The Uniqueness of OpenCitations

This document should be read together with the **OpenCitations Mission Statement**.

There are other open sources of scholarly bibliographic information and of reference lists from which citations can be inferred, including Crossref, DataCite, the NIH Open Citations Collection, OpenAlex, CiteSeerX, OpenAIRE, the Open Ukrainian Citation Index, Wikidata and Europe PubMed Central. However, OpenCitations is unique among them in having ***all*** of the following characteristics:

Data

We treat scholarly citations as first class data entities: It is the fundamental aim of OpenCitations to provide integrated access to a comprehensive high-quality corpus of the world's scholarly citations, in addition to the bibliographic metadata characterising citing, cited and uncited scholarly publications.

Public domain waiver: To ensure the greatest possible reusability, all OpenCitations data is published under a Creative Commons CC0 Public Domain Waiver that permits downloading and re-use of any nature, including added-value re-purposing and commercial exploitation.

No fees for services, data access or software: OpenCitations charges no fees for any of its services, for access to any of its data, or for reuse of its software. OpenCitations members, donors and third parties all have equal free access.

Provenance information: Every item of data within OpenCitations is accompanied by machine-readable metadata recording the provenance of the information, including its source, the nature of any subsequent curatorial modification or correction, the dates of such actions, and the identities of the agents involved (whether human or computational).

OpenCitations scope: The scope of OpenCitations' coverage is universal, not limited to a particular scholarly domain (e.g. biomedicine, computer science, or humanities), nor to the English language, nor restricted by imposed acceptance criteria.

OpenCitations coverage: As of July 2022, the OpenCitations coverage of global scholarly citations, in excess of 1.3 billion, was approaching parity with that of two of the principle proprietary citation indexes. Our plans to increase this coverage are outlined in our Mission Statement.

Semantic interoperability: We encode our data in ways that promote semantic interoperability with third-party data.

Community engagement and governance

Academia-based: OpenCitations is owned and run by the academic community, not by a commercial company. This ensures that it is not at risk of arbitrary closure for commercial reasons, as was the fate of Microsoft Academic in December 2021.

POSI principles: As an open scholarly infrastructure, OpenCitations has adopted the Principles of Open Scholarly Infrastructure (POSI).

Community engagement: Community engagement in OpenCitations has three aspects -

collaborative partnerships with like-minded infrastructures, services and data providers;

community provision of financial support for OpenCitations; and

community participation in the development and governance of OpenCitations.

OpenCitations is presently exploring how to enhance community engagement in all three areas.

Technical

Citations as first class data entities: Uniquely, OpenCitations treats citations as first class data entities, each with a unique persistent identifier, an Open Citation Identifier (OCI), rather than simply as relationships between citing and cited publications.

Citation metadata: This treatment enables us to enrich citations by assigning descriptive metadata to each citation individually, for example its citation timespan, and whether or not it is a self-citation (author self-citation, journal self-citation, institutional self-citation, etc.), encoded using CiTO, the citation typing ontology. Such information is held separately from bibliographic metadata characterising citing, cited and uncited scholarly publications.

Linked Data: Central to OpenCitations' activities is the adoption and practice of Semantic Web technologies, particularly the Resource Description Framework (RDF) and the Web Ontology Language (OWL), enabling us to publish our metadata as Linked Open Data, thereby promoting semantic interoperability with third-party data.

Ontologies: We have created the SPAR (Semantic Publishing and Referencing) Ontologies to foster semantic interoperability of all scholarly publishing and referencing metadata, and we use these ontologies to encode all our data in RDF.

Data model: The generic OpenCitations Data Model, available for use by third parties, is used to structure all our RDF data. This helps ensure unambiguous machine readability and semantic interoperability of OpenCitations data with related data from other providers of open scholarly information.

Open Source Software: The OpenCitations open source software used to run all our databases and services not only permits third parties to re-use our developments, but also ensures that the OpenCitations system could be re-created elsewhere in the event that OpenCitations itself ceased to operate, thereby ensuring the preservation and longevity of our work and data.

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