

## **If Data is Used in the Forest and No-one is Around to Hear it, Did it Happen? A Citation Count Investigation**

### **Abstract**

In this article I describe the process and results of tracking a citation from a data repository through the article publication process and trying to add a citation event to one of our DOIs. I also discuss some other confusing aspects related to citation counts as indicated in various systems, including reference managers, the publisher's perspective, aggregators, and DOI minters. I discovered numerous problems with citations. Addressing these problems is important as citations can be key to determining both the original use and reuse of a dataset, especially for repositories that do not track usage by requiring people to login or provide an email to download a dataset. The lack of transparency in some data citation systems and processes obscures how and where data is being used.

*Submitted 15 June 2022 ~ Revision received 19 December 2022 ~ Accepted 9 April 2023*

Correspondence should be addressed to: [sborda@umich.edu](mailto:sborda@umich.edu)

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



## Introduction

In my role as the Data Workflows Specialist at the University of Michigan Library, I reviewed large datasets and code deposits and developed data management workflows for researchers. I also supported various aspects of our research data repository, Deep Blue Data,<sup>1</sup> which is based on the Samvera Hyrax platform.<sup>2</sup> In this capacity, I worked with our developers to improve connections between our system and others to gather various metrics including Altmetrics<sup>3</sup> and citation counts for our datasets.

Deep Blue Data is a general institutional data repository for open-access data that does not require users to log in before downloading a dataset. While this feature reduces barriers to accessing our data, it can make it challenging to understand how, where, or if Deep Blue Data datasets are reused. Because of this issue, when I began looking into citation counts, I started with the “original use” of datasets, i.e., researchers citing their own datasets, to verify if the citation event process was working. Directly tracking these citations offered a more approachable option than using COUNTER.<sup>4</sup> Nonetheless, I learned that the process of capturing data citation events was not as straightforward as one would expect.

This paper follows my quest to understand how citation events are captured. This includes the important aspects of the citation format itself and the role that reference managers play in maintaining that format, the publisher process, interaction between the publisher and other systems such as Crossref,<sup>5</sup> the latter’s relationship with DataCite (who mints our dataset DOIs),<sup>6</sup> and DataCite’s accumulation of these events.

One method of capturing and displaying citation counts, particularly at “point-of-need,” is to use the DataCite Data Metrics Badge shown in Figure 1.<sup>7</sup> In early January 2021, we added Data Metrics Badge to our repository. A few weeks later, I checked some of our more popular datasets to see if they were showing any citations and discovered they were not. Indeed, none of our more than 400 datasets showed citations.



**Figure 1** DataCite Data Metrics badge frontend with no indication of citations.

I investigated further by spot-checking datasets and found that although there were articles related to these datasets, as indicated in our “Citations to Related Materials” field, researchers did not appear to be citing their datasets in the “References” section of their papers. In some cases, they mentioned their dataset, complete with DOI, in the abstract of the paper. In other cases, they only indicated the existence of the dataset in the “Data Availability” statement rather than citing the dataset in both the “Data Availability” section as well as the “References” section (Ball and Duke, 2015). I assumed these errant practices were the root cause of missing citations. After researching the issue, I determined they are only partially to blame. Figure 2 shows my

<sup>1</sup> Deep Blue Data: <https://deepblue.lib.umich.edu/data>

<sup>2</sup> Hyrax: <https://hyrax.samvera.org/>

<sup>3</sup> Altmetric: <https://www.altmetric.com/>

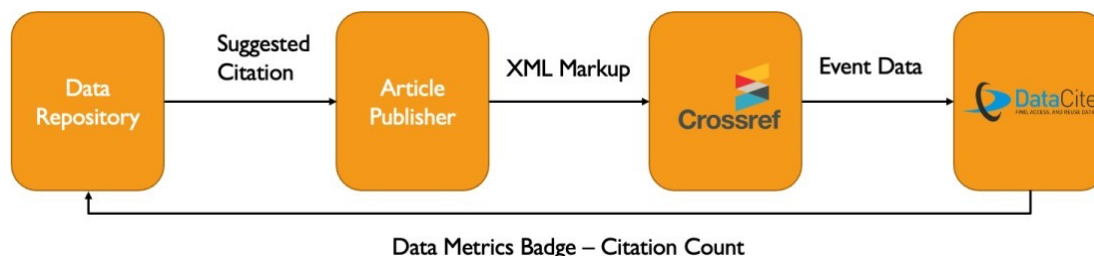
<sup>4</sup> COUNTER: <https://www.projectcounter.org/counter-release-5/>

<sup>5</sup> Crossref: <https://www.crossref.org/>

<sup>6</sup> DataCite: <https://datacite.org/>

<sup>7</sup> DataCite Data Metrics Badge: <https://support.datacite.org/docs/displaying-usage-and-citations-in-your-repository>

rough understanding at that point of the Data Citation Pipeline and the major players in my investigation.



**Figure 2** The Data Citation Pipeline.

## Problem Investigation

My investigation into the “0” citation count began in earnest when a researcher who I had been encouraging to cite his data in the “References” section informed me that he had just submitted a paper with the data cited in this manner. This paper provided me with a “known” example to track. The next step entailed waiting several months until the publication process was complete to see if the citation was recognized or not.

In case the “normal” citation count process was not working, I tried using “Contribute Citations”<sup>8</sup> to our DataCite DOIs as mentioned by Cousijn et al. in the section “Contributing data citations: data repositories” (2019) as well as by the “Make Data Count” movement.<sup>9</sup> I reasoned that it would at least indicate whether articles and datasets were linked somehow. I followed the instructions using the DataCite API with a JSON payload indicating that the DataCite dataset DOI “IsCitedBy” (Starr & Gastl, 2011) an article DOI. See Figure 3 for an example of a JSON payload.<sup>10</sup>

```
{
  "data": {
    "attributes": {
      "relatedIdentifiers": [
        {
          "relationType": "IsCitedBy",
          "relatedIdentifier": "10.1029/2019JC015306",
          "relatedIdentifierType": "DOI"
        }
      ]
    }
  }
}
```

**Figure 3** JSON payload indicating that dataset DOI 10.7302/dbfp-s644 “IsCitedBy” article DOI 10.1029/2019JC015306 submitted to DataCite via API.

<sup>8</sup> Contribute Citations: <https://support.datacite.org/docs/contributing-data-citations>

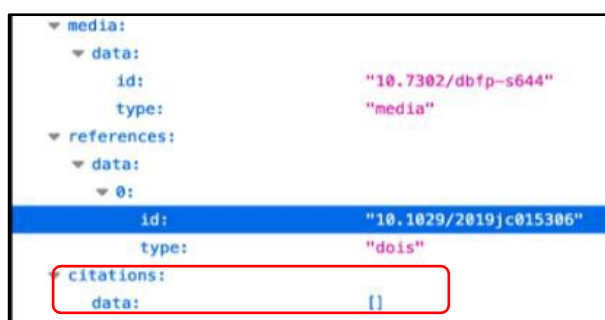
<sup>9</sup> Make Data Count: <https://makedatacount.org/data-citation/>

<sup>10</sup> Updating DataCite DOI Metadata with JSON Payload: <https://support.datacite.org/docs/updating-metadata-with-the-rest-api>

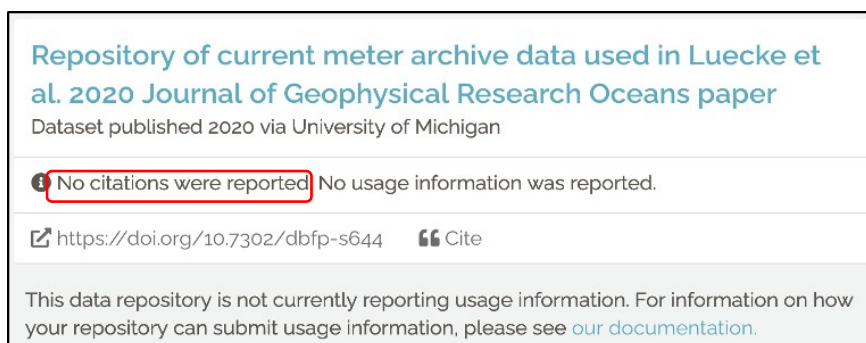
I used the following Curl command (run from the Mac Terminal):

```
curl -v -X PUT -H "Content-Type: application/vnd.api+json"
-- user[redacted user]:[redacted passwd] -d
@doi_update.json
https://api.datacite.org/doi/10.7302/dbfp-s644
```

The first problem appeared at this point. After submitting the JSON payload via the DataCite API, there were no citations listed in the “Citations” field in the API results, as shown in Figure 4.<sup>11</sup> In addition, there were no new citations displayed on the DataCite Search results page for DOI 10.7302/dbfp-s644 as shown in Figure 5.<sup>12</sup>



**Figure 4** Result of “IsCitedBy” payload in “references” not “citations.”



**Figure 5** DataCite Search results for DOI 10.7302/dbfp-s644.

Through further investigating, testing, and help from the DataCite Support staff, I discovered a bug in DataCite’s system: using “IsCitedBy” causes the new citation to end up in the “References” section of the DOI, not in the “Citations” section.

In attempt to get a citation to show up, I used different “payloads” and checked other repositories that use DataCite and indicate citation counts. When I noticed that Dryad’s DOIs in the DataCite API view were using “Cites” instead of “IsCitedBy”, I promptly modified my

<sup>11</sup> DataCite API results: <https://api.datacite.org/doi/10.7302/dbfp-s644>

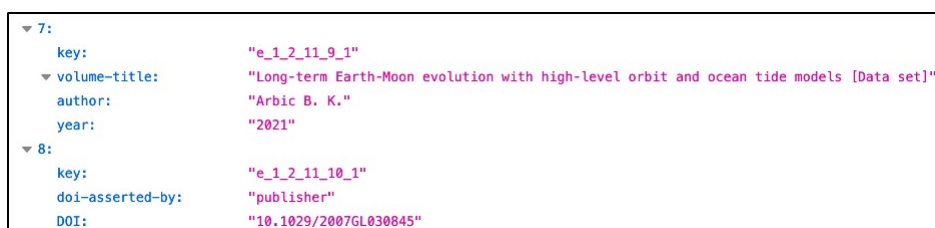
<sup>12</sup> DataCite Search: <https://search.datacite.org/works?query=10.7302%2Fdbfp-s644>

“payload” accordingly and it worked. A single citation count was indicated in the “Citations” section. DataCite Support eventually created a DataCite GitHub issue for this bug.<sup>13</sup>

The final version of the original article I had been tracking was published in early November 2021 with its reference section hyperlinked.<sup>14</sup> This allowed me to see how the dataset citation was formatted and whether it would be recognized as an actual citation of the dataset with the DataCite Data Metrics Badge in the Deep Blue repository. Much to my amazement and chagrin however, the dataset citation went from displaying the DOI in the references list, as shown in Figure 6, to completely ignoring the DOI altogether in the “Markup” view in Crossref API results, as shown in Figure 7 as number “7”. It should be noted that this was how it looked in early November 2021.



**Figure 6** Dataset citation fully hyperlinked in the article references list.



**Figure 7** Markup view the dataset citation in the Crossref API results.

To provide another interesting example of publishers not handling datasets as well as articles, the Google Scholar link (circled in Figure 5), resolves to the AGU/Wiley article itself in a circular manner, not to the dataset itself (see Figure 8).<sup>15</sup> Rather than pointing to Google Scholar, it seems that AGU/Wiley could point to DataCite Commons,<sup>16</sup> (they have links to Crossref for other citations) or better yet Google Dataset Search, (Figure 9).<sup>17</sup> It should be noted that best practices were not followed in the naming of this dataset, where “Dataset for <article title>” should have been used. In discussion with an AGU representative, I learned that Google Dataset Search was not currently one of the options for linking in the references.

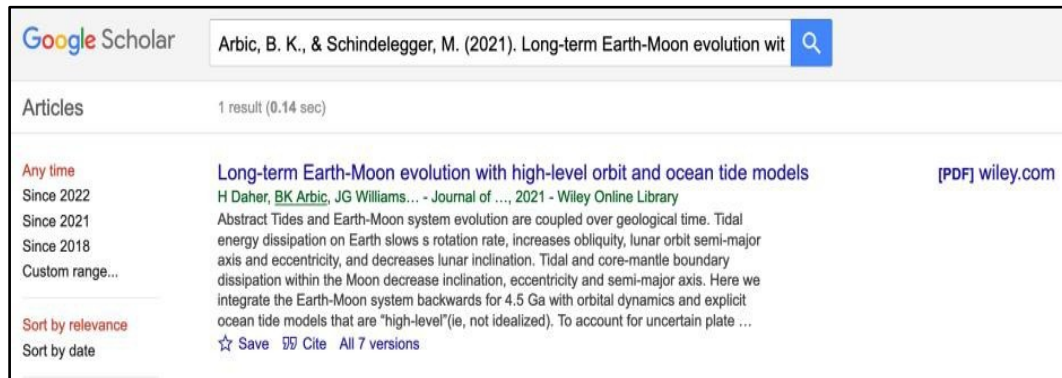
<sup>13</sup> DataCite GitHub issue #1416: <https://github.com/datacite/datacite/issues/1416>

<sup>14</sup> Long-Term Earth-Moon Evolution with High-Level Orbit and Ocean Tide Models <https://doi.org/10.1029/2021JE006875>

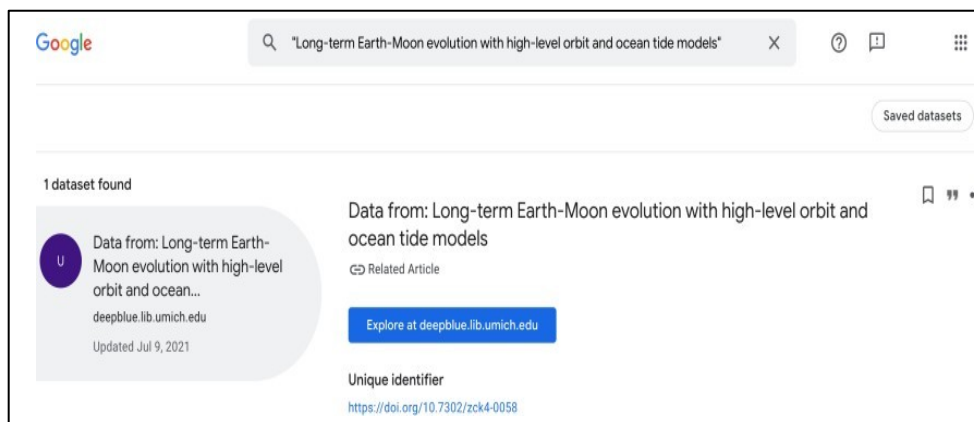
<sup>15</sup> Link to Long-term Earth-Moon article in Google Scholar: <https://scholar.google.com/scholar?hl=en&q=Arbic%2C+B.+K.%2C+%26+Schindelegger%2C+M.+%282021%29.+Long%2E%80%90term+Earth%E2%80%90Moon+evolution+with+high%2E%80%90level+orbit+and+ocean+tide+models+%5BDataset%5D.+University+of+Michigan%2E%80%93Deep+Blue+Data.+https%3A%2F%2Fdoi.org%2F10.7302%2FZCK4-0058>

<sup>16</sup> DataCite Commons: <https://commons.datacite.org/doi.org/10.7302/zck4-0058>

<sup>17</sup> Link to Long-term Earth-Moon dataset in Google Dataset Search: <https://datasetsearch.research.google.com/search?query=%22Long-term%20Earth- Moon%20evolution%20with%20high- level%20orbit%20and%20ocean%20tide%20models%22&docid=L2cvMTFwdmgYmJjEYzA%3D%3D>



**Figure 8** Google Scholar links to article in AGU not the dataset itself.



**Figure 9** Screen shot of Google Dataset Search results for dataset title “Long-term Earth-Moon evolution with high-level orbit and ocean tide models.”

Armed with information about how dataset citations appeared to be processed — at least as compared to article citations at Wiley — I reached out to Shelley Stall, Senior Director of Data Leadership at AGU, to inquire about this second issue since its markup was very different from how article citations were handled. It turned out that AGU and several other publishers were working a on fix for this very problem and planned to implement it later in the same year. Figure 10 shows the citation after the initial fix (Stall et al., 2022) in Crossref.<sup>18</sup> Unfortunately, the AGU/Wiley initial fix does not help DataCite to see this citation mention in the references as a citation to the dataset in DataCite Commons frontend (Figure 11).<sup>19</sup>

```
<citation key="e_1_2_11_9_1">
  <unstructured_citation>Arbic B. K. &Schindelegger M. (2021).Long-term Earth-Moon evolution with high-level orbit and ocean tide
  models[Dataset].University of Michigan–Deep Blue Data.https://doi.org/10.7302/ZCK4-0058</unstructured_citation>
</citation>
```

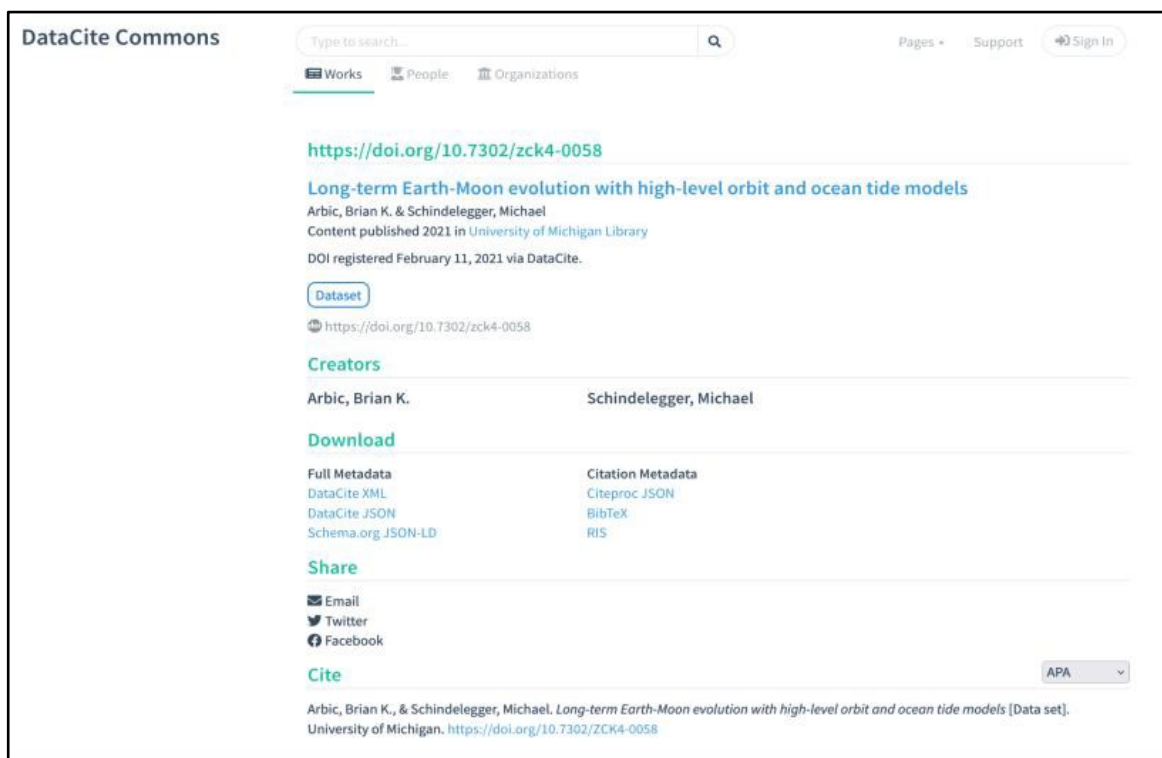
**Figure 10** The dataset citation markup after the initial fix.

<sup>18</sup> Crossref JSON results for DOI 10.1029/2021JE006875:

<https://api.crossref.org/v1/works/10.1029/2021je006875>

<sup>19</sup> DataCite Commons results for DOI 10.7302/zck4-0058:

<https://commons.datacite.org/doi/10.7302/zck4-0058>



**Figure 11** DataCite Commons frontend with 0 citations indicated.

Even though our researcher was following “Wiley’s Data Citation Policy”<sup>20</sup> by including [Dataset] in his citation it was being missed, which meant Crossref was not including it in their Event Data.<sup>21</sup> Consequently, DataCite would not know there was a citation to be counting. The following problem statement describes what was happening in this and other datasets in our repository:

‘Problem Statement:

When the journal production process validates these citations, the machine actionable version of the citation is – in present practice – not always handled optimally, and the integrity of the machine-actionable citation may be damaged. In such cases, the citation will not be able to give automated attribution and credit to the software or dataset authors, nor will it support links to the other persistent identifiers associated with the article.’ (Stall et al., 2022)

## Current state of Issues

Since my initial investigations in 2021, I am pleased to report there have been fixes and improvements made.

<sup>20</sup> Wiley’s Data Citation Policy: <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-citation-policy.html>

<sup>21</sup> Crossref Event Data: <https://www.crossref.org/services/event-data/>



## IsCitedBy

Now that DataCite fixed the “IsCitedBy” bug, we can see the results of pushing the “IsCitedBy” payload as described in Figure 1. In the DataCite API view,<sup>22</sup> “IsCitedBy” shows as the “relationType” (see Figure 12) in addition to appearing correctly in the “citations” section (see Figure 13).

```

▼ 7:
  relationType:      "IsCitedBy"
  relatedIdentifier: "https://doi.org/10.1029/2019JC015306"
  relatedIdentifierType: "DOI"

```

**Figure 12** relationType "IsCitedBy" for DOI 10.7302/dbfp-s644.

```

▼ relationships:
  ▶ client:      {...}
  ▶ provider:    {...}
  ▶ media:       {...}
  ▶ references:  {...}
  ▼ citations:
    ▼ data:
      ▼ 0:
        id:      "10.1029/2019jc015306"
        type:    "dois"

```

**Figure 13** "IsCitedBy" DOI payload now showing correctly in “citations.”

Finally, “1 citation” appears in the DataCite Search results (Figure 14).<sup>23</sup>

**Figure 14** DataCite Search results for 10.7302/dbfp-s644 now showing 1 citation.

<sup>22</sup> DataCite API view: <https://api.datacite.org/doi/10.7302/dbfp-s644>

<sup>23</sup> DataCite Search results for DOI 10.7302/dbfp-s644: <https://search.datacite.org/works?query=10.7302%2Fdbfp-s644>



The Data Metrics Badge in Deep Blue Data also shows “1” (Figure 15).



**Figure 15** Data Metrics Badge showing 1 citation.

## Publisher Fix

The publisher/metadata fix was released in late May of 2022 and in July 2022 the fix was processed for the DOIs I have been tracking, as shown in the Crossref JSON of Figure 16.<sup>24</sup>

```

  7:
    key: "e_1_2_11_9_1"
    doi-asserted-by: "publisher"
    unstructured: "Arbic B. K. &Schindelegger M.(2021).Long-term Earth-Moon evolution with high-level orbit and ocean tide models[Dataset].University of Michigan-Deep Blue Data.https://doi.org/10.7302/ZCK4-0058"
    DOI: "10.7302/ZCK4-0058"
  
```

**Figure 16** Result after Publisher “fix” in Crossref - indicating DOI field with “10.7302/ZCK4- 0058”.

With the DOI split out and the citation type indicated, Crossref can update Event Data. Event Data would then be read by DataCite and the reference should show in the “referenceType” section.

## Other issues of note

In addition to the issue with publishers that I have described, there are issues with data citations in other parts of the pipeline. One problem is in how repositories sharing the data are making the suggested citation available. Another is that reference managers do not make it easy to create or manage a data citation through to the other end of the pipeline with data aggregators, nor do they make it clear whether they are sharing data citation metadata correctly.

The Deep Blue Data repository provides a suggested citation via plain text in the metadata section of the deposit (see Figure 16). The suggested citation must be highlighted and selected with a pointing device and copied to the clipboard to be used. Deep Blue Data does not currently have integrations with reference managers, nor does it make the citation available in other formats such as Biblatex.<sup>25</sup> From the Wiley’s standpoint, having [Data set] or [Dataset] in the citation is critical for their process as it is or should be an immediate indicator that this is indeed a dataset and should be handled as such (Federer, 2020).

<sup>24</sup> Crossref JSON for DOI 10.1029/2021je006875:  
<https://api.crossref.org/v1/works/10.1029/2021je006875>

<sup>25</sup> Biblatex: <https://ctan.org/pkg/biblatex>

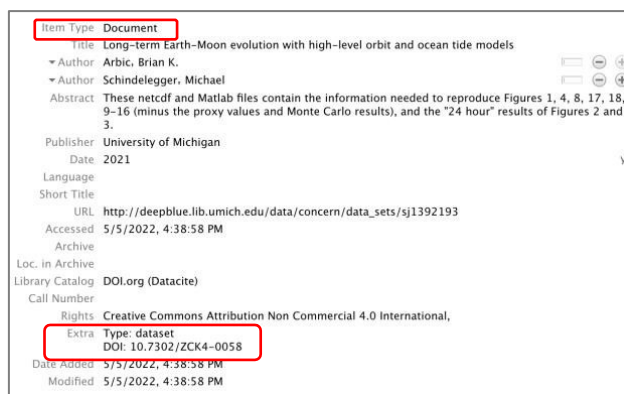
To Cite this Work:  
 Arbic, B. K., Schindelegger, M. (2021). *Long-term Earth-Moon evolution with high-level orbit and ocean tide models* [Data set], University of Michigan - Deep Blue Data. <https://doi.org/10.7302/zck4-0058>

**Figure 17** Suggested citation for dataset 10.7302/zck4-0058 in Deep Blue Data

## Reference Managers, how are they handling data citations?

As indicated previously, the citation component of [Data set] or [Dataset] is key to downstream processing. Unfortunately, not all reference managers handle this equally. With Zotero, for instance, the results can differ depending on how you use the tool. Per my exchange with Zotero support, the “Browser plug-in” option pulls metadata information from COinS (ContextObjects in Spans) first and then from “meta tags.”<sup>26</sup>

Alternatively, using the Zotero desktop application and providing a DOI via the “Add Item(s) by Identifier” option will cause the metadata to be pulled from DataCite. Figure 17 is an example that shows the results of entering 10.7302/ZCK4-0058. In this case, the dataset is imported as “Item Type: Document” because Zotero does not currently have a “Data” Item Type. To address this issue, Zotero has implemented a temporary measure in the form of a field called “Extra” that indicates “Type: dataset” and “DOI: 10.7302/ZCK4-0058.”



**Figure 18** Screen shot from Zotero desktop “Add item by Identifier.”

Using the Zotero plug-in for Microsoft Word to “Add/Edit Citation” results in a citation that includes the [Dataset] designation. For an example, see (Arbic & Schindelegger, 2021) in the “References” section of this paper. Because the metadata in DataCite differs slightly from the suggested citation in Deep Blue Data, the entry in the References section is not the same as the suggested citation which is also less than ideal (see Figure 16).

Mendeley does not handle datasets well at all. It has no option to add a dataset by entering a DOI nor does the system does not search DataCite. Mendeley does not have a “Dataset” Document Type either—a fact confirmed by Mendeley Support in this response via email: “You can add the dataset into your library first then cite it using the citation plugin. Since there's no Dataset document type, you can try the Generic document type.”

Biblatex has had an option for datasets using @dataset since at least 2019.<sup>27</sup>

It should be noted that the researcher who submitted this dataset, Dr. Brian Arbic, uses Overleaf,<sup>28</sup> a LaTeX editor supporting Biblatex, when preparing his papers. This way he avoided using Zotero or Mendeley and their issues with “type.”

<sup>26</sup> Using Zotero to Create Metadata: [https://www.zotero.org/support/dev/exposing\\_metadata/coins](https://www.zotero.org/support/dev/exposing_metadata/coins)

<sup>27</sup> Biblatex Documentation: <http://linorg.usp.br/CTAN/macros/latex/contrib/biblatex/doc/biblatex.pdf>

<sup>28</sup> Overleaf LaTeX editor: <https://www.overleaf.com/>

In my cursory examination of citation styles as options of Zotero export, only APA supports the use of [Dataset] which is pulled automatically from the “Extra” field. Using another style requires including [Dataset] to ensure it won’t be missed downstream.

The concern with some of these manual copy/paste data entry issues and data not being handled natively as a “type” is that the designation [Data set] or [Dataset] might be lost along with the DOI.

## Data Availability Statements

Although using “Data Availability” statements might suggest a possible solution, it does not solve the problem. Crossref does not currently check the “Data Availability” statement for dataset DOIs and my investigation indicates that information is not currently passed from the publisher to Crossref (at least not in the case of AGU/Wiley).

For evidence of this issue, I tracked a dataset DOI, <https://doi.org/10.7302/fn7r-hq31> mentioned in a “Data Availability” statement but not in the “References” section in an article in Nature (Figure 18).<sup>29</sup>

<b>Data availability</b>
<a href="https://doi.org/10.7302/fn7r-hq31">https://doi.org/10.7302/fn7r-hq31</a>
<b>Code availability</b>
<a href="https://doi.org/10.7302/fn7r-hq31">https://doi.org/10.7302/fn7r-hq31</a>

**Figure 19** Data availability statement for article <https://doi.org/10.1007/s00382-022-06257-6>.

Unfortunately, there is no indication that data availability statements in the metadata are shared with Crossref.<sup>30</sup>

## Dataset Aggregators

Dataset aggregators are on the other end of the dataset publishing pipeline and have their own problems because they are highly dependent on other systems earlier in the pipeline. Interestingly but not surprisingly, the Google Dataset Search appears to be using their massive computing power to mine Google Scholar for any mentions of dataset DOIs.<sup>31</sup> For example, one of our more popular datasets from the perspective of downloads and hits is the RIGA dataset.<sup>32</sup> According to Google Dataset Search, it has 50+ citations. The article that is listed as a “related resource” in our repository only mentions the dataset DOI in the “Abstract” part of the paper.

Clarivate – Data Citation Index<sup>33</sup> seems to be somewhat unreliable for citation counts. For instance, when using “Deep Blue Data” as a search term in “Data Source,” it returns 4,000+ entries (see Figure 19). This is far more than we have in our system (400+). Upon researching

<sup>29</sup> Data Availability Statement: <https://link.springer.com/article/10.1007/s00382-022-06257-6#data-availability>

<sup>30</sup> Crossref metadata: <https://api.crossref.org/v1/works/10.1007/s00382-022-06257-6>

<sup>31</sup> Google Dataset Search: <https://datasetsearch.research.google.com/>

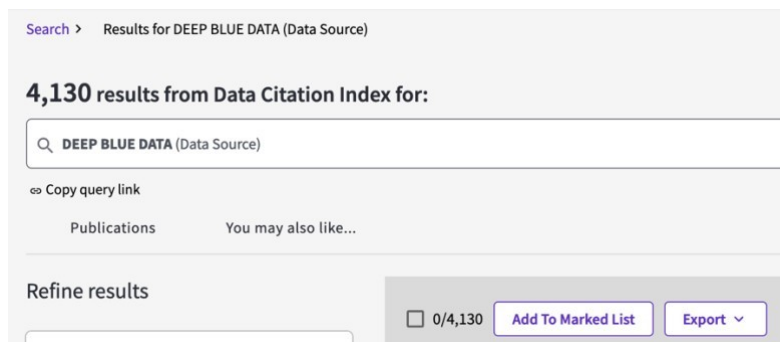
<sup>32</sup> RIGA Data:

[https://datasetsearch.research.google.com/search?src=0&query=10.7302%2FZ23R0R29&docid=L2cvMTFz\\_bXcyeTN3Nw%3D%3D](https://datasetsearch.research.google.com/search?src=0&query=10.7302%2FZ23R0R29&docid=L2cvMTFz_bXcyeTN3Nw%3D%3D)

<sup>33</sup> Clarivate Data Citation Index: <https://clarivate.com/webofsciencegroup/solutions/webofscience-data-citation-index/>

this discrepancy, I discovered most of the so-called “datasets” were actually articles in another repository, “Deep Blue Documents.”

In addition, there are some actual datasets that are showing “1” citation and, for at least some of these, that citation was in fact listed in the “References” section of the related article. Unfortunately, in this case the citation is not showing in DataCite and therefore is not actually showing in our system as a citation count. It is not clear how the Data Citation Index is getting these citation counts, there are vague mentions of OIA-PMH to fill in metadata and references for citations in their white paper.<sup>34</sup>



**Figure 20** Deep Blue Data results in the Clarivate Data Citation Index

## Conclusion

My exploration was extremely small in scope. It should be noted that the purity of my process and results were somewhat compromised in the middle of this investigation because the Deep Blue Data repository was a participant in the RADS project.<sup>35</sup> As a result, most datasets that were published before March 31, 2022 had their DOI metadata “relationType” metadata fields including “IsReferencedBy” updated via script. Also, as of Fall 2022, Deep Blue Data has yet to implement DataCite’s Data Metrics badge, so it will not be visible if visiting datasets from this paper.

Nevertheless, the number of issues I encountered even in the small sample used in this study suggest these problems may discourage researchers from citing their data or, at least, citing it properly. My research also reveals that those of us in the data repository field cannot assume these citation and reference processes are working effectively.

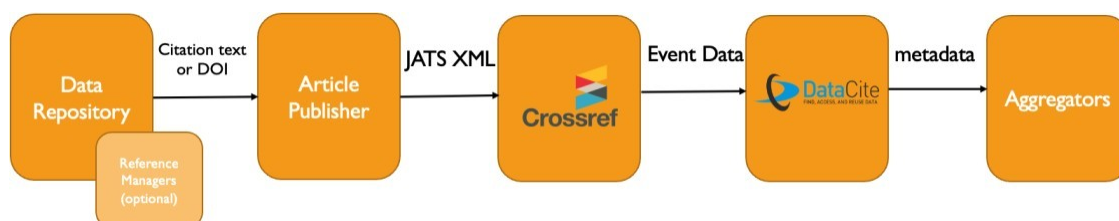
<sup>34</sup> Data Citation Index white paper:

[https://clarivate.com/wp-content/uploads/dlm\\_uploads/2023/03/XBU1005829680-Data-Citation-Index\\_Whitepaper-V4.pdf](https://clarivate.com/wp-content/uploads/dlm_uploads/2023/03/XBU1005829680-Data-Citation-Index_Whitepaper-V4.pdf)

<sup>35</sup> RADS project blog:

<https://metadatagamechangers.com/blog/2022/3/7/ivfirlw6naf7am3bvord8pldtuyqn4r>

We need to test, check, and provide feedback to publishers, DOI organizations, and other systems in the citation pipeline if we want these systems to perform effectively. I have included an updated and slightly more complete Data Citation Pipeline to include additional data sources and metadata forms, including Reference Managers, JATS XML <sup>36</sup> and Aggregators - see Figure 21.



**Figure 21** Updated Data Citation Pipeline

In addition, pushing aggregation databases such as Data Citation Index to be more transparent about their sources for dataset citation counts would help provide a more accurate picture of dataset reuse.

It would be great to see updated recommendations of steps intended “to increase data citation and develop metrics” e.g., as detailed in Federer’s 2020 paper, “Measuring and Mapping Data Reuse: Findings from an Interactive Workshop on Data Citation and Metrics for Data Reuse.” These could go a long way toward resolving these issues and improving the data citation process.

## Acknowledgements

I would like to acknowledge Shelley Stall and Chris Erdman from AGU for their valuable insights and information about this issue from the publisher’s end. And I want to thank the kind support folks at DataCite (Liz Krznarich, Mary Hirsch, and Kelly Stathis) and Martyn Rittman, and other support from Crossref and Sebastian Karcher from Zotero for answering my questions throughout this process. I would also like to thank Jake Carlson for coming up with the title for this paper and for supporting me in this rather absorbing endeavour.

## References

- Arbic, B. K., & Schindelegger, M. (2021). *Long-term Earth-Moon evolution with high-level orbit and ocean tide models* [Data set]. University of Michigan. <https://doi.org/10.7302/ZCK4-0058>
- Ball, A. & Duke, M. (2015). ‘How to Cite Datasets and Link to Publications’. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: <https://www.dcc.ac.uk/guidance/how-guides/cite-datasets>
- Cousijn, H., Feeney, P., Lowenberg, D., Presani, E., & Simons, N. (2019). Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal*, 18(1), 9. DOI: <http://doi.org/10.5334/dsj-2019-009>

<sup>36</sup> JATS XML: <https://jats.nlm.nih.gov/>

- Federer, L. (2020). Measuring and Mapping Data Reuse: Findings From an Interactive Workshop on Data Citation and Metrics for Data Reuse. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.ccd17b00>
- Starr, J., & Gastl, A. (2011). isCitedby: A metadata scheme for DataCite. *D-Lib Magazine*, 17(1–2). <https://doi.org/10.1045/JANUARY2011-STARR>
- Stall, Shelley, et al. (2022). Journal Production Guidance for Software and Data Citations. *ESS Open Archive*. <https://doi.org/10.22541/essoar.167252601.17695321/v1>