

Response to reviewer comments

Dear Dr. Callaghan,

Thank you for accepting our commentary article, “Common sense approaches to sharing tabular data alongside publication”, for publication in Patterns.

We thank the reviewers for time and detailed comments. We have addressed the final comments from the reviewers, our response to individual reviewer comments are presented in the table on page 2 of this letter.

Regards,

Nicholas Tierney & Karthik Ram

ID	comment	response
R1-1	<p>The authors have addressed the specific points in my earlier review; and I thank them for that. Despite this, I must say that I still find the overall tone of the piece to be very normative and somehow disconnected from other recommendations for data sharing, data citation and data management. This is particularly evident in the brief section in 2.7 on data citation. Specifically I found myself looking for a reference for the following sentence or wondering if this is the authors' own recommendation: "When citing data, you should cite the preferred CITATION file in the first instance, else, if there is a data publication, (for example in Scientific Data), cite this, and if none of those apply, cite the repository in which the data is deposited (e.g., a Zenodo repository)." Does this come from one of the existing data citation recommendations, e.g. from FORCE11?</p>	<p>We appreciate the reviewer catching this issue. We have clarified the role of citation files based on FORCE 11 recommendations. We specifically cite work from the Make Data Count project to support and clarify these statements. The paragraph now reads: "One critical advantage of depositing data in a trusted repository (Section 2.6) is that it comes with many benefits, such as preservation, indexing, discoverability, and most importantly, citation with a persistent identifier. Authors should add a CITATION file when possible but the real burden lies with publishers to ensure that these citations are marked up appropriately to allow for the building of open data metrics.³⁶ When citing data, you should cite the preferred CITATION file in the first instance, else, a data publication, (for example in Scientific Data, Ecology), and if none of those apply, cite the repository in which the data is deposited (e.g., a Zenodo repository) in the references section of related papers."</p>
R1-2	<p>I also struggled with the final paragraph in section 2.6. The presentation of "data papers" is still a bit off (p.8). As the authors state, data papers explain in detail the creation of archived data. What is unclear is why it would be expected that data papers should provide "guidelines on how to structure data for effective reuse and sharing." This is of course true (although some data papers, e.g. those in Scientific Data - which the authors include as an example - can have a section on how to reuse the archived data which is described in the data paper), but why mention this? Including this sentence as it is currently phrased makes it somehow seem as if the authors are not that familiar with data papers.</p>	<p>To clarify our point here, we have reworked the sentence to indicate that it is the journals, not the data papers that lack guidelines. The sentence was changed from: "Writing a data paper can be very useful for researchers, but, data papers lack guidelines on how to structure data for effective reuse and sharing." to "While writing a data paper can be very useful for researchers, **journals that publish data papers** lack guidelines on how to structure data for effective reuse and sharing."</p>

(continued)

ID	comment	response
R1-3	The final two sentences in this paragraph (2.6) also seem disconnected and almost like an advertisement for Dryad. Could the authors make a more general statement, i.e. "a trusted data repository such as Dryad can help to... "?	We appreciate the reviewers comment that this sounds like an advertisement for Dryad. We have fixed as follows: "Journals provide data papers in a formal publication explaining data that is archived (e.g. Ecology, Scientific Data). This is a familiar avenue for researchers to receive credit and citation. Writing a data paper can be very useful for researchers, but, journals that publish data papers lack guidelines on how to structure data for effective reuse and sharing. We recommend partnering with best practice data repositories that emphasize curation, like Dryad. This way the data will have long-term archiving and will also have professional curation to ensure that the published data are usable and have enough metadata to be understood. While writing a data paper can be very useful for researchers, journals that publish data papers lack guidelines on how to structure data for effective reuse and sharing. Data journals should routinely be partnering with best practice repositories to best support the underlying data."
R2-2	I have a minor concern that readers will reinvent the wheel as they are implementing these ideas, and that concern could be mitigated by the authors including more specific examples of already existing infrastructure. I think the original version of this had some more examples that aren't present in this version (which is understandable considering the changes between the versions), but I do think some of those examples were helpful. Maybe add a sentence about how a lot of infrastructure currently exists to do many of these tasks, and to look for that first before creating your own.	We have added the following sentence to the second last paragraph of the conclusion: "There might be potential for researchers to reinvent the wheel, for example by creating new platforms for data sharing, when services such as zenodo and Dryad already exist. So it is worth mentioning that researchers should not reinvent new tools without an exhaustive search."

(continued)

ID	comment	response
R2-3	I think the authors do a nice job of this with mentioning Dryad. Two specific suggestions I have are for the metadata introduction section and the avenues to share data section. In the metadata introduction text ("There have been some efforts to improve this process, such as gogetdata, the Dat Project, and the formation of a private company, Quilt."), I'd also add the Frictionless Data project, which also helps researchers prepare metadata and data for sharing (and I believe was in the original submission).	We have added a mention to the Frictionless Data project. Section now reads: "Tools that can be used to prepare datasets and metadata (the infrastructure) and minimal sets of practices (guidelines) for reuse are sparse. 19,20 Unlike code, datasets are far more diverse in terms of heterogeneity, size, and format. This makes them particularly challenging to standardize or easily "install" where the code is running. There have been some efforts to improve this process, such as gogetdata, the Dat Project, Frictionless Data, 21 and the formation of a private company, Quilt."
R2-4	For the avenues to share data section, I recommend adding in some specific journals the authors would recommend for publishing data papers (this sentence: "Journals provide data papers in a formal publication explaining data that is archived").	We have added two specific example journals in parentheses. "When citing data, you should cite the preferred CITATION file in the first instance, else, a data publication, (for example in Scientific Data, Ecology), and if none of those apply, cite the repository in which the data is deposited (e.g., a Zenodo repository) in the references section of related papers."
R2-5	For the following two sentences, I suggest adding the date so readers understand how recent the work was done as trends in data sharing are changing rapidly: "Rowhani-Farid et al.14 showed that 7 of 157 (4.5%) sampled papers in the British Medical Journal shared data..." and Similarly, Stodden et al.15 showed that only... e.g. "In 2016, Rowhani-Farid et al.14 showed..." This will improve readability, but isn't strictly necessary.	We have made these changes: "In 2016, Rowhani-Farid et al. 14 showed that 7 of 157 (4.5%) sampled papers in the British Medical Journal shared data, and only 14% of articles claiming open sharing of data actually shared their data. Similarly, in 2018, Stodden et al. 15 ..."