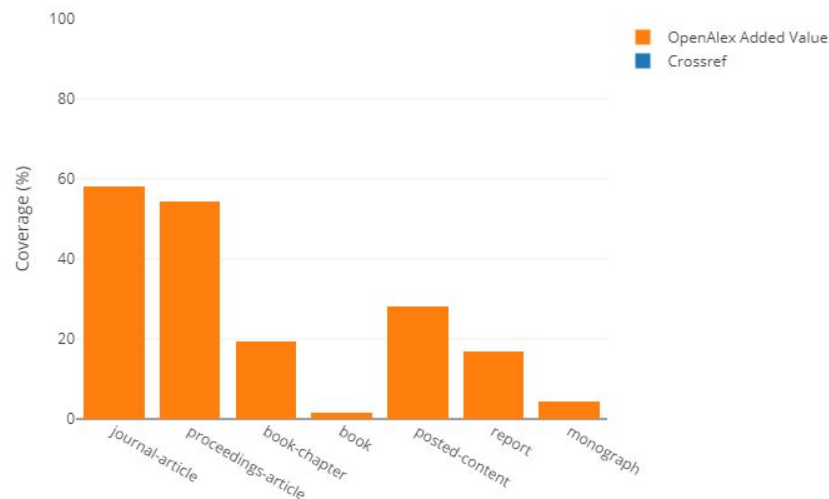
OPENALEX ADDED VALUE TO CROSSREF DATA

AFFILIATIONS



added value - 2021

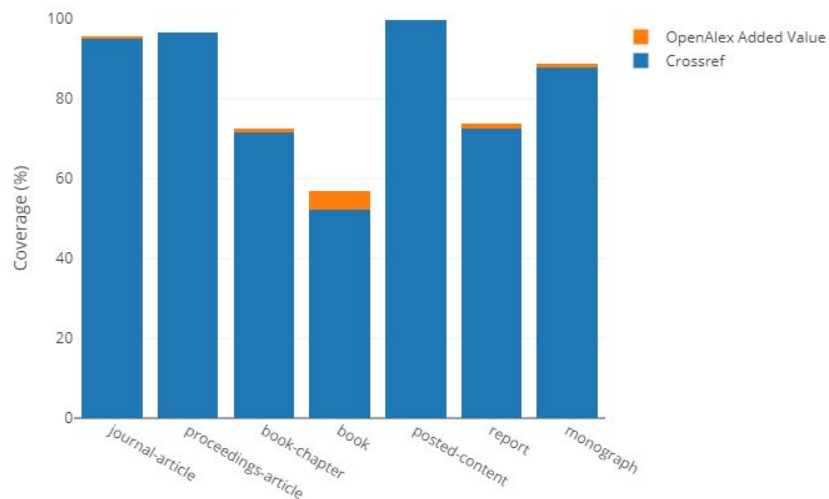
AFFILIATIONS - ROR



added value - 2021

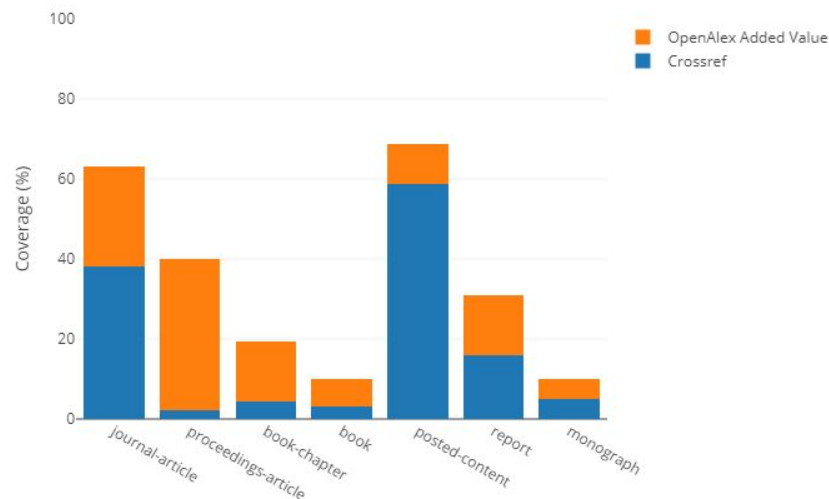
OPENALEX ADDED VALUE TO CROSSREF DATA

AUTHORS



added value - 2021

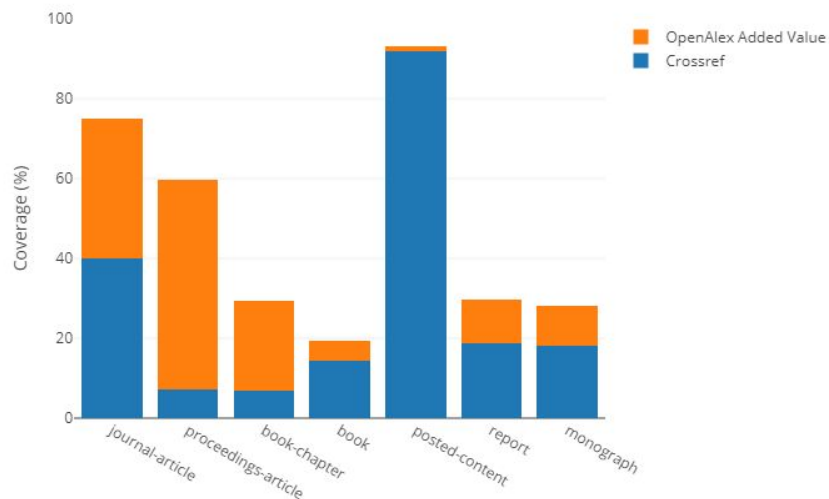
AUTHORS - ORCID



added value - 2021

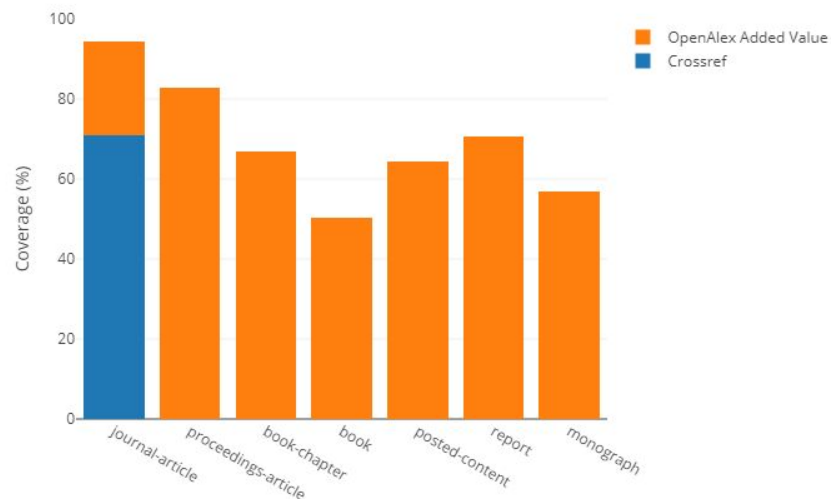
OPENALEX ADDED VALUE TO CROSSREF DATA

ABSTRACTS



added value - 2021

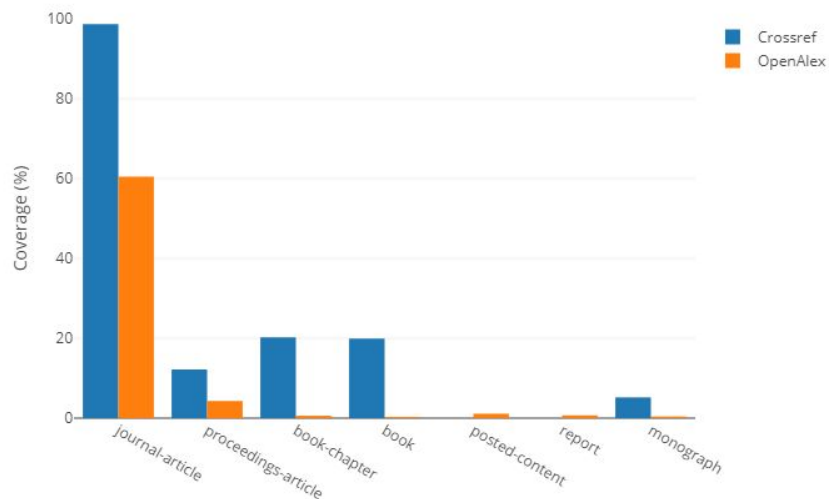
SUBJECTS



added value - 2021

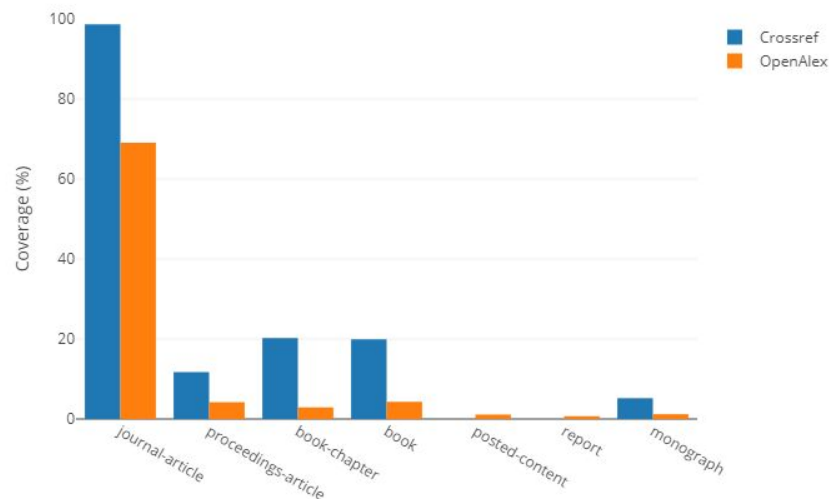
OPENALEX ADDED VALUE TO CROSSREF DATA

ISSN



coverage comparison - 2021
OpenAlex January 2022

ISSN



coverage comparison - 2021
OpenAlex March 2022

REFLECTIONS

OPEN ANALYSIS OF OPEN DATA



THE WORLD HAS CHANGED

- Greater expectations of transparency and reproducibility
- A lot more data, not all of which is being audited
- A lot more critical analysis of algorithms and analysis (not all of which is in good faith)
- A lack of investment in the systems that would allow us to deliver more transparent, community curated and validated information

We have an opportunity to do things differently...

OPEN ALEX IS A POWERFUL RESOURCE

- It is growing and changing as you might expect
 - It has gaps and issues (as you would reasonably expect)
 - It is already sufficiently powerful to be used in a range of applications and its current state is a massive achievement
-
- The fact that it exists and is open allows us to do things differently
 - It will only get better, particularly if it is critiqued and supported

PICTURE A WORLD...

- ...where open and re-usable resources are available and combined for a wide variety of community use cases
- ...where there is an ecosystem of pathways for correction and curation by the experts on a particular context
- ...where the choices on how data is gathered and processed, and how metrics are calculated is in our hands

This world is within grasp if we choose to support it

Acknowledgements

Centre for Culture and Technology

Lucy Montgomery
Katie Wilson
Chun-Kai (Karl) Huang
Niamh Quigley

Funding from

Research Office at Curtin
Faculty of Humanities
School of Media, Creative Arts & Social Inquiry
Andrew W. Mellon Foundation
Arcadia Fund

Curtin Institute for Computation

Kathryn Hosking
Rebecca Handcock
Aniek Roelofs
Jamie Diprose
Julian Tonti-Filippini
Tuan Chien
Rebecca Lange
Richard Hosking

Precipy Reporting Framework

Ana Nelson

LINKS AND MATERIALS

- This slide deck: <https://bit.ly/oax-kzb>
- [Code, data and queries](#) Github repository for this analysis
- Automatically generated reports:
 - MAG Last Dump ([2020 Focus](#), [2021 Focus](#))
 - OpenAlex January ([2021 Focus](#))
 - OpenAlex March ([2021 Focus](#), [2022 Focus](#))
- Data Integration [Platform](#) and [Workflows](#) [Zenodo]
 - Documentation: [Platform](#) and [Workflows](#)
 - Github: [Platform](#) and [Workflows](#)