

# OpenCitations

## A Short Introduction

ULITE workshop 2022  
Cologne, Germany  
June 23, 2022

Silvio Peroni

Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy  
[silvio.peroni@unibo.it](mailto:silvio.peroni@unibo.it) – [@essepuntato](https://www.unibo.it/sitoweb/silvio.peroni/en) – <https://www.unibo.it/sitoweb/silvio.peroni/en>

Director of OpenCitations

[silvio.peroni@opencitations.net](mailto:silvio.peroni@opencitations.net) – [@opencitations](https://www.opencitations.net) – [http://opencitations.net](https://www.opencitations.net)



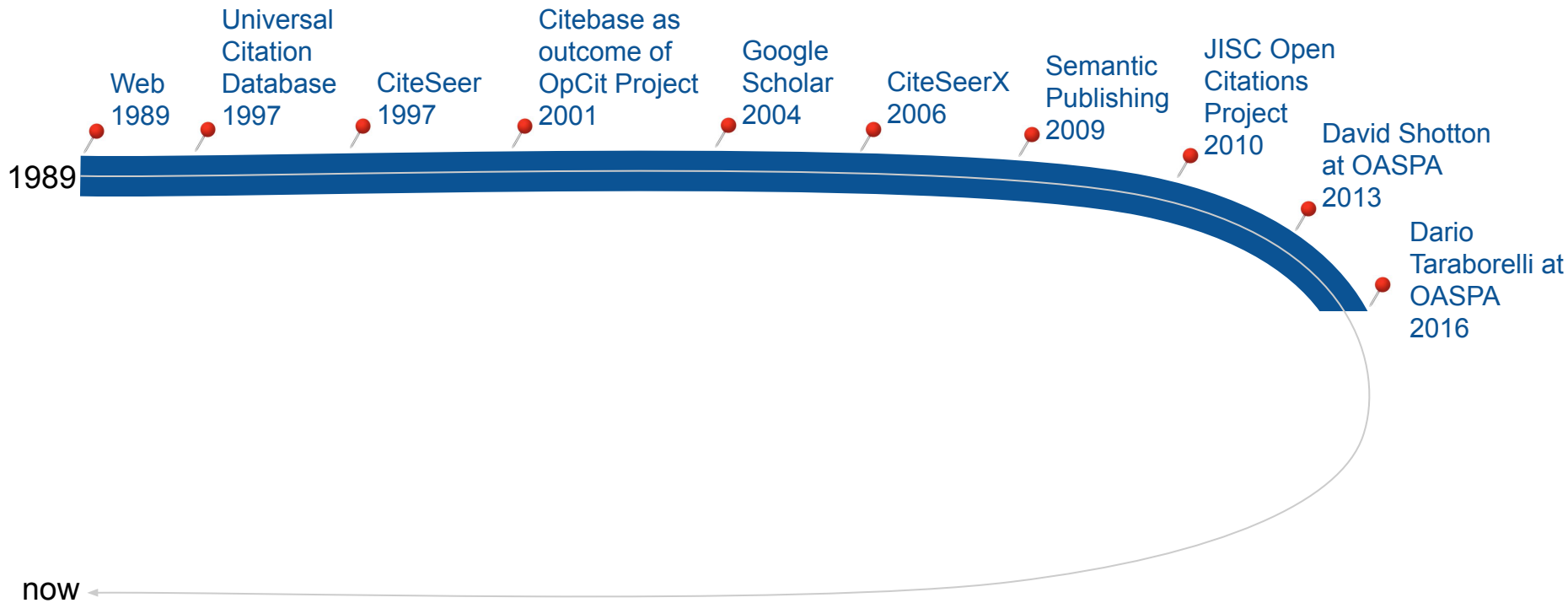
RESEARCH CENTRE  
FOR OPEN SCHOLARLY METADATA



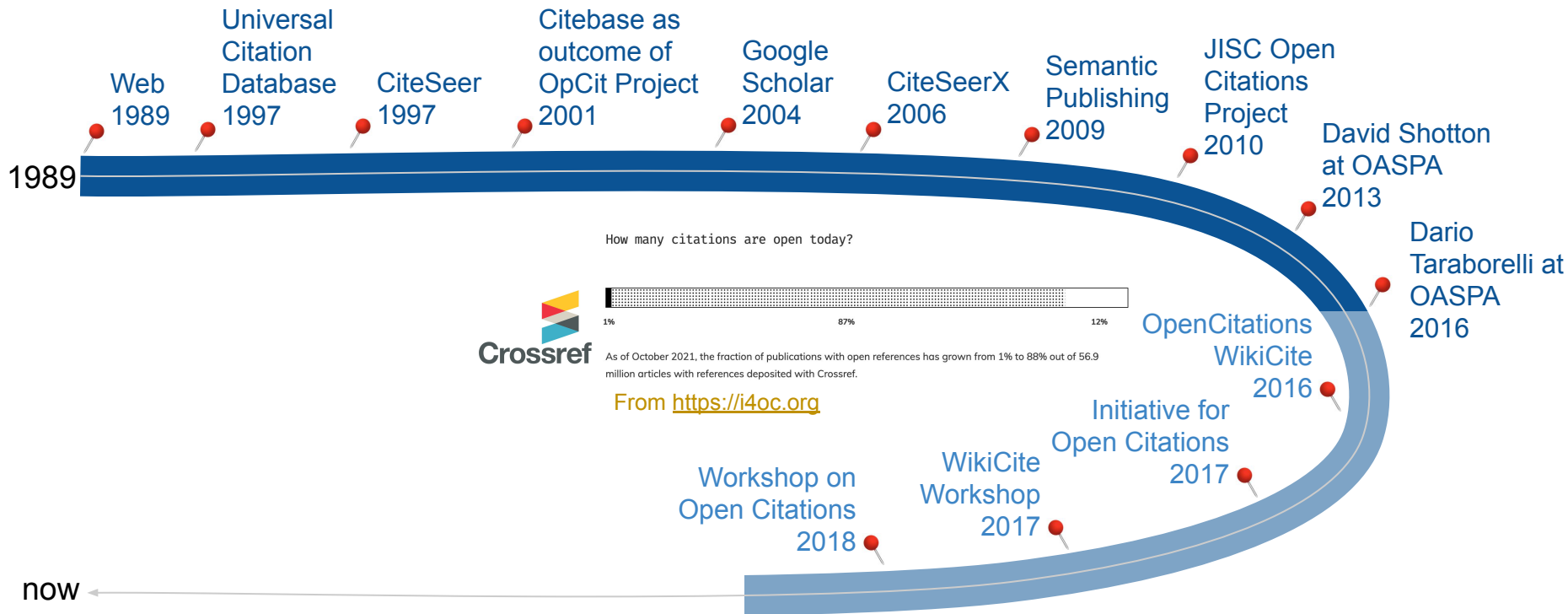
ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA  
DEPARTMENT OF CLASSICAL PHILOLOGY  
AND ITALIAN STUDIES



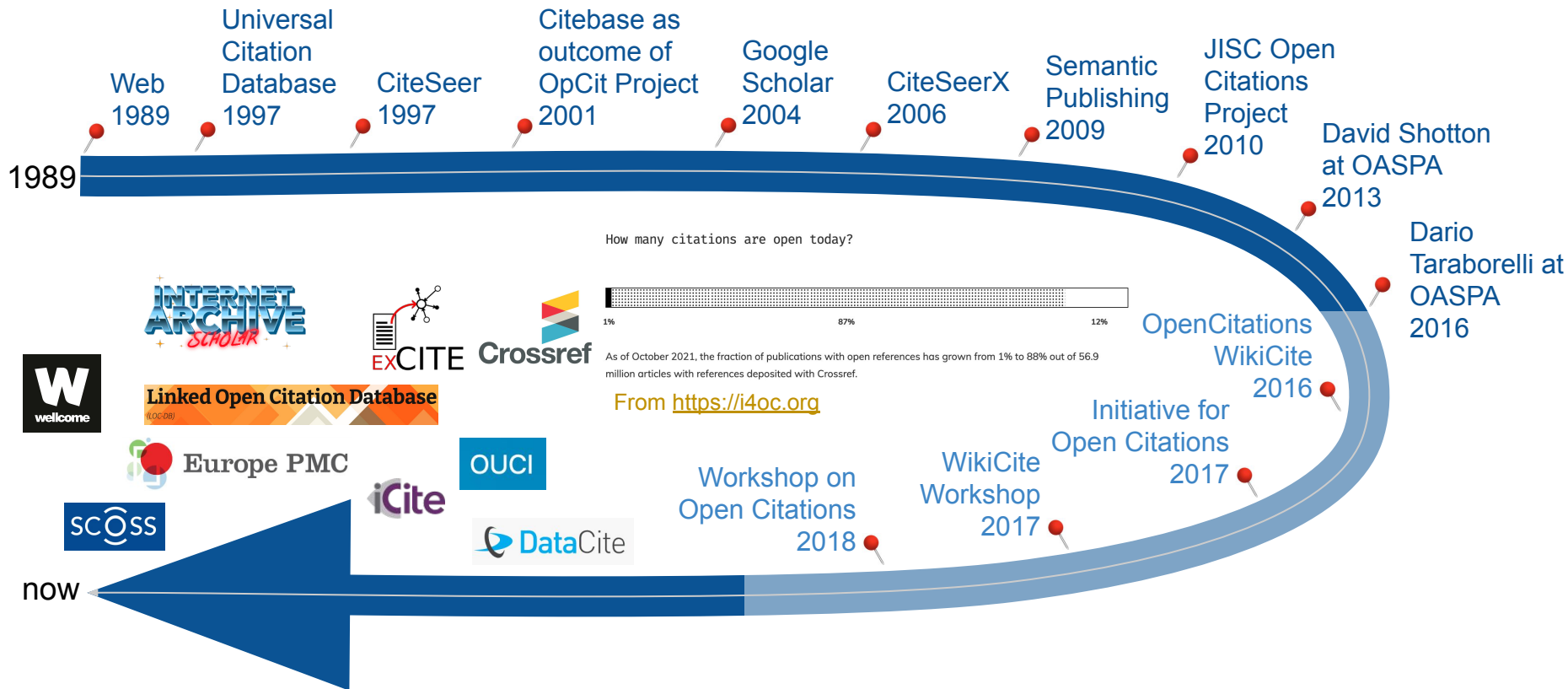
# The Web era



# The turning point



# What happened next



# What is an *open* citation



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:

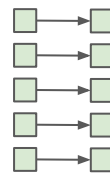
- the *representation* of such a conceptual directional link
- the *basic metadata* of the citing entity and the cited entity, i.e. sufficient information to create or retrieve textual bibliographic references

A citation is **open** when its data are compliant with the following characteristics:

structured

```
"reference": [{
  "issue": "2",
  "key": "10.7717/peerj.4375/ref-11",
  "doi-asserted-by": "crossref",
  "first-page": "237",
  "DOI": "10.1002/asi.22963",
  "article-title": "Anatomy of green open access",
  "volume": "65",
  "author": "Björk",
  "year": "2014",
  "journal-title": "Journal of the Association for
},
...
```

separate



open



without  
restrictions to  
maximise reuse

identifiable

PID

available

http://

+

PID

=

metadata

S. Peroni & D. Shotton (2018).

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

## Academic publishing



Now is the time to work together toward open infrastructures for scholarly metadata

Publishers should commit to making **complete metadata** for all their works, including **references** and abstracts, available in a suitable open infrastructures such as Crossref or DataCite

Publishers that do not yet support the Initiative for Open Citations (I4OC) and the Initiative for Open Abstracts (I4OA) should **join these initiatives**

Likewise, preprint servers and institutional repositories should also **deposit their metadata to open infrastructures**

Ginny Hendricks, Bianca Kramer, Catriona Maccallum, Paolo Manghi, Cameron Neylon, Silvio Peroni, David Shotton, Aaron Tay, Ludo Waltman (2021). Now is the time to work together toward open infrastructures for scholarly metadata. LSE Impact Blog.  
<https://blogs.lse.ac.uk/impactofsocialsciences/2021/10/27/now-is-the-time-to-work-together-toward-open-infrastructures-for-scholarly-metadata/>

## Amendments to membership terms to open reference distribution and include UK jurisdiction



[Ginny Hendricks](#), [Martyn Rittman](#), [Amanda Bartell](#) – 2022 April 04  
In [References](#), [Metadata](#), [Board](#), [Strategy](#)

Since the 3rd June 2022, [all reference lists](#) that are deposited in Crossref [are open](#) and available through their API

Note: depositing references is necessary in order to retrieve citation links from their members-only Cited-by API

Ginny Hendricks (2022). Amendments to membership terms to open reference distribution and include UK jurisdiction. Crossref Blog.

<https://www.crossref.org/blog/amendments-to-membership-terms-to-open-reference-distribution-and-include-uk-jurisdiction/>

# What is OpenCitations

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

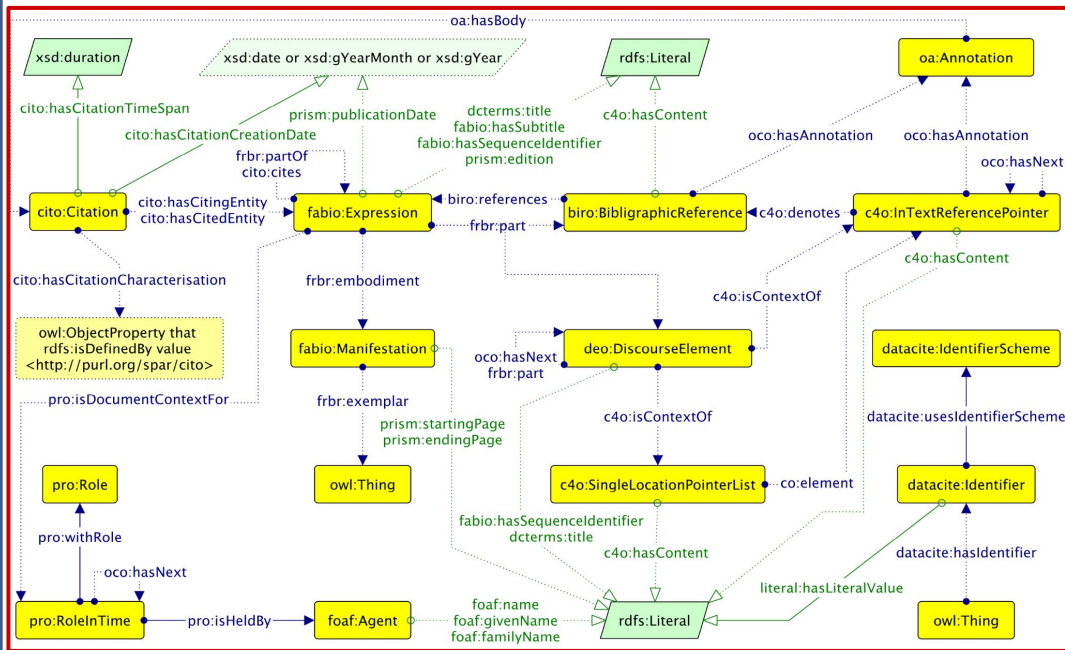
# Our values

OpenCitations espouses the following:

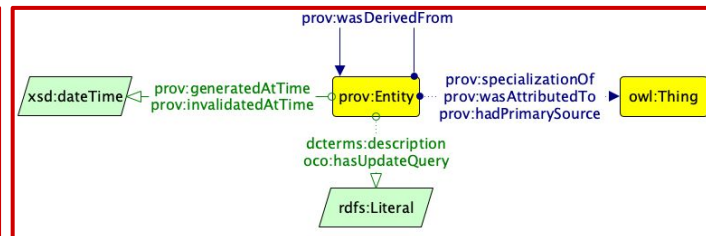
- UNESCO [principles of Open Science](#)
- [POSI principles](#)
- The [FAIR data principles](#), that data should be Findable, Accessible, Interoperable, and Reusable
- The [I4OC principles](#) that citation data should be Structured, Separable, and Open

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

# The OpenCitations Data Model



<https://opencitations.net/model>  
<https://w3id.org/oc/ontology>



provenance

It is based on  
SPAR Ontologies  
<http://www.sparontologies.net>



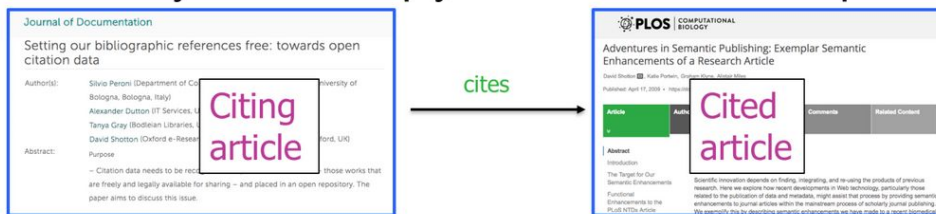
+

Other ontological models including as FOAF, DCTerms, PROV-O

data

# The citations available are treated as first-class data entities, with accompanying properties including the citations timespan, modeled according to the **OpenCitations Data Model**

Citations are normally treated simply as the links between published entities







Alternative richer view is to regard a citation as a data entity in its own right



Additional metadata: citation creation date, timespan, self-citation

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>

# Enabling trust: change tracking

-  adding data
-  removing data
-  remaining data
-  delta:  
data added & removed



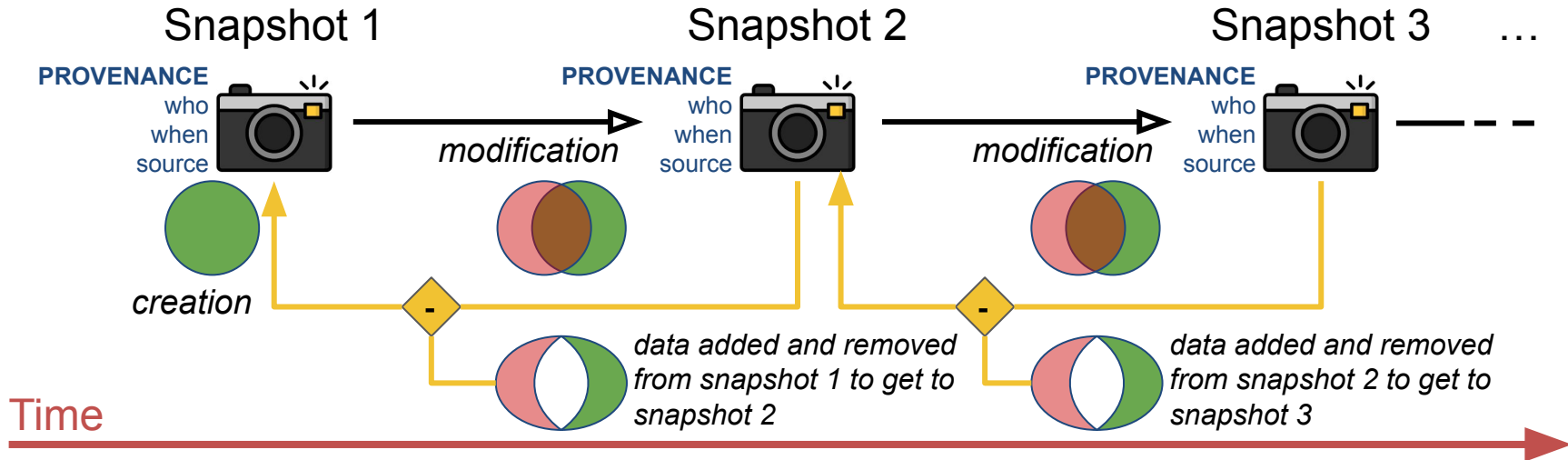
snapshot: all the data having such an entity as subject at a given time



An entity

(e.g. a journal article)

To keep track of the history of an entity we need to store all the data of its last snapshot plus all the deltas built by modifying the previous snapshots



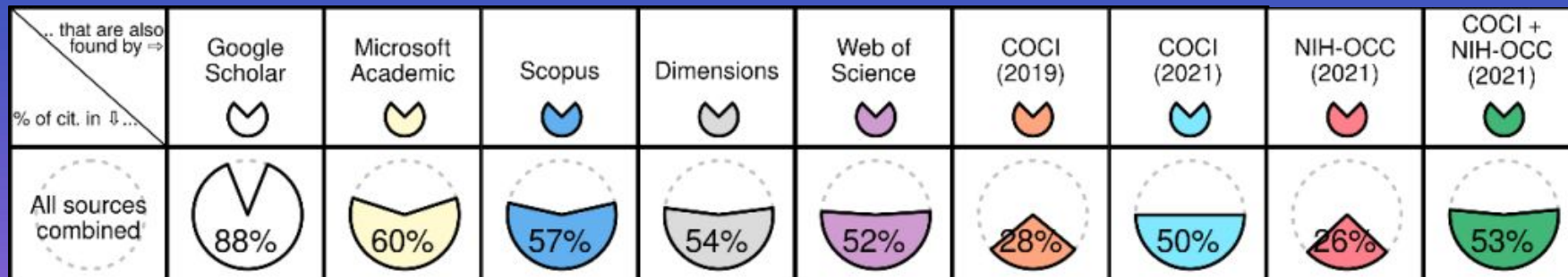
# OpenCitations Indexes

OpenCitations Indexes (<http://opencitations.net/index>), including COCI (launched in July 2018), the OpenCitations Index of Crossref open DOI-to-DOI citations, currently contains more than

**1.31 billion citations** between **73 million** bibliographic resources

COCI data and services	URL
REST API <i>results in CSV and JSON</i>	<a href="https://w3id.org/oc/index/coci/api/v1">https://w3id.org/oc/index/coci/api/v1</a>
SPARQL endpoint <i>results in several formats</i>	<a href="https://w3id.org/oc/index/sparql">https://w3id.org/oc/index/sparql</a>
Dumps <i>all data in CSV, JSON, RDF</i>	<a href="https://w3id.org/oc/download#coci">https://w3id.org/oc/download#coci</a>

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-approaches-parity-with-web-of-science-and-scopus/>

# 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, and more than 17 million accesses in 2022 so far

<https://inciteful.xyz/>

Inciteful Documentation Paper title, DOI, PubMed ID, arXiv ID, or Microsoft Academic ID

OpenCitations, an infrastructure organization for open scholarship

Silvio Peroni | David M. Shotton

OpenCitations is an infrastructure organization for open scholarship dedicated to the publication of open citation data as Linked Open 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 paper describes... [more](#)

16 Cited By 12 Citing 2020 Published Yes Open Access 8,183 Papers in Graph 29,947 Citations in Graph 2 Graph Depth


## OpenCitations, an infrastructure organization for open scholarship



Silvio Peroni  and David Shotton 


Posted Online January 23, 2020



[https://doi.org/10.1162/qss\\_a\\_00023](https://doi.org/10.1162/qss_a_00023)






© 2020 Silvio Peroni and David Shotton. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Favorite  Times Cited: 9

Download Citation  RSS TOC 

RSS Citation 

 Download Options 

Article [https://doi.org/10.1162/qss\\_a\\_00023](https://doi.org/10.1162/qss_a_00023) visualised with the Open Access Helper plugin (<https://www.oahelper.org/>)

<https://dblp.uni-trier.de/rec/conf/semweb/PeroniS18.html>



"The SPAR Ontologies." Silvio Peroni, David M. Shotton (2018)

> Home

   Silvio Peroni, David M. Shotton: The SPAR Ontologies. International Semantic Web Conference (2) 2018: 119-136

[-] References

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

   Simon Buckingham Shum, Enrico Motta, John Domingue: ScholOnto: an ontology-based digital library server for research documents and discourse. Int. J. on Digital Libraries 3(3): 237-248 (2000)


   Paolo Ciccarese, Silvio Peroni: The Collections Ontology: Creating and handling collections in OWL 2 DL frameworks. Semantic Web 5(6): 515-529 (2014)

N. van Eck and L. Waltman (2019). VOSviewer supports large number of new data sources. <https://www.cwts.nl/blog?article=n-r2v284>

<https://twitter.com/TheLensOrg/status/1225004057086967808>

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](https://lens.org)

 LENS.ORG

Release 6.6.0 New citing patents chart for individual scholarly works, new first or last author filter toggle and more!

Filters: Scholarly Works (10,000) | Filters: No items applied

Author: [input field]

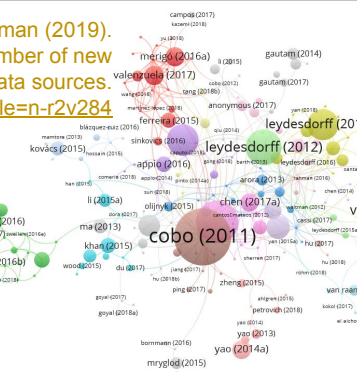
Citing Patents Citing Scholarly Works

In: Science, Issue: 4254, Volume: 193, Page: 673-675

2 26 39

The Lens @TheLensOrg

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



# Working with the community

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”

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](http://www.scoss.org).

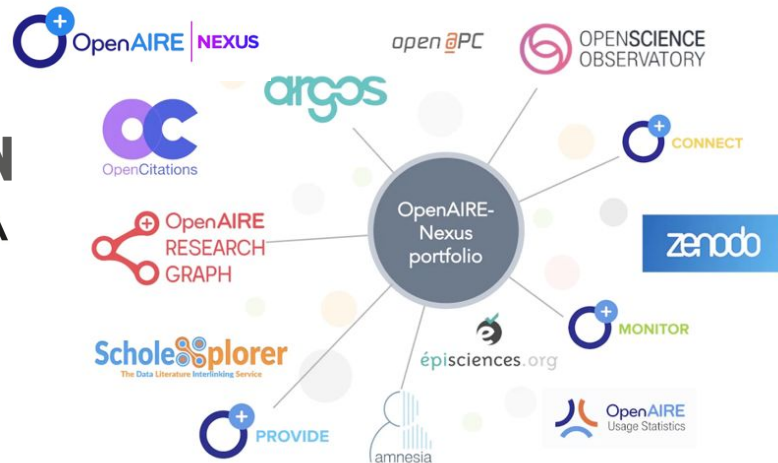
SCOSS

On our blog:

<https://opencitations.wordpress.com/2022/01/13/five-reasons-why-2021-has-been-a-great-year-for-opencitations/>  
<https://opencitations.wordpress.com/2022/03/08/opencitations-and-ec-funding-openscience-nexus-and-risis2/>  
<https://opencitations.wordpress.com/2022/05/31/strongtwo-years-of-achievements-within-the-scoss-family-and-its-not-over-yet-strong/>



**RISIS**  
RESEARCH INFRASTRUCTURE FOR SCIENCE  
AND INNOVATION POLICY STUDIES



# 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

# People and community support

## Directors

David  
Shotton



Silvio  
Peroni



## Administration and communication

Claudio  
Fabbri



Chiara  
Di Giambattista



Arcangelo  
Massari



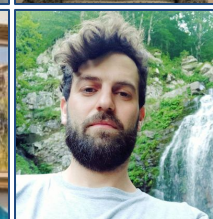
Giuseppe  
Grieco



## Development



Arianna  
Moretti



Ivan  
Heibi

### OpenCitations Members

#### Strategic members

- Scientific Interest Group of the French National Fund for Open Science (since 2020)
  - Centre National de la Recherche Scientifique (CNRS)
  - French Ministry of Higher education, research and innovation
  - French Conférence des présidents d'université (CPU)
  - Conférence des directeurs des écoles françaises d'ingénierie (CEDEFI)
  - French Conférence des grandes écoles (CGE)
  - Coordination des universités de recherche intensive françaises (CURIF)
  - Institut national de la recherche agronomique (INRA)
  - Institut national de recherche en informatique et en automatique (INRIA)
  - Institut national de la santé et de la recherche médicale (INSERM)
  - Commissariat à l'énergie atomique et aux énergies alternatives (CEA)
  - Consortium unifié des établissements universitaires et de recherche pour l'accès aux publications numériques (Couperin)
  - Agence nationale de la recherche (ANR)

#### Supporting members

- AgroParisTech (since 2020)
- European University Institute (since 2021)
- FinELib Consortium
  - Lapland University Consortium Library (since 2021)
  - Tampere University including Tampere University Hospital (since 2021)
  - University of Helsinki including Helsinki University Central Hospital (since 2021)
  - Åbo Akademi University Library (since 2022)
- Katholieke Universiteit Leuven (since 2020)
- Library and Information Centre of the Hungarian Academy of Sciences (since 2021)
- Spanish National Research Council (since 2021)
- Universitäts- und Landesbibliothek Darmstadt (since 2021)
- University of Management and Technology, Lahore, Pakistan (since 2021)

#### OpenCitations Supporters

- California Digital Library (since 2021)
- Canadian Research Knowledge Network
  - Memorial University of Newfoundland (since 2020)
  - University of New Brunswick (since 2020)
  - Université de Montréal (since 2020)
  - Brock University (since 2020)
  - Ryerson University (since 2020)
  - University of Ottawa (since 2020)
  - University of Toronto (since 2020)
  - University of Waterloo (since 2020)
  - University of Windsor (since 2020)
  - MacEwan University (since 2020)
  - University of Victoria (since 2020)
- Council of Australian University Librarians
  - Curtin University (since 2020)
  - Griffith University (since 2020)
  - Queensland University of Technology (since 2020)
  - University of Canterbury (since 2020)
  - University of New South Wales (since 2020)
- Institut Français de Recherche pour l'Exploitation de la Mer (since 2021)
- Royal Danish Library (since 2020)
- SLSP AG Konsortium der SchweizerHochschulbibliotheken
  - CERN (since 2021)
  - EPF Lausanne (since 2021)
  - ETH Zürich (since 2021)
  - HSLU (since 2021)
  - Schweizerischer Nationalfonds (since 2021)
  - Universität Bern (since 2021)
  - Universität St Gallen (since 2021)
  - Universität Zürich (since 2021)
  - Université de Fribourg (since 2021)
  - Université de Neuchâtel (since 2021)
  - Zürcher Hochschule der Künste (since 2021)
  - Université Paris Lodron (since 2021)



**Thank you  
for your attention**