

OpenScience Fair  
Oporto - September 17, 2019

# The OpenCitations Graph

Marilena Daquino, Digital Humanities Advanced Research Centre (/DH.arc), University of Bologna  
marilena.daquino2@unibo.it

# OpenCitations

OpenCitations (<http://opencitations.net>) is a **scholarly infrastructure organization**

It works on:

- the **publication of open bibliographic and citation data** by the use of **Semantic Web technologies**, and RDF for its description
- **advocacy for open citations**

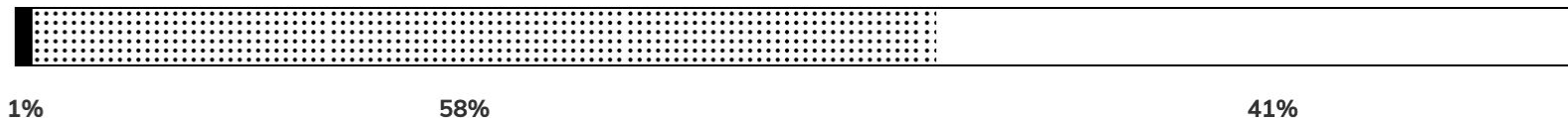
# Initiative for Open Citations (I4OC): <https://i4oc.org>

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

Founders:



As of September 2019, the fraction of publications with open references has grown from 1% to **59% out of 47.6 million articles** with references deposited with Crossref



# Citation data

**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**

# OpenCitations

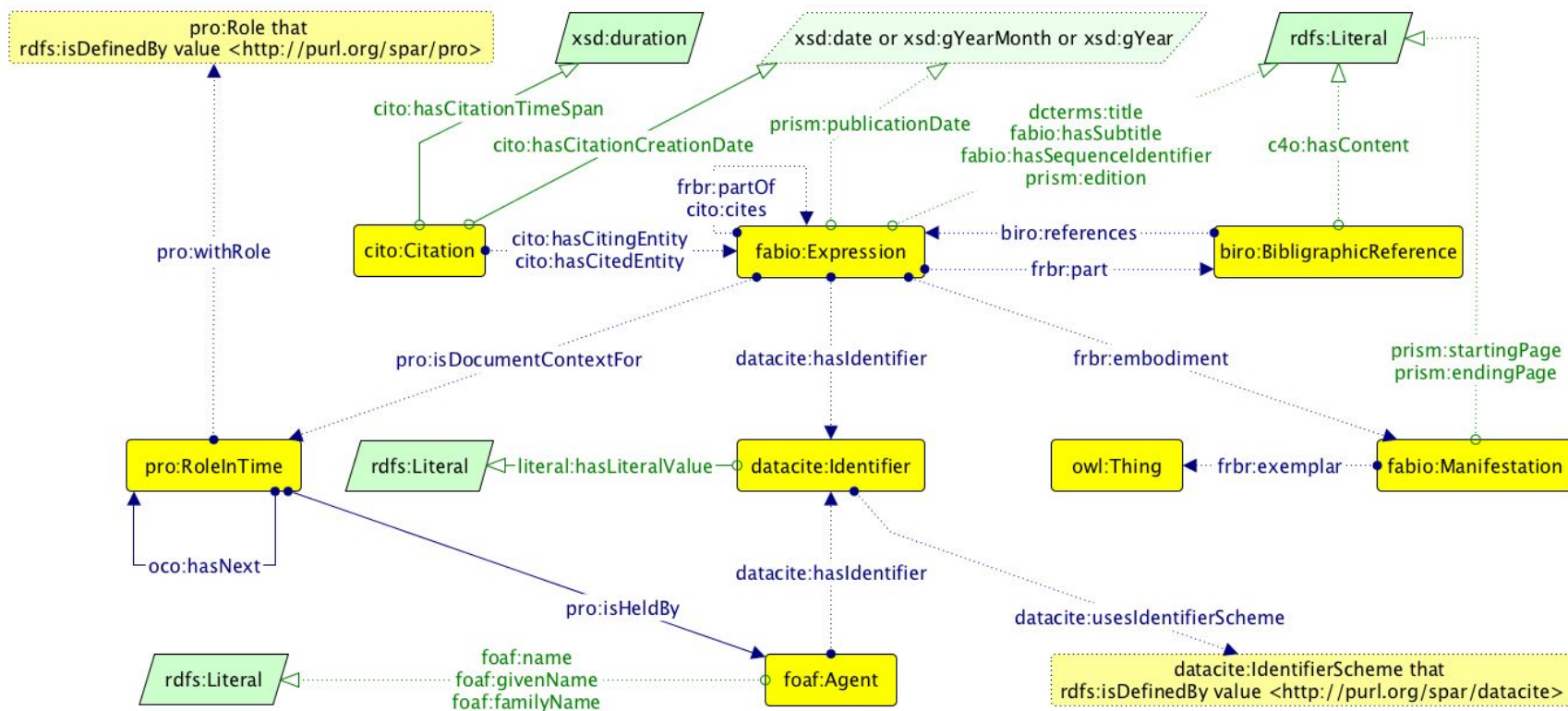
**OpenCitations** (<http://opencitations.net>) is a **scholarly infrastructure organization**

It ingests citation data from a number of sources (Crossref, ORCID, PMC)

It provides:

- **data models:** the [OpenCitations Data Model](#) (based on SPAR Ontologies)
- **Linked Open Datasets (in CCO):** [OpenCitations Corpus](#), [COCI](#), [CROCI](#)
- **software:** [GitHub repository](#) released with open source licenses
- **online services:** [dumps](#), [REST APIs](#), [SPARQL endpoints](#), and [interfaces](#)

# OpenCitations Data Model



# Datasets

- **The OpenCitations Corpus** (OCC, <http://opencitations.net/corpus>):
  - citation links to over 7.5M cited resources
  - information about different kinds of bibliographic entities such as: bibliographic resources (e.g. journals, and journal articles), identifiers (local), responsible agents, and provenance data.
- **The OpenCitations Indexes** (<http://opencitations.net/index>): bibliographic indexes recording citations between publications using the data available in particular bibliographic databases.
  - **COCI** (launch: July 2018): ~445M citations between ~46M bibliographic entities
  - **CROCI** (launch: March 2019): 76 citations between 81 bibliographic entities

# Online services

- Dumps:
  - **Corpus and Indexes data and provenance**
  - Available in **CSV, N-Triples** formats
- SPARQL endpoints:
  - **SPARQL: is the RDF Query Language**
  - Used to **Query** both the Corpus and Indexes
  - **Suitable for Semantic Web experts**
- REST APIs:
  - Implemented by means of **RAMOSE**, the Restful API Manager Over SPARQL Endpoints (<https://github.com/opencitations/ramose>).
  - Accessing/Retrieving the data of the Corpus and Indexes
    - Corpus: [opencitations.net/api/v1](https://opencitations.net/api/v1)
    - Indexes: [opencitations.net/index/api/v1](https://opencitations.net/index/api/v1)

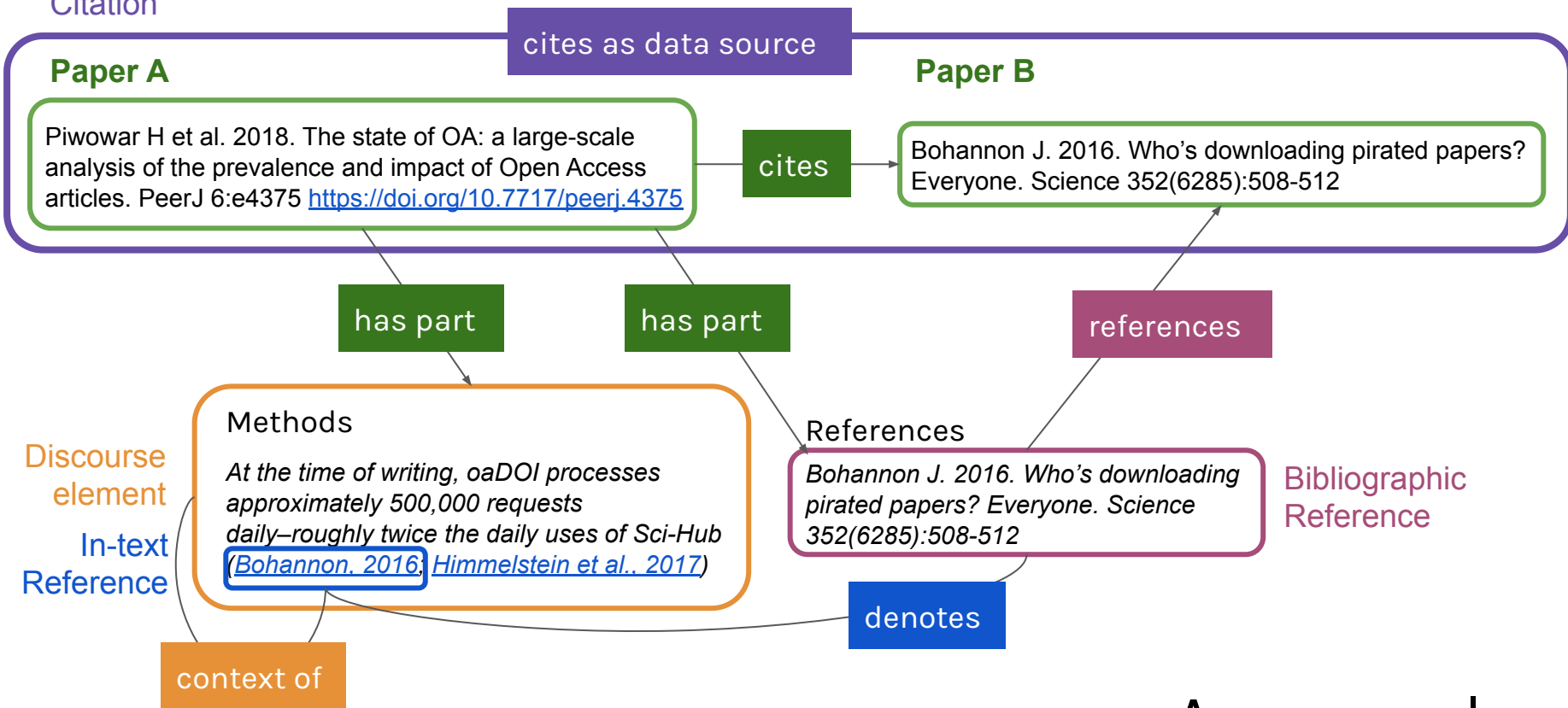


# The *Open Biomedical Citations in Context* Corpus

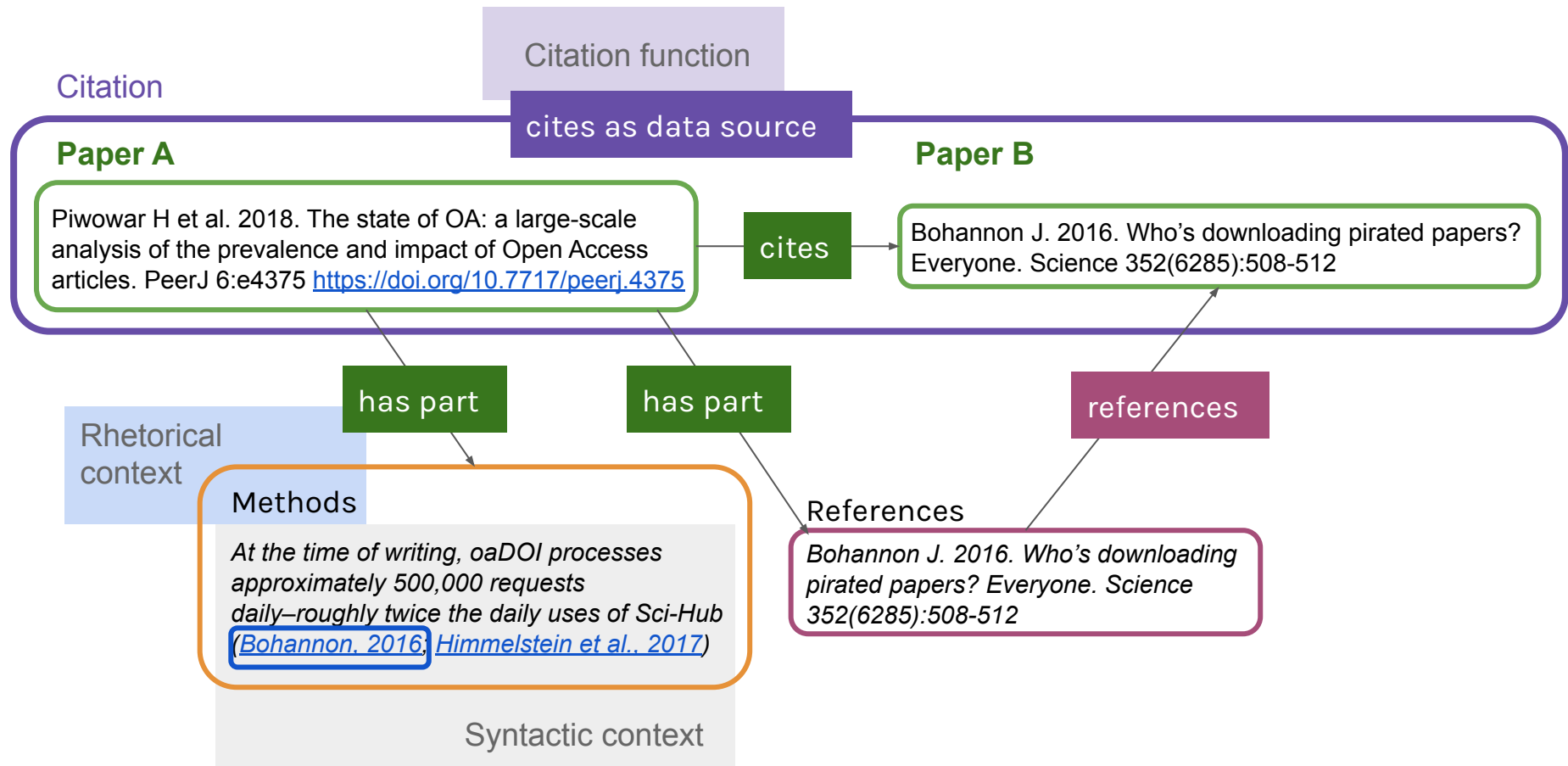
An ongoing project for mining and publish a Linked Dataset from PubMedCentral OA corpus, including:

- In-text reference pointers, e.g. [1], (Peroni, Shotton 2019), referencing a bibliographic resource
- Syntactic context of citations, e.g. sentences, paragraphs, captions, tables.
- Rhetorical context of citation, e.g. Introduction, Methods, Background, Related work
- The purpose of the citation (or *citation function*), e.g. extends, disagrees with, cites as data source.

## Citation



An example



An example



Alfred P. Sloan  
FOUNDATION



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA



A special thank to

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA

# Thanks!

Any question?

Marilena Daquino, Digital Humanities Advanced Research Centre (/DH.arc), University of Bologna  
marilena.daquino2@unibo.it