

NEXUS

OpenCitations

an infrastructure for
open bibliographical
metadata

Ivan Heibi

University of Bologna

Outline

- OpenCitations
 - Overview
 - Citations
 - OpenCitation Indexes
 - Citation coverage and Statistics
- Integration with other services
 - Integration with Nexus services
 - Integration with research-related services
 - OpenCitations in EOSC
 - Usage KPIs
- Future steps
 - OpenCitations Meta
 - COCI, NOCI, and DOCI
 - OpenCitations badge
- Querying data
 - SPARQL endpoints
 - Search interfaces
 - REST APIs
 - OpenCitations access token

OpenCitations

<https://opencitations.net/>

Overview

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, recently included in FAIRSharing.org)
- bibliographic and citation data (CC0): OpenCitations Indexes
- software: in our GitHub repository, released with open source licenses
- online services: REST APIs, SPARQL endpoints, dumps and interfaces

Citations

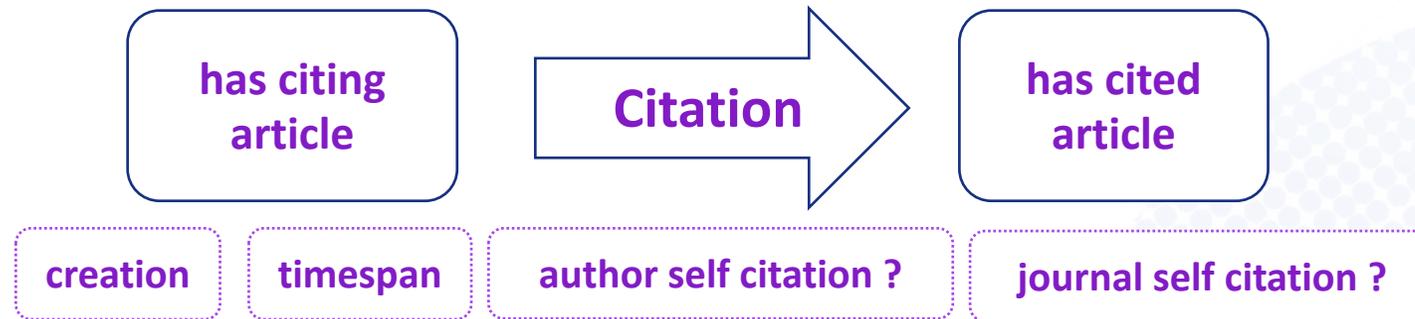
Citation: conceptual directional link from a citing entity to a cited entity



- data of a particular citation must include: (1) the representation of such a conceptual directional link; (2) basic metadata of the citing entity and the cited entity
- citation data should be structured, separable, and open
- citations in OpenCitations are hosted as Linked Open Data, defined using the [OpenCitations Data Model](#) (OCDM)

Citations in OpenCitations

Citations are defined as **First-Class Data Entities** (following OCDM)



Advantages:

- all the information regarding each citation is available in one place.
- citations become easier to describe, distinguish, count and process.

Citations are identified with the **Open Citation Identifier (OCI)**, a globally unique persistent identifier (PID) for the identification of open bibliographic citations stored in a specific database or in other kinds of storages.

OpenCitations Indexes

OpenCitations Indexes have crossed the significant threshold of one billion citations and the last release (August 2022) contains

1,363,718,366 citation links

75,030,924 bibliographic resources

OpenCitations provides citation data via: REST APIs, SPARQL endpoints, dumps and interfaces

COCI and CROCI

COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations

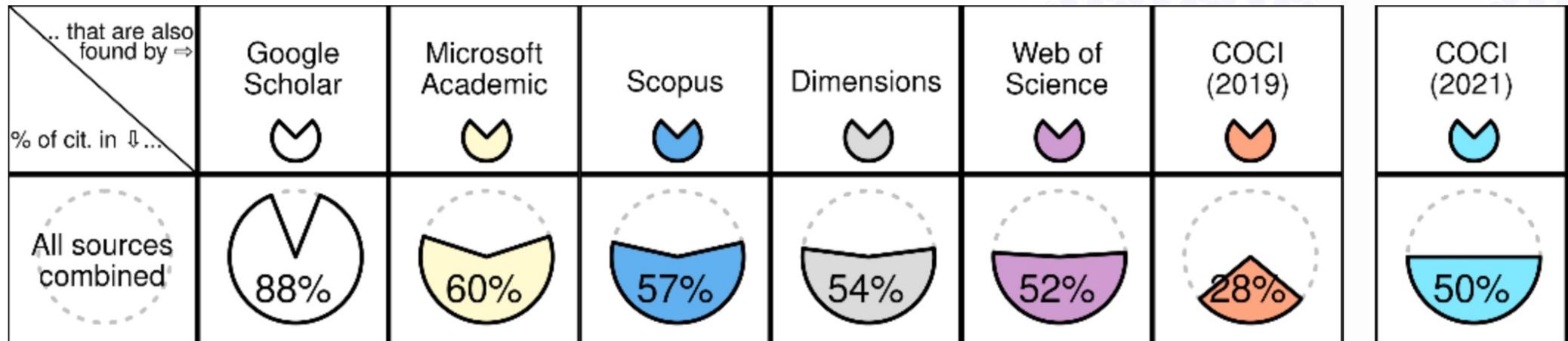
- an RDF dataset containing data of all the citations that are specified by the open references to DOI-identified works present in Crossref.

CROCI, the Crowdsourced Open Citations Index

- individuals (identified by ORCID) may deposit citation information that they have a legal right to submit, and within which these submitted citation data will be published under a CC0 public domain waiver

Citation coverage

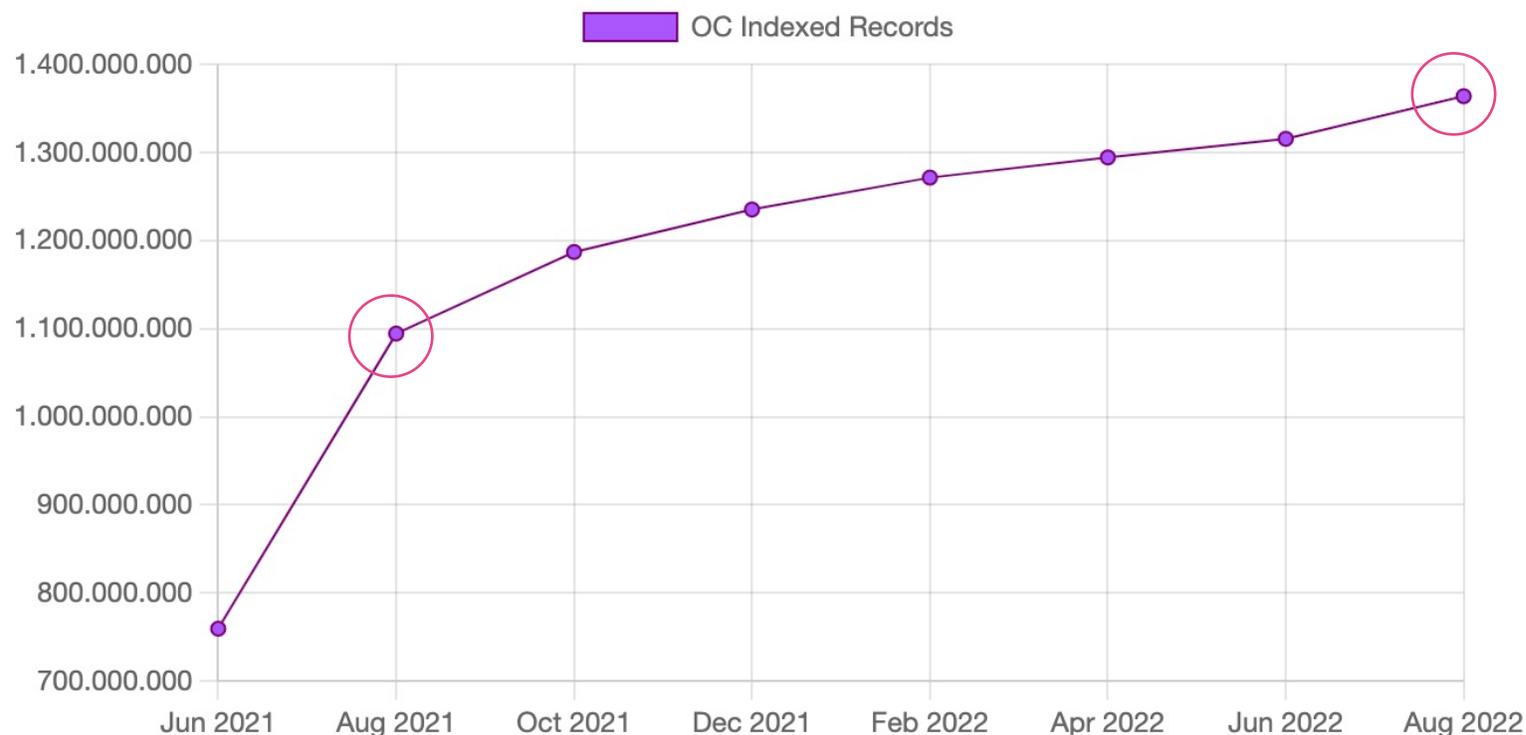
A recent independent comparison has shown that the coverage of COCI is approaching parity with those of the expensive subscription citations indexes Web of Science and Scopus



<https://opencitations.hypotheses.org/1420>

Statistics regarding the indexed citations

The number of records (citations) indexed in the COCI dataset of OpenCitations throughout the time

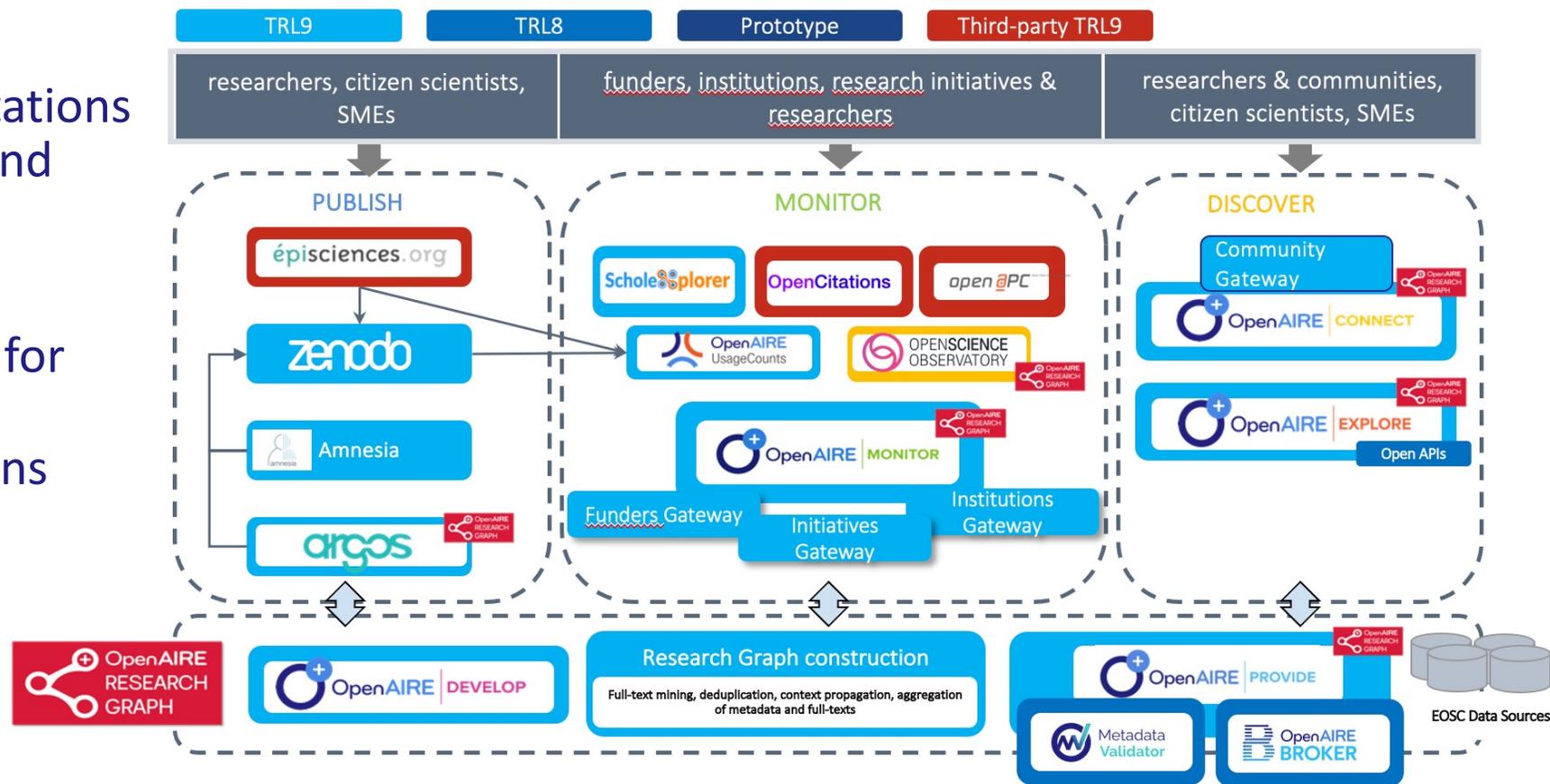


Integration with Other Services

Integration with Nexus services

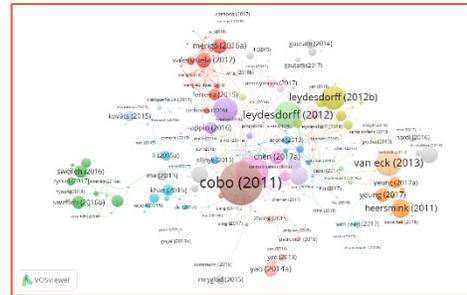
Research Graph

- Ingestion of OpenCitations data (bibliographic and citation data) to the Research Graph
- Software developed for enabling ingestion to/from OpenCitations (integration with EpiSciences is still in progress)



Integration with research-related services

- Exposed in the **EOSC Marketplace**
- Other users:** DBLP, OpenAccess Helper, Lens.org, VOSviewer, Inciteful, Citation Gecko, VisualBib, CiteCorp, Scipedia, scanR, Spark, Pure Suggest, B!SON, BIP! Finder, ...



Article Navigation <https://www.oahelper.org/>

February 01 2020

OpenCitations, an infrastructure organization for open scholarship

Silvio Peroni, David Shotton

Check for updates CC Times Cited: 33

> Author and Article Information

Quantitative Science Studies (2020) 1 (1): 428–444.
https://doi.org/10.1162/qss_a_00023 **Article history**

EUROPEAN OPEN SCIENCE CLOUD

Find resource... All resources

Resources > Sharing & Discovery > Scholarly Communication > Discovery > OpenCitations

OpenCitations
 OpenCitations for Open Science
 Organisation: Alma Mater Studiorum - Università di Bologna
 ★★★★★ (0.0/5) 0 reviews Add to comparison Add to favourites

Access the resource

B!SON

B!SON is still in beta and subject to change.

B!SON helps you to find a suitable Open Access journal for your publication by leveraging semantic and bibliometric methods. Simply enter the details of your manuscript below or [fetch the details of a paper via its DOI](#).

Title
 What do we mean by "data"? A proposed classification of data type

Abstract
 Objectives: We describe here the interviews we conducted in late 2021 with 19 researchers at the Department of Classical Philology and Italian Studies at the University of Bologna. The purpose has been to shed light on the definition of the word "data" in the humanities domain, as far as FAIR data management practices are

References
 Akers, K. J., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5–26. <https://doi.org/10.2218/ijdc.v8i2.263>
 Armeni, K., Brinkman, L., Carlsson, R., Eerland, A., Fijten, R., et al.

Check

SCHLOSS DAGSTUHL
 computer science bibliography

Stop the war!

"OpenCitations, an infrastructure organization for open scholarship."
 Silvio Peroni, David M. Shotton (2020)

Home Details and statistics

Silvio Peroni, David M. Shotton

OpenCitations, an infrastructure organization for open scholarship. Quant. Sci. Stud. 1(1): 428-444 (2020)

DOI: 10.1162/qss_a_00023
 access: open
 type: Journal Article
 metadata version: 2021-02-17
 load more data from openalea.org

References
 load references from crossref.org
 open citations.net, and semantic scholar.org

Christian Bizer, Tom Heath, Tim Berners-Lee
Linked Data - The Story So Far. Int. J. Semantic Web Inf. Syst. 5(3): 1-22 (2009)

Marco Corbatto, Antonina Dattolo
A Web Application for Creating and Sharing Visual Bibliographies. SAVE-SD@WWW 2016: 79-94

Ivan Hebl, Silvio Peroni, David M. Shotton
Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. Scientometrics 121(2): 1213-1228 (2019)

OpenCitations in EOSC

1. Openness

EOSC is based on the principle of openness



For end-users	For service providers
EXPLORE To discover open access versions of research resources 60,391,907 Open Access content	PROVIDE To increase visibility, discoverability and compliance of open repositories with the OpenAIRE Guidelines 9K+ Open Access repositories 2K+ Open Access journals 140 Open Access datasets
MONITOR To track open science activities and funding based on open indicators 21 Countries	OpenAIRE UsageCounts To track usage events and evaluate impact of usage activity 200 Repositories 185M+ Downloads 600M+ Views
OPENSOURCE OBSERVATORY To track Open Access policies and activities at national level 44 Countries	CC OpenCitations To track open bibliographic and citation data 1,284,283,603 Citations
zenodo To publish and discover open research 2.5M+ Research outputs	open@PC To track open access publishing costs 341 Repositories 731+ Journals 8.7K+ Articles
EPISciences To publish in diamond open access journals 24 Journals	<h1>OPEN</h1>
Open Access Journals To aggregate journal databases with gateways and publish them in Open Access 2955 Open Access Journals	

FAIR 2. FAIRness

EOSC resources align with FAIR principles

For end-users	For service providers
EXPLORE To discover open access versions of research resources 155,995,360 Open Access content	ScholarXplorer To make links between different research outputs 282,288,774 Open Access content
argos To create and publish datasets 700+ Datasets	Metabola Explorer To explore the richness of metadata 82 Countries
CONNECT To enhance all content for research purposes from one single entry point 12 Countries	OpenAIRE SICKBOX To search metadata 23,997,326 Open Access content
DOI To publish and manage FAIR metadata and research outputs 6710 Countries	CC OpenCitations To track maximum share of citation data 72,268,850 Citations
EPISciences To publish scientific publications and their metadata 4779 Published articles	open@PC To provide research open access fees 147K+ Journals € 282M+ Revenue
OpenAIRE UsageCounts To provide standardized open usage metrics 4,371,166 Open Access content	
Open Access Journals To track richness of scientific journals 13 Journals 7 Countries 7 Journals	
OPENSOURCE OBSERVATORY To track Research data Management and FAIR policies and activities at national level 44 Countries	

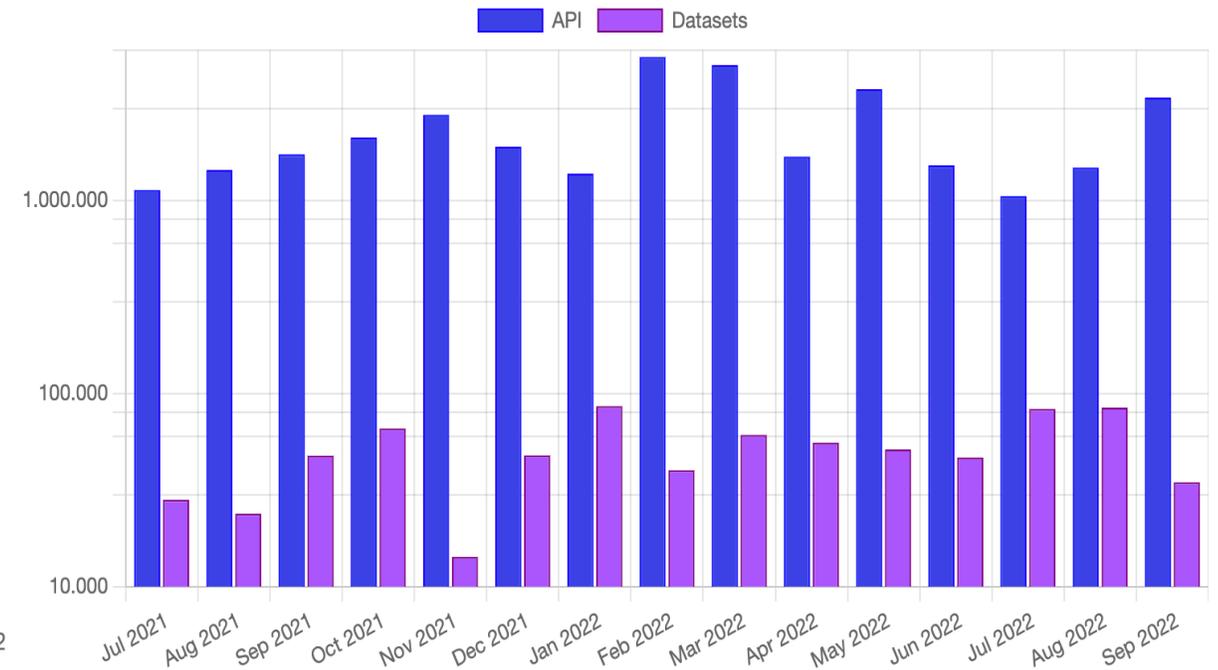
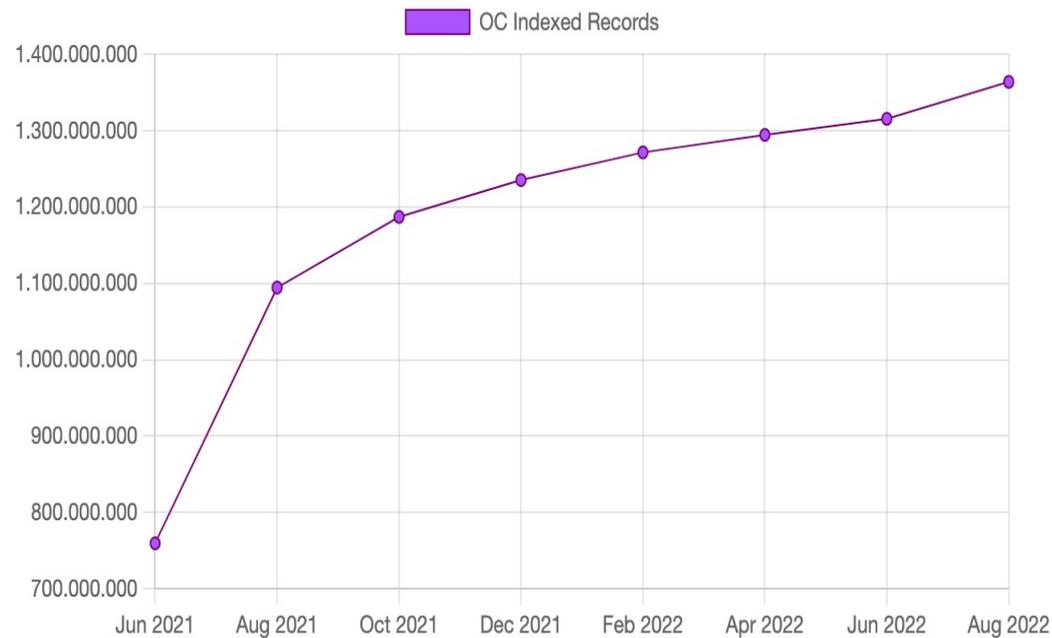
3. EOSC architecture

OpenAIRE services used at the Core of EOSC



For end-users	For service providers
OPENSOURCE OBSERVATORY To track Open Science Maturity in the EU open findability accessibility interoperability reusability	OpenAIRE ResearchGate To explore EOSC research outputs, resources and relationships open findability accessibility interoperability reusability
OpenPlato To foster modular Open Science training for capacity building open findability accessibility interoperability reusability	PROVIDE To enable interoperability of metadata across content providers open interoperability
	AAI To authenticate and authorize users to access the EOSC Portal open accessibility

Usage KPIs



- On September 2022, 12 different Tokens made API requests – with an average of 1414 calls per token
- Currently we have registered 36 different tokens

Future Steps

Learn more about current/future activities of OpenCitations on its public roadmap at <https://opencitations.hypotheses.org/1519>

OpenCitations Meta

Enables to store in house bibliographic metadata for the citing and cited publications involved in all our indexed citations

Advantages:

- Index citations involving entities lacking DOIs
- Improving the performance of some API operations
- Enables text search (e.g. in titles)

Expected by the
end of 2022

COCI, NOCI and DOCI

A new release of COCI based on the October dump of Crossref

A coverage expansion with the integration of new citation indexes of open citations, based on the holdings of DataCite (DOCI) and the NIH Open Citation Collection (NOCI)

TODO – SOME NUMBERS FROM ARIANNA

Expected by the
end of 2022

OpenCitations badge

A badge widget to be integrated in any webpage by simply adding a precise HTML tag. The hosting service can decide what kind of information to display (mainly number of citations)



Citations
163

Expected by the
end of 2022

Querying Data

<https://opencitations.net/querying>

SPARQL endpoints

OpenCitations maintains a SPARQL endpoint for all its datasets. When such a SPARQL endpoint is accessed with a browser, it shows an editor GUI generated with YASGUI

Requires technical background on semantic web technologies and SPARQL language

OpenCitations Indexes SPARQL endpoint

```
1 PREFIX cito:<http://purl.org/spar/cito/>
2
3 SELECT DISTINCT ?citing_entity ?cited_entity ?creation_date ?timespan WHERE {
4   GRAPH <https://w3id.org/oc/index/coci/> {
5     ?citation a cito:Citation ;
6             cito:hasCitingEntity ?citing_entity ;
7             cito:hasCitationCreationDate ?creation_date ;
8             cito:hasCitationTimeSpan ?timespan ;
9             cito:hasCitedEntity ?cited_entity
10    }
11 } LIMIT 10
```



Raw Response

Table

Pivot Table

Google Chart

Search:

Show 50



entries

Search Interfaces

User-friendly text search interfaces that can be used to search data in all the OpenCitations datasets and filter the outcoming results.

Developed using OSCAR,
The OpenCitations RDF
Search Application

(<https://github.com/openCitations/oscar>)

Limit to 36/36 results

< Fewer More >

All Show only Exclude

Select Creation ^

Select Timespan ... ^

Number of rows per page: 10 Export results Sort: None ^

OCI	Citing DOI	Citing reference	Cited DOI	Cited reference	Creation	Timespan (days)
020010009	10.1097/pai.	Kim, K., & Lee, H. W.	10.1186/175	Liu, L., Wu, N., & Li,	2017-05	1790
073625101	00000000000	(2018). Expression of	6-8722-5-31	J. (2012). Novel		
837000000	00524	Phosphoinositide 3-		targeted agents for		
000000000		Kinase p110 α and		gastric cancer. In		
000000000		p110 β Subunits and		Journal of		
00050204-		PIK3CA Mutation in		Hematology &		
020010108		Patients With		Oncology (Vol. 5,		
063601070		Advanced Gastric		Issue 1). Springer		
506630807		Carcinoma. In		Science and Business		
020263056		Applied		Media LLC.		
30301		Immunohistochemistr		https://doi.org/10.11		
		y & Molecular		86/1756-8722-5-31		
		Morphology (Vol. 26,				
		Issue 10, pp. 740-748).				
		Ovid Technologies				
		(Wolters Kluwer				
		Health).				
		https://doi.org/10.10				
		97/pai.0000000000000				
		524				

REST APIs

Data in any of the OpenCitations datasets can be retrieved by using an HTTP REST API. This solution represents a convenient access to the data included in the OpenCitations datasets for Web developers and users who are not necessarily experts in Semantic Web technologies.

Implemented by means of
RAMOSE, the Restful API
Manager Over SPARQL
Endpoints

(<https://github.com/opencitations/ramose>)

The unifying REST
API for all the
OpenCitations
Indexes

DESCRIPTION

PARAMETERS

OPERATIONS

[/references/{doi}](#)

[/citations/{doi}](#)

[/citation/{oci}](#)

[/metadata/{dois}](#)

[/citation-count/{doi}](#)

[/reference-count/{doi}](#)

HOME

THE UNIFYING REST API FOR ALL THE OPENCITATIONS INDEXES

VERSION: Version 1.1.0 (2020-03-25)

API URL: <https://w3id.org/oc/index/api/v1>

CONTACT: contact@opencitations.net

LICENSE: This document is licensed with a [Creative Commons Attribution 4.0 International License](#), while the REST API itself has been created using [RAMOSE](#), the *Restful API Manager Over SPARQL Endpoints* created by [Silvio Peroni](#), which is licensed with an [ISC license](#).

DESCRIPTION

[↑ BACK TO TOP](#)

This document describe the REST API for accessing the data stored in all the [OpenCitations Indexes](#) hosted by [OpenCitations](#). This API implements operations to retrieve the citation data for all the references to other works appearing in a particular bibliographic entity, or the citation data for all the references appearing in other works to a particular bibliographic entity, given the DOI of a bibliographic entity, or to retrieve citation data about a particular citation identified by means of its [Open Citation Identifier \(OCI\)](#).

OpenCitations access token

Before accessing our services, we encourage getting the OpenCitations Access Token.

An opaque string that anonymously identifies a unique user of the OpenCitations' APIs.

Its usage is not compulsory; however, it will help OpenCitations incredibly, by enabling us to monitor the number of the unique users accessing our data and services.

OpenCitations Access Token

Email:

Get token



Non sono un robot



reCAPTCHA
Privacy - Termini

<https://opencitations.net/accesstoken>



THANKS

Email

Ivan.heibi2@unibo.it

Twitter

@ivanhb
