

# Analytics data – New options, quality and coverage

**Ludo Waltman**

Centre for Science and Technology Studies (CWTS), Leiden University

**OPERA workshop**

Lyngby, Denmark

March 28, 2019

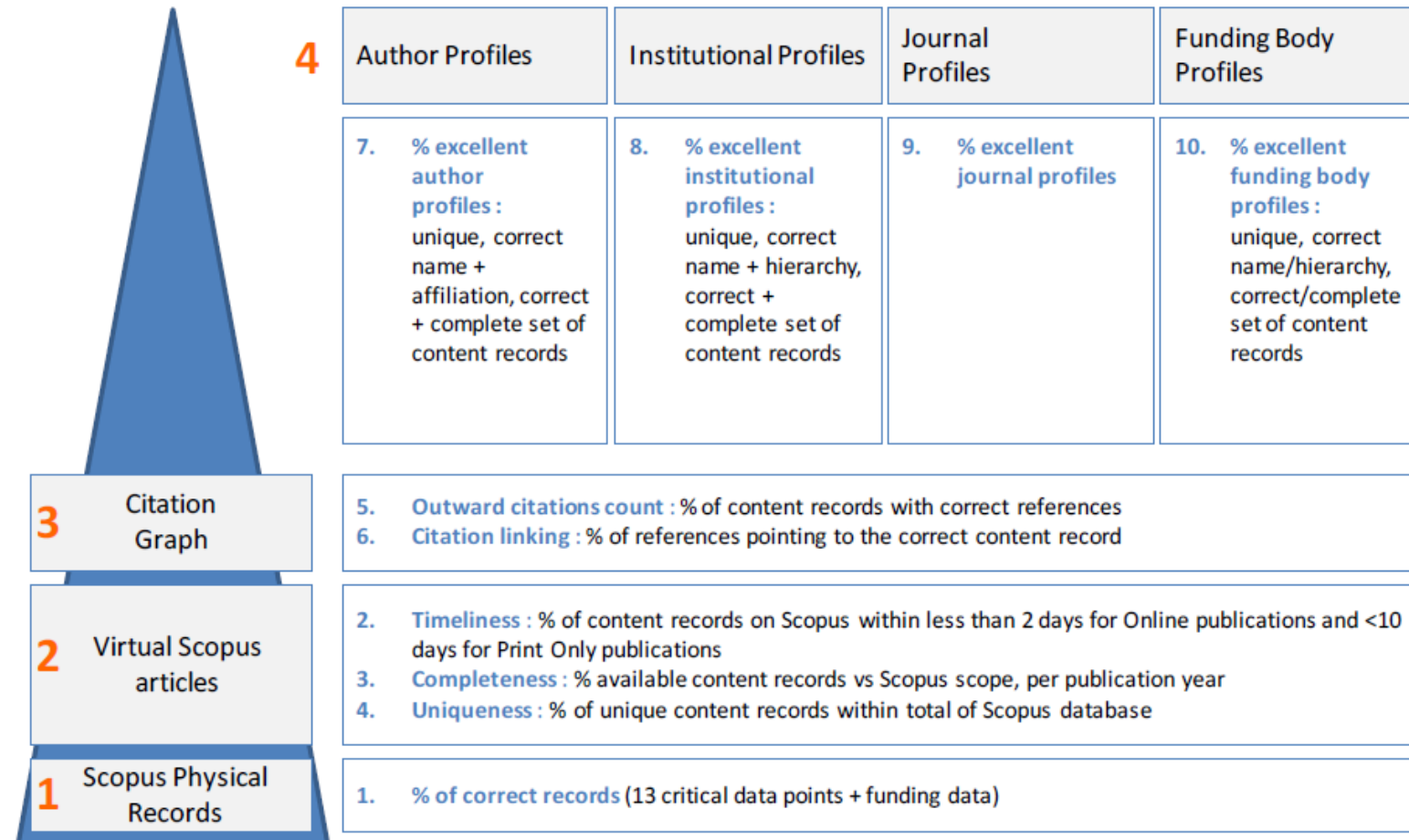
# Outline

- Developments in scientometric data sources
- Open metadata

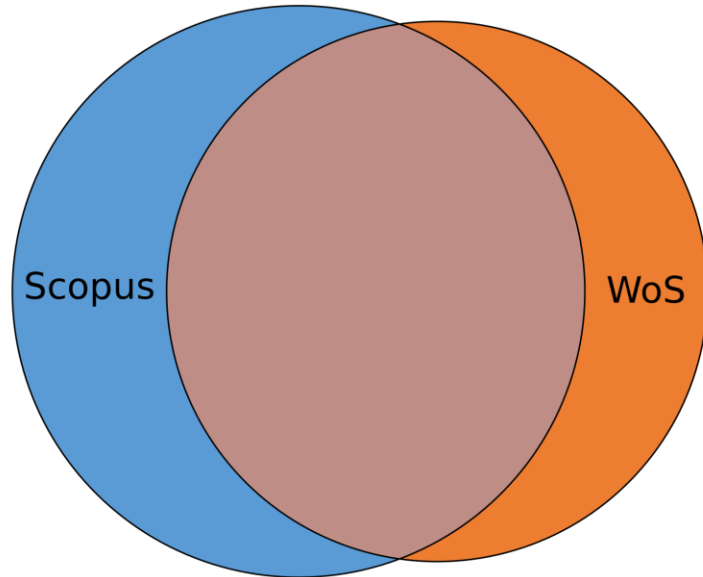
# Developments in scientometric data sources

# Scopus data quality program

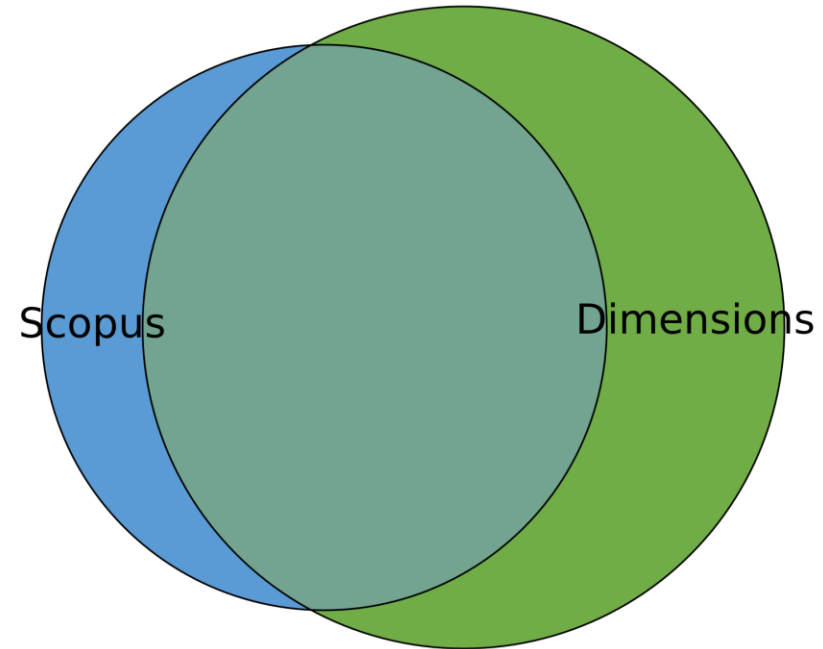
We measure Scopus Data Quality in a structured KPI framework



# Coverage of documents

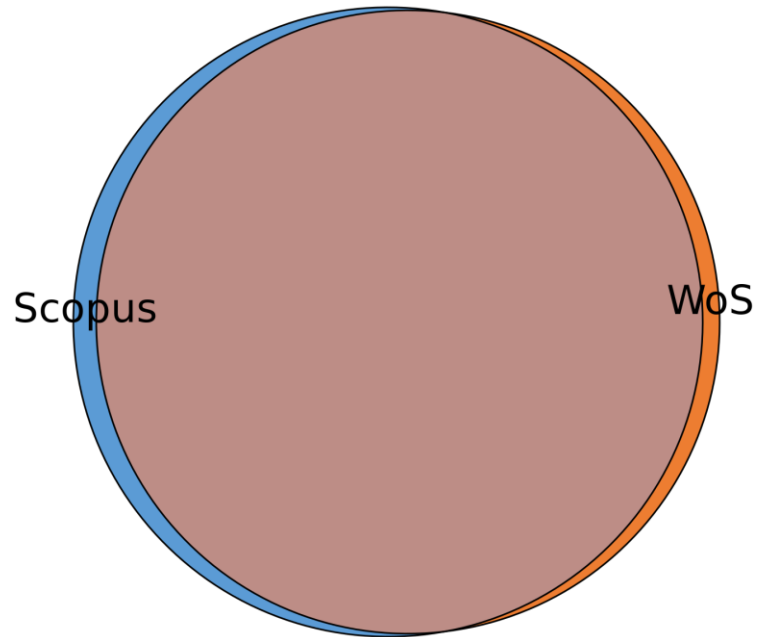


Scopus: 44.9M  
WoS: 40.1M  
Overlap: 29.1M

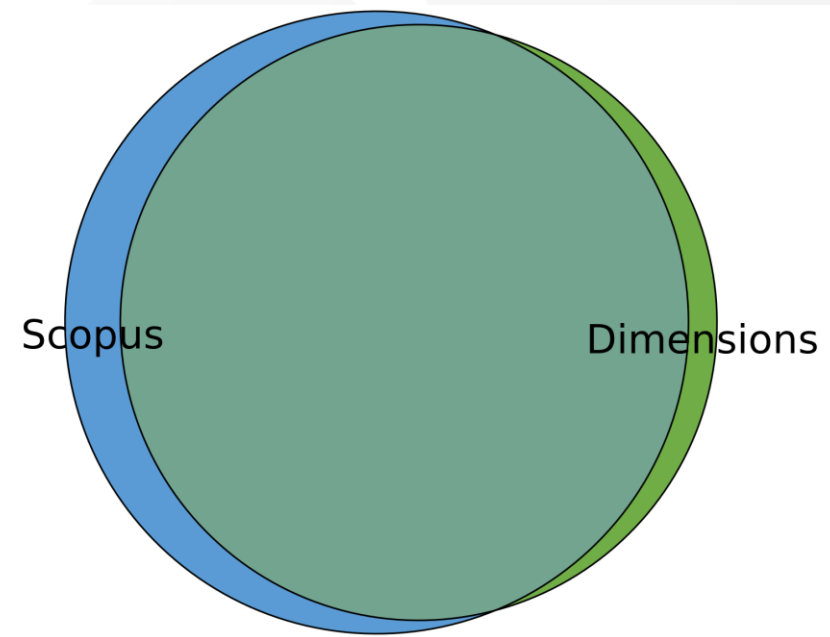


Scopus: 44.9M  
Dimensions: 57.9M  
Overlap: 35.9M

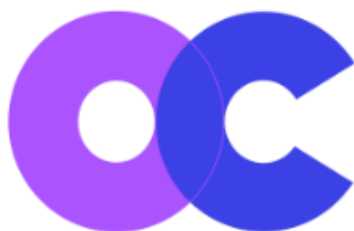
# Completeness and accuracy of citation links



Scopus: 481.2M  
WoS: 472.1M  
Overlap: 457.1M



Scopus: 562.9M  
Dimensions: 515.5M  
Overlap: 489.7M



## Welcome to the OpenCitations homepage!

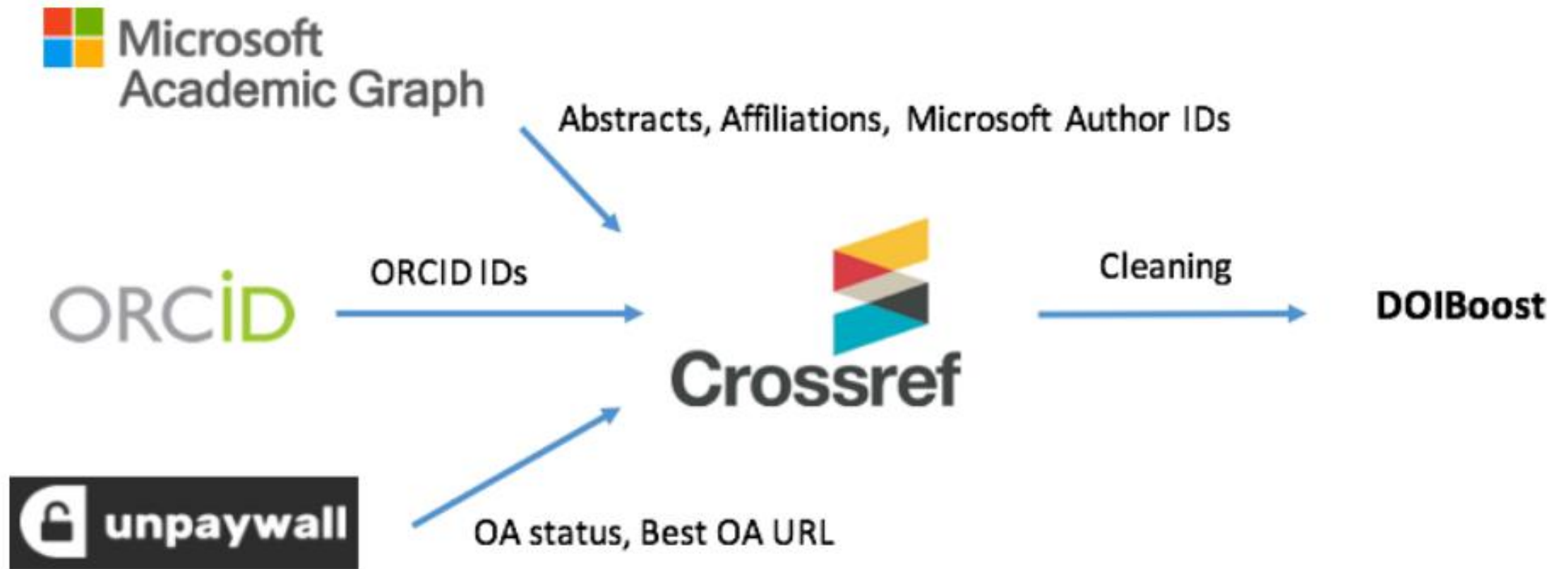
OpenCitations is a scholarly infrastructure organization dedicated to open scholarship and the publication of open bibliographic and citation data by the use of Semantic Web (Linked Data) technologies, and engaged in advocacy for semantic publishing and open citations. It provides the OpenCitations Data Model and the SPAR (Semantic Publishing and Referencing) Ontologies for encoding scholarly bibliographic and citation data in RDF, and open software of generic applicability for searching, browsing and providing APIs over RDF triplestores. It has developed the OpenCitations Corpus (OCC) of open downloadable bibliographic and citation data recorded in RDF, and a system and resolution service for Open Citation Identifiers (OCIs), and it is currently developing a number of Open Citation Indexes using the data openly available in third-party bibliographic databases.

OpenCitations is currently working to expand and improve the supporting infrastructure of the OpenCitations Corpus (OCC), our open repository of scholarly citation data made available under a Creative Commons public domain dedication, which provides in RDF accurate citation information (bibliographic references) harvested from the scholarly literature. These are described using the SPAR Ontologies according to the OpenCitations Data Model, and are made freely available so that others may freely build upon, enhance and reuse them for any purpose, without restriction under copyright or database law.

The OCC is being continuously populated from the scholarly literature.

As of March 27, 2019, the OCC has ingested the references from  
326743 citing bibliographic resources  
and contains information about 13964148 citation links  
to 7565367 cited resources.

# DOIBoost



Source: La Bruzzo et al. (2018), <https://doi.org/10.5281/zenodo.1446848>



# Open metadata

# Initiative for Open Citations (I4OC)

I4OC

About

Goals

Publishers

Stakeholders

Founders

FAQ

News

Press

How many citations are open today?



As of February 2019, the fraction of publications with open references has grown from 1% to 55% out of 43.2 million articles with references deposited with Crossref.

We encourage all other scholarly publishers to follow the example of these trail-blazing publishers by making their reference metadata publicly available. Please contact Crossref Support ([support@crossref.org](mailto:support@crossref.org)) for more information, or to let them know that you are ready to open up your reference metadata now. See also our list of responses to [frequently asked questions](#).

# Large publishers not supporting I4OC



ELSEVIER

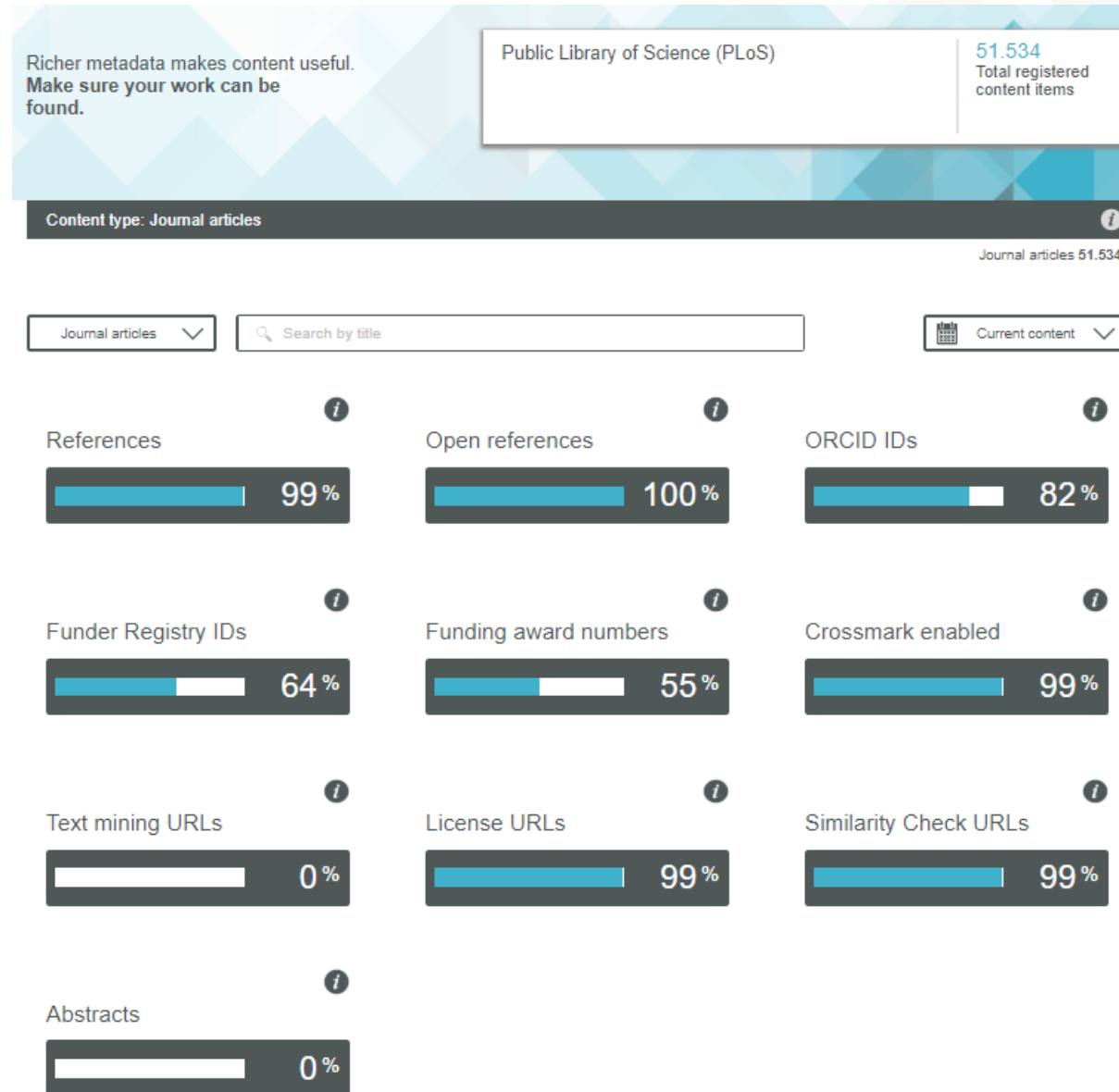


ACS  
Chemistry for Life®

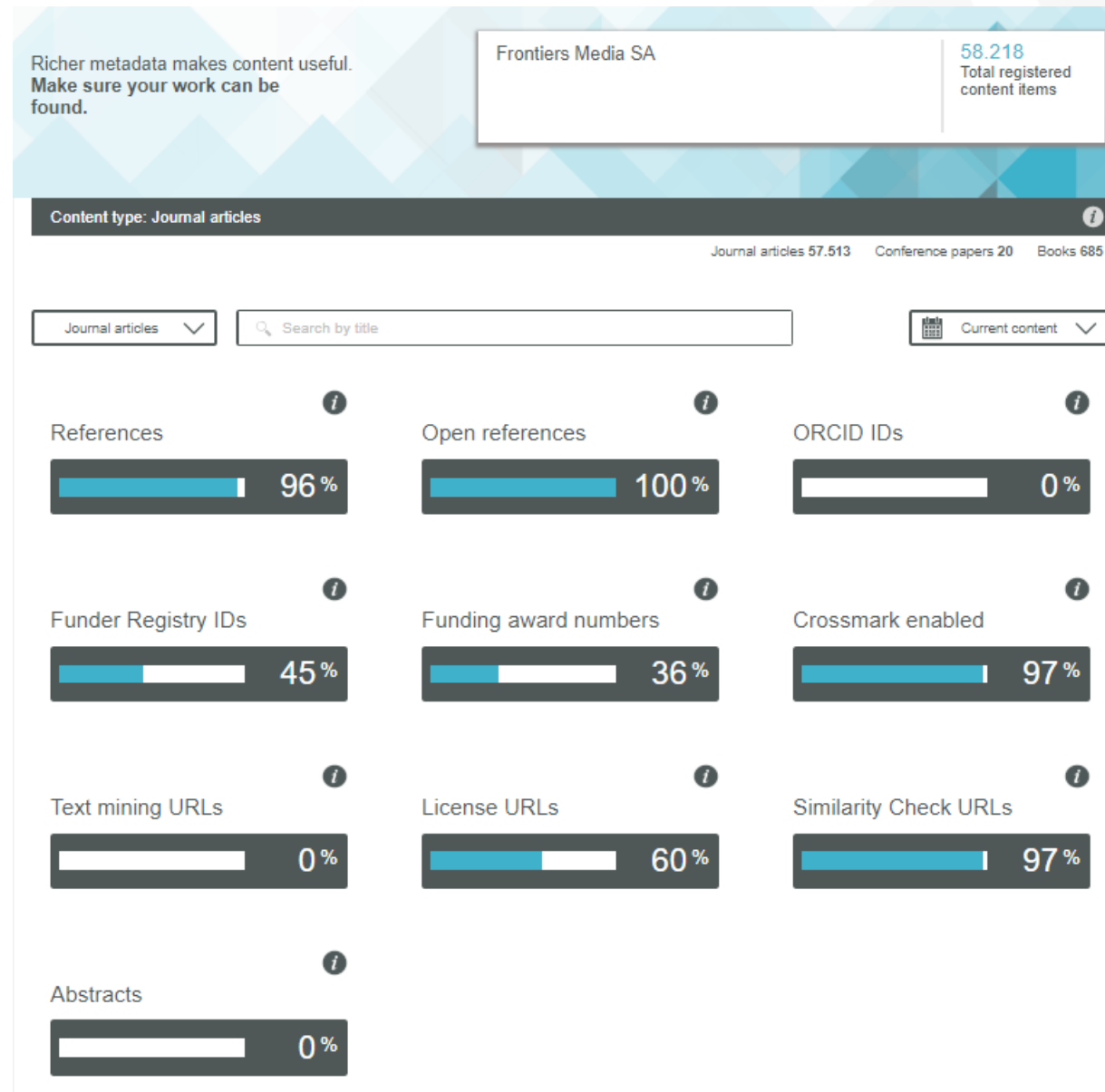


Wolters Kluwer

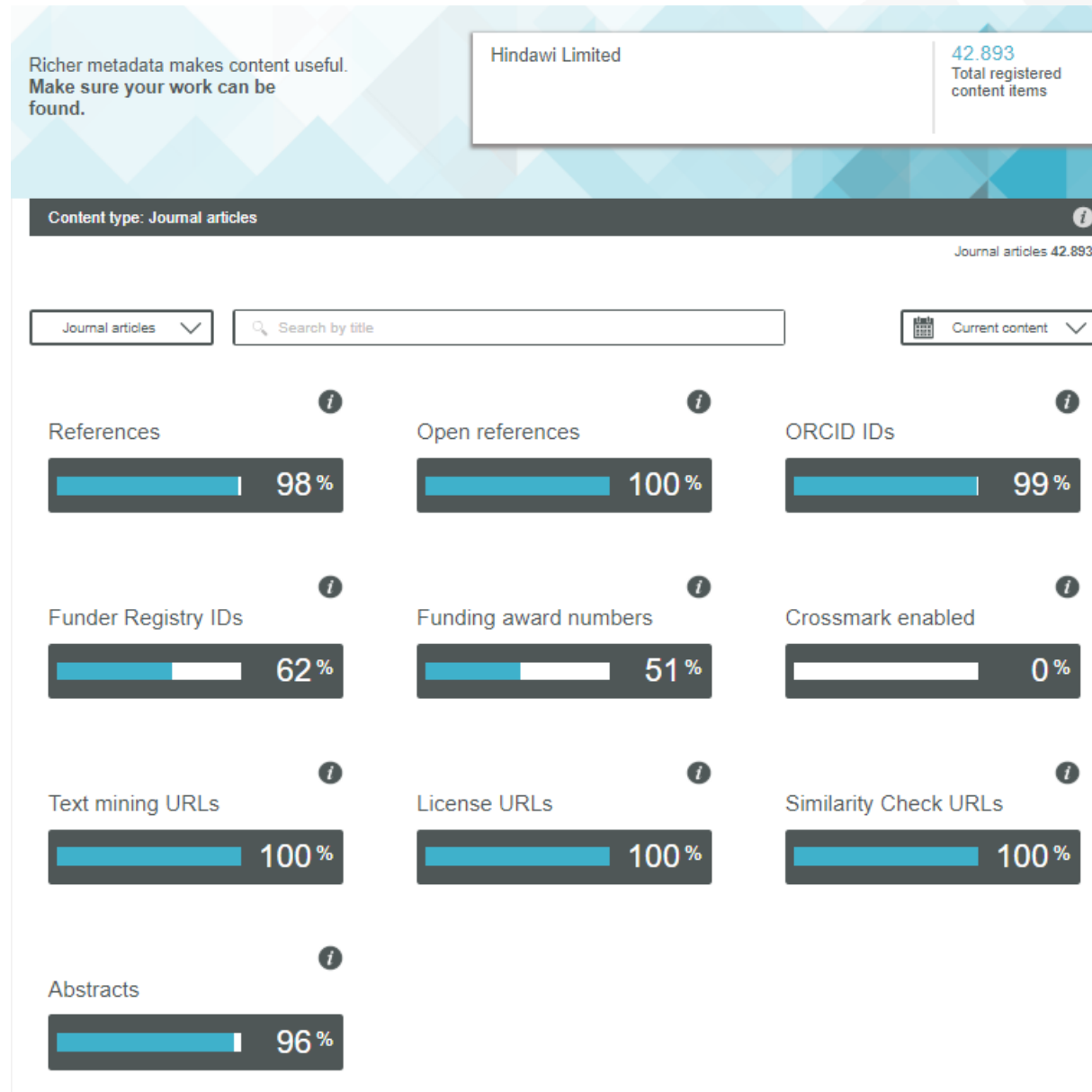
# Crossref participation report: PLOS



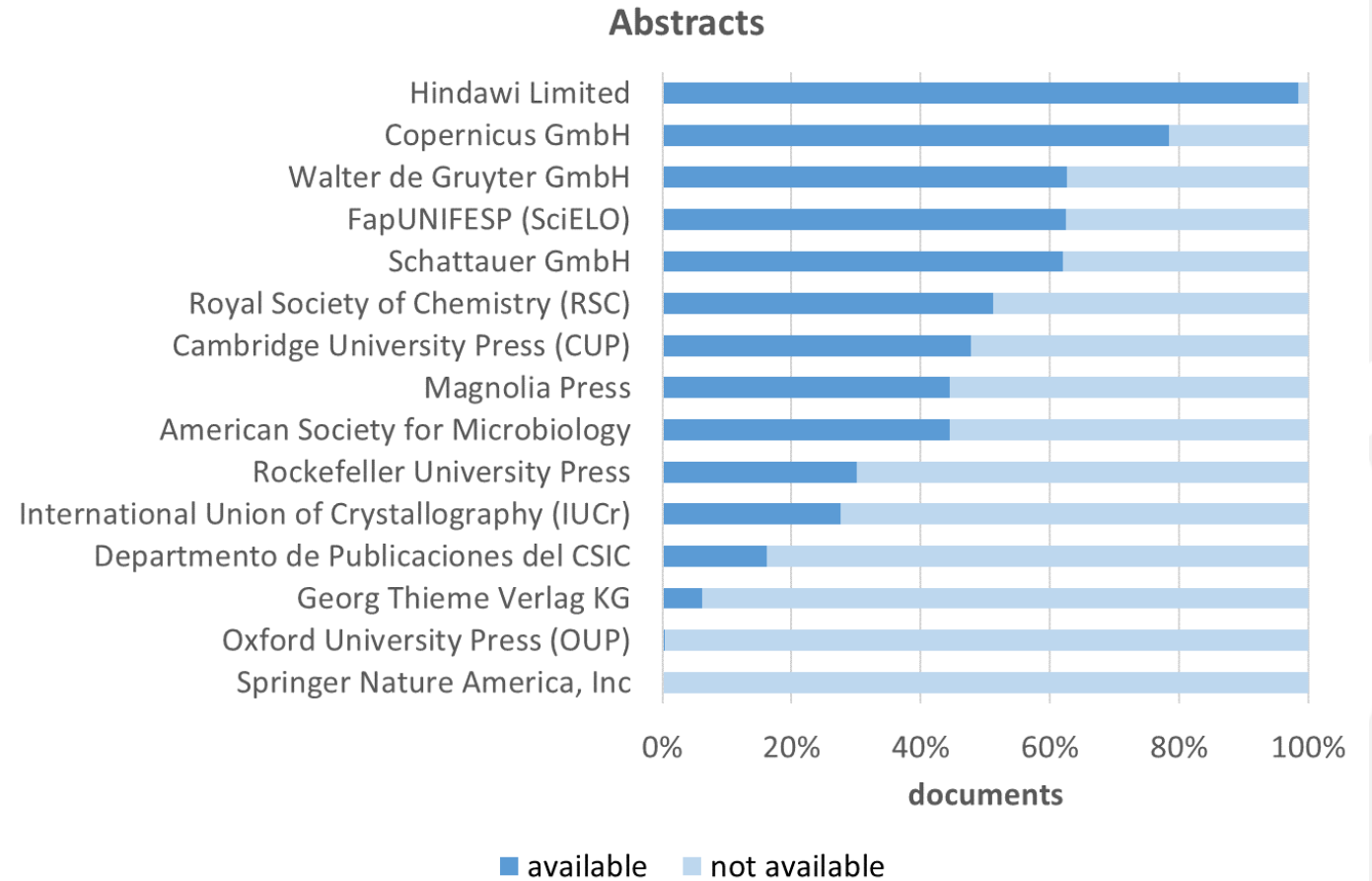
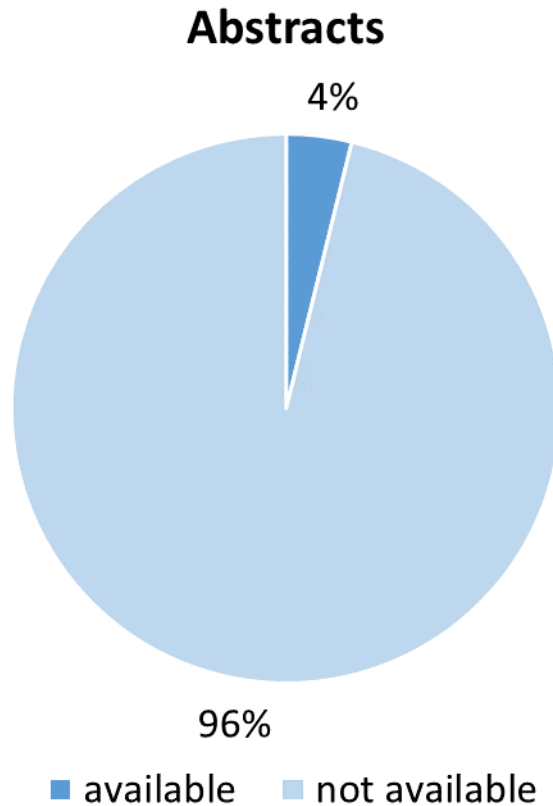
# Crossref participation reports: Frontiers



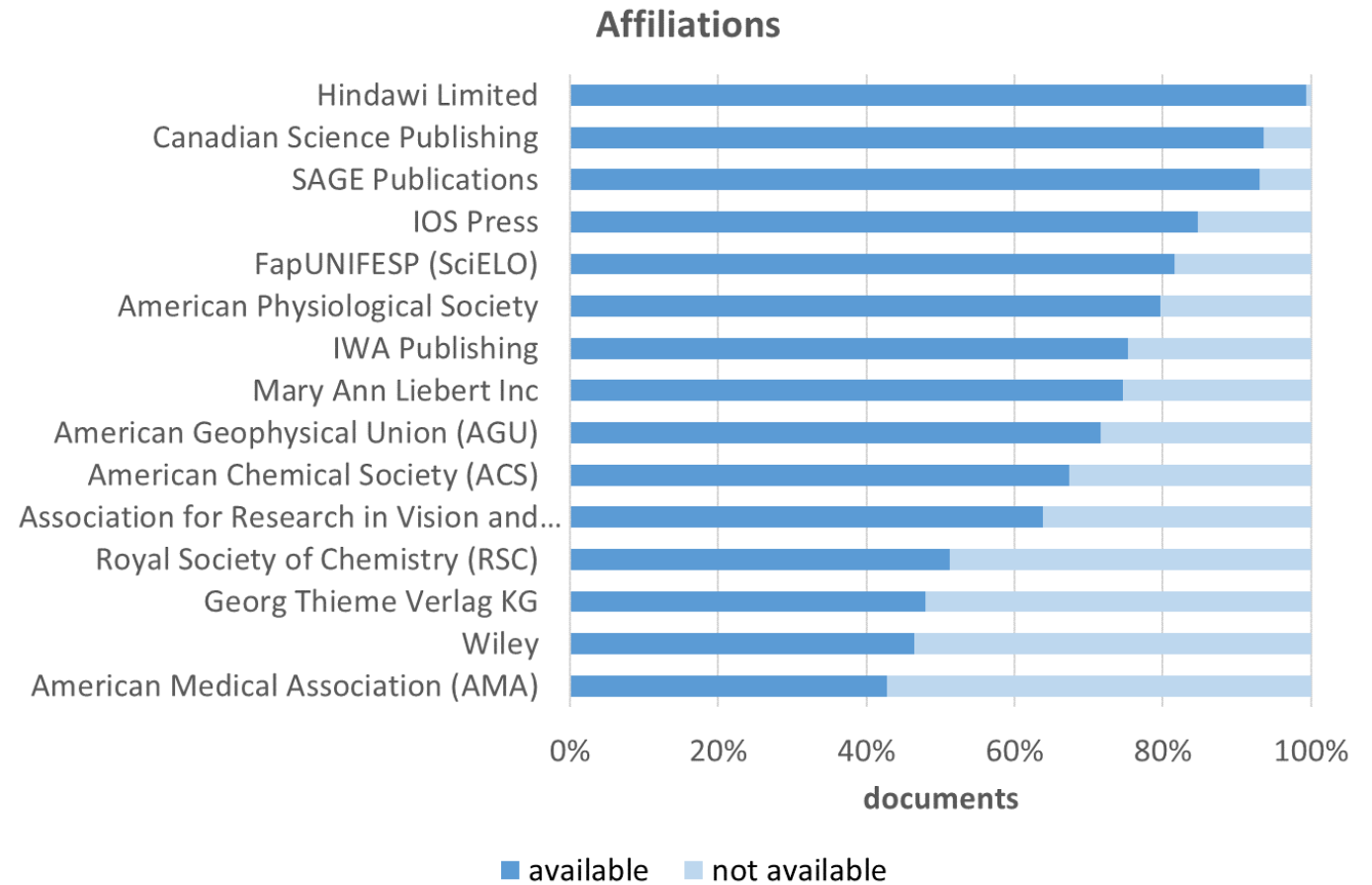
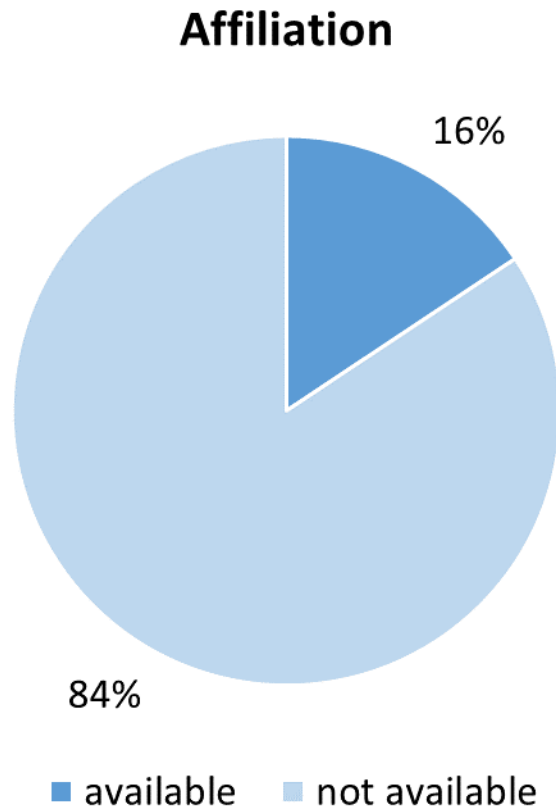
# Crossref participation report: Hindawi



# Availability of abstracts in Crossref (2008-2017)



# Availability of affiliations in Crossref (2008-2017)






# Taking concrete action!

## Editorial Mutiny at Elsevier Journal

Following in the footsteps of linguistics journal *Lingua*, the editorial board of the Elsevier-owned *Journal of Informetrics* has resigned and launched a rival journal that will be free for all to read.

By **Lindsay McKenzie** // January 14, 2019

8 COMMENTS 

The entire **editorial board** of the Elsevier-owned *Journal of Informetrics* **resigned** Thursday in protest over high open-access fees, restricted access to citation data and commercial control of scholarly work.

Today, the same team is launching a new fully open-access journal called **Quantitative Science Studies**. The journal will be for and by the academic community and will be owned by the International Society for Scientometrics and Informetrics (ISSI). It will be published jointly with MIT Press.

The editorial board of the *Journal of Informetrics* said in a statement that they were unanimous in their decision to quit. They contend that scholarly journals should be owned by the scholarly community rather than by commercial publishers, should be open access under **fair principles**, and publishers should make citation data freely available.

Elsevier said in a statement that it regretted the board's decision and that it had tried to address their concerns.

"Since hearing of their concerns, we have explained our position and made a number of concrete proposals to attempt to bridge our differences," Tom Reller, vice president of global communications at Elsevier, said in a statement. "Ultimately they decided to step down and we respect that decision and wish them the best in their future endeavors."



**Thank you for your attention!**