

Review of: "Coverage of open citations in DOAJ journals v1"

Giulia Venditti¹

¹ University of Bologna

Potential competing interests: The author(s) declared that no potential competing interests exist.

Introduction

The DMP of "Coverage of open citations in DOAJ journals" describes data gathering and future dataset use to investigate the coverage of DOAJ articles in terms of citations according to OpenCitations indexes. Researchers asks how many citations DOAJ journals receive and don't; how many of these involve open access articles; and finally, the presence of trends over time of the availability of citations involving these articles.

This DMP describes the availability and reusability of data and relevant metadata about the two tools used in the research: one software, used for gathering, manipulating and analysing data, and one dataset in JSON format containing statistical informations for journals and the citations they do and receive. The DMP follows the Horizon 2020 layout and gives a DOI identifier obtained through Zenodo.

Since the research is at its infancy, these descriptions do not go into detail but are limited to setting the project according to the FAIR principles. For this motive reader should take into account that I'm reviewing the actual project at the state of the art.

Flaws

For what concerns software dataset, team choose to collect primary data of canonical type to obtain and share informations, and to keep on record. Instead, the dataset on DOJA coverage is composed by derived or compiled data, analysed and extracted from OpenCitation and DOAJ data. Team, as evident, considers and evaluates the existence of secondary resources form Zenodo to follow-up research on a specific area, deciding to reuse existing data in order to compare and combine them with others data.

The information on data to be produce dare adequate and realistic according to the methodology applied by the research including all data that is planned to be generated from the research.

Data will be described by free-of-charge Dublin Core metadata describing their quality. Provided with clear version numbers and persistent identifiers (DOI).

While for the data related to the software is specified that will be manipulated and stored in .py format, the dataset on DOJA coverage dataset will have JSON format. Moreover, software data are searchable through linked open data and accessible through Zenodo. On the contrary, the OpenCitation and DOAJ data are also searchable through OpenAIRE but have no open accessing tool.

No information about data quality assurance is present for Software data description. But that's also obvious given the state of the research. The dataset on DOJ coverage, on the other hand, has documented procedures for quality assurance of the described data. They will use tools for automatic checks, data conform to format specification, consistency verification with data models and standards.

Both datasets will be available on GitHub repositories. No other back-up information is present. A suggestion would be to think about multiple media and multiple copies for back-up and to add information on institutional and Github back-up policy.

Data collected will be openly accessible without encountering sharing difficulties. No ethical nor legal issue can impact on sharing. Moreover, the access doesn't required particular tools or methods. Copyright and data sharing plan of existing data are clarified. Software dataset will use Creative Commons Zero v1.0 Universal licence, the dataset on DOJ coverage will take the Creative Commons Attribution 4.0 International one. Reuse is made immediately available but support is given up to 5 years.

Data management responsibilities have been allocated to named people , also specified by their ORCID. However, at the end of the project, the ensure will be taken by an institutional archive. Another suggestion would be to clarify data management procedures, considering the variety of data management tasks that may be required for the research.

Conclusion

In general, also considering the state of the research, the DMP is well formulated and, although very generic, manages to provide the substantial information to describe the nature of the datasets. Suggestion to improve the current state, could be the addition of more descriptive information on the contents represented in order to clarify the different type of data needed for the search. In particular, it might be useful to document collection method, origin, circumstances, processing and analysis of data.