Introduction to the project

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato

Open Science (A.Y. 2021/2022)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA

Groups of the project

You have to form two groups of people – balancing the number of members so as to have an almost equal number of people per group

Important notice: a group is not in competition with the other bur rather it complements the other

You must decide a name to assign to your group – please, be creative

Setting up a GitHub space

Each member of a group must have a GitHub account – in case you do not have it yet, please create one

Each group will be assigned to a GitHub team I will create using the name of your group

I will create a specific folder on the GitHub repository of the course for each of the groups, to allow you to store all the material collected for the project

Digital Object Identifier

The Digital Object Identifier (DOI, https://doi.org) system provides an infrastructure for persistent unique identification of objects of any type (shape of the id: **`10.xxxx/xxxxxxxxxx`**)

A DOI is a digital identifier of an object rather than an identifier of a digital object, that means that it can be used to identify objects that are not born-digital, such as print books and articles

The DOI system is designed to work over the Internet, and a DOI is permanently assigned to an object to provide a resolvable persistent network link to current information about that object

A DOI can be resolved within the DOI system to values of one or more types of data relating to the object identified by that DOI, such as descriptive metadata

A Rest API is provided to query the system

International DOI Foundation. (2019). DOI® Handbook. https://doi.org/10.1000/182

Crossref

Crossref is a not-for-profit membership association which aims at promoting the development and cooperative use of new and innovative technologies to speed and facilitate scientific and other scholarly research

Crossref is one of the ten International DOI registration agencies, and allows its members to register the DOIs of their publications

Each DOI registered in the Crossref system is associated with a URL to the publication's webpage and accompanied with the metadata of the publications

Crossref provides a REST API to retrieve data about the entities

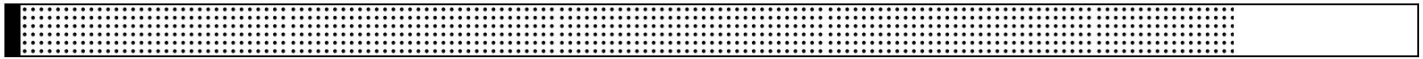
Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. Quantitative Science Studies, 1(1), 414–427. https://doi.org/10.1162/qss_a_00022

Initiative for Open Citations (I4OC)

The aim of the Initiative for Open Citations (I4OC, https://i4oc.org) is to promote the availability of data on citations

How: publishers ask to open their references, along with the other bibliographic metadata, that they send to Crossref

How many citations are open today?



1% 86% 13%

As of March 2021, the fraction of publications with open references has grown from 1% to 87% out of 54.2 million articles with references deposited with Crossref.

What is an open citation

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

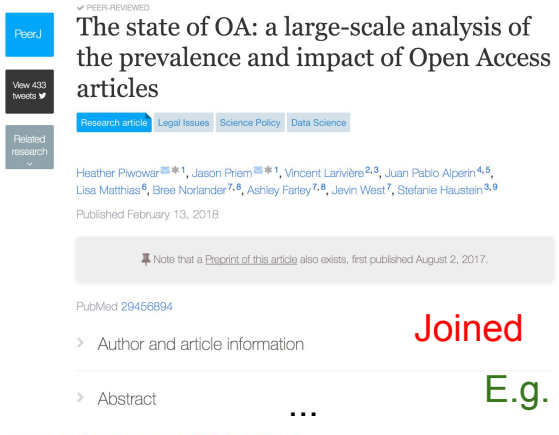
The citation data related to a particular citation must include:

- 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

Peroni, S., & Shotton, D. (2018). Open Citation: Definition. Figshare. https://doi.org/10.6084/m9.figshare.6683855

Open citations: characteristics


PeerJ

View 433 tweets

Related research

PEER-REVIEWED

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 Piwowar*1, Jason Priem*1, Vincent Larivière[2,3], Juan Pablo Alperin[4,5], Lisa Matthias[6], Bree Norlander[7,8], Ashley Farley[7,8], Jevin West[7], Stefanie Haustein[3,9]

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

...

Identifiable

Structured (JSON; machine readable)

```
"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
},
...
```

Available

E.g. HTTP + ID = metadata

https://api.crossref.org/works/10.7717/peerj.4375

REFERENCES

Joined

Unstructured

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.

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

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

"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

Closed

Open

"No claims of ownership to individual items of bibliographic metadata"

https://api.crossref.org

OpenCitations' COCI

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

All the data available in COCI are derived from those accessible through Crossref

All the citation data are stored using Semantic Web technologies (RDF and OWL) and are compliant with a specific data model (see http://opencitations.net/model)

Currently COCI contains more than 1.27 billion DOI-to-DOI citation links made available under that can be accessed and queried through a REST API

Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. Scientometrics, 121(2), 1213–1228. https://doi.org/10.1007/s11192-019-03217-6

DOAJ

The Directory of Open Access Journals (DOAJ, https://doaj.org/) is an independent index containing more than 17,500 peer-reviewed, open access journals covering all areas of science, technology, medicine, social sciences, arts and humanities

In addition to journals, it also contains basic metadata of more than 7 million journal articles by several publishers from 130 countries

All its data can be queried via their REST API or downloaded as a full dump

Research questions

1. Which is the coverage, in terms of citations, of the open access journals in DOAJ according to the data available in OpenCitations? How many citations DOAJ journals receive? How many citations DOAJ journals do? How many of these citations involves articles published in open access journals as both citing and cited entities? What is the trend (increasing? decreasing?) in time of the availability of such citations involving journals in DOAJ?
2. How many articles published in the open access journals in DOAJ are included in Crossref? How many of these articles include information about their reference lists? How many references have a DOI specified? How many of these DOIs have been specified by the publishers? And how many by Crossref?

Action items

The groups must agree on which research question to address, since they cannot address the same research question

You have to provide a structured abstract presenting your work – yes, even if it is not yet completed! It will be updated by you daily everytime you need

A structured abstract is just a very brief document describing your research

This exercises oblige you to think about your research before addressing it

Please follow the template proposed by Emerald Publishing to sketch the structured abstract, to upload a first version of it in your GitHub folder in a file named “abstract.md”

End

Introduction to the project

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato

Open Science (A.Y. 2021/2022)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA