Opening citation data to the public domain

Silvio Peroni

Digital Humanities Advanced Research Centre (DHARC), 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
open citations
–
OpenCitations
To cite

Oxford dictionary: “refer to (a passage, book, or author) as evidence for or justification of an argument or statement, especially in a scholarly work”

Isaac Newton (1675): “If I have seen further, it is by standing on the shoulders of giants”

Citations are unanimously recognised as crucial for knitting together our scientific and cultural knowledge
What is an open citation?

A bibliographic citation is a conceptual directional link from a citing entity to a 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 bibliographic citation is an open citation when the data needed to define the citation are: structured, separate, open, identifiable, available
Open citations: characteristics

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

https://doi.org/10.5539/ass.v9n5p18

“No claims of ownership to individual items of bibliographic metadata”
https://api.crossref.org

Available
E.g. HTTP + ID = metadata
https://api.crossref.org/works/10.7717/peerj.4375

Separate (e.g. via REST call to external services)

Open
https://api.crossref.org

Closed

Structured (JSON; machine readable)

Identifiable

Unstructured

Join	ed
Initiative for Open Citations (I4OC): [https://i4oc.org](https://i4oc.org)

Launched in April 2017 to persuade the major academic publishers to open their deposited references all at once

Founders:

As of February 2019, the fraction of publications with open references has grown from 1% to **55% out of 43.2 million articles** with references deposited with Crossref – and more than **500M citations** are now open
ACM joined I4OC since the very beginning (April 2017) and continues to release its article metadata and the related references in Crossref.

Unluckily, IEEE is in good company...

It is included in the five publishers among the top 20 DOI depositors are not distributing open references:

- Elsevier
- Wolters Kluwer Health
- IOP Publishing
- American Chemical Society

These are scholarly societies, that should be dedicated to the “benefit of humanity.”

Tried to contact multiple times from different channels, no positive answers received.
Producers of open citation data

Disclaimer: this is a very incomplete list of producers of open citation data

Linked Open Citation Database (LOC-DB)
OpenCitations

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 data (CC0): OpenCitations Corpus, COCI, CROCI
- software: GitHub repository released with open source licenses
- online services: dumps, REST APIs, SPARQL endpoints, and interfaces

About the datasets

**OpenCitations Corpus (OCC, [http://opencitations.net/corpus](http://opencitations.net/corpus)):** new instance was set up at the University of Bologna in early July 2016, and currently contains ~14M citation links to over 7.5M cited resources

**OpenCitations Indexes ([http://opencitations.net/index](http://opencitations.net/index)):**
- **COCI** (launch: July 2018): ~445M citations between ~46M bibliographic entities
- **CROCI** (launch: March 2019): 76 citations between 81 bibliographic entities

<table>
<thead>
<tr>
<th>Service</th>
<th>OCC</th>
<th>Indexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>REST API</td>
<td><a href="https://w3id.org/oc/api/v1">https://w3id.org/oc/api/v1</a></td>
<td><a href="https://w3id.org/oc/index/api/v1">https://w3id.org/oc/index/api/v1</a></td>
</tr>
<tr>
<td>SPARQL endpoint</td>
<td><a href="https://w3id.org/oc/sparql">https://w3id.org/oc/sparql</a></td>
<td><a href="https://w3id.org/oc/index/sparql">https://w3id.org/oc/index/sparql</a></td>
</tr>
<tr>
<td>Textual search</td>
<td><a href="https://w3id.org/oc/search">https://w3id.org/oc/search</a></td>
<td><a href="https://w3id.org/oc/index/search">https://w3id.org/oc/index/search</a></td>
</tr>
</tbody>
</table>
The rise of the REST APIs (April 2018 - March 2019)
World view

Number of accesses to the OpenCitations services and website per country
Studies using OpenCitations data

Bibliometric analysis on the open citation data of PLOS ONE articles available in the OpenCitations Corpus – https://doi.org/10.5281/zenodo.1066316

Investigate possible cross-discipline and cross-department potential collaborations – https://tinyurl.com/man-new-research-opportunities

Replicate part of the outcomes of the Italian evaluation exercise for the Computer Science research field using only open scholarly data – https://arxiv.org/abs/1902.03287

Explore the roles of books in scholarly communication – https://arxiv.org/abs/1906.06039

Integration in VOSviewer (http://www.vosviewer.com), Citation Gecko (http://citationgecko.com), VisualBib (http://visualbib.uniud.it)
A special thank to
Thank you for your attention

OpenCitations
Opening citation data to the public domain
Silvio Peroni

Digital Humanities Advanced Research Centre (DHARC),
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

Meeting of the Informatics National Associations - On Research Evaluation
9 July 2019, Zurich, Switzerland