OpenCitations: present status and future plans
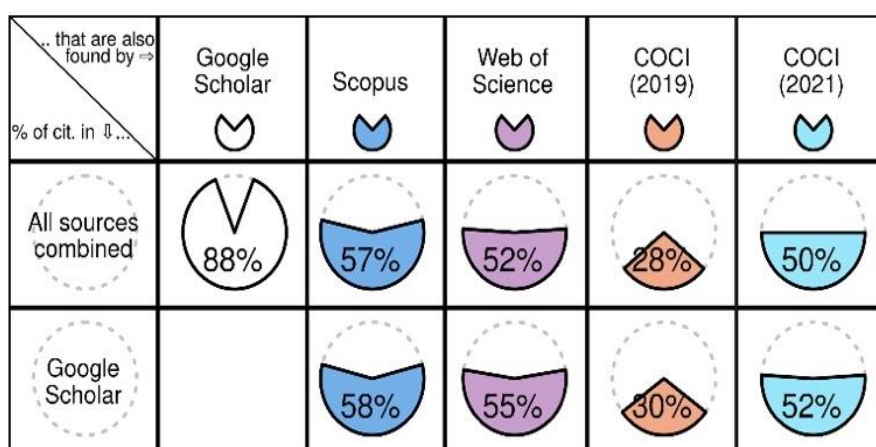
Purpose

OpenCitations is an independent global not-for-profit infrastructure organisation for open scholarship that provides free access to comprehensive open bibliographic metadata describing the world's academic publications and the scholarly citations that link them, accompanied by full provenance information, and that preserves ongoing access to this information by secure archiving. Our data, together with full provenance information, is made available in both in human-readable form and in interoperable machine-readable Linked Open Data formats, and is published under an open license so that others can re-use and add value to it. This ensures the transparency and reproducibility of analyses based on our open data. Further details are provided in the **OpenCitations Mission Statement** and in the accompanying document entitled **The Uniqueness of OpenCitations**.

Coverage

Currently, the principal database maintained by OpenCitations is COCI, the OpenCitations Index of Crossref Open DOI-to-DOI Citations, which presently contains over 1.3 billion citation records, approaching parity in terms of coverage with Web of Science and Scopus.

.. that are also found by ⇒ / % of cit. in ⇓ ...	Google Scholar	Scopus	Web of Science	COCI (2019)	COCI (2021)
All sources combined	88%	57%	52%	28%	50%
Google Scholar		58%	55%	30%	52%

It should be pointed out that our COCI index stores only *citation* metadata, together with DOI identifiers for the citing and cited publications. *Bibliographic* metadata for those publications is presently obtained on-the-fly by live API calls to external authorities such as Crossref and ORCiD, but will in future be stored within the OpenCitations Meta database.

Present objectives

Given appropriate resources, OpenCitations' present objectives are:

to broaden our citation coverage, by creating additional citation indexes over other sources of open bibliographic references, particularly (a) in the domains of the humanities, social sciences and computing, (b) from publications in non-English languages, thereby breaking away from the present preponderance of citations involving English language publications, and (c) from publications in the Global South, starting with those in Latin America;

to permit the recording of citations involving publications that lack conventional DOI identifiers, such as books, conference papers, UN and WHO reports, government policy documents, and many journal articles, particularly those from small publishers and non-anglophone countries;

OpenCitations

to host citations involving datasets, software, patents and other non-conventional reference targets, in addition to those involving textual publications;

to host in-house the bibliographic metadata relevant to those publications whose citations are recorded in the OpenCitations indexes, and to uncited publications; and

to further develop our unified OpenCitations API to perform federated information retrieval over all OpenCitations indexes and, where appropriate, also over third party indexes that store information in accordance with the generic OpenCitations Data Model.

Future technical development plans

We are starting our coverage expansion by developing two new citation indexes of open references, based on the holdings of DataCite (DOCI) and the NIH Open Citations Collection (NOCI), which, together with COCI, will be cross-searchable through our unified OpenCitations API.

Our intention is then to ingest the Internet Archive's newly released Refcat: The Internet Archive Scholar Citation Graph, the references from arXiv eprints, and open data from other DOI registration agencies such as the Japan Link Centre. These are presently on hold due to our lack of resources, but are set to proceed once funding permits, following our very positive discussions with colleagues at those institutions.

Additionally, as mentioned above and as anticipated in our February 2020 canonical paper on OpenCitations in Quantitative Science Studies (QSS), *OpenCitations, an infrastructure organization for open scholarship*, we have started work to create a new database entitled OpenCitations Meta. OpenCitations Meta is a key part of our development plan, and will provide three major benefits:

It will permit us to store in-house bibliographic metadata for the citing and cited publications involved in all our indexed citations, including author identifiers using ORCID iDs, institutional affiliations and funding information where available. This will provide better query performance than our present system, which obtains bibliographic metadata on-the-fly by live API calls to Crossref and ORCID.

It will permit us to index citations involving publications lacking DOIs, by providing Meta Identifiers (MIDs) for the citing and cited publications that will enable the citations between them to be assigned Open Citation Identifiers and thus to be included in an OpenCitations Index.

It will permit us to store in-house bibliographic metadata for publications that at present lack incoming citations.

OpenCitations development working groups

To guide the non-technical development of OpenCitations, members of the OpenCitations International Advisory Board are serving in two working groups that are meeting several times over the coming year and will report their recommendations back to the Directors and the Advisory Board.

The Community Building Working Group will address the following issues:

- Which OC stakeholders are not presently being involved in OpenCitations, and what can we do in terms of outreach to improve that situation?
- How can the OC activities and membership models better include and reflect diverse stakeholder groups?
- How can we promote partnerships and initiate community data provision and curation of OpenCitations data?
- How can we best involve the community to increase its financial support of OpenCitations,

to enable us to realise our ambitious plans to expand our activities?

The Governance Evolution Working Group will address the further issues:

- Clarification of the present representation of stakeholders in OpenCitation governance bodies and their actual power.

- How can we involve our stakeholder community more fully in the governance of OpenCitations, in accord with POSI principle?

- What is the future role of the University of Bologna in OpenCitations?

'Business model'

OpenCitations resembles Wikipedia by providing open access to fundamental factual information, and also by charging nothing for such access or for related services. That sets us apart from many other OpenScience infrastructures, for example Crossref that charges members for issuing DOIs and for services such as its CitedBy service. This means that OpenCitations' activities generate no income to sustain our operations. While this poses an obvious financial problem, it also provides our unique 'selling point' in terms of charitable and institutional stakeholder support.

The alternative viewpoint, that OpenCitations should sell some of its services, or provide access to data in exchange for fees, really does not make sense. If we sold access to our data, which itself is derived from Crossref and other open reference sources, that would make us no different from commercial citation indexes.

One great advantage to providing free access to all our data and services is that we cannot easily be bought out – there is nothing to buy that is not already openly available!

Present funding

During the first ten years of its existence, OpenCitation was a very low-budget research operation supported primarily by voluntary effort, supplemented by three small grants from Jisc, the Sloan Foundation and the Wellcome Trust. Since 2020, OpenCitations has benefited greatly by crowdfunding from the scholarly community that resulted from the 2019 selection by SCOSS of OpenCitations as a scholarly infrastructure worthy of support. However, partially due to the COVID-19 pandemic and the ensuing difficulty of meeting key people from potential supporting institutions, the amount donated after two years of fund-raising has been only one third of the three-year budget we submitted to SCOSS to enable our planned developments. Indeed, were it not for the very generous support from the French Government's Open Science Plan, we would have raised only one sixth of our requested developmental funding.

The community funding we *have* received permitted the appointments during 2021 of Giuseppe Grieco as our new Software and Systems Developer, dedicated to maintaining and improving the OpenCitations software and the computational infrastructure on which it runs, together with Claudio Fabbri as Administrator and Research Manager and Chiara Di Giambattista as Communications Director and Community Development Manager. We are most grateful for this generous support.

In addition, in January 2021 we started our involvement in the EC-funded OpenAIRE Nexus project, within which we are providing open bibliographic citations as part of the open data components of OpenAIRE and the European Open Science Cloud (EOSC) "for the benefit of the open science community worldwide". This has permitted us to appoint Arcangelo Massari as our new software engineer to start the development of a new database for bibliographic metadata relevant to our indexed citations, OpenCitations Meta.

Future financial support

Although our recent research grants and the support we have received as a result of our selection by SCOSS are of great benefit, we have concluded that what OpenCitations additionally needs quite soon is a major benefaction to move us from being a 'sustainable infrastructure' (in POSI terms) to being a *financially sustained infrastructure* upon which the global scholarly community can rely for open bibliographic and citation metadata for many years to come. Sustaining financial support should never be expected from grants awarded for specific developmental projects, valuable as these are, and is proving difficult to achieve solely from community crowdfunding. We thus plan to submit applications to suitable funders during 2022, benefitting from input and advice on these applications from members of the OpenCitations International Advisory Board.

Blog posts

Recent news and information about OpenCitations is to be found on our blog, https://opencitations.hypotheses.org/. Recent posts include the following:

Five reasons why 2021 has been a great year for OpenCitations

Open citations in Informatics: current status and lines of research

OpenCitations receives the Open Publishing Award in Open Data

Performing live time-traversal queries on RDF datasets

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

The OpenCitations Roadmap is now publicly available on Trello.

Two years of achievements within the 'SCOSS family'

Document date: 20th July 2022. Comments and enquiries to contact@opencitations.net.