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
Myriad of new classes of discovery tools

New Scholarly search citation indexes (Cross Disciplinary)
- OpenAlex
- Dimensions
- Semantic Scholar
- Mendeley
- Scopus
- LENS.org
- SciVerse Scopus
- Crossref
- Scimem

New mega citation indexes
- VOSviewer
- CiteSpace
- HistCite
- CiteNetExplorer
- SciMAT
- ScientoPy

Science mapping tools
- CitNet Explorer
- OpenAIRE
- ORCID
- ORCID.org

Research Graphs
- OpenAIRE
- ORCID
- ORCID.org

Citation based literature mapping services
- Litmaps
- ResearchBunny
- Connected Papers
- SciVerse Scopus

Citation sentiment tools
- SCITE
- Semantic Scholar
- IRIS.AI

Full-text extraction
- Yewno
- Elicit
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
Responsible research assessment: Advocacy

Our vision: To advance practical and robust approaches to research assessment globally.

The Leiden Manifesto for research metrics

Use these ten principles to guide research evaluation, urge Diana Hicks, Paul Wouters and colleagues.
Responsible research assessment: Policy initiatives in the Netherlands and Europe

Room for everyone’s talent
towards a new balance in the recognition and rewards of academics

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

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
Open research information: Infrastructures
Initiative for Open Citations (I4OA)

The Initiative for Open Citations (I4OC) is a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data.

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

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.

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 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
{ecknpvw, waltmanl}@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)

Welcome to the Research Organization Registry Community

ROR is a community-led project to develop an open, sustainable, usable, and unique identifier for every research organization in the world.
Open funding data in Crossref
Open funding data

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

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)

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/0000-0001-4448-4521)²
Ludo Waltman (https://orcid.org/0000-0001-8749-1752)²

¹ European Commission, DG RTD, Unit G2
² Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands
Keeping the scholarly record connected

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

By Michele Avissar-McIntosh | Jun 5, 2021 | 3 Comments

AUTHORITY | INFRASTRUCTURE | PEER REVIEW | TECHNOLOGY

![Graph showing the growth of preprints over years](image)
Publications should be FAIR

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.
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 welcome – but further steps must follow.

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, H4OA 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

Openness Profile
Modelling research evaluation for open scholarship
Published March 2021

Contributions to Open Scholarship
- structured content with PID
- (DOI, ORG ID, Grant ID)

- manual entry, text + URL
- without PID (events, blog posts, etc.)

- manual entry, descriptive text
- for items without PID or URL
- see OS-CAM for examples
TECHNICAL VISION OF THE FAIRer ASSESSMENT INFRASTRUCTURE

THE VISION FOR FAIR AND MACHINE-ACTIONABLE ASSESSMENT INFRASTRUCTURE:

GLOBAL

- A Generic Global PhD Recruitment Architecture (G3RA)
- ORCID, DOI, FunderID

- Platform for making descriptions of assessment indicators and algorithms FAIR

- Register of Academic Assessment Indicators
- Register of CV/Portfolio Indicators

- Platform for discovering linked academic entities
- The repository can support auto-assessment with automated classifications

LOCAL

- Local academic infrastructures

- Local academic assessment platforms
- E.g. at national, institutional or funding organization, will support responsible academic assessment

STEPS FOR REALISING THE VISION FOR FAIRer ASSESSMENTS

1. Make it meaningful
   - Recognize and value diversity and disciplinary differences of academic work
   - Outputs
   - Missions
   - Impacts

2. Make it possible
   - Diversity needs to be represented in information supporting assessment
   - Data models and structures
   - FAIR and transparent data
   - Integrated eInfrastructure

3. Make it rewarding
   - Diversity of outputs, activities and missions need to be included among assessment criteria
   - Recruitment
   - Promotion
   - Funding

FAIRer ACADEMIC ASSESSMENTS
Seven Guiding Principles for Open Research Information

Open research information: Guiding principles
Seven Guiding Principles for Open Research Information

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

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

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

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

“Knowledge institutions and third parties must engage in open collaboration where innovation, competition, and public value are recognised and respected cornerstones.”
“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.”
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!
Thank you for your attention!