

Interactive comment on “Technical note: A prototype transparent-middle-layer data management and analysis infrastructure for cosmogenic-nuclide exposure dating” by Greg Balco

Sebastian Kreutzer (Referee)

sebastian.kreutzer@aber.ac.uk

Received and published: 15 April 2020

Dear Dr Balco,

Thank you for this exciting and detailed response. It was more than I had expected, and it lays out compelling arguments and thoughts. To have a little bit of a discussion, I may respond to four points:

Printer-friendly version

Discussion paper



1. I still do not quite get the vulnerability argument that would require a cybersecurity audit first before the source code can be publicly available. This is not a NASA project. So, I guess this is more related to the way resources are hosted on external servers instead of servers at a research institute, which then imposes private liability you do not want to take (for understandable reasons). I am not part of the cosmogenic nuclide dating community, but this is obviously something the community should take up if they are interested in your project in the long run. I can only recommend that you try to host your project on servers part of a public research unit. Whether this is technically possible at all, of course, I cannot tell, and my comment should not give the impression that I hold this against you. Besides, maybe you can link to the 'API description' of the calculator? Ok, it is a blog entry, but it serves its purpose.
2. *[...] careful versioning of middle-layer code to facilitate reconstructing past calculations is fundamentally not consistent with the basic concept of storing only observational data and performing all calculations dynamically. If the middle-layer code is updated, then by definition the results of calculations using the new code are better than the results of calculations using the old code.*

Yes and no. I have no doubt that every new version is released only with the best intentions to make the calculation better in terms of age accuracy and precision. However, it neglects the 'human factor'. Newer \neq better and old \neq wrong. Even with thorough manual and automated testing before the release, and I have no doubt that this is done, mistakes happen. The more it is essential to document and communicate this as soon as possible. Here transparent versioning helps a lot. Probably I spelt out the most apparent arguments you have heard a thousand times before. Perhaps our understanding collides simply because of a different perception of what users should do and what they actually do. You do believe,

[Printer-friendly version](#)[Discussion paper](#)



and the manuscript leaves no doubt, that calculations should be done on the fly, always using the most updated version, which is immanent in the “middle-layer” concept. You have my full consent. However, my very personal experience with some software projects is that this is only remotely true to reality. People use data available and then add newly calculated data and start comparing them (without realising that they actually introduce a systematic error). Versioning would indicate whether this is safe to do so, or not. Probably this is my only point where I tend to say: It should be done, in particular given the minimum amount of additionally needed effort. I also think that this is not really off-topic when it comes to the manuscript, although such thing can be only implemented in the software prototype itself.

3. Regarding the rules for how data can be used. I realised that I was not clear enough. My intention was to provide some kind of “legal clarity” to users, data donators, and of course, the maintainer, you. Data can be harvested from published work but is it “safe” to store them again in an open repository? As long as a project does not raise too much attention, my best guess is that no one cares. But of course, the project should get attention, and it should be used. Such a project, as you have it presented, should take this into account by applying standard licences that allow legally safe reuse and recycling of data (e.g., creative commons licence) so far possible and applicable. Implementing such “rules” is not overly complicated and it does not restrain the user but avoids that a project with the best intentions lashes back to people who started it and who use it. You use the word “rules”. I do agree, creative minds don’t like too many rules (and researchers are usually creative minds), and the feeling of being chained to something. However, what I was talking about was a minimum set of principles that helps to safeguard the valued degrees of freedom.
4. Yes, the FAIR-guidelines paint an ideal world and adhering them in whole might be impossible. I was asking for two to three additional lines of a reflection to

show that you are aware of it (it does not mean that you have to agree to all or endorse them), which I still think is justified and would strengthen the manuscript. Basically, a summary of your thoughts from the response would do. Why do I consider this as necessary? For example, the multi-billion euro European research framework programmes Horizon 2020 and Horizon Europe (hopefully from next year on), do not fail to emphasis over and over again the importance of those guidelines and make it even a prerequisite for projects' data management plans. The main reason for it might be that they have a lot of good ideas summarised and guidelines (in general) provide only guidance but do not set strict rules with the potential to lead to the opposite. Besides, I may have a distorted perspective because of my current position. Since I was never in favour of demanding the citation of a particular reference, I will finally leave this up to you to add a few lines (or not).

Finally, and just for the case that my comments leave doubts, my original recommendation did not and does not change, and I am pleased about this manuscript and the related discussion.

Sebastian Kreutzer – Bordeaux – April 15, 2020

Interactive comment on Geochronology Discuss., <https://doi.org/10.5194/gchron-2020-6>, 2020.

Printer-friendly version

Discussion paper

