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

Chiara Catizone[1]

1 University of Bologna

Introduction

This is the peer review of team DonTLockUpprotocol, for the final project elaborated for the course of Open Science held by prof. Silvio Peroni in A.y. 2021/2022 at the University of Bologna.

Authors of the reviewed protocol are: Alessandro Bertozzi, Constance Dami, Umut Kucuk and Chiara Manca.

PLOS ONE guidelines have been followed for performing this review. Please, note that at as the protocol was written, the research work was still at an early stage of development, thus there is no expectancy of all the PLOS ONE guidelines to be completely met by the reviewed protocol.

Rationale

The protocol is introduced by three research questions aiming to investigate the coverage of open citations in DOAJ journals as citing and cited articles, their citations (done and received) and the involvement of open access journals, the trends over time of DOAJ journals citations' availability.

More context could be provided in a next version of the protocol, introducing for example some references to previous researches on the same topic.

The protocol

The protocol is divided in four main sections covering the various stages of the research. These sections provide a general overview on the work covered by the team, that I judge to be exhaustive in terms of data management, spanning to a first gathering stage, to publication further analysis and visualization of results.

Remarkably, in this first version of the protocol the group had already settled down how data will be manipulated in order to answer their main research questions.

In section 1 they provide the source of the citations' data as well as a subsection describing the manipulations performed on the gathered data for answering each research question.

In data publishing they specify open source formats in which the data will be stored allowing this first outcome's reusability and first-stage's testing.

Data analysis has not been clearly defined, which is totally understandable at this stage of the research, but they envision to study the future of the research, which could be a good point to be stated in their abstract as an addiction of the original set of research questions they provide.

Finally, in the last section they mention data visualization, without specifying which kind of data will be used ,as it is impossible to determine yet. Note that the group has already chosen Python libraries as the tool to perform such task.

The data source and library do not present any limitation to third party testing.

Data availability

In terms of data availability the protocol provides the publication format of the data gathered from the OpenCitations API. It is missing specifications on where this – an further – data will be stored and there is no mention on how (format) and where (storage) section 3 results will be made available for reuse and comparison.
At this stage of the research we do not expect to have all of the above mentioned information provided by the protocol, but we recommend the team to add this information in a next version.

Methodology

As already stated, in points 1 and 2 of this review, also under a methodological point of view, this protocol provides the main steps ideated by the team for answering their research questions and for managing predicted outcomes. The workflow follows a logical order and each section – and subsection – is briefly described introducing the intentions of the researchers.

An issue here is that there is no strong methodological justification to the choice of using the OpenCitations API (or their SPARQL endpoints) for retrieving DOAJ articles, which could be a misleading lack of information to non-expert readers. For example: what are the benefits of the selected API, do you envision any possible empish,ent to your work?
Also, it could be useful to briefly explain the differences between the COCI and CROCI indexes as well as providing further specification on the approaches used for extracting the data from the API's results.

Further analysis on data envisioned in section 3 is briefly mentioned as well as data visualization. But, as it has been previously stated, this kind of gaps are justified by the early stage of the research the protocol refers to.

Conclusions

The reviewed protocol enables the reader to have a general understanding of the state of the research carried out by DonTLockUp team, and to envision the main structure of the related workflow its members have created.
The intentions of the research team and their final research questions have been made clear, maybe the context description could be expanded in a later version of the protocol.
The elaborate is missing the explanations on how data will be made available after the second study (Section 3) and on extraction of the data needed for carrying further analysis, which is totally comprehensible as some experiments must be usually carried out before making this kind of statements.
Even if at an early stage, the protocol is structured in a way that allows the testing of the work the research team has

done until that point, meeting PLOS ONE's criteria.

References

Coverage of open citations in DOAJ journals V.1

PLOS ONE Guidelines for Reviewers