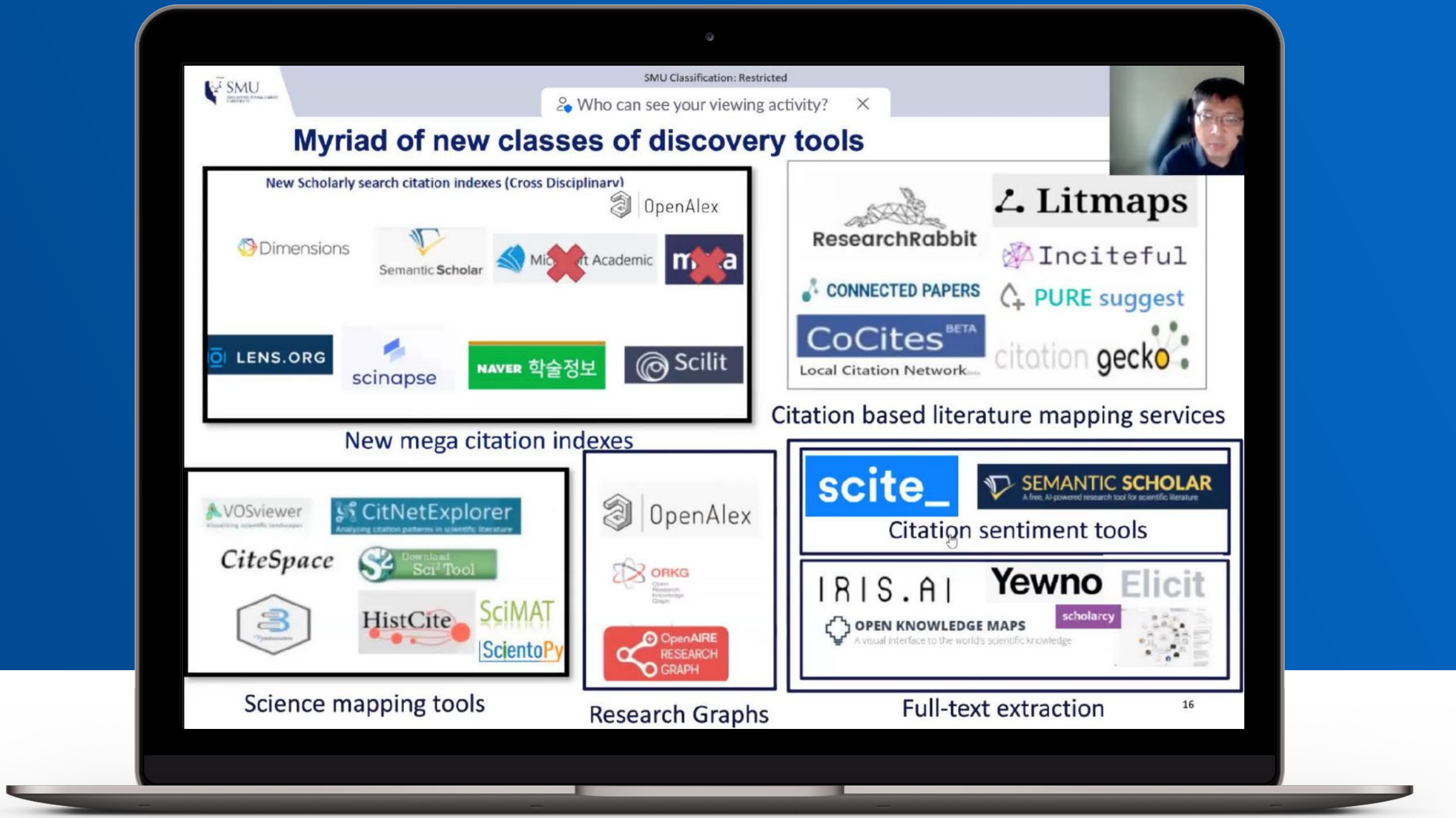
CWTS
Meaningful metrics

Open metadata of scientific publications: Ongoing developments, new opportunities and next steps

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

Hong Kong University of Science and Technology
November 9, 2022

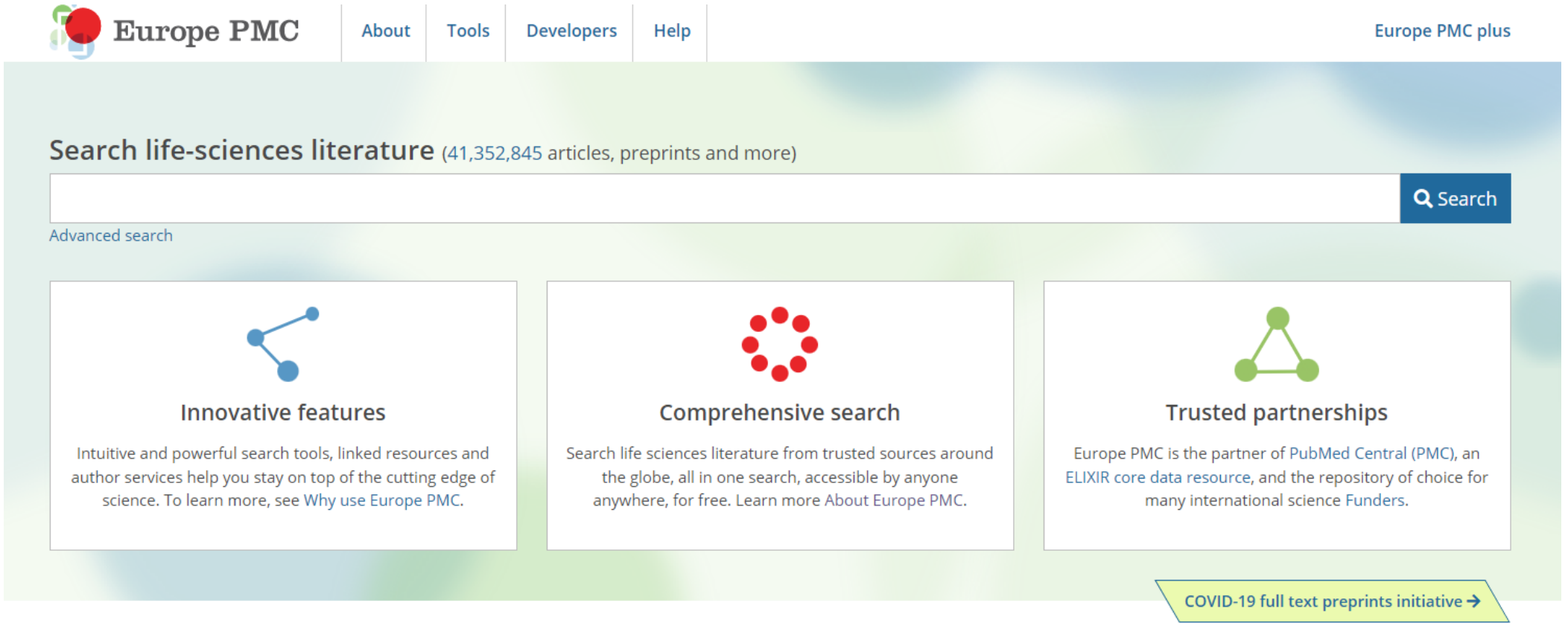
Universiteit Leiden


SMU
SMU Classification: Restricted
Who can see your viewing activity?
Myriad of new classes of discovery tools
New Scholarly search citation indexes (Cross Disciplinary)
OpenAlex
Dimensions
Semantic Scholar
Academic
LENS.ORG
scinapse
NAVER 학술정보
Scilit
New mega citation indexes
ResearchRabbit
Litmaps
Inciteful
CONNECTED PAPERS
PURE suggest
CoCites BETA
Local Citation Network
citation gecko
Citation based literature mapping services
VOSviewer
CitNetExplorer
CiteSpace
Download Sci² Tool
HistCite
SciMAT
ScientoPy
Science mapping tools
OpenAlex
ORKG
OpenAIRE RESEARCH GRAPH
Research Graphs
scite_
SEMANTIC SCHOLAR
A free, AI-powered research tool for scientific literature
Citation sentiment tools
IRIS.AI
Yewno
Elicit
scholarcy
OPEN KNOWLEDGE MAPS
A visual interface to the world's scientific knowledge
Full-text extraction
16

Which tools or databases do you typically use to search for scientific literature?

https://ahaslides.com/HKUST

Europe PMC - Can we also have something like this outside the biomedical fields?



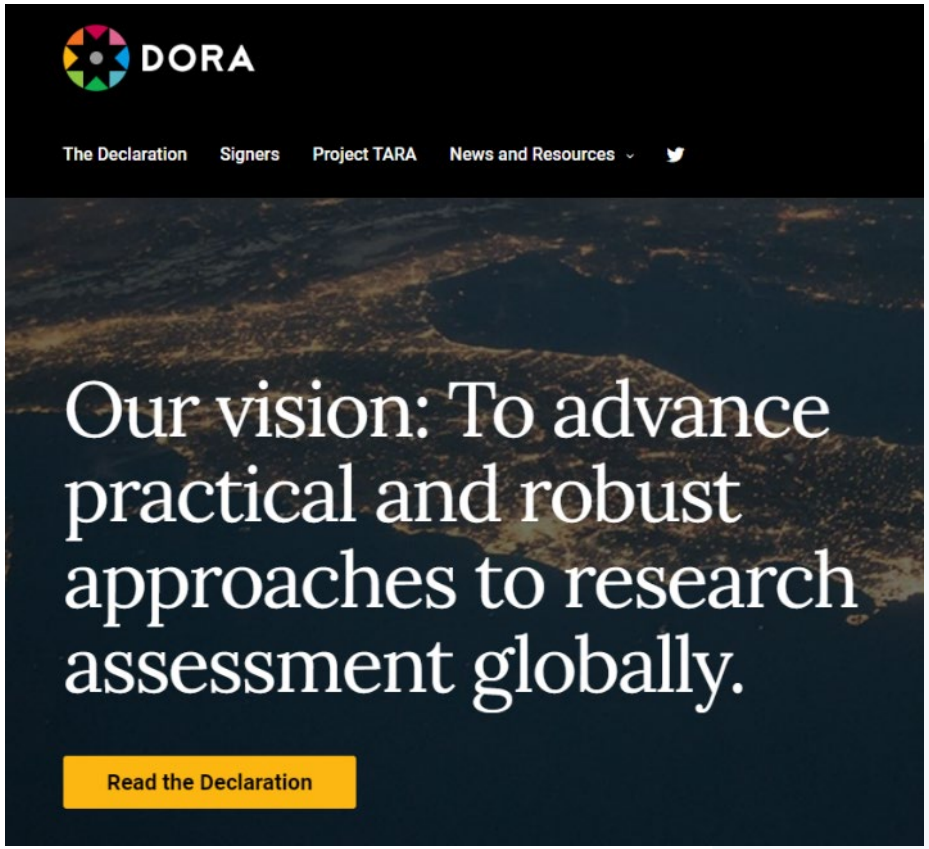
Outline

- Responsible research assessment
- Open research information: Infrastructures
- Open research information: Guiding principles
- Conclusion: How we can all contribute and benefit

CWTS
Meaningful metrics

Responsible research assessment

Responsible research assessment: Advocacy



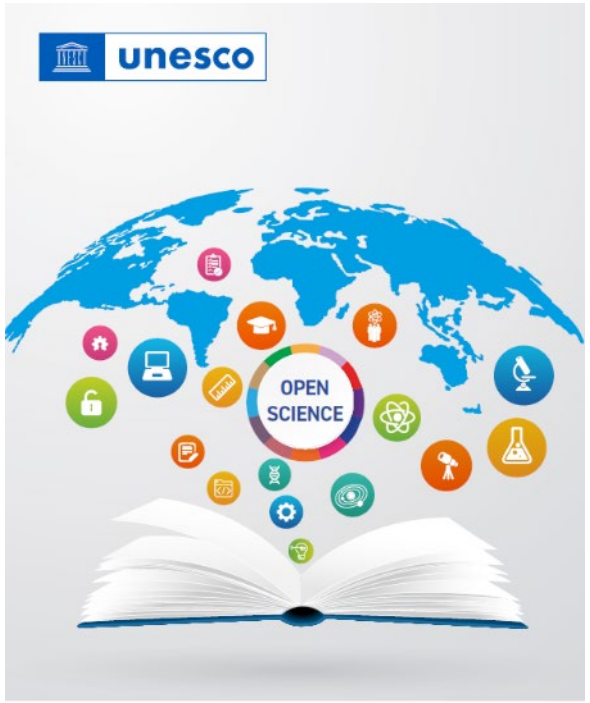
Responsible research assessment: Policy initiatives in the Netherlands and Europe

CoARA

Coalition for Advancing Research Assessment

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators.

UNESCO Recommendation on Open Science: Responsible research assessment



UNESCO Recommendation on Open Science

Reviewing research assessment and career evaluation systems in order to align them with the principles of open science. Considering that a commitment to open science requires time, resources and efforts that cannot be automatically converted into traditional academic output, such as publications, but which can have a significant impact on science and society, evaluation systems should take into account the wide breadth of missions within the knowledge creation environment. These missions come with different forms of knowledge creation and communication, not limited to publishing in peer reviewed international journals.

Encouraging responsible research and researcher evaluation and assessment practices, which incentivize quality science, recognizing the diversity of research outputs, activities and missions.

How to facilitate responsible research assessment

To facilitate responsible research assessment, we need research analytics that are

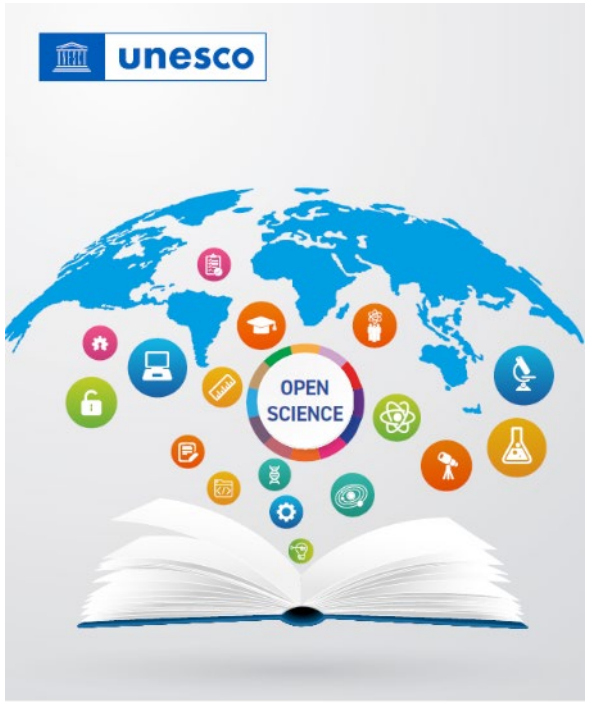
Transparent

Pluralistic

Democratic

This requires openness of research information

UNESCO Recommendation on Open Science: Infrastructures for open research information

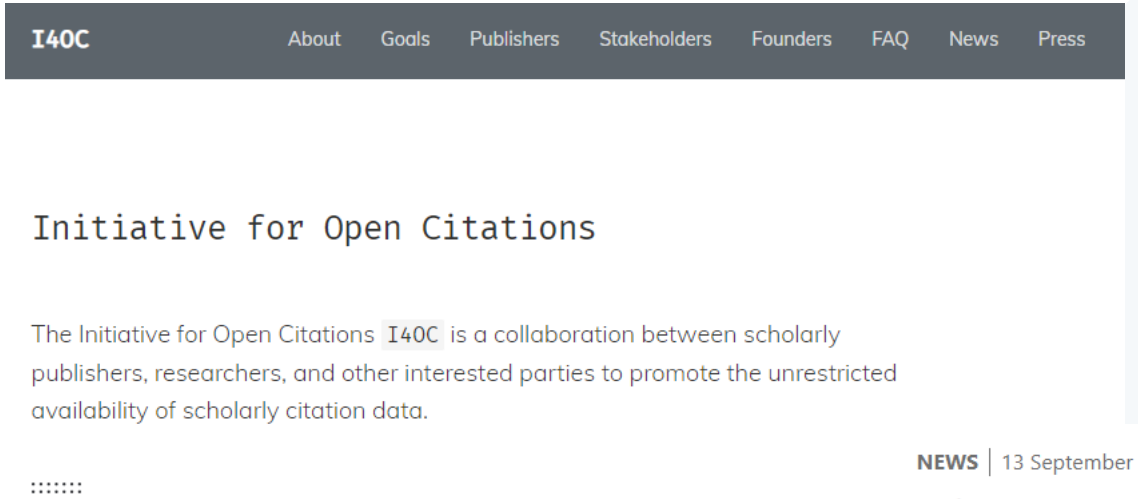


Open science infrastructures refer to shared research infrastructures (virtual or physical, including major scientific equipment or sets of instruments, knowledge-based resources such as collections, journals and open access publication platforms, repositories, archives and scientific data, current research information systems, open bibliometrics and scientometrics systems for assessing and analysing scientific domains, open computational and data manipulation service infrastructures that enable collaborative and multidisciplinary data analysis and digital infrastructures) that are needed to support open science and serve the needs of different communities. Open labs, open science platforms and repositories for publications, research data and source codes, software forges and virtual research environments, and digital research services, in particular those that allow to identify unambiguously scientific objects by persistent unique identifiers, are among the critical components of open science infrastructures, which provide essential open and standardized services to manage and provide access, portability, analysis and federation of data, scientific literature, thematic science priorities or community engagement. Different repositories are adapted to the

CWTS
Meaningful metrics

Open research information: Infrastructures

Initiative for Open Citations (I4OA)



Coverage of open citation data approaches parity with Web of Science and Scopus

Posted on October 27, 2021 by David Shotton

Guest blog post by Alberto Martín-Martín, Facultad de Comunicación y Documentación, Universidad de Granada, Spain <albertomartin@ugr.es>

In this post, as a contribution to Open Access Week, Alberto Martín-Martín shares his comparative analysis of COCI and other sources of open citation data with those from subscription services, and comments on their relative coverage.

NEWS | 13 September 2022

Five-year campaign breaks science's citation paywall

Reference lists for more than 60 million journal studies in Crossref are now free to view and reuse.

Dalmeet Singh Chawla

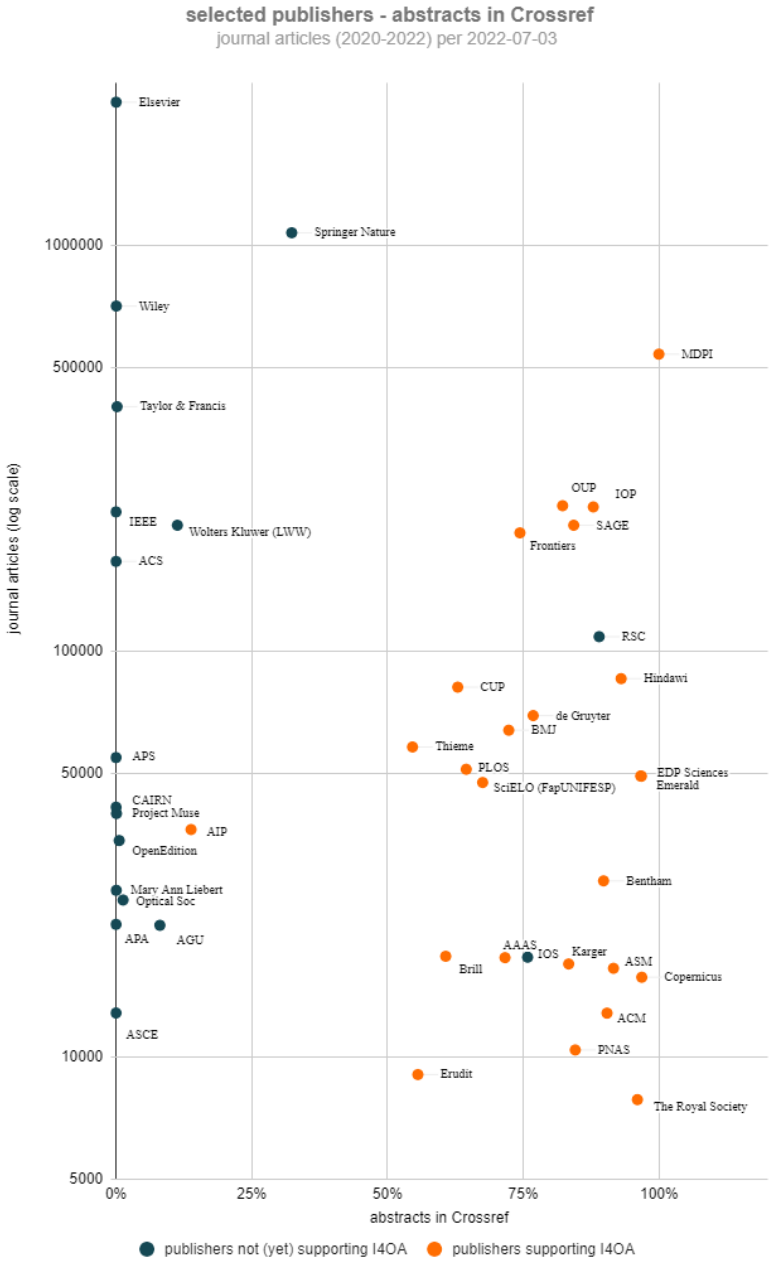
Initiative for Open Abstra

ABS TRA CTS

About | Open abstracts | Publishers | Crossref | Stakeholders | Founders | FAQ | Press

Initiative for Open Abstracts

The Initiative for Open Abstracts (I4OA) is a collaboration between scholarly publishers, infrastructure organizations, librarians, researchers and other interested parties to advocate and promote the unrestricted availability of the abstracts of the world's scholarly publications, particularly journal articles and book chapters, in trusted repositories where they are open and machine-accessible. I4OA calls on all scholarly publishers to open the abstracts of their published works, and where possible to submit them to Crossref.



Availability of open metadata in Crossref

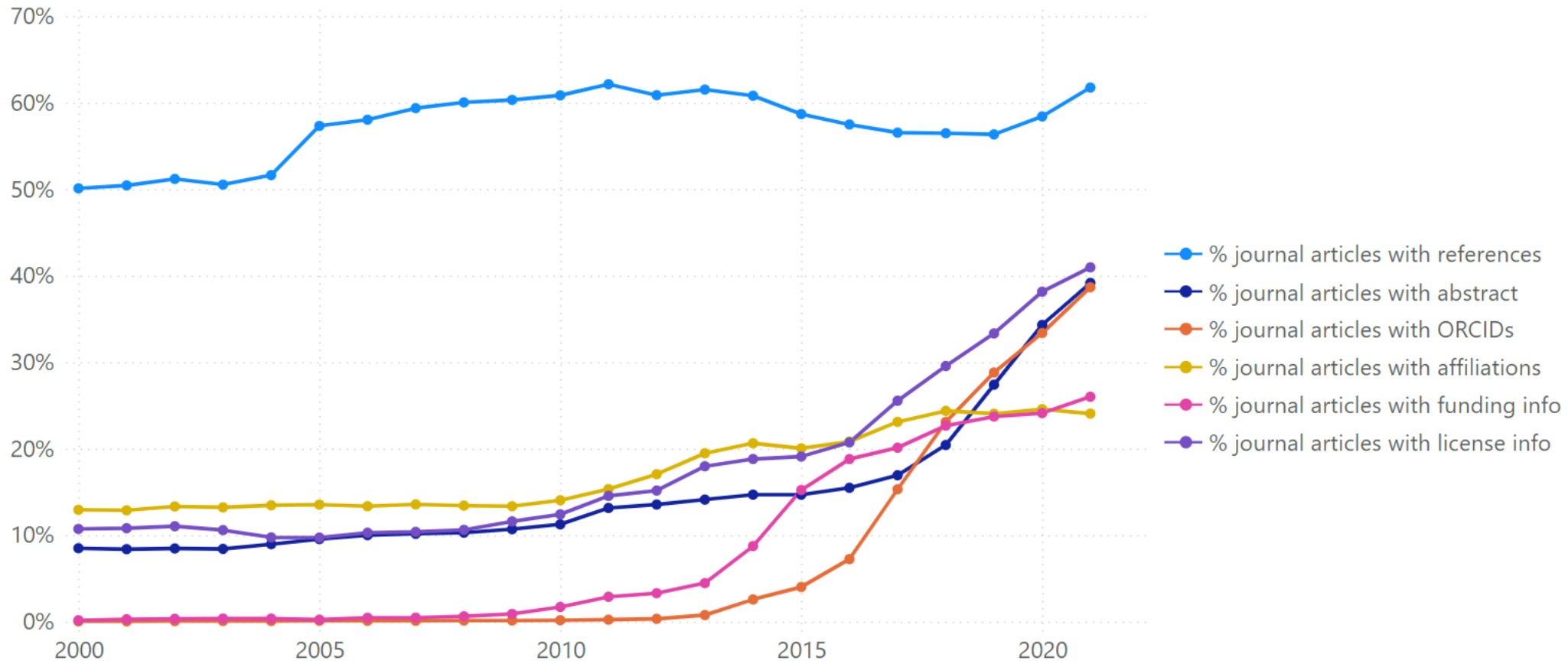
Crossref as a source of open bibliographic metadata

Nees Jan van Eck and Ludo Waltman

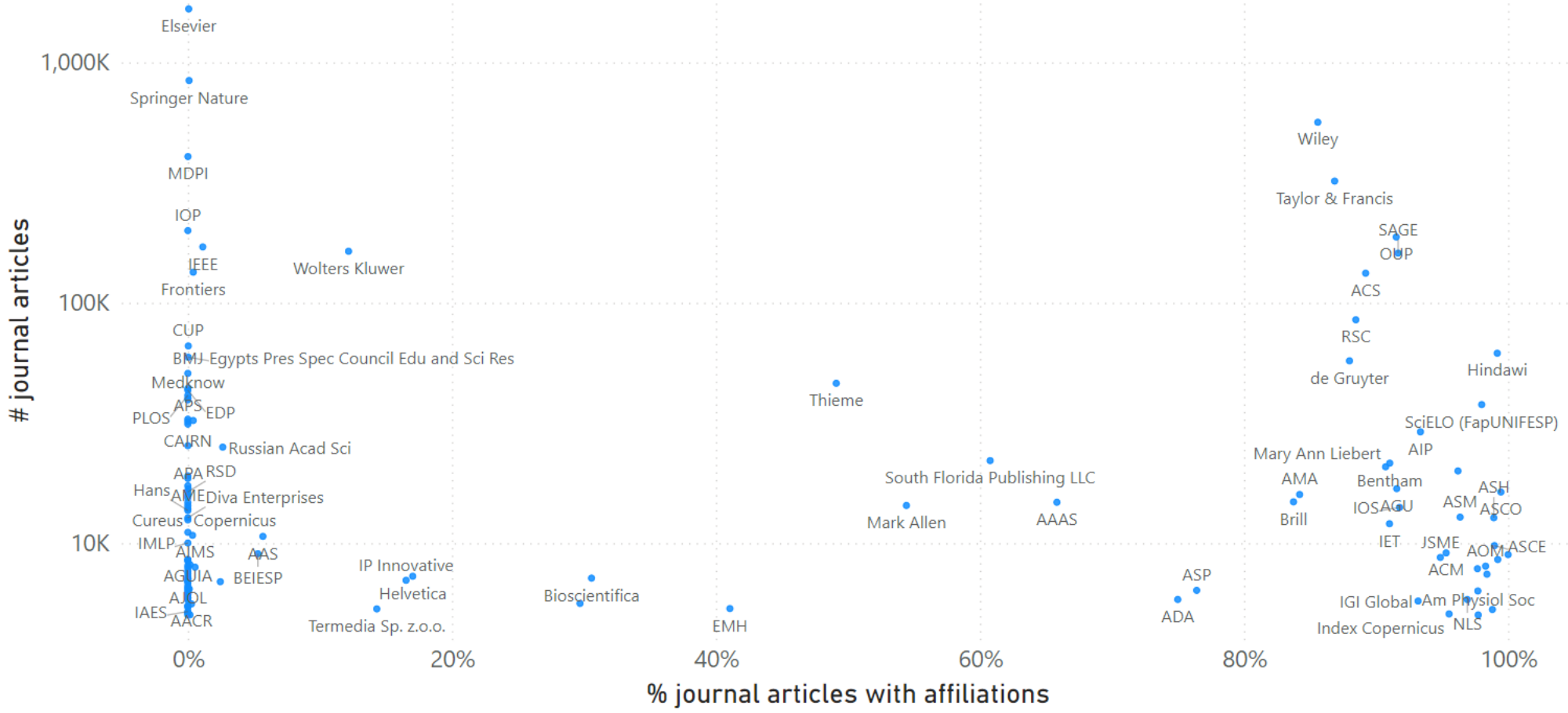
Centre for Science and Technology Studies, Leiden University, The Netherlands
{ecknjpvan, waltmanlr}@cwts.leidenuniv.nl

Several initiatives have been taken to promote the open availability of bibliographic metadata of scholarly publications in Crossref. We present an up-to-date overview of the availability of six metadata elements in Crossref: reference lists, abstracts, ORCIDs, author affiliations, funding information, and license information. Our analysis shows that the availability of these metadata elements has improved over time, at least for journal articles, the most common publication type in Crossref. However, the analysis also shows that many publishers need to make additional efforts to realize full openness of bibliographic metadata.

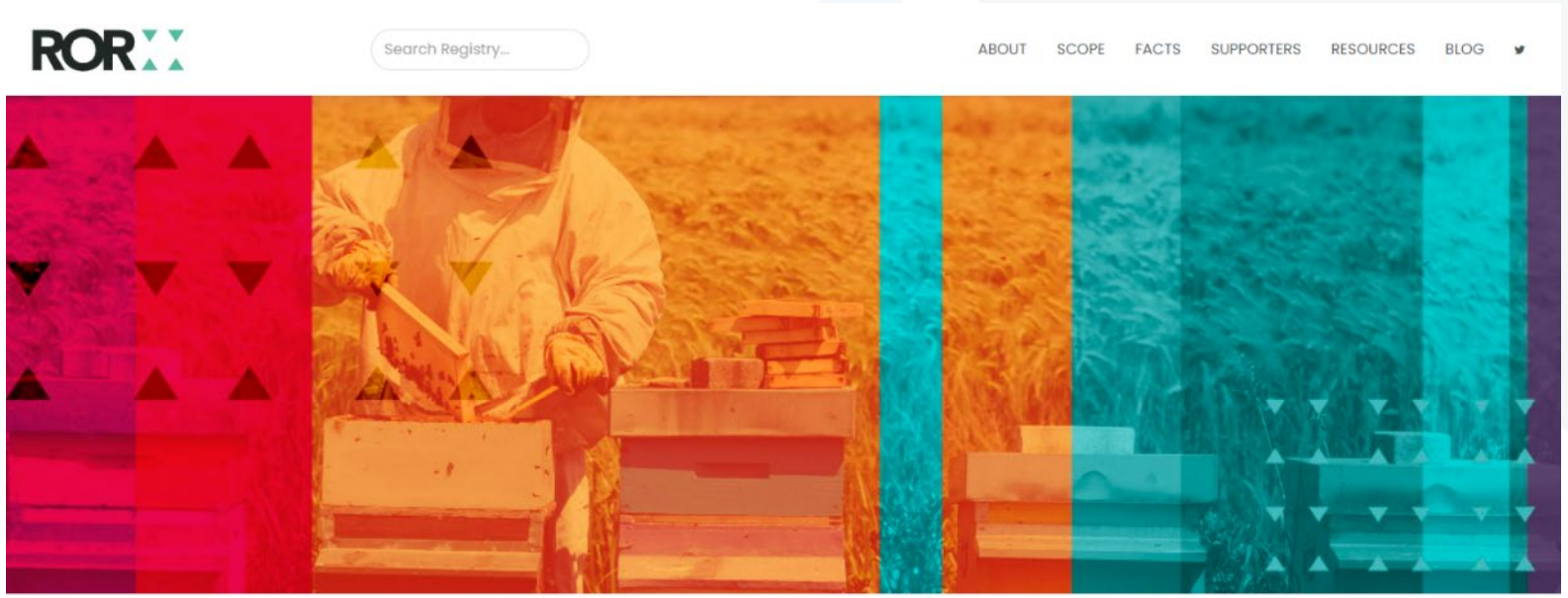
Availability of open metadata in Crossref



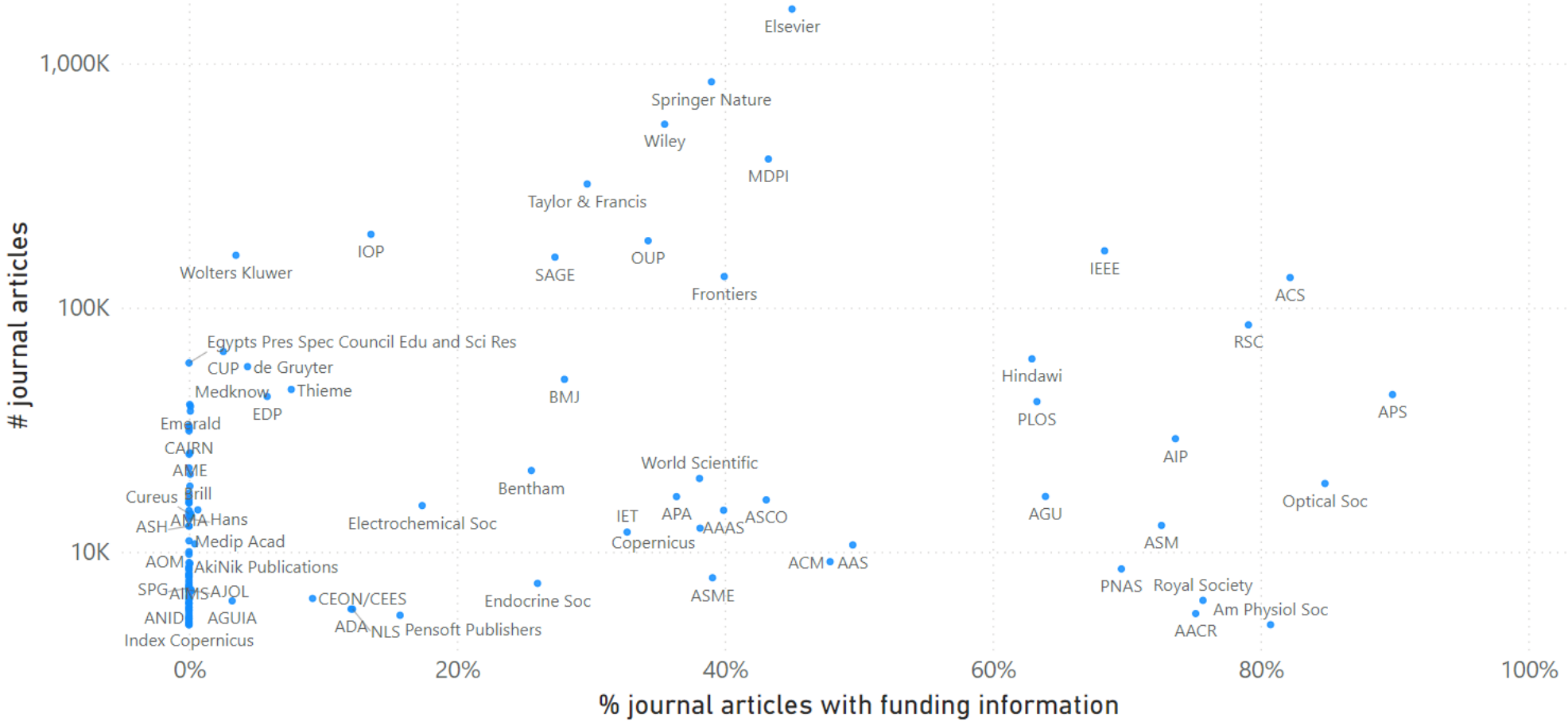
Open affiliation data in Crossref



Research Organization Registry (ROR)



Open funding data in Crossref



Open funding data

Funding Covid-19 research:
Insights from an exploratory analysis using open data infrastructures

Alexis-Michel Mugabushaka (https://orcid.org/0000-0003-4624-568X)[1]
Nees Jan van Eck (https://orcid.org/0000-0001-8448-4521)[2]
Ludo Waltman (https://orcid.org/0000-0001-8249-1752)[2]

[1] European Commission, DG RTD, Unit G2[1]
[2] Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands

Figure 7: Overlap of Crossref, Scopus, and WoS in terms of Covid-19 publications with funding data (considering only publications indexed in all three databases)

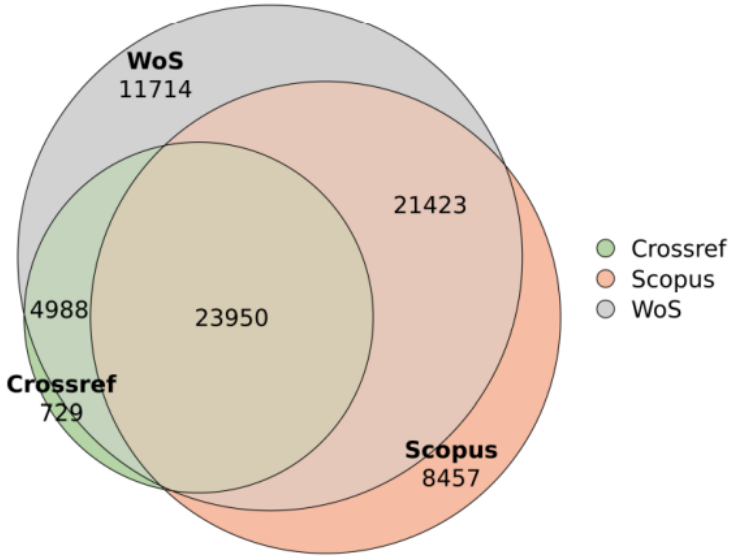
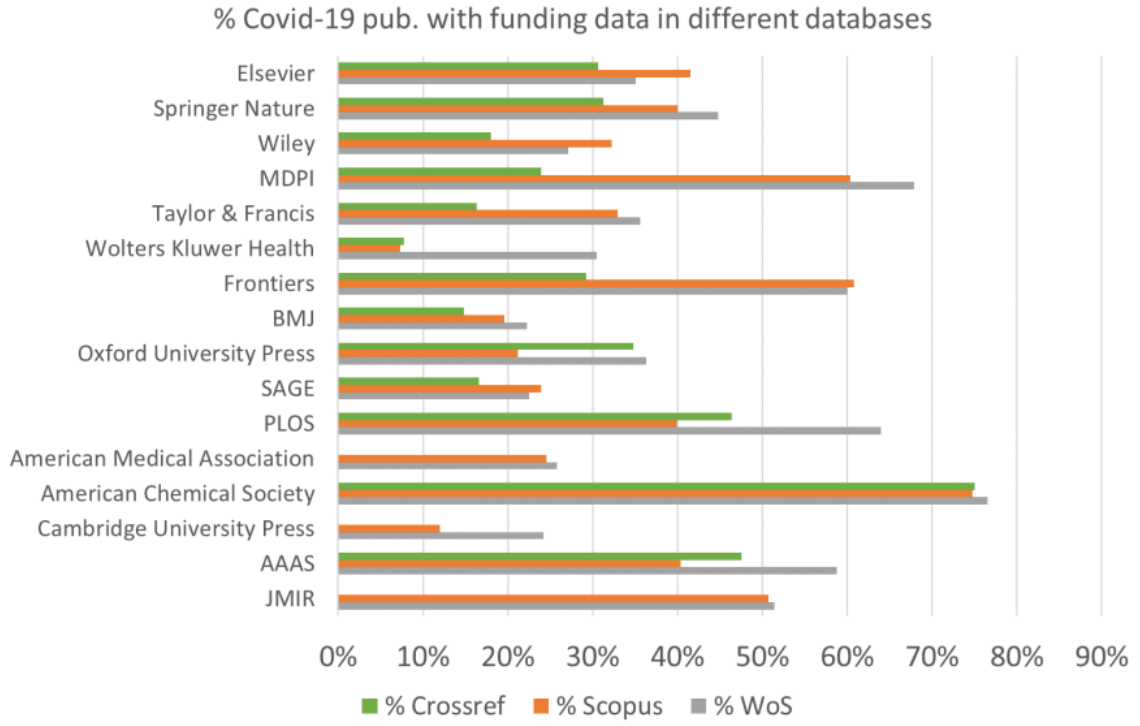


Figure 8: Percentage of Covid-19 publications with funding data, breakdown by publisher and database (considering only publications indexed in all three databases)

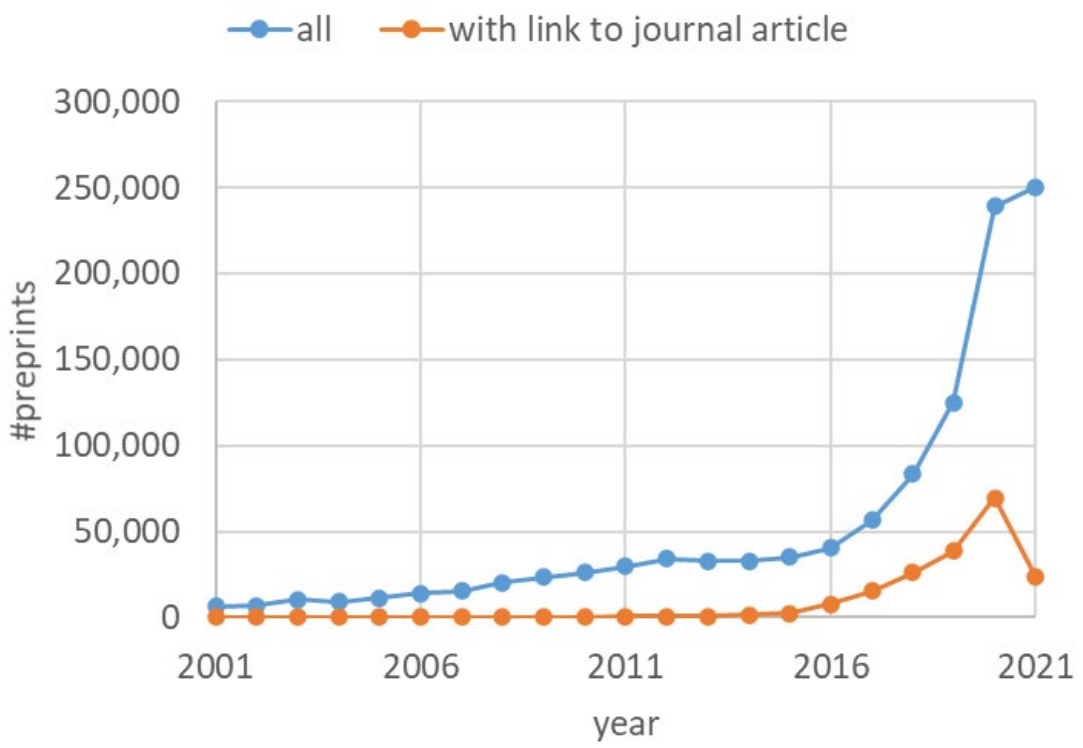


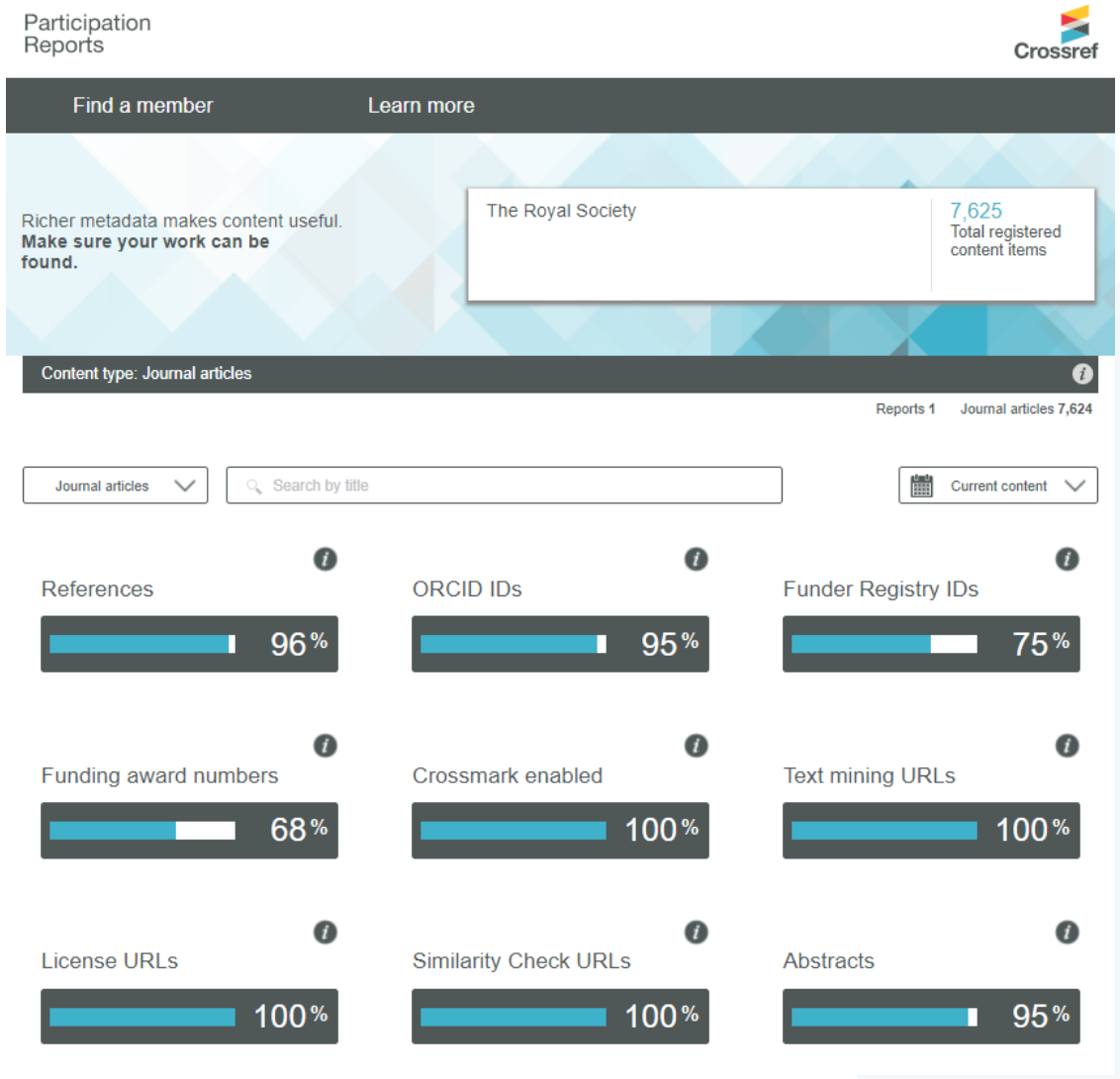
Keeping the scholarly record connected

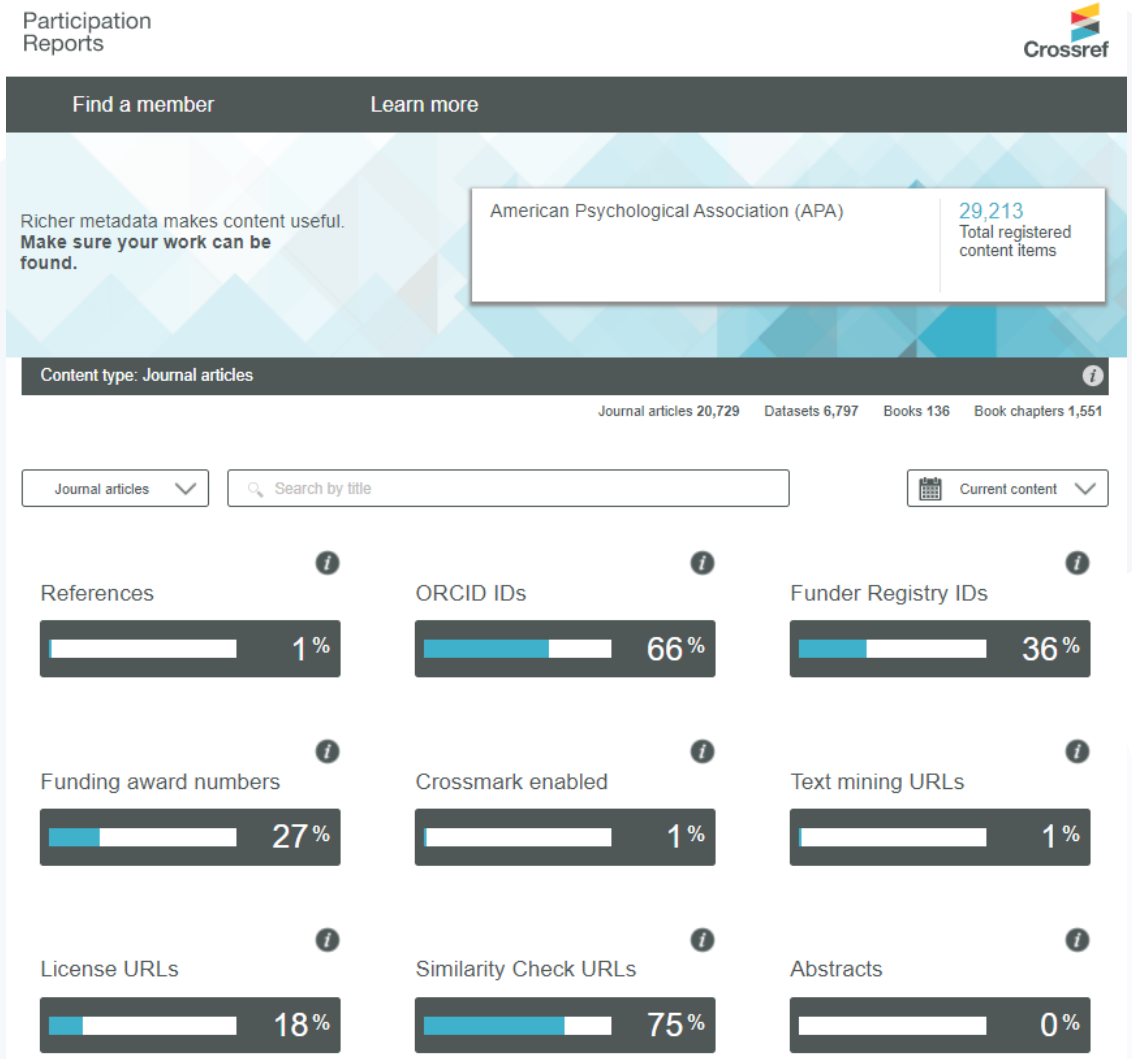
Building Stronger Chains Together: Keeping Preprints Connected to the Scholarly Record

By MICHELE AVISSAR-WHITING | JUN 7, 2022 | 3 COMMENTS

AUTHORITY | INFRASTRUCTURE | PEER REVIEW | TECHNOLOGY




Participation Reports
Crossref
Find a member
Learn more
Richer metadata makes content useful.
Make sure your work can be found.
The Royal Society
7,625
Total registered content items
Content type: Journal articles
Reports 1
Journal articles 7,624
Journal articles
Search by title
Current content
References
96%
ORCID IDs
95%
Funder Registry IDs
75%
Funding award numbers
68%
Crossmark enabled
100%
Text mining URLs
100%
License URLs
100%
Similarity Check URLs
100%
Abstracts
95%


Participation Reports
Crossref
Find a member
Learn more
Richer metadata makes content useful.
Make sure your work can be found.
American Psychological Association (APA)
29,213
Total registered content items
Content type: Journal articles
Journal articles 20,729
Datasets 6,797
Books 136
Book chapters 1,551
Journal articles
Search by title
Current content
References
1%
ORCID IDs
66%
Funder Registry IDs
36%
Funding award numbers
27%
Crossmark enabled
1%
Text mining URLs
1%
License URLs
18%
Similarity Check URLs
75%
Abstracts
0%

Ludo Waltman
Professor of Quantitative Science Studies

Publications should be FAIR

October 26, 2020 • Opinion & Commentary • 3 min read

Scholarly data sets are increasingly expected to be FAIR (findable, accessible, interoperable, and reusable). To fully realize the benefits of open access to the scholarly literature, Ludo Waltman argues that publications should be FAIR as well.

EDITORIAL | 13 September 2022

Citation data are now open, but that's far from enough

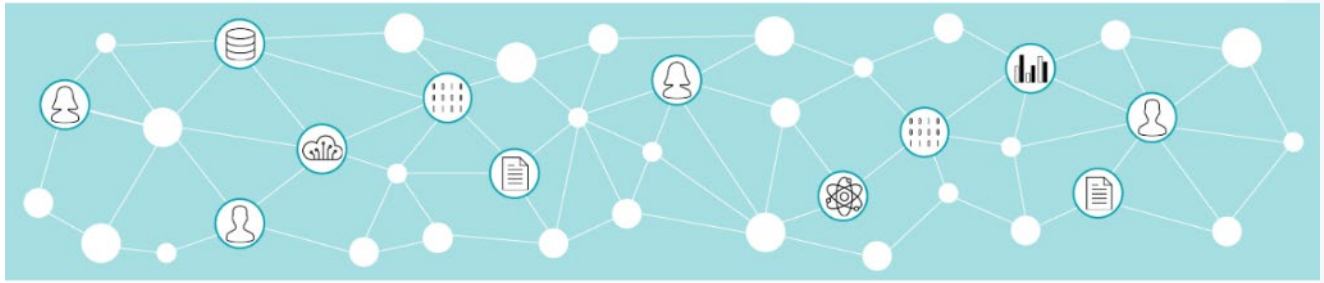
The reference lists of more than 60 million papers on the linking site Crossref are now openly available. That is is welcome – but further steps must follow.

Assessing how papers cite each other has been a painful business until now. Credit: Getty

Depositing all relevant metadata on Crossref should become the norm in scholarly publishing, as should generating DOIs for every paper. For those publishers that don't have the time or resources to do this, I4OC, I4OA and others in the open-science community have declared themselves ready to offer assistance.

Ultimately, all these moves must be only steps towards the goal of having all research papers openly available in their entirety. But until we arrive at that point, they are key to the transparency and reproducibility of research. They should be supported by all.

PID graphs



Introducing the PID Graph

Author: Martin Fenner (DataCite) & Amir Aryani (Swinburne University)

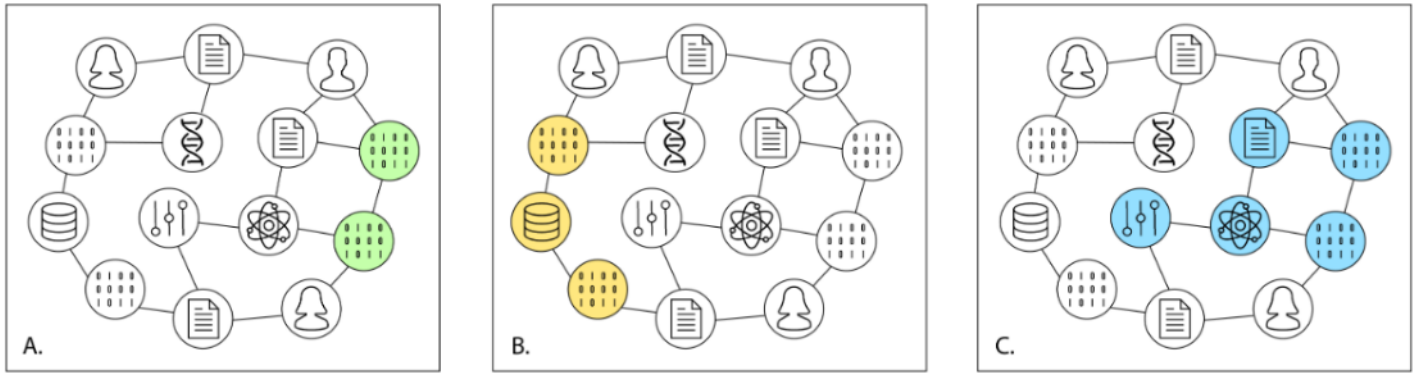
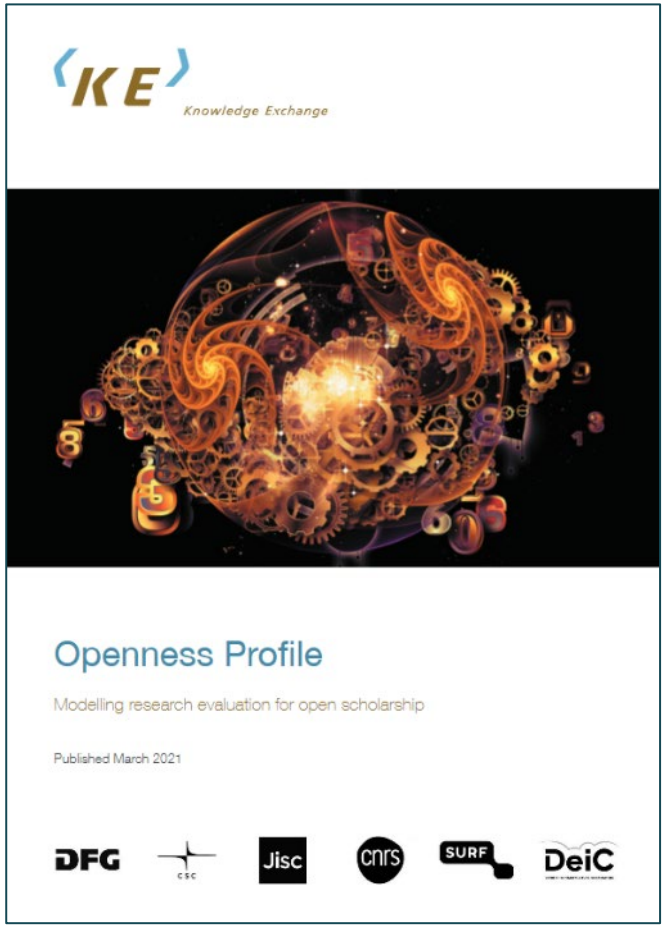
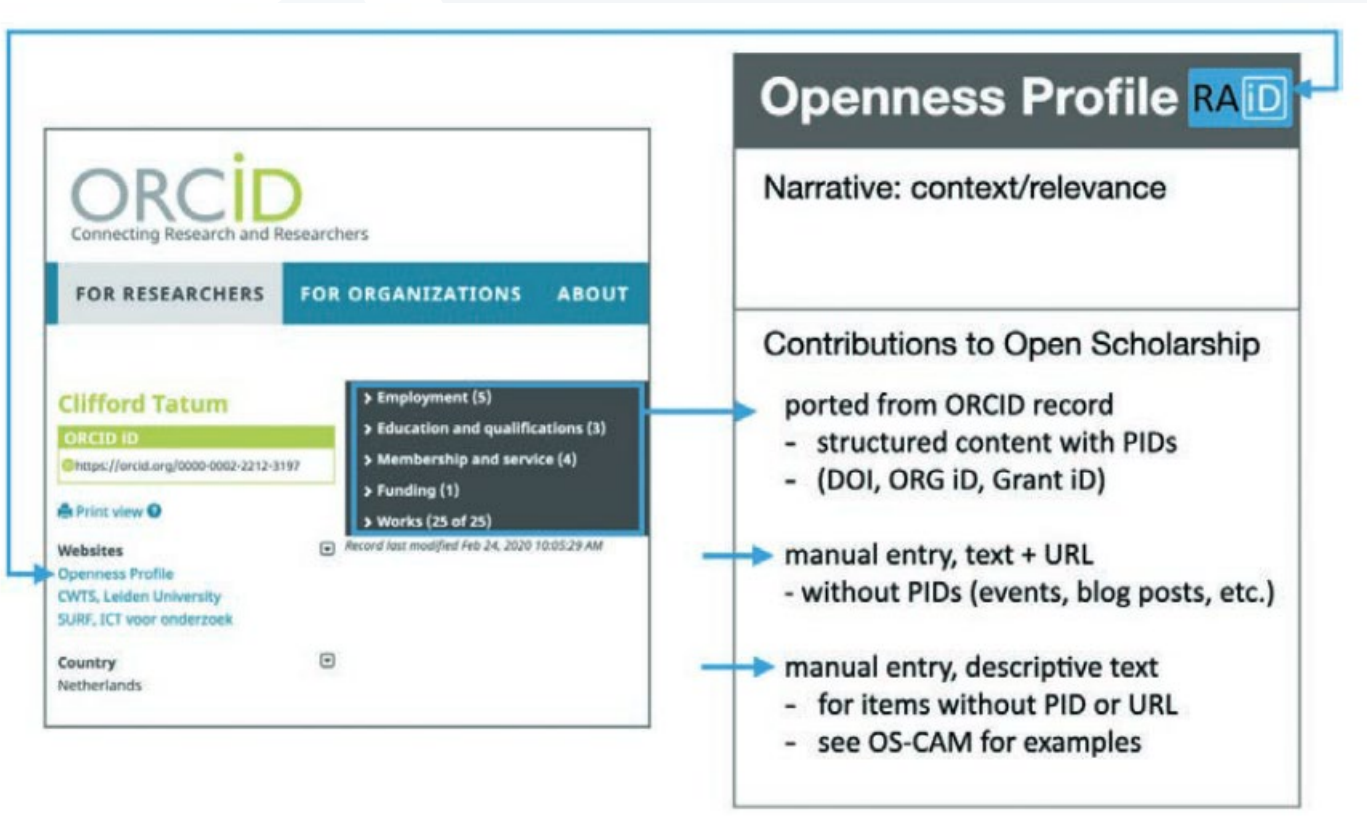
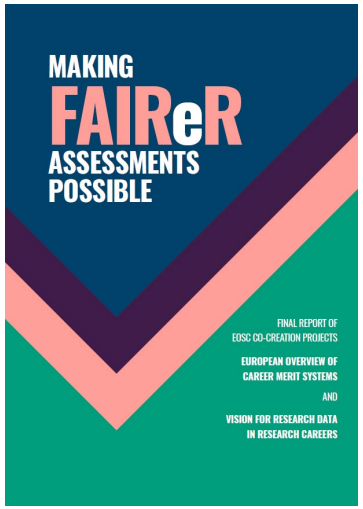


Fig 1. A schematic representation of the PID graph with digital objects connected by PIDs, showing three use cases: A: Different versions of software code, B: Datasets hosted by a particular repository, C: All digital objects connected to a research object.

Openness profiles





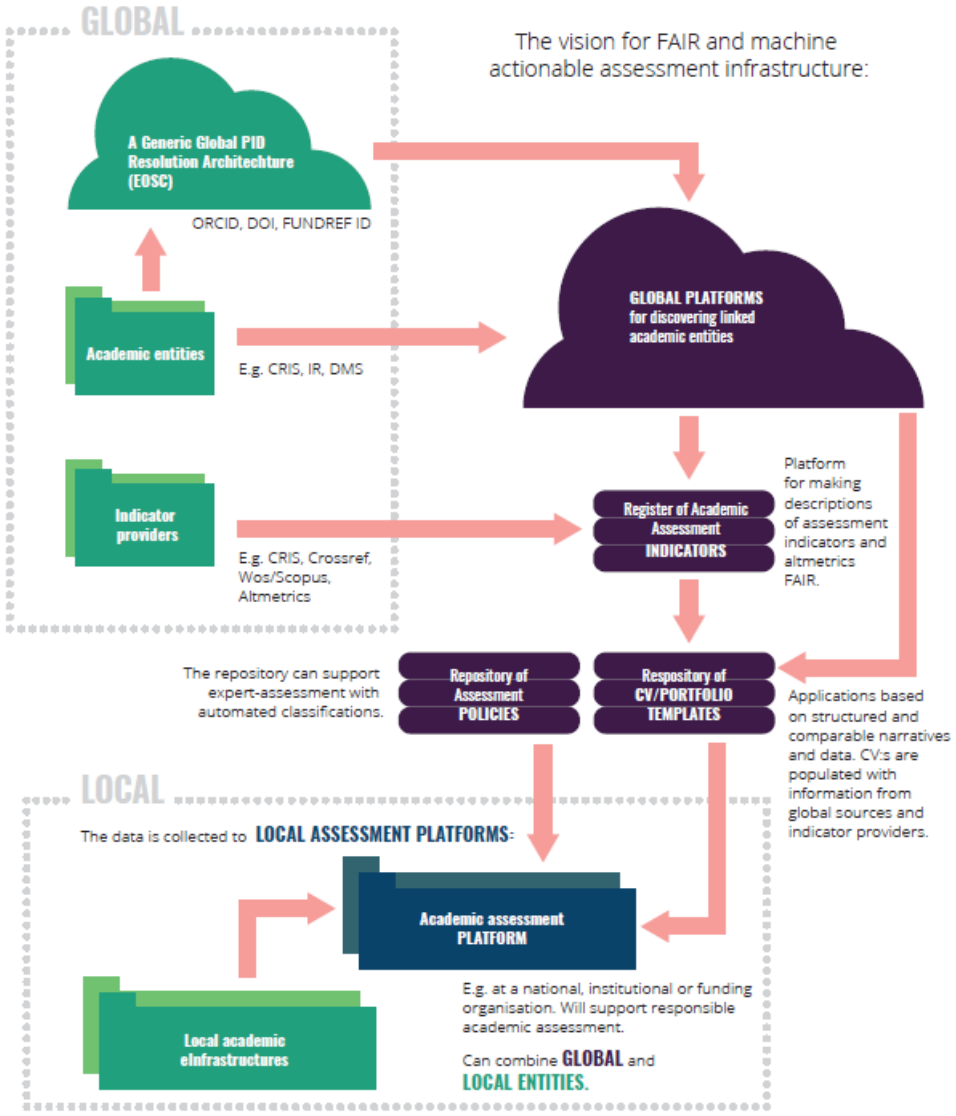


STEPS FOR REALISING THE VISION FOR FAIReR ASSESSMENTS

1	2	3
MAKE IT MEANINGFUL	MAKE IT POSSIBLE	MAKE IT REWARDING
FAIReR ACADEMIC ASSESSMENTS		
Recognise and value diversity and disciplinary differences of academic work ◆ Outputs ◆ Missions ◆ Impacts	**Diversity needs to be represented in information supporting assessment** ◆ Data models and structures ◆ FAIR and transparent data ◆ Integrated eInfrastrucure	**Diversity of outputs, activities and missions need to be included among assessment criteria** ◆ Recruitment ◆ Promotion ◆ Funding

TECHNICAL VISION OF THE FAIReR ASSESSMENT eINFRASTRUCTURE

◆ ALREADY EXISTING PLATFORMS ◆ RESEARCH INFRASTRUCTURE ECOSYSTEM ◆ LOCAL ASSESSMENT PLATTFORMS



Seven Guiding Principles for Open Research Information

Open research information: Guiding principles

Seven Guiding Principles for Open Research Information

Magchiel Bijsterbosch (SURF)
Alastair Dunning (Delft University of Technology)
Darco Jansen (Universiteiten van Nederland, UNL)
Max Haring (University of Amsterdam)
Sarah de Rijcke (Leiden University)
Maurice Vanderfeesten (Vrije Universiteit Amsterdam)

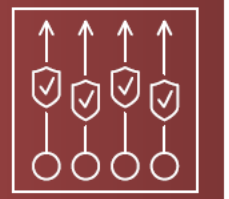
February 2022

Introduction

Picture this:

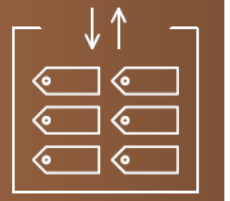
- A postdoctoral researcher in marine biology is hired on the basis of her impressive h-index and citation count;
- A university committee decides which NWO Gravity proposal to submit, based on a predictive analytics tool that utilises global trends in grant awards;
- A government panel for the Dutch *Nationaal Groeifonds* makes its selection based on metrics provided by a commercial company;
- A journal editor publishes controversial research, hoping to raise the impact factor of her journal.

But what if not all publishing venues for marine biology are equally well covered by the underlying data sources? And what if her high scores resulted from choosing a large commercial publisher over an academic society to publish the work? And how about potential biases included in the algorithms that shaped the decision of the university committee? And did the metrics of the commercial company



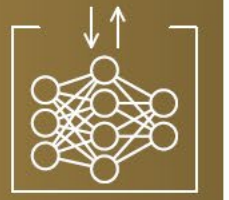
GP1.
Trusted and transparent provenance

"Within any infrastructure or service for research metadata, the provenance of the metadata, and the related algorithms, must be clear."



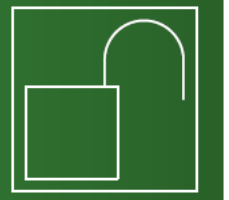
GP2.
Openness of Metadata

"Knowledge institutions must release metadata related to research output as openly as possible, ideally as CC0."



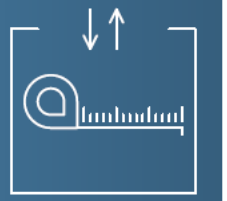
GP3.
Openness of Algorithms

"Algorithms and other techniques and methodology used to analyse and report on scholarly outputs must be available for public inspection."



GP4.
Enduring access and availability

"Knowledge institutes and third-party services must facilitate complete, non-discriminatory and enduring access to primary metadata and enriched metadata without functional, technical, legal, or financial limitations."



GP5.
Open Standards & Interoperability

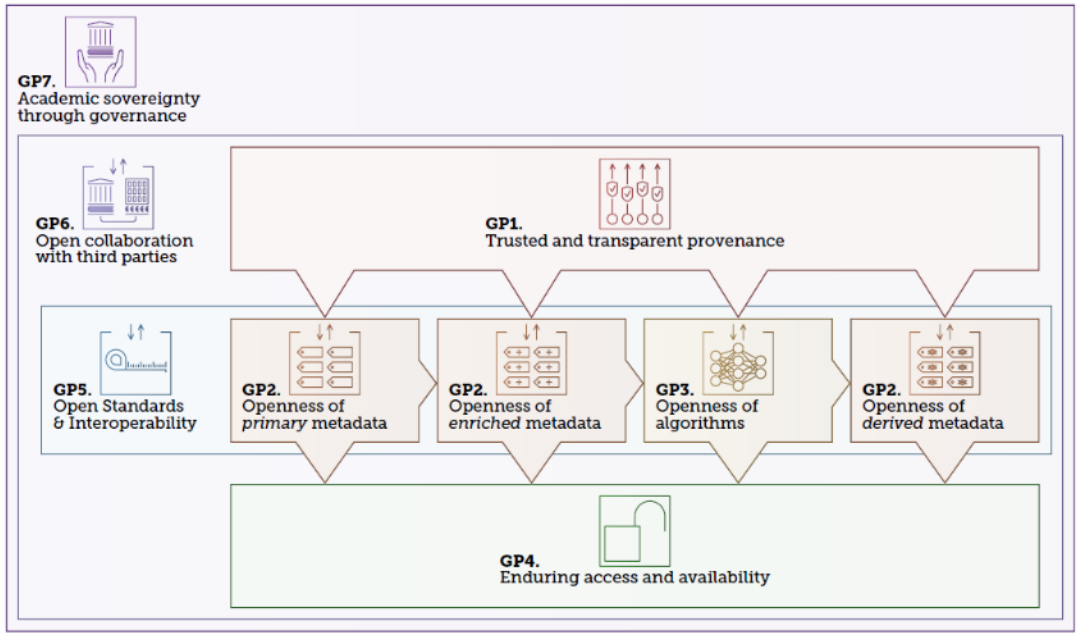
"All stakeholders must agree to work towards common definitions and open standards for exchanging and describing both metadata and algorithms."

GP6.
Open collaboration with Third parties

"Knowledge institutions and third parties must engage in open collaboration where innovation, competition, and public value are recognised and respected cornerstones."

GP7.
Academic sovereignty through governance

"A suitable governance structure must be established in order to fully implement the principles, and to ensure that stakeholders remain engaged and share accountability towards the community goals and values."



CWTS
Meaningful metrics

Conclusions and recommendations

Conclusion: How we can all contribute and benefit

- **Students and researchers:** Publish your research in venues that are open access and that support open metadata
- **Students and researchers:** Take advantage of open scholarly infrastructures for literature discovery
- **Librarians and administrators:** Support open scholarly infrastructures
- **Librarians and administrators:** Require open access and open metadata in your negotiations with publishers
- **Librarians, administrators and evaluators:** Use open research information to promote transparency, plurality and democracy in research assessments

- **Students:** Lots of very exciting career opportunities in this area!

CWTS
Meaningful metrics

Thank you for your attention!