Review of: "Protocol OCPRIP"

Salvatore Di Marzo[1]

1 University of Bologna

Potential competing interests: No potential competing interests to declare.

The protocol presented outlines a comprehensive methodology for augmenting the OpenCitations Index by systematically extracting, aligning, and analyzing peer review data sourced from Crossref. The process is delineated into four distinct phases: data gathering, alignment, data management, and data analysis. Each phase is meticulously described, elucidating the procedures involved in extracting, organizing, and analyzing the peer review data. Additionally, the protocol emphasizes the importance of adherence to the OpenCitations Data Model and continuous synchronization with Crossref for ensuring the currency and accuracy of the dataset.

The methodology proposed in the manuscript provides a structured approach to harnessing peer review data from Crossref to enrich the OpenCitations Index. The protocol is well-structured and clearly articulated, guiding researchers through each phase of data extraction, alignment, management, and analysis. The integration of existing methodologies, such as the Ingestion workflow developed for COCI, enhances the credibility and reliability of the proposed approach.

 The protocol demonstrates a rigorous approach to data extraction, alignment, and analysis, encompassing detailed procedures for each phase. The utilization of established libraries and methodologies, such as crossref_commons_py and the COCI Ingestion workflow, enhances the methodological robustness of the protocol. The incorporation of SPARQL queries for data management and continuous synchronization with Crossref ensures that the OpenCitations Index reflects the most up-to-date information available. By enriching the OpenCitations Index with peer review data, the protocol contributes significantly to the advancement of scholarly infrastructure.

The protocol outlines data analysis tasks such as isolating venues, counting DOIs, and verifying article DOIs. However, a more detailed discussion on insights that can be derived from these analyses would enrich the protocol. Providing examples of potential research findings would illustrate the utility of the enhanced OpenCitations Index.

In the data alignment part, specifically in step 3.3, there is this passage that seems to be missing something: "The DOIs of the reviews and the reviewed entities become DOI URLs, while the IRI of the Citations." This can be easily fixed by providing further clarifications.

Besides this, this is the first version of the workflow of an on-going project, and being such, it is very well structured and exhaustive, allowing the user to replicate and to see first-hand all of the steps necessary for the completion of the project.