

# Review of: "DMP Investigating Missing Citations in COCI (zenodo.4671487)"

Ricarda Boente

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

This data management plan specifies information on the data used to investigate the publishers responsible for missing citations in COCI. Following the Horizon 2020 template, the authors supply mostly sufficient information on their data strategies. They supply two datasets, one for the source code they create and one for the generated CSV file that will contain the results of the research. However, there are also some aspects that could be improved in order to create a more comprehensive plan.

One formal issue can be noticed in the mentioning of the authors. In the pdf printed version of the DMP, one author remains without a last name, being called "Alessia null".

Concerning the assessment of existing data, the source of the CSV file that will be used could be made explicit in point 2 (Reusable data). By providing a link to the openly accessible source, the reader of the data management plan could be informed easily and effectively about what type of data is used in the research.

By dividing the data management plan in two datasets, all data generated by the research seems to be covered. The dataset "Missing Citations in COCI" could be renamed, though, in order to make the purpose of this dataset clearer to the reader. Also, in 1.2 there is a typo "exiting" instead of "existing" source. In the second dataset, the introduction specifies well the expected outcome format. However, it would be recommendable to make this point also clear in 1.3, where there is only a very generic sentence at the moment. Also in this section, there are a lot of data formats mentioned (text file, etc.) without being explained on how or why they will be generated.

The authors state to use standards like PEP8 for their python code and plan to store their data with backup and recovery. They plan to work in compliance with the FAIR principles. Although they don't assure to comply to every point mentioned in this section, in my opinion, they are giving a realistic and sufficiently FAIR picture for a research project of this size.

Like this, it also seems realistic that the data management plan will be followed. In point 4.2, the responsibilities could maybe be clearer divided, as it might be easier to have one data manager than sharing the responsibility. In a group of four people though, sharing can maybe be more realistic, so a specification of the responsibilities is more an optional improvement in this case in my opinion.

By resolving the mentioned issues, this DMP could become even more informative and clear than it already is.