## **OpenCitations**

### **OpenCitations: an independent infrastructure** organization for open citation data

### EAHIL Evaluation and Metrics SIG Webinar February 2, 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 **Director of OpenCitations** silvio.peroni@opencitations.net - @opencitations - http://opencitations.net







## General information and services

- OpenCitations (<u>http://opencitations.net</u>) is an <u>independent infrastructure organization</u> 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): <u>OpenCitations Corpus</u>, <u>COCI</u>, <u>CROCI</u>, <u>CCC</u>
  - software: in our <u>GitHub repository</u>, released with open source licenses
  - online services: <u>REST APIs</u>, <u>SPARQL endpoints</u>, <u>dumps</u> and <u>interfaces</u>

## What is an open citation



cites

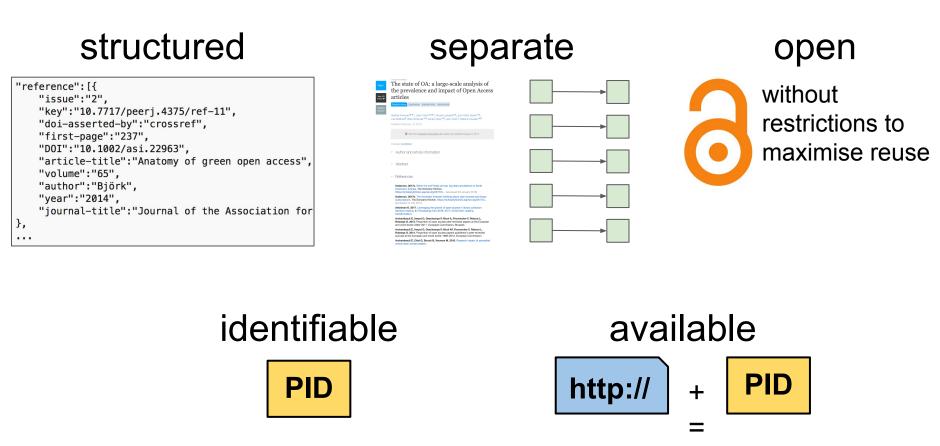
Entity B

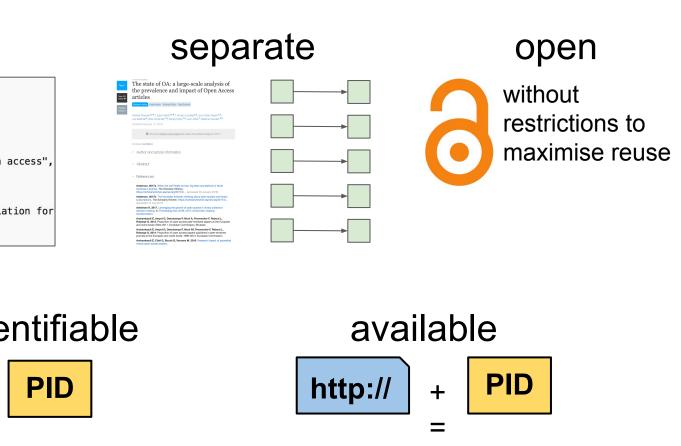
Citation: a conceptual directional link from a citing entity to a cited entity, for the purpose of acknowledging or ascribing credit for the contribution made by the author(s) of the cited entity.

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:

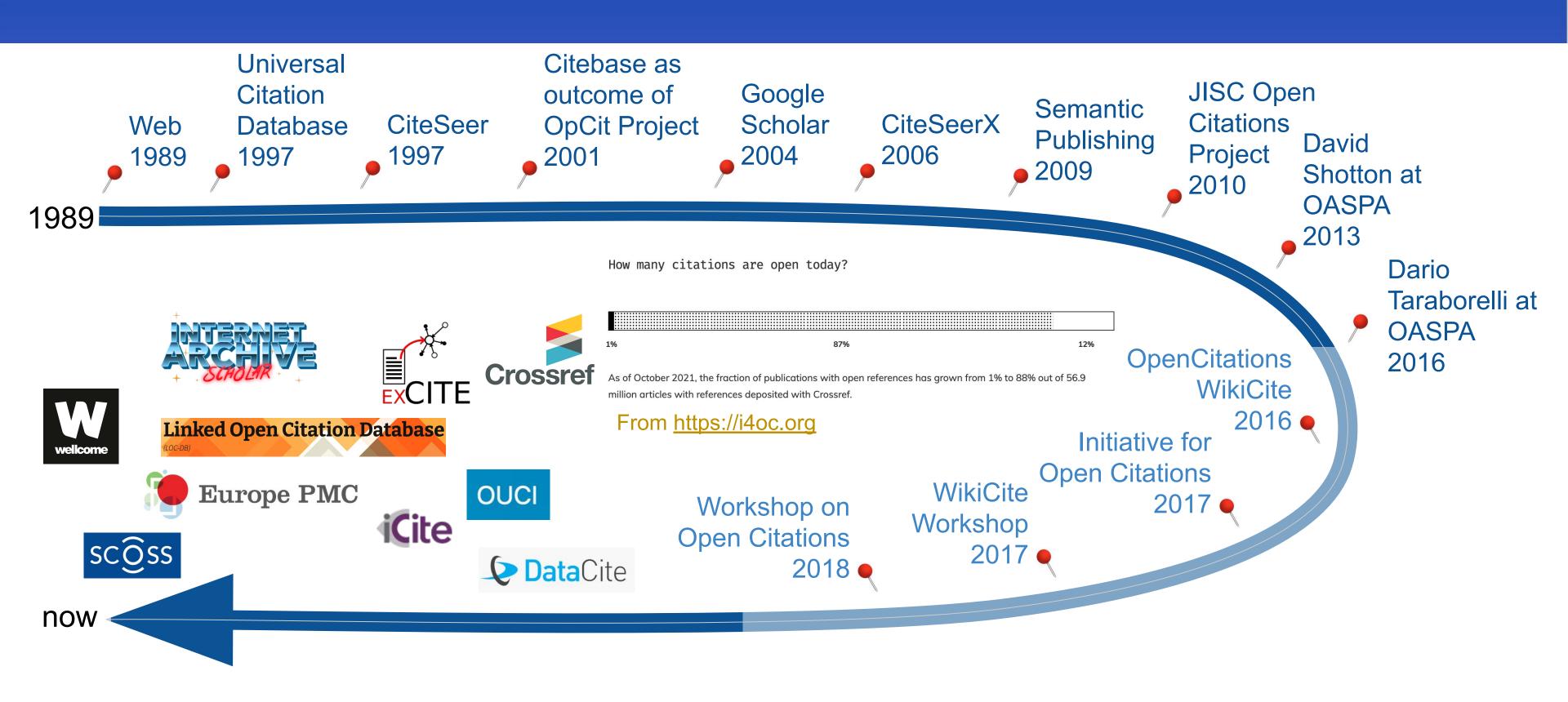




metadata

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

### History of open citations, in brief (3/3)



## Citation coverage

Our main citation index, COCI (the OpenCitations Index of Crossref open DOI-to-DOI citations, <u>http://opencitations.net/index/coci/</u>), 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)

## 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

https://inciteful.xvz/ M Inciteful OpenCitations, an infrastructure organization for open scholarship Silvio Peroni | David M. Shottor 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 Ope 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 pape describes... 💥 more 16 2 12 Yes 8.183 29.947 **Open Access** Cited By Citino Papers in Graph **Citations in Graph** Graph Depth

### **OpenCitations**, an for open scholarship

Silvio Peroni () and David Shotton ()

Posted Online January 23, 2020 https://doi.org/10.1162/qss a 00023

4.0 International (CC BY 4.0) license.

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

Rele	ase 6.6.0 New cit	ng patents chart for	IOI https://www.lens.org/scholar/					
individual scholarly works, New first or last author filter toggle and more!			Filters 💿	Scholar Results				
			Date Range	~	Scholarly Works (184,922) = malaria 🗢			
			Flags	~	Filters: No filters applied			
			Institution	~	Works in S	et	Works Cited by I	
			Institution Country/Region	• •	184,922		9,529	
	Citing Patents	Citing Scholarly Works	Author	^	-			
	-		Search Author ID	•	Scholarly Works	Citing Pat	ents	
	1980 1985 1990 1995 2000 2005 2010 2015		Al Authors First Author Only Last Author Only Author Identifiers Nicholas J White Francois Nosten Brian Greenwood  Kevin Markh a	(844) (605) (515)	Colpand All Blanc Qury Elawa Realt Global and regional mortality from 2: 2010: A systematic analysis for the Gl Rafeel Lozano <sup>1</sup> , Mohaen Naghavi, Kjel J Por Pubraham <sup>1</sup> , Timothy Adair, Sakesh Aggaran H Boss Anderson <sup>1</sup> , Laurie M Anderson, Rat Barker-Colfo <sup>1</sup> , Duivid Blanche <sup>1</sup> , Michelle L The Lancet, Issue S89, Volume: 180, Page Additional Info [Control Of Lines 180]			
GIF	In: Science, Iss Page: 673-675	ue: 4254, Volume: 193,	Chris Drakeley 🥸 Robert W Snow 🚳 Sornchai Looareesuwan 🍏	(446) (366) (353) (341)	Journal Article	e a 101 023-1	8 Scholarly Citatio 641-106-024-532 xt at your institution	
Q 2		① 26	♡ 39		≏			

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



# 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

Setting of citation	our bibliographic references free: towards open data
4,444	Ross Annas Department of Computer Science and Expressing, University of Beinges, Respire, Sep) Appendix Science Of Section, University of Onlong, Onlong, UNI

Band (puttor clother) - Assarch (units, unserb) of lother), (other), (

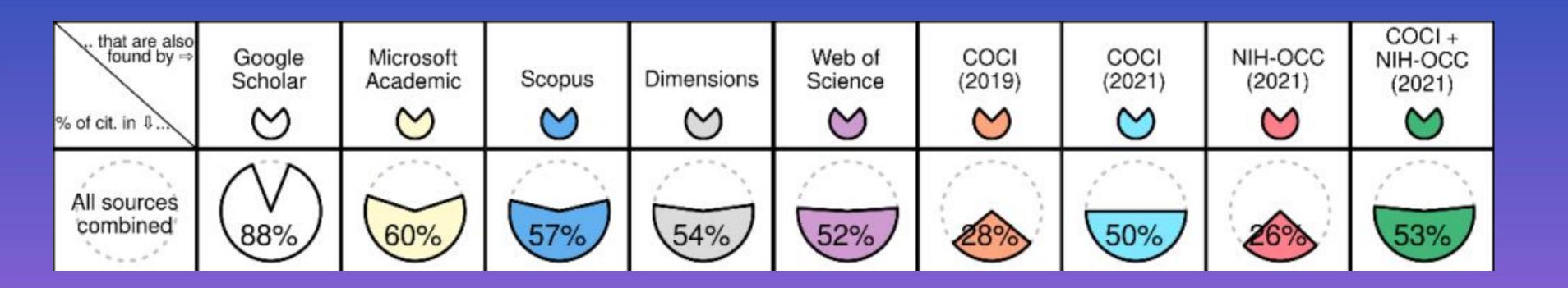
ant Testy and highly available for charting - and parent in an ownit reposition. The appendence of the second first many





nantic	
	Robbiel Cornect
files technological price of approximation of approximati	products of previous qp. particularly Pose was to providing semantic totarty pound publishing, do b a second burneling

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



*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-appr oaches-parity-with-web-of-science-and-sco pus/

## Why OpenCitations?

There are other open sources of scholarly citation information and of bibliographic reference lists from which citations can be inferred – however OpenCitations is unique among them in having *all* of the characteristics shown in the following scheme, covering data, governance and technical infrastructure

### DATA

- Primary focus is on citations
- Public domain waiver
- No fees for services or data access
- Full provenance information
- Extensive scope
- Non-selective coverage

### **GOVERNANCE**

- Academia-based
- POSI principles
- Community engagement

### **TECHNICAL**

 Citations treated as first class data entities

- Citation metadata
- Linked Data (RDF)
- Ontologies (SPAR)
- Data model (OCDM)
- Open Source Software

## 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, 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

## **OpenCitations for Open Science**

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"

OpenAIRE NEXUS

### **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.





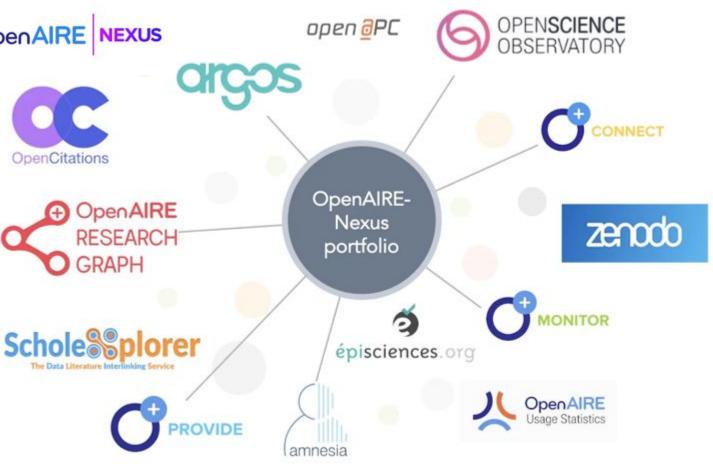












During Force2021 conference, OpenCitations received the **Open Publishing Award** in the Open Data category

> "It is clear that OpenCitations is one of the fundamental projects in open publishing, specifically in open scientific publishing"





## Our values and priorities

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 <u>I4OC principles that citation data should be Structured</u>, 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

## 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 <u>membership</u> and thus support us financially

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

Otherwise, ask your institution to fund us via a donation More info at <a href="http://opencitations.net/membership">http://opencitations.net/membership</a>

**Support OpenCitations** 

**Support Open Science** 

# Thank you for your attention

