

## ***Interactive comment on “Technical Note: Open-paleo-data implementation pilot – The PAGES 2k special issue” by Darrell Kaufman and PAGES 2k special-issue editorial team***

**GL Simpson**

gavin.simpson@uregina.ca

Received and published: 27 February 2018

This Technical Note raises a number of issues regarding the Open Data movement and data stewardship in the palaeosciences.

I wish to raise two additional points; i) the appropriate time in the data cycle to consider data management and stewardship, and ii) the use of embargoes to assuage concerns regarding data sharing and archival.

Kaufman et al suggest that “[p]ublication is the ideal stage in... a research project for concerted data management and stewardship.” This, they claim, is because

C1

“[s]cientists are most familiar with the specific details of their data when submitting a manuscript for publication and therefore are better able to prepare the data for transfer to a public repository at that moments rather than following publication when the familiarity and incentive have faded.” I agree that archiving data post-publication is far from ideal, but I disagree with the statement that the point of submission is the ideal time to address data archival and options for stewardship.

The time point when researchers are most familiar with data is, contrary to Kaufman et al's suggestion, at the point of data capture or collection into a digital format. If the data product is the result of subsequent processing of raw data, then it is at this data processing stage when scientists are most familiar with the data product and how it was produced. At this point, not the some later period closer to publication, it is appropriate to prepare data for subsequent archival via the addition of metadata, description of the data, and preparation of any data-processing code. Whilst archival itself may happen at a later stage, “concerted data management” is best conducted as the data are collected, produced, and prepared for analysis (see, for example, Strasser et al “Primer on Data Management: What you always wanted to know” at [https://www.dataone.org/sites/all/documents/DataONE\\_BP\\_Primer\\_020212.pdf](https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf)).

The most important aspect of data stewardship during the transition to Open Data within the palaeosciences is to actively pursue and promote data archival in standard formatss with appropriate metadata and computer code. Indeed, it is good data management practice, as discussed above, to prepare data for archival at the point of creation not publication. However, there is no reason why data cannot be archived in an approved repository prior to publication and held under embargo until a predetermined time. It is not clear from the Discussion Paper whether such an option was considered by the PAGES 2K special-issue editorial team? Such an embargo would allow for approved reuse within a synthesis activity yet retain some control over availability until other activities are complete. Once the agreed embargo period is complete the data should automatically be made publicly available. In the interim, the data could be refer-

C2

enced via a DOI or similar identifier, but not accessed without permission from the data generator. Journals and their editorial boards should determine, in consultation with the community, suitable time limits and other conditions for the embargo, the situations in which embargoes may be used, and the technical arrangements that insure timely, automatic public access following lifting of an embargo.

This should go a long way to satisfying concerns regarding the impacts on ECRs or research projects of participating in open synthesis activities — there are time lines for completion of degrees and projects that would set reasonable limits on any embargo — as well as those pushing for better data stewardship — the data would be archived to a required standard prior to their use in syntheses, for example.

I believe Pangaea is set-up in such a way that this may already be possible, though I suspect some work may be required at the backend to provide the protection of an automatic lifting of an embargo — although the Data submission page of the Pangaea wiki ([https://wiki.pangaea.de/wiki/Data\\_submission](https://wiki.pangaea.de/wiki/Data_submission)) speaks of “publication” in the described workflow, the process seems sufficiently flexible enough to accommodate a general embargo period. To this end, engaging all members of the community, not just the data generators and those organizing large collaborative data syntheses, would be productive in finding workable solutions to the concerns of colleagues in the short-term.

I am no fan of embargoed scientific outputs in general, however it is vitally important that we bring the community with us, not kicking and screaming, but because they want to make the journey. The concerns among some members of the palaeo community could be addressed through appropriate and limited use of embargoes during this period of transition.

---

Interactive comment on Clim. Past Discuss., <https://doi.org/10.5194/cp-2017-157>, 2017.