<table>
<thead>
<tr>
<th>項目</th>
<th>内容</th>
</tr>
</thead>
<tbody>
<tr>
<td>タイトル</td>
<td>講演 OPEN CITATIONS 101 : Historical Background and Current Developments 2019年度 京都大学図書館機構講演会「オープン・サイテーションと機関リポジトリの展開」</td>
</tr>
<tr>
<td>著者</td>
<td>Peroni, Silvio</td>
</tr>
<tr>
<td>言及</td>
<td>(2019)</td>
</tr>
<tr>
<td>発行日</td>
<td>2019-05-20</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/241636">http://hdl.handle.net/2433/241636</a></td>
</tr>
<tr>
<td>型式</td>
<td>Presentation</td>
</tr>
<tr>
<td>出版社</td>
<td>Kyoto University</td>
</tr>
<tr>
<td>言語</td>
<td>日語</td>
</tr>
<tr>
<td>発行者</td>
<td>京都大学</td>
</tr>
<tr>
<td>論文種類</td>
<td>研究小論</td>
</tr>
</tbody>
</table>
Open Citations 101
Historical Background and Current Developments

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

Current Developments of Open Citations
and Institutional Repository
20 May 2019 – Kyoto University Library
Kyoto, Japan
Who am I

Bachelor, Master, Ph.D., Postdoc in Computer Science

Most significative research appendices (in chronological order)

Tag cloud from the abstracts of my published papers in 2017-2018

Assistant Professor
Department of Classical Philology and Italian Studies
University of Bologna
Acknowledgements

I prefer to thanks all the people and collaborators behind this presentation and the Open Citations movement at the very beginning, since their contributions have been crucial for preparing what I am going to tell you in this talk.

Here the list (in alphabetic order): Geoff Bilder, Jonathan Dugan, Martin Fenner, Ivan Heibi, Ginny Hendricks, Vincent Larivière, Jennifer Lin, Catriona MacCallum, Jo McEntyre, Daniel Mietchen, Cameron Neylon, Mark Patterson, David Shotton, Gianmarco Spinaci, Cassidy Sugimoto, Dario Taraborelli, Nees van Eck, Ludo Waltman… and apologies for any other I’ve forgotten to thank!
citations
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
References and citations

Everything is a reference

Semantic overload!

A citation is a **conceptual directional link** from a citing entity to a cited entity
Unambiguous vocabularies

Development of a **set of complementary and orthogonal models** that can be used for the description of the main areas of this publishing domain, from the metadata of scholarly artefacts to the specification of the workflow processes that result in the publication of a scholarly product.

**SPAR (Semantic Publishing and Referencing) Ontologies**

http://www.sparontologies.net – @sparontologies

- Available with a Creative Commons Attribution License 4.0
- Short descriptive page on the SPAR website, diagrams, examples of usage, publication information, HTML documentation

SPAR Ontologies: a diagram

17 OWL 2 DL ontologies

Reuse of several existing models and design patterns

Bibliographic resources, their identifiers, and their internal components (paragraphs, sections, results, methods, etc.)

Components involved in the citation process

Contextual aspects of a publication, namely agents’ roles, document statuses, steps in the publishing workflow, contributors’ roles, and related academic administrative information

Specification of qualitative and quantitative evaluations of a bibliographic resource or an agent (impact factor, h-index, e-index, etc.) and for the encoding of a Five Stars rating for articles
open citations
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**


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> .
```
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.
Citation data must be freely accessible and reusable **without restrictions**, for example by publication under the **CC0 1.0 Universal waiver/license**.

"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**"  

"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)
Towards open citations

2010: Open Citations Corpus released as main outcome of Jisc project

2013: David Shotton’s talk COASP 2013

2016 (May): First WikiCite workshop

2016 (July): launch of new instance of the OpenCitations Corpus by OpenCitations

2016 (September): Dario Taraborelli’s talk at COASP 2016

2016 (around November): first conf call of I4OC (without a formal name)
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

stakeholders

758 publishers participating (as of February 2019)
How many citations are open now

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.

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

See the **Workshop on Open Citations** ([https://workshop-oc.github.io](https://workshop-oc.github.io)) and **WikiCite 2018** ([https://meta.wikimedia.org/wiki/WikiCite_2018](https://meta.wikimedia.org/wiki/WikiCite_2018)) and the related hashtags (**#WOOC2018** and **#wikicite**) 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 May 2019):

- 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, part 1

Open citations: A letter from the scientometric community to publishers

December 5th, 2017

Openness is central to the research endeavor. It is essential for reproducibility and appraisal of research, reduce misconceptions, and equitable access to and participation in science. Yet, challenges to openness in science are often met with initial resistance by publishers and societal organizations. Open access repositories, open data repositories, and open access publishers are initially resisted, but eventually adopted without adverse consequences and benefits to the wider scientific community. The launch of the Initiative for Open Citations is an important step forward in this area, with many similar initiatives taking place around the world. This initiative has campaigned for scholars to make openly available the references found in articles they publish, including most of the large ones, support the open access movement and opened their references. However, the initiative still lacks support from the majority of the large publishers.

Funders should mandate open citations

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

David Shotton

Over the past two decades, open access to journal articles, data, 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 reuse. Publishers and funders should include open citation policies in their mandates and agreements.
Uptake, part 2

DORA: https://sfdora.org/read/

9. Whether a journal is open-access or subscription-based, it should provide unlimited access to references or bibliographies in research articles and not impose any limitations on reference lists in research articles and not impose any limitations on reference lists in research articles and not impose any limitations on reference lists in research articles and not impose any limitations on reference lists in research articles and not impose any limitations on reference lists in research articles.

Plan S:

Plan S
Making full and immediate Open Access a reality

9.3 Recommended additional criteria for journals and platforms:

› Support for PIDs for authors (such as ORCID), funders, funding programmes and grants, institutions, and so on.

› Direct deposition of publications by the publisher into Plan S compliant author designated or centralised Open Access repositories.

› Openly accessible data on citations according to the standards by 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), 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
A ongoing study on Humanities and Social Sciences with Crossref open citation data
ERIH-PLUS journal and disciplines

The European Reference Index for the Humanities and the Social Sciences (ERIH PLUS) makes available a list of journals in the Humanities and Social Sciences domain.

ERIH PLUS includes 30 disciplines: we have asked experts to categorise them as Arts & Humanities (AH) or Social Science (SS) – some of them have been categorise both AH and SS, e.g. “Media Studies and Communication”

<table>
<thead>
<tr>
<th>Number of journals</th>
<th>Total</th>
<th>At least one AH discipline</th>
<th>At least 50% AH disciplines</th>
<th>Only AH disciplines</th>
<th>At least one SS discipline</th>
<th>Only SS discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7,226</td>
<td>4,920</td>
<td>4,386</td>
<td>2,511</td>
<td>4,714</td>
<td>2,306</td>
</tr>
</tbody>
</table>
ERIH PLUS: Crossref vs. WoS and Scopus

We looked at how many journals in ERIH PLUS are actually mentioned in three databases, i.e. Crossref (open), Web of Science Core (WoS Core, closed) and Scopus (closed).

Crossref has wider coverage.

<table>
<thead>
<tr>
<th>Source and dump date</th>
<th>Total number of journals</th>
<th>Number of journals included in ERIH-PLUS</th>
<th>Number of journals included in ERIH-PLUS with &gt;= 50% disciplines in AH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossref (August 2018)</td>
<td>95,367</td>
<td>4,977</td>
<td>2,841</td>
</tr>
<tr>
<td>Scopus (May 2018)</td>
<td>37,452</td>
<td>3,328</td>
<td>2,026</td>
</tr>
<tr>
<td>WoS Core (13th week 2018)</td>
<td>19,226</td>
<td>2,133</td>
<td>1,238</td>
</tr>
</tbody>
</table>
Identify macro-areas in Crossref

We used the Leiden algorithm to identify clusters of publications (via open citations) and to group them in 5 areas: Physical sciences and engineering, Biomedical and health sciences, Social sciences and humanities, Mathematics and computer science, and Life and earth sciences.

Question: what are the Humanities?

Propose a strategy to define Arts and Humanities in Crossref

Methodology:

1. Create the clusters of publications by means of the Leiden algorithm using the citation network extracted from Crossref open citation data
2. For each cluster, record the percentage of publications that were published in any of the AH journals as defined in ERIH PLUS
3. All the publications in the cluster that have at least an X% (to define) of publications in ERIH PLUS AH journals are recorded as AH publications
4. Visualise the data to show the shape of the AH domain via VOS viewer
Humanities: what are they
Humanities clusters

Clusters

Life and Heart Sciences

Mathematics and Computer Science

Percentage of humanities publications

Clusters
Preliminary results

The majority of clusters with the highest percentage of publications in ERIH PLUS journals are contained in the main field Social Sciences and Humanities.

The labels of the clusters are coherent with its Arts and Humanities content – this is true even for clusters outside Social Sciences and Humanities main field.
Beyond Crossref: OpenCitations
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 data (CC0): OpenCitations Corpus, COCI, CROCI
- software: GitHub repository released with open source licenses
- online services: dumps, REST APIs, SPARQL endpoints, and authoring / querying / browsing interfaces
OpenCitations Data Model

Citations as first-class data entities

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.

Open Citation Identifier (OCI)

New persistent identifier scheme for citations contained in bibliographic databases

Simple structure: $oci:\text{number-number}$, where “oci:” is the identifier prefix

Examples:

- $oci:01027931310-01022252312$ (citation in Wikidata, identified by “010”)
- $oci:02001010806360107050663080702026306630509-02001010806360107050663080702026305630301$ (citation in Crossref, identified by “020”)
- $oci:0302544384-0307295288$ (citation in the OCC, identified by the “030”)

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

Peroni S, Shotton D (2019). Open Citation Identifier: Definition. Figshare. DOI: https://doi.org/10.6084/m9.figshare.7127816
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 ~14M citation links to over 7.5M cited resources

OpenCitations Indexes (https://w3id.org/oc/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>
COCI

COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations ([https://w3id.org/oc/index/coci](https://w3id.org/oc/index/coci)), is the first of the indexes proposed by OpenCitations ([https://w3id.org/oc/index](https://w3id.org/oc/index)), in which citations are exposed as first-class data entities with accompanying properties.

Why CROCI?

Analysis of COCI + Crossref shows that number of closed citations is of great significance.

One third of these are references to entities published by Elsevier, that refuses to open its references (see I4OC – [https://i4oc.org](https://i4oc.org)).

CROCI uptake (limited so far)

Goal: individuals identified by ORCiD identifiers 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 to emphasize and ensure their openness for every kind of reuse without limitation.

Scholarly: Read less. Learn more. @scholarly · Feb 25
Initial implementation of CROCI for @opencitations @i4oc_org is now live at ref.scholarly.com. Select the download endpoint and reference_format=croci. Somewhat slow and currently limited to docs below 2MB at a few requests per day. Will improve matching process over time.

OpenCitations @opencitations · Mar 4
#CROCI (opencitations.net/index/croci) launched with citations by @aarontay @libsun @essepuntato. Its data are queryable using #REST #API (w3id.org /oc/index/croci...) or the unifying API for all indexes (w3id.org/oc/index/api/v1). To contribute: github.com/opencitations/... #openscience #I4OC

Blenca Kramer @MsPhelps · Mar 15
Just submitted a proposal for a pratical session for Monday’s Barcamp Open Science 2019 @lfvopenscience #oscibar - to contribute citation data to @opencitations Crowdsourced Open Citation Index (#CROCI) etherpad.wikimedia.org/p/oscibar2019 ... (see also opencitations.wordpress.com/2019/02/07/croci...)

https://twitter.com/konradfoerstner/status/1107657296396673024
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
- **OCI Graphe** developed by Dominique Rouger is a Web tool to search articles in COCI
The rise of the REST APIs (April 2018 - March 2019)
World view

Number of accesses to the OpenCitations services and website per country
A special thank to
Conclusions
Extending the data

While there are available several open citation data in Crossref and “sister” initiatives and projects – e.g. OpenCitations (overall, 450M citations) and Wikidata (more than 170M citations) – there are several citation data that are not covered yet.

Users (e.g. researchers, libraries, institutions) can contribute directly to fill this gap in different ways – e.g. see CROCI and Wikidata and their crowdsourced approaches for gathering additional open citation data.
Towards broad adoption

Open citation data can be reused, due to the freedom provided by the license / waiver associated to them.

Making available REST APIs is crucial to maximise the reuse of citation data, since they are the primary mechanism for allowing others to create applications upon citations.

In addition to the project mentioned previously, we are talking to the University of Liège Institutional Repository (ORBi, https://orbi.uliege.be) and to DBLP (https://dblp.uni-trier.de/), which aim at integrating open citation data available in OpenCitations and Crossref in their Web services.
The way to help

It is crucial to keep such very permissive license / waiver to metadata and citation data so as to maximise their reuse.

Institutional repositories should ask explicitly to their authors the permission to release metadata and citation data of deposited papers as CC0 material independently from the license that is associated to the papers itself, which may be different.
Thank you for your attention

Open Citations 101
Historical Background and Current Developments

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

Current Developments of Open Citations and Institutional Repository
20 May 2019 – Kyoto University Library
Kyoto, Japan