Introduction to the project

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

Open Science (A.Y. 2023/2024)
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

What is an open citation

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

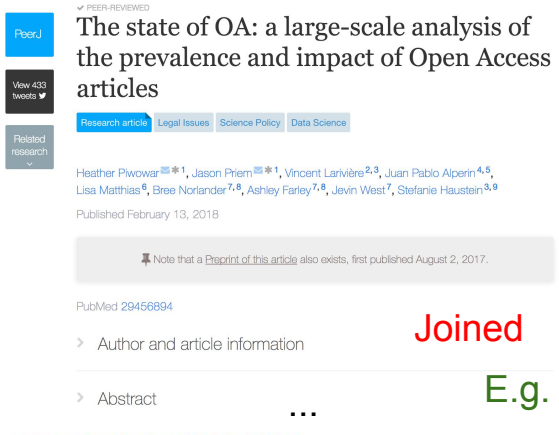
The citation data related to a particular citation must include:

- 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 Index

A citation index is a bibliographic index recording citations between publications, allowing the user to establish which later documents cite (i.e. contain references to) earlier documents

OpenCitations, as an infrastructure organization for open scholarship, has built the OpenCitations Index, a totally open citation index that ingest data from Crossref, DataCite, NIH Open Citation Collection, OpenAIRE, and the Japan Link Center

Currently, OpenCitations Index contains more than 1.98 billion citation links made available under that can be accessed and queried through a REST API and downloaded as a full dump (http://opencitations.net/download#index)

Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. Quantitative Science Studies, 1(1), 428–444. https://doi.org/10.1162/qss_a_00023

OpenCitations Meta

The OpenCitations Meta database (http://opencitations.net/meta) stores and delivers bibliographic metadata for all publications involved in the OpenCitations Index

For each publication, the metadata exposed by OpenCitations Meta includes the publication's title, type, venue (e.g. journal name), volume number, issue number, page numbers, publication date, and identifiers such as Digital Object Identifiers (DOIs) and PubMed Identifiers (PMIDs)

Currently OpenCitations Meta contains more than 90 million bibliographic entities that can be accessed and queried through a REST API and downloaded as a full dump (http://opencitations.net/download#meta)

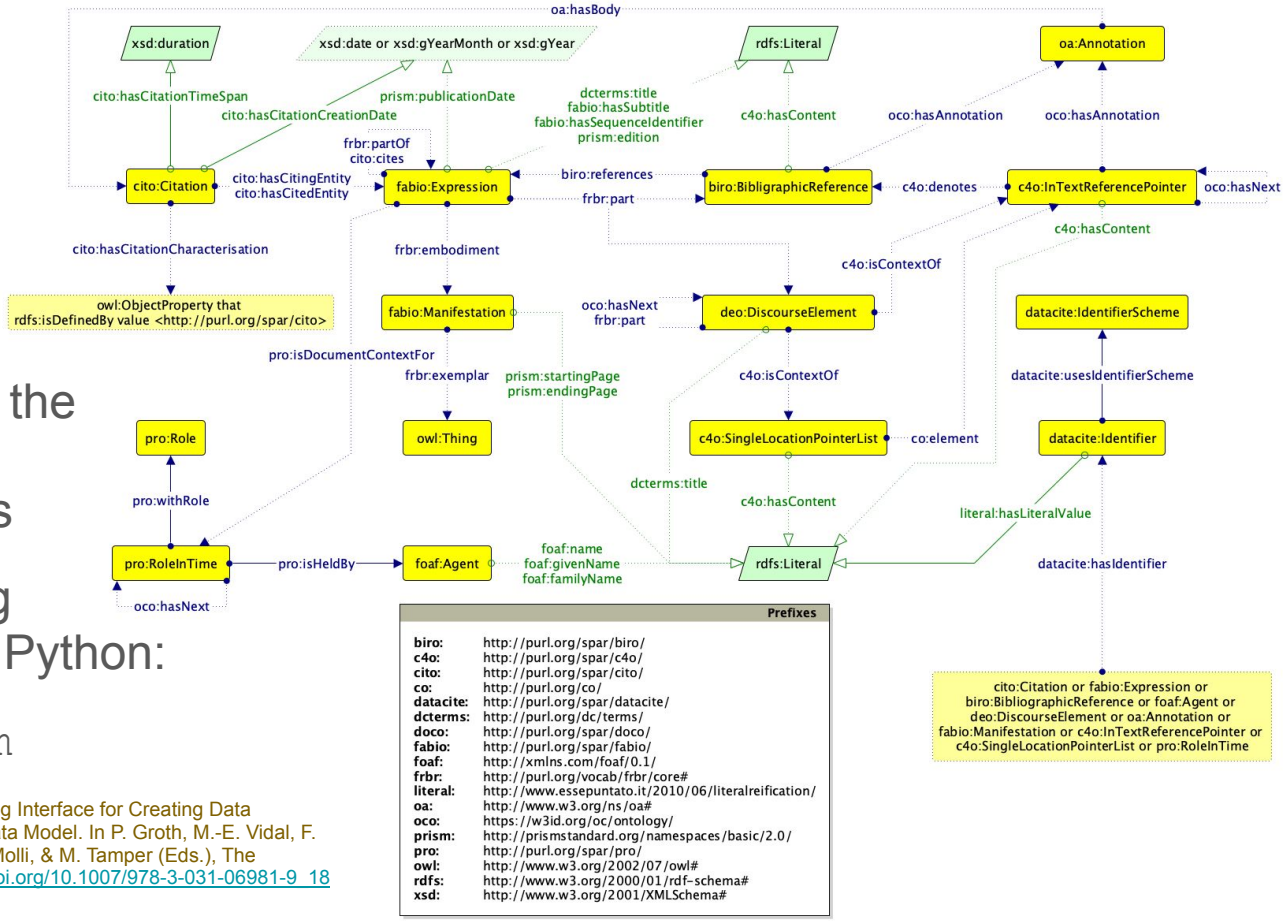
Massari, A., Mariani, F., Heibi, I., Peroni, S., & Shotton, D. (2024). OpenCitations Meta. Quantitative Science Studies. https://doi.org/10.1162/qss_a_00292

OpenCitations Data Model

The OpenCitations Data Model (OCDM) is the metadata model used for the data stored in all the OpenCitations' databases

Python library for creating OCDM-compliant data in Python:

```
pip install oc_ocdm
```



Persiani, S., Daquino, M., & Peroni, S. (2022). A Programming Interface for Creating Data According to the SPAR Ontologies and the OpenCitations Data Model. In P. Groth, M.-E. Vidal, F. Suchanek, P. Szekley, P. Kapanipathi, C. Pesquita, H. Skaf-Molli, & M. Tamper (Eds.), The Semantic Web (Vol. 13261, pp. 305–322). Springer. https://doi.org/10.1007/978-3-031-06981-9_18

Daquino, M., Peroni, S., Shotton, D., Colavizza, G., Ghavimi, B., Lauscher, A., Mayr, P., Romanello, M., & Zumstein, P. (2020). The OpenCitations Data Model. The Semantic Web – ISWC 2020, 12507, 447–463. https://doi.org/10.1007/978-3-030-62466-8_28

IRIS

IRIS is the implementation of the Current Research Information System of the University of Bologna, which permits gathering data related to research activities and outcomes

It offers several interfaces and tools to researchers, administrators, evaluators to document and monitor research outputs and to increase their visibility

It offers several REST APIs to gather public data from it

A dump with public bibliographic data has been recently made available on the University of Bologna's institutional repositories for research data, i.e. AMS Acta

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 and provides annual dumps

The entities Crossref comprises include also peer reviews of other articles, as shown in the exemplar result of the of the following API call:
http://api.crossref.org/works/10.7287/peerj-cs.266v0.1/reviews/1 (see the field "relation")

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

Research questions

1. What is the coverage of the publications available in IRIS, that strictly concern research conducted within the University of Bologna, in OpenCitations Meta? What is the types of publications that are better covered in OpenCitations Meta? What is the amount of citations (according to OpenCitations Index) the IRIS publications included in OpenCitations Meta are involved in (as citing entity and as cited entity)? How many of these citations come from and go to publications not included in IRIS, and how many of these citations involves publications in IRIS as both citing and cited entities?
2. Is that possible to create a new index of citations which contains typed citations where a peer review (citing entity) reviews (specific citation function) a publication (cited entity)? What is the necessary transformation of the Crossref dump necessary to create such an index to be compliant with the OpenCitations Data Model? What are the top publication venues in terms of the number of peer reviews received? How many peer reviews in Crossref are included in OpenCitations Meta? How many articles that have been reviewed by a peer review are included in OpenCitations Meta?

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. 2023/2024)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA