

OpenCitations

Enabling transparency in research assessment

KB Webinar Series “Future Perspectives of Open Citation Data”
Competence Centre for Bibliometrics
March 16, 2022

Silvio Peroni

Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy
silvio.peroni@unibo.it – [@essepuntato](https://www.unibo.it/sitoweb/silvio.peroni/en) – <https://www.unibo.it/sitoweb/silvio.peroni/en>

Director of OpenCitations

silvio.peroni@opencitations.net – [@opencitations](https://www.opencitations.net) – <http://opencitations.net>



A starting point

“In this open-access age, it is a scandal that reference lists from journal articles — core elements of scholarly communication that permit the attribution of credit and integrate our independent research endeavours — **are not readily and freely available for use by all scholars.**”

David Shotton (2013). Open citations. *Nature*, 502 (7471): 295-297.
<https://doi.org/10.1038/502295a>

Replication crisis in research assessment

The difficult access to citation data is part of the so-called Replication Crisis

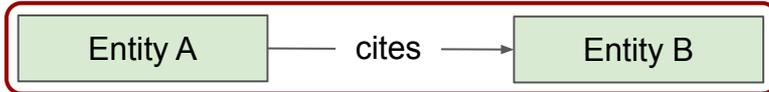
“The replication crisis [...] is an ongoing methodological crisis in which it has been found that the results of many scientific studies are difficult or impossible to reproduce.”

from https://en.wikipedia.org/wiki/Replication_crisis

The reproducibility of research is limited by **lack of access** to open data and an overall **lack of transparency** of the methods used

To improve this situation, the scholarly community is striving to put into the commons scholarly data, including bibliographic citations

What is an *open* citation



Citation: a conceptual directional link from a citing entity to a cited entity, for the purpose of acknowledging or ascribing credit for the contribution made by the author(s) of the cited entity.

The **citation data** related to a particular citation must include:

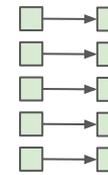
- the *representation* of such a conceptual directional link
- the *basic metadata* of the citing entity and the cited entity, i.e. sufficient information to create or retrieve textual bibliographic references

A citation is **open** when its data are compliant with the following characteristics:

structured

```
"reference": [{
  "issue": "2",
  "key": "10.7717/peerj.4375/ref-11",
  "doi-asserted-by": "Crossref",
  "first-page": "237",
  "DOI": "10.1002/asi.22963",
  "article-title": "Anatomy of green open access",
  "volume": "65",
  "author": "Björk",
  "year": "2014",
  "journal-title": "Journal of the Association for"
},
...]
```

separate



open



without restrictions to maximise reuse

identifiable



available



= metadata

S. Peroni & D. Shotton (2018).

Open Citation: Definition. <https://doi.org/10.6084/m9.figshare.6683855>

General information and services

OpenCitations (<http://opencitations.net>) is an [independent infrastructure organization](#)

- dedicated to open scholarship and the publication of open bibliographic and citation data by the use of Semantic Web technologies
- engaged in advocacy for open citations and open bibliographic metadata, particularly via **I4OC** (<https://i4oc.org>) and **I4OA** (<https://i4oa.org>)

It provides:

- a data model: the [OpenCitations Data Model](#) (based on the [SPAR Ontologies](#))
- bibliographic and citation data (CC0): [OpenCitations Corpus](#), [COCI](#), [CROCI](#), [CCC](#)
- software: in our [GitHub repository](#), released with open source licenses
- online services: [REST APIs](#), [SPARQL endpoints](#), [dumps](#) and [interfaces](#)

Our values

OpenCitations espouses the following:

- UNESCO [principles of Open Science](#)
- [POSI principles](#)
- The [FAIR data principles](#), that data should be Findable, Accessible, Interoperable, and Reusable
- The [I4OC principles](#) that citation data should be Structured, Separable, and Open

In compliance with these values, one of OpenCitations' main priorities is to **keep its services, software and data always without charge under open licenses** (CC0 for data ISC for software) for fostering maximum reuse

A fundamental project in open publishing

During FORCE 2021 conference, OpenCitations received the [Open Publishing Award](#) in the Open Data category

*“At the time of writing this review, the largest database provided by OpenCitations contains more than 1.23 billion citations. Compiling this database in a license-friendly way is a feat on its own, but combine that with OpenCitations’ persistence (established 11 years ago), their active and consistent involvement with the community, and the number of works that were made possible by their effort (Google Scholar lists 1440 results), it is clear that OpenCitations is **one of the fundamental projects in open publishing**, specifically in open scientific publishing”*



Indexes

OpenCitations Indexes (<http://opencitations.net/index>), including [COCI](#) (launched in July 2018), the OpenCitations Index of Crossref open DOI-to-DOI citations

COCI Service	URL
REST API <i>results in CSV and JSON</i>	https://w3id.org/oc/index/coci/api/v1
SPARQL endpoint <i>results in several formats</i>	https://w3id.org/oc/index/sparql
Dumps <i>all data in CSV, JSON, RDF</i>	http://w3id.org/oc/download#coci

Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics*, 121(2), 1213–1228. <https://doi.org/10.1007/s11192-019-03217-6>

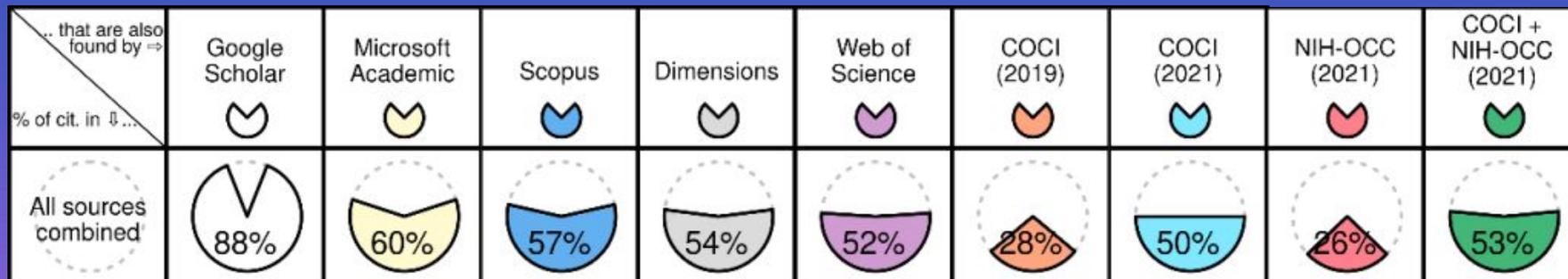
Citation coverage

COCI has recently crossed the significant threshold of one billion citations, and the last release (January 2022) contains

1,271,360,867 citation links

- Derived from open references within Crossref
- Includes the references within journal articles published by Elsevier and the American Chemical Society that recently been made openly available
- Citations hosted as Linked Open Data and published under CC0
- Citations accessible via a flexible API using DOIs (Digital Object Identifiers)

A recent independent comparison has shown that the coverage of COCI is approaching parity with those of the expensive subscription citations indexes Web of Science and Scopus



Percentage of citations found by each database, relative to all citations (first row), and relative to the number of citations found by the other databases (subsequent rows).

<https://opencitations.wordpress.com/2021/10/27/coverage-of-open-citation-data-approaches-parity-with-web-of-science-and-scopus/>

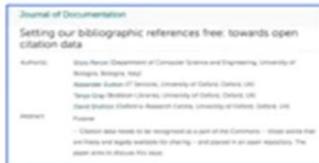
The citations available in COCI are treated as first-class data entities, with accompanying properties including the citations timespan, modeled according to the OpenCitations Data Model

Conventional citation



Citation as first-class data entity

has citing article



The Citation



has cited article

Heibi, I., Peroni, S. & Shotton, D. Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *Scientometrics* 121, 1213–1228 (2019). <https://doi.org/10.1007/s11192-019-03217-6>

Why OpenCitations?

The availability of the citation data via REST APIs has fostered new applications, from the description of references and citations of a given article in metadata libraries to the visualisation of networks of publications using scientometrics tools

In 2021 we had more than 22 million accesses to the REST API, and more than 10 million accesses in 2022 so far

<https://inciteful.xyz/>

Inciteful Documentation

OpenCitations, an infrastructure organization for open scholarship

Silvio Peroni | David M. Shotton arXiv: Digital Libraries, vol 1 no 1.

OpenCitations is an infrastructure organization for open scholarship dedicated to the publication of open citation data as Linked Open Data using Semantic Web technologies, thereby providing a disruptive alternative to traditional proprietary citation indexes. Open citation data are valuable for bibliometric analysis, increasing the reproducibility of large-scale analyses by enabling publication of the source data. Following brief introductions to the development and benefits of open scholarship and to Semantic Web technologies, this paper describes... [more](#)

16 Cited By 12 Citing 2020 Published Yes Open Access 8,183 Papers in Graph 29,947 Citations in Graph 2 Graph Depth

OpenCitations, an infrastructure organization for open scholarship

Silvio Peroni and David Shotton

Posted Online January 23, 2020
https://doi.org/10.1162/qss_a_00023

© 2020 Silvio Peroni and David Shotton. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Favorite Times Cited: 9

Download Citation RSS TOC

RSS Citation

Download Options

Article https://doi.org/10.1162/qss_a_00023 visualised with the Open Access Helper plugin (<https://www.oahelper.org/>)

<https://dblp.uni-trier.de/rec/conf/semweb/PeroniS18.html>

dblp computer science bibliography

SCHLOSS BAGSTUHL Leibniz Center for Informatics

home | browse | search | about

search dblp

"The SPAR Ontologies." Silvio Peroni, David M. Shotton (2018)

> Home

Silvio Peroni , David M. Shotton
 The SPAR Ontologies. International Semantic Web Conference (2) 2018: 119-136

References

- load references from crossref.org, open citations.net, and semantic scholar.org
- Simon Buckingham Shum , Enrico Motta , John Domingue : *ScholarOnTo: an ontology-based digital library server for research documents and discourse*. Int. J. on Digital Libraries 3(3): 237-248 (2000)
- Paolo Ciccarese, Silvio Peroni : *The Collections Ontology: Creating and handling collections in OWL 2 DL frameworks*. Semantic Web 5(6): 515-529 (2014)

The Lens @TheLensOrg - Jan 20

Release 6.6 is now live with new citing patents chart for individual scholarly works, new first or last author filter toggle and more! Check it out at lens.org

LENS.ORG

Release 6.6.0 New citing patents chart for individual scholarly works, new first or last author filter toggle and more!

Citing Patents Citing Scholarly Works

Filters: No filter applied

Scholar Results

Filters: No filter applied

26 39

The Lens @TheLensOrg

Update: Release 6.6 now includes updated citation data from the January 2020 COCI release, thanks @OpenCitations!

N. van Eck and L. Waltman (2019). VOSviewer supports large number of new data sources. <https://www.cwts.nl/blog?article=n-r2v284>

<https://twitter.com/TheLensOrg/status/1225004057086967808>

VOSviewer

Working with the community

At the end of 2019, **OpenCitations** was selected by the [Global Sustainability Coalition for Open Science Services \(SCOSS\)](#) for their second round of crowd-funding support

SCOSS stated that OpenCitations

“aligns well with open science goals, is an innovative service, and if successful could be a game changer by challenging established proprietary citation services”

DIRECTORY OF OPEN ACCESS BOOKS
OPEN ACCESS PUBLISHING FOR EUROPEAN NETWORKS
PUBLIC KNOWLEDGE PROJECT
OPENCITATIONS

SCOSS LAUNCHES SECOND FUNDING CYCLE

Read about these essential services, their funding goals and how your institution can help support them at www.scoss.org.



RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES



On our blog:

- <https://opencitations.wordpress.com/2022/01/13/five-reasons-why-2021-has-been-a-great-year-for-opencitations/>
- <https://opencitations.wordpress.com/2022/03/08/opencitations-and-ec-funding-openaire-nexus-and-risis2/>

Plans in 2022

New Indexes

The NIH-OCC underpins the iCite database that includes data from PubMed Central (PMC), MedLine, and Entrez from the National Library of Medicine (NLM), and reference data harvested from full-text scientific papers that have been made freely available on the internet

We intend to create **a new OpenCitations index of NIH-OCC (NOCI)**, enabling this entire corpus to be searched via a single API

In addition, we plan to index the citations in **DataCite** and other sources of open references

OpenCitations Meta, a new database of bibliographic metadata

At present, OpenCitations indexes include only DOI-to-DOI citations and the performance of some of its API operations is slowed by having to make on-the-fly API calls to Crossref, DataCite and ORCID

These two limitations will be removed by the creation of Meta, a new database containing in-house metadata relating to scholarly publications

Aims and goals

To **provide high-quality metadata** with full provenance relating to scholarly publications and the citations that link them, including those in areas such as the humanities and social sciences, the global south, and non-English publications

To **expand our coverage** into the 'grey literature' of reports, patents, datasets, software, etc.

To **surpass in terms of coverage and quality**, and thereby provide an open and free alternative to, the major commercial citation indexes

To **provide the open data** crucial for research in bibliometrics and scientometrics, and the creation of transparent and reproducible metrics for research assessment

To **continue to develop and make public free** and open-source software (FOSS) with relevant functionality, and our own open services built over our data

How to help

OpenCitations services and data are free and will always be free – but this leads to an acknowledged sustainability issue, principally in terms of salaries and technical infrastructure costs, which makes the support from the scholarly community crucial

You can apply for [membership](#) and thus support us financially

- Supporting membership (€500 - €8,000 per year)
- Development membership (€8,000 - €30,000 per year)
- Strategic membership (more than €30,000 per year)

Otherwise, ask your institution to fund us via [a donation](#)

More info at <http://opencitations.net/membership>

Support OpenCitations

Support Open Science



**Thank you for your
attention**