

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2021-392-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-392

Koen Hufkens (Referee)

Referee comment on "A hydrologist's guide to open science" by Caitlyn A. Hall et al.,
Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-392-RC2>, 2021

I enjoyed reading this manuscript. The manuscript argues for the use of open science practices in hydrology and does this admirably in a very structured and transparent way. This makes for a helpful document, as are reference for practisioners but also for teaching purpose. It should facilitate discussions surrounding open science both in hydrology but also in other fields (where similar issues exist).

The manuscript is structured along four principles, addressing various stages of research and their potential to be more open. Although there is overlap between them I think this approach works well to address some of the discussions surrounding open science. Further "scenarios" as listed at the end of the manuscript help put some of these principles in context.

As such I'm willing to accept the manuscript as is, as most comments are minor and deal with very nuanced language. Comments listed below should be considered if possible.

General comments:

When looking at the larger picture, it might be very important to note somewhere that open access is not a technical challenge anymore but mostly a socio-cultural one. When looking at table A1 and the challenges discussed only 3 out of 13 are technological, while the remaining are mostly political / socio-cultural. Despite the tools listed in the manuscript, I fear open science practices are not governed by the lack of these tools, the access to them (most of them are free), or even the use of them by some.

Detailed comments:

[Line 83] "Open hydrologists intentionally plan for, describe, and share the entire research process and approach from motivation to the final product"

I would be careful about leaning too strongly on either defining output as products or the fact that a pre-defined motivation should be provided in order to keep things open. From a very practical point of view dealing with science output as products is often very efficient, as it sets clear expectations. However, language matters and managerial language creep is sometimes very toxic as it often debases research (favouring short term returns, products, over slow generation of bodies of knowledge). I would suggest to shy away from defining research as products, and use research outputs / research results instead.

[Line 96] I think asking for a reasonable explanation of methods is ok. However, more often than not the answer might come down to a lack of funding. I wonder to what extent such a motivation will be accepted in review, and if this will stigmatize those who have less resources if there is the expectation that you are really explicit about these methodological choices (limited by funding). This dynamic already exists, but at least the expectation doesn't exist that it is written in full.

[Line 135] Requirements for version control shouldn't be grounded in the ability to peruse through previous in silico experiments. It is a tall order to ask people to use version control in the first place, it is another barrier to do this consistently in the way of a lab notebook. I heavily use git and to be honest my commits aren't clean. It is important that people save the state of the software used in particular experiments (either through a release/checkpoint on zenodo, github or similar). Actual commits or even branches probably have less value, and make things difficult if not less transparent. i.e. I would stress the deposition of code in repositories, rather than front loading additional computational skills (which are for some hard to acquire - saving data/code in a repo isn't). In general, I will take dirty code over no code and a static release tied to a manuscript over a dynamic repo (with recent changes).

[Line 159] This section focusses on software documentation, mostly for the end user. However, true open development is often hampered by not only the lack of user end documentation but proper code comments in the software itself. The lack of clear documentation of the code functioning (not the code use) by inline comments is something that is often forgotten and limits code re-use within different contexts. It also limits learning opportunities by seeing how computational problems are solved within a real world situation, not a classroom setting. A line on this could go a long way.

[Line 160] Include a link to ReadTheDocs, people might not be familiar

[Line 179 (and the section below)] I would suggest an open license. Permissive is the prerogative of the researcher. It must be acknowledged that permissive licenses, on open software and data, in the past have been abused by industry by offloading tasks to OSS volunteers while not giving back proportionally to the gains made (e.g. recent more

restrictive licensing on the openstreetmap data) and this risk is to be assessed by researchers. I think this is echoed by some other reviewers as well. However, the spirit of open science would suggest a permissive license and I value the sentiment.