Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2018-32-SC4, 2018

Open Access Earth System Science Data Discussions

Interactive comment on "OCTOPUS: An Open Cosmogenic Isotope and Luminescence Database" *by* Alexandru T. Codilean et al.

T. A. Ehlers

todd.ehlers@uni-tuebingen.de

Received and published: 8 June 2018

I am supportive of the need for the surface processes community to have a database for uploading and downloading cosmogenic radio nuclide sample information. This is sorely needed. Thus, I greatly appreciate the authors efforts to do this, and think this paper could eventually make a highly valuable contribution.

An important aspect of the production of OCTOPUS is that if this paper is published, then it will likely be the de facto database. This means - the authors need to get this right - and make is useful. Once this exists, I don't anyone else will undertake this effort, or be able to secure funding to do it better, so it's important to get it right in this attempt.

From a practical standpoint (as someone who has set up data bases himself) there are a couple of serious issues that need to be addressed prior to publication.

1. Web interface is incredibly clunky and limited in functionality. This should not currently be on-line, and is not in a form that the community can use. Example issues are:
a. Following the intended instructions - I was unable to download data and have them emailed to me (using the latest Safari browser version on Mac OX 10.13.5. I had to have a colleague submit the data request. The issues what the 'submit request' button would not become active. I was also unable to select subsets of data either..

b. When the data arrived with links via email - my selection of data for the entire world came in with about ∼160 links that had to be individually downloaded. Awkward. Stream line this into a single file download.

c. The ∼160 folders that were delivered contain way too much information. Yes - different parts of it will be useful to different people, but the actual data files (in the CSV folder) were split into two files for each folder (therefore about 320 total csv files I have to merge into a single one for processing / plotting on my own). Please 1) combine searches made into a single file, and 2) allow users to select what is provided to them (e.g. I don't want/need all the arc files). The total file size was ∼42 GB, when all I needed was <1 MB of CSV files. I don't see this database as useful if every user will have to merge individual files together after they get them.

2. Longevity of the database. Data storage and archiving does not seem to be through a long-term data warehouse. This is not good. Does the database disappear in 5-10 years when the PI(s) loose interest/momentum on this, or funding runs out. Please archive through a national/international warehouse and run the data download web interface through this also. The manuscript should have a section added to it that addresses long-term storage and maintenance plans. It is unclear from the current web sites mentioned if the web interface and user upload options will be available for the very long term. I'm guessing Australia has long-term data infrastructure system for

geosciences, but if not - I can recommend using the German Pangaea system which is been available for a long time for geoscience data storage, and has a national funding scheme (long-term) behind it.

3. User submission of data. This is not clear from the web interface how users can submit data they've published. It also not clear form the publication who will approve and maintain the database. Is perpetual funding in place for this? Without a clear addition of this, I don't see the utility of database. It will become outdated in a couple years (like the Portenga and Bierman compilation) and would then only provide an incremental improvement.

4. The source code for the web interface should also probably be archived and available. This would allow (for example) someone else to take over responsibility of the database in the event the PIs stop doing it (and before the source code is lost).

I'll refrain from making any comments on the data processing and quality of the compilation - other highly qualified reviewers in this area have submitted comments.

Based on the above concerns - I do not recommend publication at this time. After the authors can address these issues, then the manuscript should most definitely be considered for publication and this contribution would very useful.

Thank you to the authors for starting this initiative - please finalise your approach so people can actually use this.

Best wishes, Todd Ehlers - Univ. Tuebingen.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-32, 2018.