

This document is a rebuttal letter to the reviews received on the protocol: Nooshin Shahidzadeh, Alessia Cioffi, Arianna Moretti, Sara Coppini 2021. [Investigating Invalid DOIs in COCI](https://dx.doi.org/10.17504/protocols.io.bt5xnq7n). protocols.io <https://dx.doi.org/10.17504/protocols.io.bt5xnq7n>.

The reviews considered in this document are:

- 1) Deniz Tural. (2021). [Review of: "Investigating Invalid DOIs in COCI v1 \(protocols.io.bt5xnq7n\)"](https://doi.org/10.32388/WHWOI8). Qeios. doi:10.32388/WHWOI8.
- 2) Arcangelo Massari. (2021). [Review of: "Investigating Invalid DOIs in COCI"](https://doi.org/10.32388/X2DX81). Qeios. doi:10.32388/X2DX81.

Answers and comments provided to the suggestions and notes of the reviews are written by the very same authors of the protocol: Nooshin Shahidzadeh, Alessia Cioffi, Arianna Moretti and Sara Coppini.

Deniz Tural. (2021). Review of: "Investigating Invalid DOIs in COCI v1 (protocols.io.bt5xnq7n)". Qeios. doi:10.32388/WHWOI8

[...] since this research is based on previous studies, the importance and necessity of this have not been adequately stated in the abstract. It is just as important to understand the significance and necessity of the research as it is for the reader to understand the research.

We do understand the importance for the reader to be oriented in the best possible way within the document. For this reason, we agree on the need to explain the academic and research context in which our project fits and thus we expanded the abstract with an introductory explanation about what OpenCitations and COCI consist of.

The language used to describe the COCI Rest API is a bit ambiguous. The reader wants to be sure whether this method is used precisely.

Also in our opinion, it is evident that the explanation of the use of the API services was confusing and lacking. For this reason, we corrected the cases in which we have referred to API services other than those actually used, and specified more accurately the methods and purposes of use, also distinguishing the use of different services of the same API. Our purpose, also determined by the nature of the protocol as a document, is to clarify the readers' doubts and therefore let them have a perfect understanding of the resource. Thus, it is right to be more precise, especially on crucial steps such as the use of an API, and better explaining that this method is used precisely and consciously.

Before coming to the first section, "Reading the CSV data" after the abstract, it would be a preliminary preparation for the reader to mention how the dataset was obtained and whether it is sufficient for this research.

We greatly appreciated the precise and justified suggestion proposed, and we willingly implemented it. In fact we agreed on the necessity to apport some corrections and integrations also with respect to the first suggestion put forward in this review since it is noticeable that they refer to the same lack: that is, an accurate description of the background of the project, which is both the research context and the provider of our input data. Accordingly, we added a new section between the abstract and the first section explaining how the dataset was obtained and whether it is sufficient for this research.

In section 2 (Creating the output JSON file), the preferred steps are presented clearly and logically, but it is not clear why this method is preferred. It could be beneficial to be more descriptive for a reader who is unfamiliar with the topic.

After taking into consideration this suggestion, we recognised some inconsistencies in making explicit the motivations behind some of our choices. Indeed, it is necessary to consider also the users who are not familiar with the data formats. For this reason, we agreed on better explaining the choices in file formats.

We opted for JSON file format to store all the data answering the research questions, because this file format allows us to store more heterogeneous information in a more complex and structured way. Indeed, we stored information not directly requested by the research questions, such as the number of missing citations for each single publisher, to provide a more precise answer. For what concerns the visualizations, the JSON file format allows us to separate different kinds of data and save also data that is relevant for the graphic representations of our findings but not strictly related to the research questions (e.g. which publishers were involved in citations that were originally invalid and then validated with the DOI API request). We clarified this distinction in our updated protocol.

My recommendations are; authors should clarify why they choose this approach and measures for those who are not familiar with the topic in a short statement.

Broadly speaking, we consider this a useful suggestion to keep in mind throughout our updates and future work. Methods and the general approach are certainly among the most relevant elements to make clear in a document such as a protocol, so we clarified both of them in a more precise way in each step of the resource where they were not given sufficient attention.

Arcangelo Massari. (2021). Review of: "Investigating Invalid DOIs in COCI". Qeios. doi:10.32388/X2DX81.

The research question appears clear and justified, but there seems to be a discrepancy with the answers obtained: the findings, in fact, are not limited to reporting the names of the publishers, but compose some sort of ranking of the most deserving publishers, based on how many correct metadata were sent, thus answering a question that was not originally asked. Therefore, it is advisable either to change the question or to change the answer, in order to make it clear to the reader what to expect from the research work.

We recognise the relevance of the highlighted lack, since a discrepancy between premises and outputs could be first of all misleading in the presentation of the results of our research and also confusing for any possible further user of the protocol. However, we decided to include some additional output files for the following reasons:

1. To give more complete answers to our research questions;
2. To organize the collected data in a way that would be suitable to visualize the information in a comprehensive way, showing all the aspects of the findings of our research and the evolution of the data from the input file creation and the moment of our second check of the DOIs' validity;
3. To exploit the additional data we retrieved beyond the ones that were strictly required in the research questions.

However, we recognised the inconsistency outlined in this passage of the review of our protocol, and thus we clarified this aspect in the introductory part of the resource. In any case, we sincerely welcomed the comment: it was very useful to understand how to update the protocol in order to avoid confusion.

In the abstract it is said that the COCI REST API will be used to obtain information about the publishers, while the actual protocol uses exclusively the Crossref API for this purpose.

About this aspect, we rearranged the abstract by replacing COCI REST API with Crossref's one, since we realized that referring to this latter was structurally more correct, since COCI is based on Crossref data. Indeed, we appreciated this relevant note, which is really similar to that advanced by Deniz Tural. As already stated, we amended the incorrect references to APIs.

The research does not offer any bibliographic references about previous works on this problem, and there is no mention of what further contribution is given by the protocol.

We agreed with the critics proposed about the lack of references, since these latter would help the readers know the background of our research. In the new version of our protocol this problem was addressed and we added the necessary bibliography. We also specified what further contribution is given by this research and this specific resource in particular in the abstract section of the updated protocol.

The workflow turns out to be just partially technically valid, as it seems to lead to significant results regarding the validation of DOI names and the recovery of publishers in the case of a valid prefix, while it appears incomplete in the case of an invalid prefix. The procedure described consists in searching Crossref for the references of the citing DOI and using the relative metadata to conduct a bibliographic query on Crossref itself. However, it is not clear how this can lead to the correction of the original wrong DOI.

We willingly welcomed these observations, since they provided us an occasion to re-analyse in detail our research scenario, with the aim of finding an optimal compromise solution between not exceeding the limits of our research questions and achieving optimal results in terms of utility.

First of all, we'd like to clarify that the attempt of correcting invalid DOIs was never explicitly mentioned among the tasks to be accomplished, according to our research questions. In fact, our primary aim was that of finding out, where possible, the publishers to which the referenced DOIs of the invalid citations point to. In other words, the focus was on the identification and retrieval of the publishers, while the work on DOIs was supposed to be limited to their validation through a REST API service.

However, the idea of a local correction of the DOIs which result to be still invalid at the time of our API request represents an interesting possible development of our project. Further, as suggested by Arcangelo Massari, this improvement could be achieved by exploiting the algorithm projected by our colleagues (see: *Protocol: "Investigating DOIs classes of errors" V.(bufwntpe)*¹), which explicitly addresses the problem of DOIs' fixation. We believe that this integration might represent a significant added value for the following reasons:

1. It could allow the identification of the publishers which would be otherwise excluded from our results because of an error in the prefix of the cited DOI of a given publication, since the mechanism we rely on for the publishers' retrieval is based on their identification throughout a request to a dedicated Crossref API

¹ Ricarda Boente, Deniz Tural, Cristian Santini, Arcangelo Massari . Protocol: "Investigating DOIs classes of errors". protocols.io

<https://protocols.io/view/protocol-34-investigating-dois-classes-of-errors-3-bufwntpe>

- service which exploits DOIs' prefixes.
2. It may provide some qualitative information about the reasons why some citational data were invalid at the time of the creation of the original input file, i.e.: "invalid_dois.csv".
 3. As an effect of the previous two points, it could enhance the quality of the results of our primary aim research both qualitatively and quantitatively.
 4. It would enhance an effective collaboration among the authors of two projects which are profitably conceivable as complementary.

In addition to that, we took into account the fact that the DOIs that resulted to be invalid both at the time of the creation of the input file and at the time of the DOI API request performed throughout our algorithm should be double checked through another API request after the correction. At this point, it is likely that a number of the previously invalid DOIs will be validated. About this aspect, we collectively discussed the implications of the corrections in terms of publishers' responsibility attributions in the communication of citational data and we finally concluded that — in the likely case the integration will be included in our software — it would be necessary to clearly distinguish the DOIs — and consequently the citations — that will have become valid after our first check through a DOI API service from the ones that will have been validated only after the correction.

Finally, for all the aforementioned reasons, we decided to address this integration in the section in which we are going to present the further developments of our project. However, all of us authors agreed on the necessity to give priority to the main objectives we are supposed to accomplish within the purpose of the current research, and then focus on enriching and improving the quality of our results by realizing the integration mentioned above.

References

- 1) Deniz Tural. (2021). [Review of: "Investigating Invalid DOIs in COCI v1 \(protocols.io.bt5xnq7n\)"](#). Qeios. doi:10.32388/WHWOI8.
- 2) Arcangelo Massari. (2021). [Review of: "Investigating Invalid DOIs in COCI"](#). Qeios. doi:10.32388/X2DX81.
- 3) Nooshin Shahidzadeh, Alessia Cioffi, Arianna Moretti, Sara Coppini 2021. [Investigating Invalid DOIs in COCI](#). protocols.io
<https://dx.doi.org/10.17504/protocols.io.bt5xnq7n>.
- 3) Ross-Hellauer, T. (2017). [What is open peer review? A systematic review](#). F1000Research, 6, 588. <https://doi.org/10.12688/f1000research.11369.2>