February 11, 2022

Dear Dr. Cook,

Please find attached the "Add-on package" submission to The R Journal entitled "**knitrdata: A Tool for Creating Standalone Rmarkdown Source Documents**" by **David M. Kaplan**. The paper describes the knitrdata package that allows integrating arbitrary text and/or binary data into Rmarkdown source documents. This allows one to create entirely self-contained Rmarkdown source documents, providing a powerful tool for communicating the complete set of data, methods and computational tools associated with a scientific publication or report.

The manuscript has been significantly revised to address the editor's and the anonymous reviewer's concerns. In particular, the manuscript has been reorganized and expanded to clarify the utility of the package for ensuring computational reproducibility and more clearly distinguish it from a package vignette. This has included clarifying that the package is useful for producing standalone Rmarkdown *source* documents, as opposed to standalone final output documents after the knitting process, as well as inclusion of additional material on potential security concerns when using knitrdata as suggested by the editor. These changes have greatly enhanced the manuscript. I sincerely thank the editor and reviewer for their suggestions.

Detailed responses to the individual reviewer comments on the manuscript are provided below. I thank you in advance for your time and effort.

Sincerely,
David M. Kaplan

——— Editor review ————

In it's current form, the paper reads more like a vignette than an R Journal paper. Papers need to have substantial intellectual content placing the work in the field. For your paper you need to make better arguments for why embedding everything into the document is necessary, where and when, and pros and cons.

The work looks interesting, and I'd like to give you a chance to get it into suitable form for the journal. There have definitely been some times when I would have liked documents to be more self-contained.

The motivation currently is very broad and vague referring to open science, but Rmarkdown itself would be considered a contribution to open science. You have a very specific variation of a reproducible document, so you should focus on the difference with the conventional way of working.

> The manuscript has now been significantly expanded to more clearly distinguish it from a package vignette and clarify under what circumstances the package is particularly useful. I have also focused the discussion primarily around the topic of computational reproducibility, for which knitrdata can provide significant benefits in appropriate cases (that are discussed in the manuscript).

Specific points to address:

- Are there potential security issues with embedding everything?

> This is a very useful suggestion and in response a new section entitled "Are there security concerns when using knitrdata?" has been added to the manuscript.

- How does this compare with using the "self-contained" in the YAML?

> This is also a valuable clarification that I did not see as a potential source of confusion in the initial draft of the manuscript. In response, an additional paragraph discussing the difference between the functionality of the knitrdata package and the "self_contained" YAML header option has been added to the end of the "Conceptual overview of knitrdata" section. Briefly, knitrdata and the "self_contained" YAML header option are addressing entirely different issues. Knitrdata is concerned with self-contained Rmarkdown source documents that include everything necessary for knitting a Rmarkdown source document (i.e., it is on the input side of the knitting process). The "self_contained" YAML header option is concerned with the portability of the output of the knitting process for HTML output formats (i.e., it is on the output side of the knitting process). As such, these two functionalities address independent and unrelated issues. This has now been clarified in the revised manuscript.

——— Reviewer 1 ————

General comments:

This paper goes over the use of the knitrdata package, which allows a user to input data into R Markdown chunks directly. As it is written now, the paper looks like it would be more appropriate as a vignette and there are at least three areas that need to be developed before it is considered for publication. First, both the title and desription are vague and it took

me until the "knitrdata installation and usage" section to figure out what the package does. Second, I'm not convinced of the need create new chunk types for data. Third, there is not a convincing use case. Why might I do this? Are there cases where this allows me to do something beyond what could be done before - beyond putting the data and its encoding in the document? I certainly don't claim that the package is not useful, but the manuscript doesn't convince me that it is.

> The paper has now been reorganized and expanded to more clearly and quickly explain what the package does and include more examples of its use. The title has been modified to include the term "Rmarkdown source documents" to clarify that the package is about generating standalone source documents and not standalone output documents or inclusion of data in output documents. Additional material has been included discussing what knitrdata can do that is beyond or different from existing functionality in R or other R packages.

Specific issues:

1. I don't understand the connection between this package and open science.

> This has been clarified by expanding the Introduction to better explain how open science and more specifically computational reproducibility is achieved now and what benefits knitrdata can provide for computational reproducibility.

2. I don't understand this sentence:

"Until now, however, it has been difficult or impossible to integrate the data itself directly into Rmarkdown documents."

It is easy enough to show the data in a table or data table and I can link to the data via hyperlink if my target is html. Can you clarify?

> I thank the reviewer for this comment as it made clear a confusion that I did not anticipate in the initial version of the manuscript. In a nutshell, the reviewer is talking about presentation of data in outputs from the knitting of Rmarkdown source documents. This indeed already exists in Rmarkdown and knitrdata does not change or enhance this. Knitrdata is about facilitating inclusion of data in the Rmarkdown source document, i.e., within the initial text document that is knit to produce the final output (PDF, HTML or Word) document. This integration of data in source documents is difficult or impossible in the absence of knitrdata, justifying the need for the package. This has now been clarified by significant expansion of the introductory material in the paper and consistent use of the term "Rmarkdown source documents" throughout the manuscript.

3. I don't understand this sentence:

The R package described below, knitrdata (Kaplan, 2020b), provides a simple mechanism for integrating arbitrary complex data directly into Rmarkdown documents, thereby contributing to our capacity to ensure computational reproducibility of scientific research.

Can you be more specific?

> The relevant text has been significantly modified to more clearly state how knitrdata contributes to computational reproducibility.

4. I don't understand the relevance of the "Text data chunks" example. If I'm willing to write out the values of a csv file, why would I do it in a separate chunk, rather than in a regular chunk or in a .csv file? The same goes for the .RDS file example.

Again, I believe the confusion here is related to the difference between including data in input source documents as opposed to output documents from the knitting process. This has now been clarified by including additional material in the manuscript discussing how knitrdata relates to other more conventional approaches that might be used to include data in Rmarkdown source documents.