

## Response Letter to the reviewers of “[Investigating DOIs’ Classes of Errors](#)”

Dear reviewers,

Thank you for your observations and your constructive criticism. We carefully read all the points that you have made in your reviews [1][2] and revised the protocol according to the remarks and the suggestions that you have thereby stated.

First, according to the review written by Alessia Cioffi [1], we added a “Before Starting” section at the beginning of our protocol in which we state the computational requirements to reproduce our procedure (e.g. which programming language and libraries required) and some rules about the notation that we used to provide verbose description of the source code of our software.

In addition, we followed the advice made in the same review to put further information related to the provenance of the data used in the project, by adding the citation to the Zenodo repository of our input csv file in section “1. Data Import”. Moreover, to enhance the readability of the protocol, we added a tabular visualization of a sample of the CSV file which we use as input in our procedure.

Following another remark made by Cioffi, we added the regular expressions that we used for cleaning prefix-type and suffix-type errors in section “2. Error Analysis”. The reason why they were not present in the first release of the protocol was because the software was not entirely available yet.

Additionally, according to Cioffi, in the first version of our protocol it was not clear how the classes of errors were identified and stored in our procedure. In addition, in the review written by Arianna Moretti [2], the author suggested adding source code illustrations as well as verbose explanations of our data cleaning procedure to test the soundness of our methodology.

As a consequence, we revised the whole section 3, now called “3. Data Cleaning Procedure” in order to provide an extensive description of our data cleaning methodology. Precisely, in this section we provide:

1. a graphic visualization of the workflow involved in the cleaning procedure,
2. a textual description of each sub-step involved,
3. we present the algorithmic functions used to process data, the lines of code to be executed in order to reproduce each sub-step, and how the output of each sub-step, from start to finish, is expected to look like.

We expect, with this protocol, to explain clearly how we classify DOIs which already became valid at the time of our research, how we classify the DOIs with respect to the type of errors that were cleaned if they are not valid, how we deal with corrected DOIs, once they are cleaned, and how we store our data in a standardized format (e.g. CSV). We expect this section also to be a guide on how to reuse our software and reproduce our experiment for other developers and researchers, as also stated in our Data Management Plan [3].

Besides, to enhance the access to the data generated by our work, in the same section we provide a link to the Zenodo repositories in which we make available the first release of our software and a sample of our output data. Moreover, also in section 3, we provide the link to our Data Management Plan in order to link the protocol to further information about data sustainability and data reuse, as suggested by both Cioffi and Moretti.

As A. Moretti has pointed in her review, we added missing information about the statistics and visualization tools used in our project to present the results in section “4. Statistics and Visualizations”.

As Moretti has further suggested, we added an additional column in our CSV for storing information related to the validity of the cited DOIs and of the cleaned DOIs obtained from our procedure (see section 3 of the protocol for further details).

As advised in the same review, we added further information about the context of our research, the source of data (e.g. Crossref), and the institutions involved in the generation of the data re-used by our project (e.g. COCI) in the “Materials” section.

Finally, following the suggestions in this review, we insert the word “Protocol” in the title of our work, we chose a unique way to entitle each section of our protocol by using nominals, and we uniformed punctuation across the whole text.

We hope that our second version of the protocol meets your expectations and satisfies the requirements of reproducibility and transparency which characterize our Open Science community.

Sincerely,

Team Grasshoppers

## References

1. Alessia Cioffi. (2021). Review of: "Investigating DOIs' classes of errors". Qeios. doi:10.32388/1PV0Y2.
2. Arianna Moretti. (2021). Review of: "Investigating DOIs' Classes of Errors". Qeios. doi:10.32388/RWNSQC.
3. Boente, Ricarda, Massari, Arcangelo, Santini, Cristian, & Tural, Deniz. (2021). Classes of errors in DOI names. Zenodo. doi:10.5281/zenodo.4665853.