# OpenCitations

## OpenCitations: Providing Enduring Open Access to Global Scholarly Citations

## OpenCitations Members Webinar - December 2021

### Silvio Peroni and David Shotton

Directors, OpenCitations – @opencitations – http://opencitations.net
silvio.peroni@opencitations.net – @essepuntato – https://essepuntato.it
david.shotton@opencitations.net - @dshotton

OpenCitations

# The present state of OpenCitations

# 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

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**

# OpenCitations for Open Science

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"*

OPEN
PUBLISHING
AWARDS
2021

OPEN
DATA

# 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<br>*results in CSV and JSON* | https://opencitations.net/index/coci/api/v1 |
| SPARQL endpoint<br>*results in several formats* | https://opencitations.net/index/sparql |
| Dumps<br>*all data in CSV, JSON, RDF* | http://opencitations.net/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

Our database COCI (the OpenCitations Index of Crossref open DOI-to-DOI citations; http://opencitations.net/index/coci/ ) has recently crossed the significant threshold of one billion citations, and the last release (November 2021) contains

## 1,235,170,583 citation links

derived from open references within Crossref, including the references within journal articles published by Elsevier and the American Chemical Society recently been made openly available at Crossref

The citations are published under the CC0 public domain waiver and can be accessed via a flexible API using DOIs and OCIs (Open Citation Identifiers)

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.
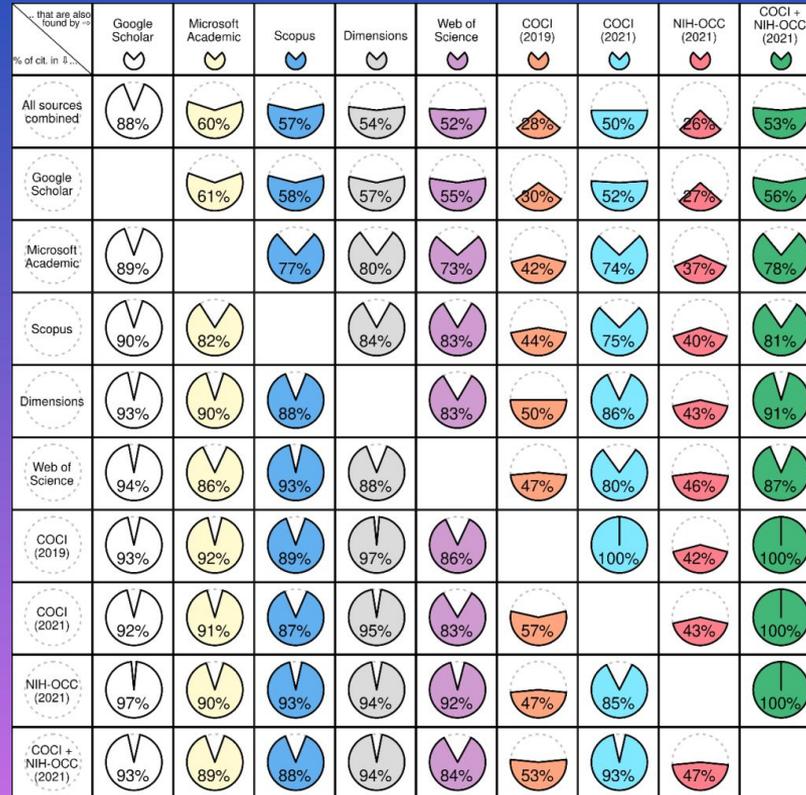


Heibi, I., Peroni, S. & Shotton, D. Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. Scientometrics 121, 1213–1228 (2019). https://doi.org/10.1007/s11192-019- 03217-6

# 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.



| ... that are also found by → / % of cit. in ↓ | Google Scholar | Microsoft Academic | Scopus | Dimensions | Web of Science | COCI (2019) | COCI (2021) | NIH-OCC (2021) | COCI + NIH-OCC (2021) |
|---|---|---|---|---|---|---|---|---|---|
| All sources combined | 88% | 60% | 57% | 54% | 52% | 28% | 50% | 26% | 53% |
| Google Scholar | | 61% | 58% | 57% | 55% | 30% | 52% | 27% | 56% |
| Microsoft Academic | 89% | | 77% | 80% | 73% | 42% | 74% | 37% | 78% |
| Scopus | 90% | 82% | | 84% | 83% | 44% | 75% | 40% | 81% |
| Dimensions | 93% | 90% | 88% | | 83% | 50% | 86% | 43% | 91% |
| Web of Science | 94% | 86% | 93% | 88% | | 47% | 80% | 46% | 87% |
| COCI (2019) | 93% | 92% | 89% | 97% | 86% | | 100% | 42% | 100% |
| COCI (2021) | 92% | 91% | 87% | 95% | 83% | 57% | | 43% | 100% |
| NIH-OCC (2021) | 97% | 90% | 93% | 94% | 92% | 47% | 85% | | 100% |
| COCI + NIH-OCC (2021) | 93% | 89% | 88% | 94% | 84% | 53% | 93% | 47% | |

*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/

# New staff members

Since last summer OpenCitations made four staff appointments to support OpenCitations' development

**Claudio Fabbri** (Administrator and Research Manager) has responsibility for all the day-to-day administrative activities of OpenCitations, including financial arrangements with supporters

**Chiara Di Giambattista** (Communications Director and Community Development Manager) takes care of all communications and social interactions made on behalf of OpenCitations

**Giuseppe Grieco** (Software and Systems Developer) has taken charge of all the technical development related to the OpenCitations services, data acquisition, and bibliographic metadata and citations provision

**Arcangelo Massari** (Software Developer) has taken charge of the development of OpenCitations Meta, a new database containing bibliographic metadata of all citing and cited entities available in OpenCitations collections, and tools for handling and querying provenance information

# Promotion and community outreach

**- Online communication:**
- OpenCitations Blog: blogging as academic practice makes it possible to reach a broad range of people interested in a specific topic. OpenCitations Blog publishes new posts weekly
- LinkedIn and Twitter
- Blogging collaborations (i.e. DIFF Blog; OpenAIRE Blog)

**- Participation in events and conferences:** OpenCitations participates in numerous SCOSS-related events and in other events and conferences, thus promoting OpenCitations mission and goals and interacting with other actors from the Open Science community.

# Governance

**RESEARCH CENTRE FOR OPEN SCHOLARLY METADATA**

University of Bologna
Administrative and accounting management

**DIRECTORS**

Ultimate responsibility for the purpose, development, operation, and financial management of OpenCitations

**OPENCITATIONS MEMBERS**

**COUNCIL**

- Elects new members of the International Advsory Board.

- Offers advice to the International Advisory Board and the Directors.

**ADVISORY BOARD**

- Guides the future developments of OpenCitations in terms of its policies, strategies and activities

- Builds the OpenCitations collaborations, membership and support community

- Ensures financial sustainability

- Assists the OpenCitations Directors in their role

**TEAM**

- Supports both the Directors and the Board by undertaking administrative and secretarial tasks

- Prepares the biannual OpenCitations progress reports and the annual OpenCitations financial report

# Business model

- For OpenCitations, "open" is the crucial value and the final purpose
- Everything we provide – data, services, software, etc. – is open and free, and **will always remain so**

**This means that our activities generate no income to sustain OpenCitations**

While this poses an obvious financial problem, it also provides our unique 'selling point' in terms of charitable and institutional support

# Finances and Funding

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

The financial support for OpenCitations from the scholarly community can be provided in two ways:

- **Donation**: the administratively simplest option
- **Membership**: becoming an OpenCitations Member brings with it privileges and opportunities to participate in the governance of OpenCitations
Three distinct tiers of Membership
Members pay an annual Membership Fee, initially for a three-year period, along the lines that SCOSS suggests



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.
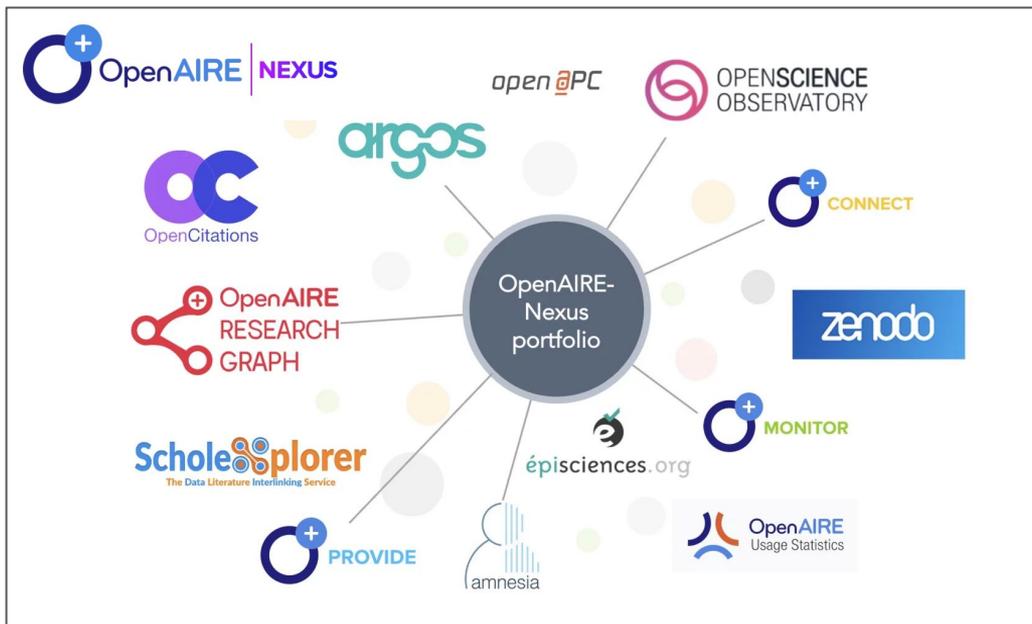
SCOSS

open Open Access Publishing in European Networks

doab directory of open access books

PKP PUBLIC KNOWLEDGE PROJECT

OC OpenCitations

The involvement of OpenCitations in different funding applications throughout 2020 has led to the successful development of three projects.

- **OpenAIRE Nexus**
- **Open Biomedical Citations in Context Corpus,**
- **Wikipedia Citations in Wikidata**

The support received from OpenCitations' donors and members during 2020/2021 is making it possible to pay for necessary services and the salaries of personnel to support the daily activities of OpenCitations thus enhancing

- Administration + Communication and community development
- Technical development

However, despite the very large contribution from the French Open Science Plan, and generous donations from others, such income so far amounts to only about **one third** of the SCOSS budget

# Plans for the future

# New citation indexes

Together, COCI and the NIH Open Citations Collection provide access to the largest corpus of CC0 citation data. We intend to create **a new OpenCitations index of NIH-OCC**, 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, by design, COCI only indexes DOI-to-DOI citations, and its performance is slowed by having to make on-the-fly API calls to Crossref, DataCite, and ORCID for article metadata. These two limitations will be removed by the creation of Meta, a new database containing in-house metadata relating to scholarly publications. We plan to populate and release Meta, which is currently under development, in 2022.

OpenCitations Meta will provide two major benefits:

- It will permit us to store in-house the basic bibliographic metadata for the citing and cited publications involved in all our indexed citations, including author identifiers using ORCiDs and VIAF, where available. This will provide better query performance than our present system, which obtains such bibliographic metadata on-the-fly by live API calls to Crossref, DataCite and ORCiD.

- It will permit us to index citations involving publications lacking DOIs, that for intentional technical reasons cannot presently be handled by the structure of the Open Citation Identifiers that underpin our citation indexes.

In principle, OpenCitations Meta will allow us, in the future, to ingest other kinds of data such as the recently released Internet Archive General Index, a collection of keywords and text-mined words and phrases from 107,233,728 journal articles, since these will be invaluable in permitting us to assign domain and scientific topic identification to our citations.

Our future goals are:

- 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 that are presently under-represented

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

- To **surpass in terms of coverage and quality**, and thereby provide an open alternative to, the major commercial citation indexes for access to which academic institutions presently spend fortunes

- Thereby 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

# Future hopes: Governance

According to the monitoring of OpenCitations' compliance to the Principle for Open Scholarly Infrastructure, the governance point is still to be fully accomplished

- Coverage across the research enterprise
- Stakeholder governed
- Non-discriminatory membership
- Transparent operations
- Cannot lobby
- Living will
- Formal incentives to fulfill mission & wind-down

The current Advisory Board currently lacks executive power and was not elected
**During 2022, a Working Group develop a strategy to identify OpenCitations stakeholders and to integrate the executive power in OpenCitations' governance**

# Future hopes: Funding

Despite the benefits we have received as a result of our selection by SCOSS, what OpenCitations really needs quite soon is **a major benefaction** to move us

from sustainable infrastructure

to *financially sustained infrastructure*

which the global scholarly community can rely for open bibliographic and citation data for many years to come

We hope in a broader participation of the community not only to support us, but also

- to participate in the shared governance of OpenCitations
- to improve our software and services
- for the provision, annotation, curation and correction of OpenCitations data
- to work with us to broaden the scope of scholarly metadata that is openly available, beyond citations and abstracts

**OpenCitations is a plural: together we are OpenCitations**

Thank you for your attention