

# The Open Citations Movement

## The Story So Far

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://opencitations.net) – <http://opencitations.net>



The 5th Conference on Scholarly Publishing in the  
Context of Open Science – PUBMET 2018  
20-21 September 2018, Zadar, Croatia



DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA

# Summary

open citation

Initiative for Open Citations (I4OC)

OpenCitations

open citation

# What is an open citation?

A **bibliographic citation** is 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 for each of the entities

A bibliographic citation is an **open citation** when the data needed to define the citation are compliant with the following principles

# Structured

Citation data must be expressed in one or more **machine-readable formats**

## Unstructured REFERENCES

**Björk BC, Laakso M, Welling P, Paetau P. 2014.** Anatomy of green open access. *Journal of the Association for Information Science and Technology* **65(2)**:237–250.

## Structured (JSON)

```
"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".
```

## Structured (RDF)

```
coci:020070701073625141427193704030705-0200100000236102818370202090603  
a cito:Citation ;  
rdfs:label "oci:020070701073625141427193704030705-0200100000236102818370202090603" ;  
cito:hasCitationCreationDate "2018-02-13"^^xsd:date ;  
cito:hasCitationTimeSpan "P4Y3M7D"^^xsd:duration ;  
cito:hasCitedEntity <http://dx.doi.org/10.1002/asi.22963> ;  
cito:hasCitingEntity <http://dx.doi.org/10.7717/peerj.4375> .
```

ology"

# Separate

Citation data must be available without the need to access the source bibliographic entity (e.g. the article or book) in which the citation is defined

## joined

✓ PEER-REVIEWED



View 433 tweets

Related research

### The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles

Research article Legal Issues Science Policy Data Science

Heather Pwower<sup>1</sup>, Jason Priem<sup>1</sup>, Vincent Larivière<sup>2,3</sup>, Juan Pablo Alperin<sup>4,5</sup>, Lisa Matthias<sup>6</sup>, Bree Norlander<sup>7,8</sup>, Ashley Farley<sup>7,8</sup>, Jevin West<sup>7,8</sup>, Stefanie Haustein<sup>4,9</sup>

Published February 13, 2018

Note that a Preprint of this article also exists, first published August 2, 2017.

PubMed 29456894

Author and article information

Abstract

Main article text

#### Introduction

The movement to provide open access (OA) to all research literature is now over fifteen years old. In the last few years, several developments suggest

...

References

Anderson. 2017a. When the wolf finally arrives: big deal cancellations in North American Libraries. The Scholarly Kitchen. <https://scholarlykitchen.sspnet.org/2017/0...> (accessed 09 January 2018)

Anderson. 2017b. The forbidden forecast: thinking about open access and library subscriptions. The Scholarly Kitchen. <https://scholarlykitchen.sspnet.org/2017/0...> (accessed 15 July 2017)

Antelman K. 2017. Leveraging the growth of open access in library collection decision making. In: Proceeding from ACRL 2017: at the helm: leading transformation.

Archambault É, Amyot D, Deschamps P, Nicol A, Provencher F, Rebout L, Roberge G. 2013. Proportion of open access peer-reviewed papers at the European and world levels-2004-2011. European Commission, Brussels

Archambault É, Amyot D, Deschamps P, Nicol AF, Provencher F, Rebout L, Roberge G. 2014. Proportion of open access papers published in peer-reviewed journals at the European and world levels-1996-2013. European Commission

Archambault É, Côté G, Struck B, Voorons M. 2016. Research impact of paywalled versus open access papers.

## separate

<https://w3id.org/oc/index/coci/api/v1/references/10.7717/peerj.4375/>

```
{
  {
    "oci": "020070701073625141427193704030705-0200100000236102818370202080506",
    "cited": "10.1002/asi.22856",
    "timespan": "P4Y8M21D",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-0200100000236102818370202090603",
    "cited": "10.1002/asi.22963",
    "timespan": "P4Y3M7D",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-0200100000236211410253701000201",
    "cited": "10.1002/leap.1021",
    "timespan": "P1Y1M12D",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-0200100000236211410253701000906",
    "cited": "10.1002/leap.1096",
    "timespan": "P1Y6D",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-020010000073628010109026300010563010706056305",
    "cited": "10.1007/s11192-015-1765-5",
    "timespan": "P2Y3M25D",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-020010000073628010109026300010663020202056306",
    "cited": "10.1007/s11192-016-2225-6",
    "timespan": "P1Y1M10D",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-020010000073628010109026300010763020209016304",
    "cited": "10.1016/j.joi.2007.04.001",
    "timespan": "P11M28D",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-02001000106361937192418370200000737000437000001",
    "cited": "10.1016/j.joi.2007.04.001",
    "timespan": "P10Y7M",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  {
    "oci": "020070701073625141427193704030705-02001000106361937192418370200010537010137000008",
    "cited": "10.1016/j.joi.2015.11.008",
    "timespan": "P2Y",
    "creation": "2018-02-13",
    "citing": "10.7717/peerj.4375"
  },
  ...
}
```

# Open

Citation data must be freely accessible and reusable **without restrictions**, for example by publication under the [CC0 1.0 Universal waiver/license](#)

## close

"databases pricing is a complex issue which depends on the size of the organization, discounts which they negotiated and other elements as well [...] the estimation of **WOS costs is about \$100,000 per year** for large organizations [...] the cost of Scopus database is about 85-95% of the cost of WOS for the same organizations [...] **Scopus pricing is set according to annual subscription fee** with unlimited usage"

Chadegani et al. (2017). A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases. Asian Social Science, 9 (5). DOI: <https://doi.org/10.5539/ass.v9n5p18>

## open

"Crossref asserts **no claims of ownership** to individual items of bibliographic metadata and associated Digital Object Identifiers (DOIs) acquired through the use of the Crossref Free Services"

<https://github.com/CrossRef/rest-api-doc>

"The data held in the OpenCitations Corpus is made freely available under a **Creative Commons public domain dedication (CC0)**"

<http://opencitations.net/licenses>

"The content of Wikidata is available under a **free license**"

<https://www.wikidata.org>

# Identifiable and Available

The citing and cited entities must be clearly **identified** by using a specific persistent identifier scheme (e.g. a DOI) or a URL, and...

... it must be possible, by resolving the identifiers of the citing and cited entities, to **obtain the basic metadata** of both the entities, sufficient to create or retrieve textual bibliographic references for each of them

(Such basic entity metadata **must also be** structured, separate and open)

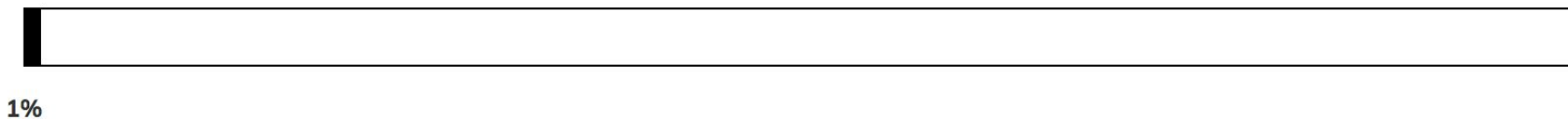


# Initiative for Open Citations (I4OC)

# How many citations were open before April 2017

Before April 2017, the fraction of publications with open references was 1% out of millions of articles with references deposited with Crossref

(Crossref is a non-for-profit official DOI Registration Agency launched in early 2000 that enables persistent cross-publisher citation linking in online academic journals. Publishers may deposit reference lists to Crossref, together with other metadata deposited for each work with a DOI)



Several publishers did not know they could release them as open data by means of the Crossref API – <https://api.crossref.org>

# April 2017: the Initiative for Open Citations (I4OC)

Goal: to persuade the major academic publishers to open their deposited references all at once

Key benefits (as listed in the website, <https://i4oc.org>):

- the establishment of a **global public web of linked scholarly citation data** to enhance the discoverability of published content, both non-OA and OA
- the ability to build new services over the open citation data, for the **benefit of publishers, researchers, funding agencies, academic institutions and the general public**, as well as enhancing existing services
- the creation of a public citation graph to explore connections between knowledge fields, and to **follow the evolution of ideas and scholarly disciplines**

# Founders and stakeholders

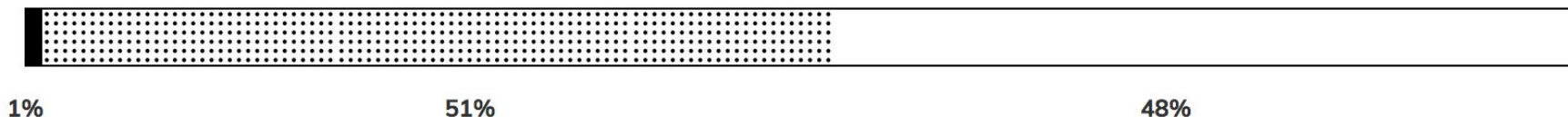
## founders



361 publishers participating (to date)

# How many citations are open now

As of September 2018, the fraction of publications with open references has grown from 1% to **52% out of 40.8 million articles** with references deposited with Crossref – and more than **500M citations** are now open



Several services have been developed using the citation data released so far

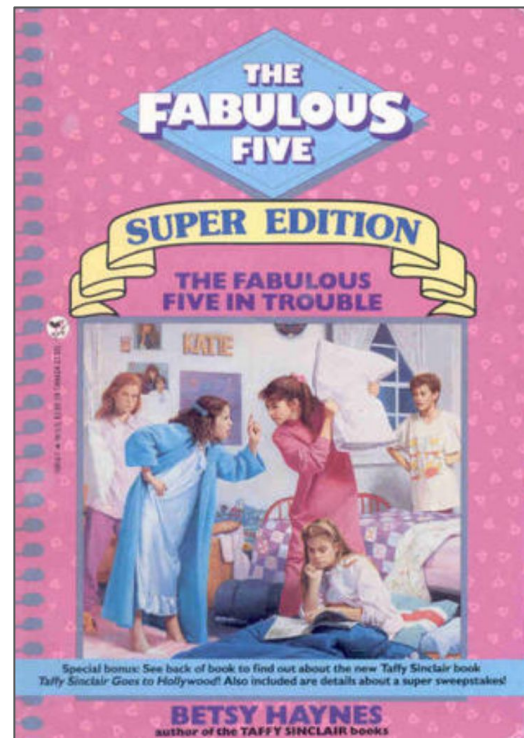
See the **Workshop on Open Citations** website (<https://workshop-oc.github.io>) and the related Twitter hashtag ([#WOOC2018](#)) for the latest highlights from the community

# Who is missing

Just five publishers among the top 20 DOI depositors are not distributing open references (as of Sept. 2018):

- Elsevier
- Wolters Kluwer Health
- IEEE “dedicated to advancing technology for  
**the benefit of humanity**”
- IOP Publishing “to advance physics for  
**the benefit of all**”
- American Chemical Society “to advance the broader chemistry enterprise  
and its practitioners for **the benefit of Earth  
and its people**”

The last three are  
**scholarly societies...**



# Uptake

ISSI:

<http://www.issi-society.org/open-citations-letter/>

## Open citations: A letter from the scientometric community to publishers

December 5th, 2017

Openness is central to the research endeavor. It is essential for the reproducibility and appraisal of research, reduce miscommunication, equitable access to and participation in science. Yet, calls for openness in science are often met with initial resistance. Open print servers, open access repositories, and open data have been initially resisted, but eventually adopted without adverse effects on the ecosystem. The launch of the **Initiative for Open Citations** aims to overcome similar obstacles. This initiative has campaigned for scientists to make openly available the references found in articles from publishers, including most of the large ones, support them to open their references. However, the initiative still lacks the support of the large publishers.

David Shotton (2018). Nature 553 (129).  
DOI: <https://doi.org/10.1038/d41586-018-00104-7>

## Funders should mandate open citations



*All publishers must make bibliographic references free to access, analyse and reuse, argues David Shotton.*

David Shotton 



 [PDF version](#)

Over the past two decades, open access to journal articles and software and research data has changed from aspirational to commonplace. However, truly open scholarship also requires that bibliographic references be freely available for analysis and

French National Plan for Open Science:  
<http://bit.ly/FranceOpenScience>

## FIRST COMMITMENT: ACCESS TO PUBLICA



Open scientific publishing must become the standard approach as soon as possible. To drive this dynamic, research publications resulting from calls for projects that receive public funding must be disseminated through open access platforms, whether in journals or books or through an open public repository such as HAL.

To sustain these practices over time, the assessment system for researchers and research institutions must be updated to reflect the principles and practices of open science. Changes in the way researchers are assessed will seek to give greater weight to quality rather than quantity, as outlined in the **San Francisco Declaration on Research Assessment (DORA)** proposals and the **Leiden Manifesto** principles, and make better use of open citations, in keeping with the **Initiative for Open Citations (I4OC)**.

# Crossref participation report

“Participation reports give a clear picture for anyone to see the metadata Crossref has”

<https://www.crossref.org/participation/>

You can check whether your favourite publisher release open references – and, if it does not, please write it an email asking to release them!



<https://www.crossref.org/members/prep/>



# How

If you are a publisher that already submits article metadata to Crossref as a participant in their [Cited-by](#) service, opening your reference data can be achieved in a matter of days, either:

1. by contacting Crossref by e-mail ([support@crossref.org](mailto:support@crossref.org)), asking them to turn on reference distribution for all of the relevant DOI prefixes, or
2. by setting the `<reference_distribution_opt>` metadata element to "any" for each DOI deposit

If not already a participant in Cited-by, a Crossref member can register for this service free-of-charge

OpenCitations

# Back to 2010

**OpenCitations** formally started in 2010 as a one-year project funded by JISC (with a subsequent extension), with David Shotton as director, who at that time was working in the Department of Zoology at the University of Oxford

Goal: to publish open bibliographic citation information in RDF and to make citation links as easy to traverse as Web links

Main outcomes – full details [in the official blog](#):

- open repository of scholarly citation data named the **OpenCitations** Corpus
- [SPAR \(Semantic Publishing and Referencing\) Ontologies](#), for describing such citation data in RDF

# OpenCitations today

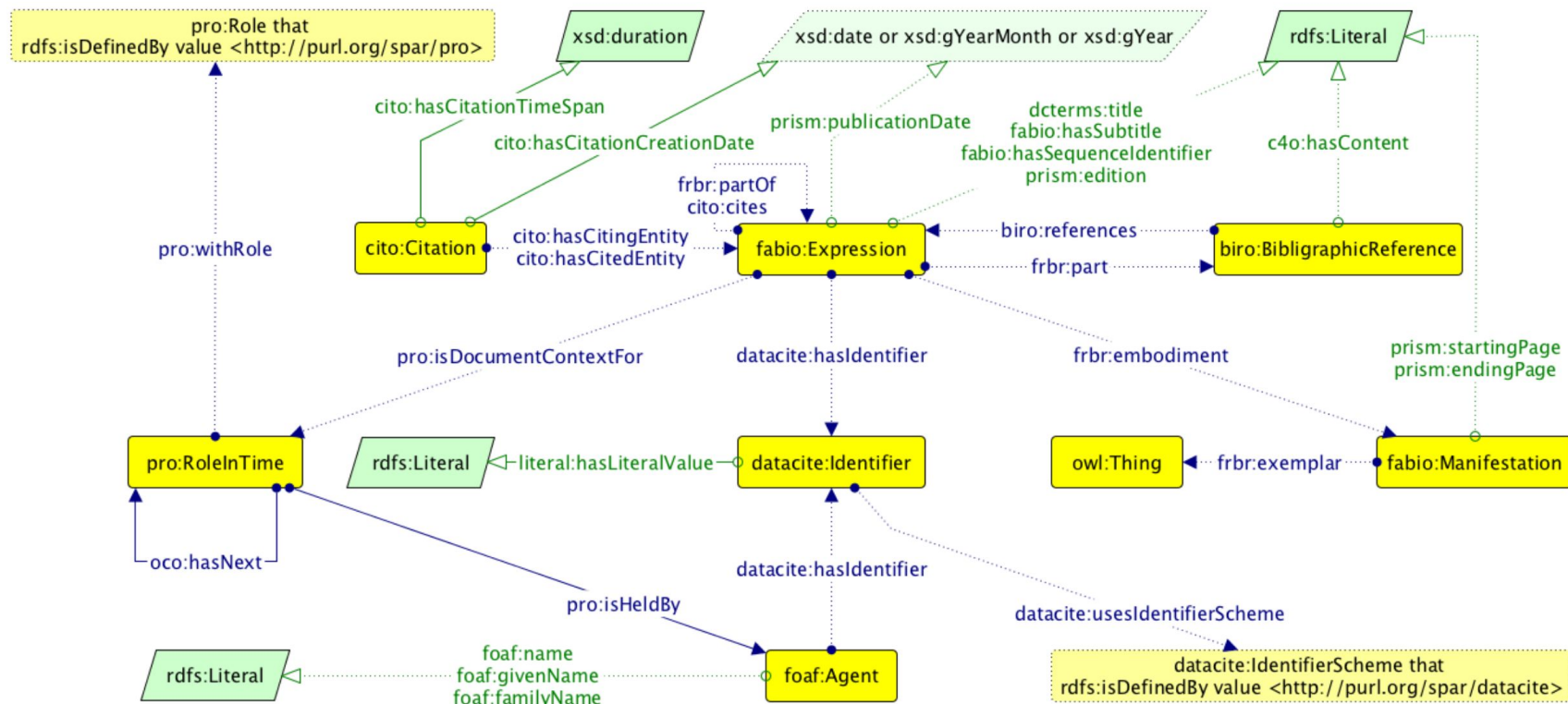
OpenCitations (<http://opencitations.net>) is a scholarly infrastructure organization

- 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**

It provides:

- data models: the [OpenCitations Data Model](#) (based on SPAR Ontologies)
- bibliographic and citation datasets: the [OpenCitations Corpus](#) and [COCI](#)
- software: [GitHub repository](#) released with open source licenses
- online services: [dumps](#), [REST APIs](#), [SPARQL endpoints](#), and [authoring](#) / [querying](#) / [browsing](#) interfaces

# OpenCitations Data Model



# About the datasets

**OpenCitations** Corpus (OCC, <https://w3id.org/oc/corpus>): new instance was set up at the University of Bologna in early July 2016, and currently contains **~13M citation links** to over **6.5M cited resources**

COCI (<https://w3id.org/oc/index/coci>): launched in July 2018, the **OpenCitations** Index of Crossref open DOI-to-DOI references contains **>316M citation links** between **~45M bibliographic resources**

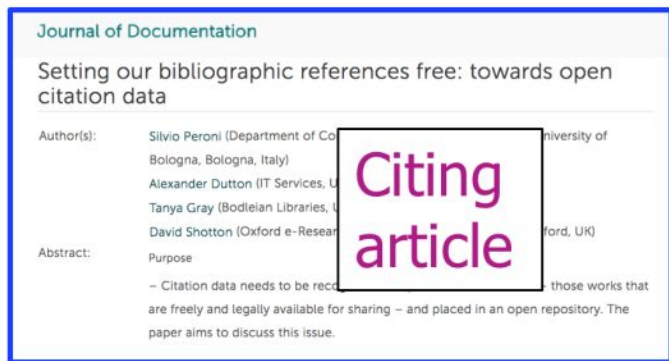
Service	OCC	COCI
REST API	<a href="https://w3id.org/oc/api/v1">https://w3id.org/oc/api/v1</a>	<a href="https://w3id.org/oc/index/coci/api/v1">https://w3id.org/oc/index/coci/api/v1</a>
SPARQL endpoint	<a href="https://w3id.org/oc/sparql">https://w3id.org/oc/sparql</a>	<a href="https://w3id.org/oc/index/coci/sparql">https://w3id.org/oc/index/coci/sparql</a>
Textual search	<a href="https://w3id.org/oc/search">https://w3id.org/oc/search</a>	<a href="https://w3id.org/oc/index/coci/search">https://w3id.org/oc/index/coci/search</a>

# Collaborations and adoptions

- **Wikidata**: several bibliographic entries have been aligned with OCC resources
- **OpenAIRE**: is importing OCC metadata about articles into their database
- Daniel Ecer and Lisa Knoll of eLife are performing **analytics** on the OCC data
- Ontotext demonstrated **SPARQL query federation** between Springer Nature LOD and OCC
- Anna Kamińska published a bibliometrics **case study of PLOS ONE** articles in OCC
- Daniel Himmelstein is processing OpenCitations data to create **DOI-to-DOI citation tables**
- Thiago Nunes and Daniel Schwabe are using OCC to exemplify their **XPlain framework**
- Antonina Dattolo and Marco Corbatto are using the OCC as source for **VisualBib framework**
- **LOC-DB**, **EXCITE**, the **Venice Scholar Index**, and **CitExCyr** adopt the OpenCitations Data Model
- Nees Jan van Eck and Ludo Waltman extended **VOSviewer** so as to use data in the OCC + COCI
- The **data of the EXCITE Project** has been just added to the OCC (+ ~1M citations)
- **Citation Gecko** uses the OCC + COCI for retrieving citation data about the papers
- Philipp Zumstein developed a **Zotero plugin** that gives information about open citations using COCI

# Citations as first-class data entities

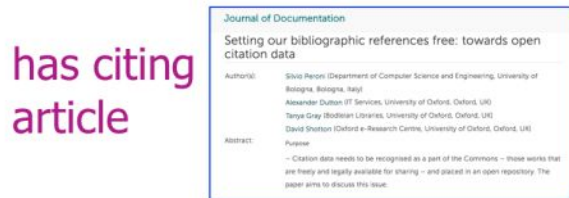
Citations are normally treated simply as the links between published entities



cites



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



**has cited article**

What we are pushing, particularly in COCI



# Open Citation Identifier (OCI)

New persistent identifier scheme for citations contained in bibliographic databases

Simple structure: `oci:number-number`, where “oci:” is the identifier prefix

Examples:

- `oci:2544384-7295288` (citation in the OCC)
- `oci:01027931310-01022252312` (citation in Wikidata, identified by “010” prefix)
- `oci:02001010806360107050663080702026306630509-02001010806360107050663080702026305630301` (citation in COCI, identified by “020” prefix)

OCI resolver: <https://w3id.org/oc/oci>

More info at <https://opencitations.wordpress.com/2018/03/12/citations-as-first-class-data-entities-open-citation-identifiers/>

# OpenCitations: people



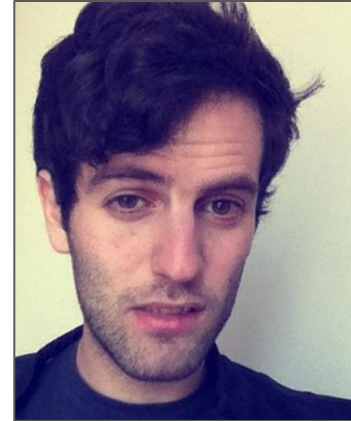
David  
Shotton

*Director*



Silvio  
Peroni

*Director*



Ivan  
Heibi

*Developer*

# International Advisory Board

We thank our OpenCitations International Advisory Board for their advice:

- Geoffrey Bilder, Crossref
- Jonathan Dugan, Berkeley Institute for Data Science
- Catriona MacCallum, Hindawi
- Johanna McEntyre, Europe PubMed Central
- Cassidy Sugimoto, Indiana University
- Dario Taraborelli, Wikimedia Foundation

# Acknowledgements



Alfred P. Sloan  
FOUNDATION



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA



DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA



# Thank you for your attention

## The Open Citations Movement: The Story So Far

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://opencitations.net) – [http://opencitations.net](https://opencitations.net)



The 5th Conference on Scholarly Publishing in the  
Context of Open Science – PUBMET 2018  
20-21 September 2018, Zadar, Croatia



DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA