

Introduction to the project

Silvio Peroni

silvio.peroni@unibo.it – <https://orcid.org/0000-0003-0530-4305> – [@essepuntato](#)

Open Science (A.Y. 2022/2023)

Second Cycle Degree in Digital Humanities and Digital Knowledge

Alma Mater Studiorum - Università di Bologna



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 but 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/xxxxxxxxxxx**)

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

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**

Open citations: characteristics

PeerJ
View 433 works
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

Hesther Piwowar¹, Jason Priem¹, Vincent Larivière^{2,3}, Juan Pablo Alperin^{4,5}, Lisa Matthias⁶, Bree Norlander^{7,8}, Ashley Farley^{7,8}, Jevin West⁷, 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

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

Structured
(JSON;
machine
readable)

Joined

Available

E.g. HTTP + ID = metadata

REFERENCES

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*.

“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



<https://api.crossref.org>

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

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 1.46 billion DOI-to-DOI 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#coci>)

OpenCitations Meta

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

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>)

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

ERIH-PLUS

ERIH PLUS (<https://kanalregister.hkdir.no/publiseringskanaler/erihplus/>) is an academic journal index for the Social Sciences and Humanities (SSH)

They provide information about more than 10 thousands journals in detail

Data about journals can be accessed using their search interface, and the full dump with journal information is also freely available online

(<https://kanalregister.hkdir.no/publiseringskanaler/erihplus/periodical/listApproved>)

Research questions

1. What is the coverage of publications in Social Science and Humanities (SSH) journals (according to ERIH-PLUS) included in OpenCitations Meta? What are the disciplines that have more publications? What are countries providing the largest number of publications and journals? How many of the SSH journals are available in Open Access according to the data in DOAJ?
2. How many citations (according to COCI) involve, either as citing or cited entities, publications in SSH journals (according to ERIH-PLUS) included in OpenCitations Meta? What are the disciplines that cites the most and those cited the most? How many citations start from and go to publications in OpenCitations Meta that are not included in SSH journals?

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. 2022/2023)

Second Cycle Degree in Digital Humanities and Digital Knowledge

Alma Mater Studiorum - Università di Bologna

