

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/hess-2021-392-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-392

Francesca Pianosi (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-RC1>, 2021

The article gives an interesting contribution to the discussion on Open Science in hydrology. I especially appreciate the practical focus on helping researchers to get started with OS, and the linking to an online repository where new materials and resources will be shared beyond the article publication.

The manuscript needs to be revised to include the section on Principle 2, which was made available as a separate file. I do not have other substantial revisions to recommend but some points for improvement and further discussion.

In particular, given the "practical guide" angle of this article, one suggestion could be to complement the text with Tables listing in a very concise manner the various recommendations / tips / dos & donts for each Principle. For example, for Principle 2 (open software) the list may start like:

- use open-source software such as R, python or QGIS to develop your analysis
- use open-source version control system (e.g. Git) to manage changes to your code
- include documentation as comments embedded in the code as much as possible
- etc.

I think this would help reinforce key messages and help readers navigate the (numerous) points made in the text

While reading the paper I noted down several other comments. I am not sure they are all worth including/addressing in the paper but I'll report them and leave to the authors to decide if they want to take them onboard.

[1] L. 19 - Social challenges to embracing OS. The authors essentially mention one, the fear of being scoped, but I think others are as important. For example some researchers may be reluctant to share their software as this may bring further scrutiny and criticism of their work. Some seem to feel a sort of "jealousy" for their software, which they don't want to see modified (maybe improved!) by others. Maybe the point here is how we

perceive and value intellectual ownership. If I make my software available to others so they will (unavoidably) find bugs to fix and weaknesses to improve, does this diminish or increase the value of my original contribution?

[2] Line 94: "sharing the entire research process and approach (e.g., failed attempts and lessons learned that impacted research outcomes)".

I totally agree although there is a tension here between conciseness (which is needed for readability) and completeness (needed for OS). I think a good way to resolve the conflict is by having unlimited "Supplementary materials" along with a paper - as some journals now allow - so that authors can keep the main article focused on key findings, while giving detailed documentation of all the research process in the SMs.

[3] Line 98: I think the point about "minimising the use of jargon" is very important. We use a lot of academic writing cliches in our articles, perhaps thinking it makes them sound more technically solid, but often it only makes them more difficult to read! Another issue is the recourse to hyper-specialised terms that are only understood within our small research niche - and often take different meanings across sub-communities even within the same broad discipline (a good example: the diverse uses of the term "bottom-up approach" across sub-communities in hydrology and water resource management). Every now and then initiatives are launched to build glossaries that should help researchers navigate each other jargon, but my impression is they are quickly abandoned (for example years ago I was involved in a project on uncertainty and risk in natural hazard assessment and such glossary was one of the project outputs... I don't think it was ever delivered!). Maybe rather than building glossaries we should just do more to use a common and simpler language. This should include avoiding the creation of new terms for concepts that may be easily described with existing maths terminology. I write and review a lot of modelling/methodological papers, and I have the impression that new terms (and acronyms!) are often created under the pressure to "demonstrate novelty" - authors may be afraid that if their proposed methodology does not have a new name but is described using standard terms from a statistic textbook, reviewers will dispute its novelty. This way though we make our papers unnecessarily obscure and in the long-run we collectively contribute to fragmentation of knowledge and duplication of efforts. This links back to the general tension I find between OS and the way we reward (over-emphasise?) "novelty" of individual contributions (see also comment [1] above).

[4] L. 193 About green OA. Maybe this sounds naive but I really wonder what are the drawbacks of "green OA" (and how it may be sustainable in the long-term for publishing companies)? Authors do not pay publication fees, readers do not pay subscription fees (after the embargo period) as they can access the non-typeset version... sounds very convenient - but for the publishing companies! Am I missing something?

[5] L. 226: "For software, we suggest the authors start by declaring a permissive license because it improves transparency and reduces downstream licensing conflicts."

I personally agree with this suggestion but I think many (including some at the R&D team of the University I work at!) would find it controversial. A permissive license (say for example the MIT license) implies that the software developed by publicly-funded researchers will be freely available also to users that may make a commercial use of it - and hence may be willing to pay for a licence. So, should universities "give away" a

potential source of revenues? Is this fair to tax-payers who funded the research and related software development?

Some more general thoughts:

[6] A lot of the activities for OS mentioned in this paper take time - for self-learning and training and to make data/software accessible to others. For example, documentation is key for open software to be meaningfully used by others but developing good documentation is very time-consuming (this was a key lesson I learnt in my own projects - see Pianosi et al 2020). So I think there is a tension here between OS and the general "publish or perish" attitude. I wonder if Open Science is necessarily also Slow Science (e.g. Frith, 2019)?

[7] Open teaching. I am all in favour for it - like most of us last year I developed lots of materials for on-line learning and I am keen to make it open access (as soon as I find out what's my University policy on this!). This said, I wonder what the long-term implications of open teaching will be for university life. If excellent study materials become available online to all and for all subjects, then what is the reason for enrolling in a university programme instead of self-learning? Will students only attend university to clarify doubts and get assessed - or, in other words, will the main role of universities become accreditation rather than delivering contents? I am not saying this is necessarily a bad thing (maybe it'll give academics more time for research or other type of engagement with students, such as mentoring or research-based teaching) - just highlighting it would be a very substantial change to the way higher education works today.

References

Pianosi et al (2020) How successfully is open-source research software adopted? Results and implications of surveying the users of a sensitivity analysis toolbox, EMS. Available at: https://research-information.bris.ac.uk/ws/portalfiles/portal/215604556/Paper_SAFE_Survey_accepted.pdf

Frith (2019) Fast Lane to Slow Science, Trends in Cognitive Sciences. Available at: https://discovery.ucl.ac.uk/id/eprint/10091940/1/Frith_Fast%20Lane%20to%20Slow%20Science%20Prefinal.pdf