

## **OpenCitations: structured open citation data as a part of the Commons**

Silvio Peroni

Digital and Semantic Publishing Laboratory, Department of Computer Science and Engineering, University of Bologna, Bologna, Italy, Orcid 0000-0003-0530-4305

OpenCitations (<http://opencitations.net>) employs Semantic Web technologies to publish scholarly citation data in an open repository. The main service developed by OpenCitations is the OpenCitations Corpus (OCC), in which RDF-based scholarly citation data (i.e. bibliographic references) are stored and made freely available under a Creative Commons public domain dedication so that others may use and build upon them. The other resources of OpenCitations – namely the ontologies describing the data and the software developed to build the OCC – are also available with open licenses. OpenCitations follows the FAIR principles for data-intensive science, namely that the data should be findable, accessible, interoperable and re-usable, and employs the three basic criteria promoted by the Initiative for Open Citations, namely that the citation data must be structured, separable and open.

In this presentation, we will explain how OpenCitations harvests citations and associated metadata from several external APIs (those of Europe PubMed Central, Crossref and ORCID), describes them using the SPAR Ontologies, the Provenance Ontology and FOAF according to the OCC metadata model, and stores all the data in JSON-LD. The stored citations are made accessible via the OCC SPARQL interface, by means of the HTTP requests (using content negotiation, e.g. <https://w3id.org/oc/corpus/br/1>), and as monthly data dumps. As of 5th August 2017, the OCC had ingested the references from 211,196 citing bibliographic resources and contained information about 8,986,030 citation links to 5,106,542 cited resources.

We will exemplify how OpenCitations can benefit the scholarly community at large, being used by international initiatives and projects including Wikidata/WikiCite, OpenAIRE, LOC-DB and eLife, and will describe future plans, made possible by funding from the Sloan Foundation, to increase the rate of citation ingest thirty-fold to 15 million citations per month, and to develop novel user interfaces to access and visualize the citation data.