

Comment on “Open-paleo-data implementation pilot – The PAGES 2k special issue”

David J. Karoly^{1,2}, Kathryn J. Allen³ and Patrick J. Baker³

¹School of Earth Sciences, University of Melbourne, VIC 3010, Australia

²ARC Centre of Excellence for Climate System Science, University of Melbourne, VIC 3010, Australia

³School of Ecosystem and Forest Sciences, University of Melbourne, Richmond, VIC 3121, Australia

Introduction

The authors of the Technical Note manuscript have sought to explain the procedure they claim was used to attain a consistent level of data stewardship across a special issue of the journal *Climate of the Past*. We are the senior authors of one of the papers published in this special issue (Freund et al., 2017) and we disagree with a number of statements in the manuscript, believe that there are important omissions in describing the procedures that were used, and disagree with the interpretation of the journal’s Data policy. The early career researchers who are authors on our paper chose not to be authors of this Comment because they are concerned that the feedback provided in this Comment might damage any future interactions that they wish to have with some members of the PAGES 2k research community.

We very much appreciate the need for input data to be made publicly available, but have a number of concerns about the procedure used for this special issue, its inflexibility, apparent inconsistency and inefficiency. While we agree with the authors that open-data sharing in paleo sciences is likely to accelerate, it is important to discuss the disadvantages as well as the potential advantages of open-data sharing. We are taking this opportunity to comment on the procedure used to attain a high level of data stewardship more generally and the challenges around this.

Specific concerns

Page 1, line 16: The first statement of the Introduction states “... the practice of making data readily available is rarely embraced”. This is an inaccurate assessment of the practice of making scientific data available in the palaeoclimate and wider climate communities. NOAA has very extensive palaeoclimate datasets made publicly available (<https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets>). In our study, almost but not all the proxy data we used were from tree-ring analysis and had been publicly through this NOAA database for many years.

Page 1, line 23: “... top-down mandates alone are unlikely to foster the necessary cultural changes in scientific communities.” We completely agree with this statement, but a top-down approach was exactly what was used to implement the data stewardship procedures for this special issue.

Page 1, line 24: “Bottom-up motivation from the research community, including early-career scientists, is needed to drive the open-data revolution” Again, we completely agree with this statement. However, there appears to have been little attempt to have bottom-up involvement of potential authors for the special issue in the preparation of the data stewardship procedures, as far as we know. In fact, exactly the opposite occurred, with some early career researchers given no flexibility when the data stewardship procedures

were imposed by Darrell Kaufman and the PAGES 2k special-issue editorial team (hereafter referred to as ‘the Team’).

Page 2, line 27: The statement that a notice was sent to prospective authors of the special issue to explain the process for data stewardship is misleading. It does not mention the time at which this notice was sent relative to the deadline for submission of papers. The Rules for the special issue were distributed to all planned authors in March 2016.

There was no mention of special data stewardship procedures in these Rules and the only information was that all manuscripts must “comply with the same quality standards as regular submissions”. The deadline for submission of manuscripts was 31 December 2016. The data notice was sent not only to prospective authors, but also to authors of submitted manuscripts, in March 2017, more than two months after the original deadline for submissions. The content of this notice is reproduced below. Providing information on data stewardship after the deadline for submissions is inappropriate and a clear demonstration of the top-down process imposed for data stewardship in this special issue.

Dear PAGES 2k Special Issue corresponding authors:

We wanted to remind all authors about Climate of the Past's data policy, which you can find [here](#). PAGES is dedicated to best practices in [data stewardship](#), one of its four integrative activities, and the editors are committed to ensuring that the Special Issue models these practices. More specifically, all of the essential data used in the analyses and generated as results must be made available through a public data repository where they can be tracked using a "Data Citation," described [here](#). A Data Citation is in addition to a conventional reference to the publication where the data are first described.

We realize that some papers have already been submitted that do not fully meet this expectation, and we don't want to discourage on-time submissions on Friday. Details of the data citations and other aspects of data stewardship will be addressed as part of the regular open manuscript review period, with input from the editorial team. If, however, you or your co-authors foresee any issues with complying with the CP data policy, we regretfully suggest that you withdraw the paper.

Please let know if you have any questions at this stage.

*Thank you,
Darrell
on behalf of the PAGES 2k Special Issue Editorial Team*

This notice refers to the Climate of the Past Data policy, from which an extract is reproduced below, from https://www.climate-of-the-past.net/about/data_policy.html

Climate of the Past Data policy

“Copernicus Publications recommends depositing data that correspond to journal articles in reliable (public) data repositories, assigning digital object identifiers, and properly citing data sets as individual contributions.” ... “To foster the proper citation of data, Copernicus Publications requires all authors to provide a statement on the availability of underlying data as the last paragraph of each article (see section [data availability](#)).”

Statement on the availability of underlying data

Authors are required to provide a statement on how their underlying research data can be accessed. ... If the data are not publicly accessible, a detailed explanation of why this is the case is required.

While the notice from the Team to authors states that all essential data “*must be made available through a public data repository*”, the Climate of the Past (CoP) Data policy only “*recommends depositing data that correspond to journal articles in reliable (public) data repositories*”. There is a clear inconsistency between the journal’s Data policy and what was implemented by the Team for the special issue. No option was provided by the Team to allow for publicly accessible data sets with a doi through journal supplementary material, which is allowed by the journal. No option was provided by the Team to allow for a detailed explanation of why the data are not publicly accessible, which is allowed by the CoP Data policy. Hence, we believe that there were clear and demonstrable inconsistencies between the journal’s established Data policy and the data stewardship procedures implemented by the Team for the special issue, despite the claims in this manuscript.

Page 3, line 10: The statement that “Authors were asked to transfer all datasets that were not easily accessible to a community-recognized public data repository” is misleading. Authors were required, not “asked”, to transfer all datasets under threat that their manuscript would have to be withdrawn if they did not comply, based on the data notice distributed by the Team. This top-down approach, with the threat of withdrawal of the manuscript, could be interpreted as intimidation by Darrell Kaufman (author of all the emails) of the graduate student leading our paper, given the inconsistencies between the journal’s Data policy and the data stewardship procedures imposed.

Page 3, Section 2.3: This section provides another misleading description of the procedures used to impose the data stewardship procedures on our paper in the special issue. Our paper is a multi-proxy synthesis involving primarily publicly available proxy data in public data repositories. In addition, a small number of the proxies used were new and provided by other researchers, or were publicly available but not in data repositories. Two examples of the interactions between Darrell Kaufman and two data providers are contained in an Appendix to this Comment. As can be seen, this was neither a helpful nor a bottom-up process.

Page 7, Section 4 Outlook: The most important lesson to learn from this special issue is that it is vitally important to distribute the planned data stewardship procedures at the same time as the Call for Papers. It is also vital to make sure that the planned procedures follow the general interpretation and practice of the journal’s Data policy, not an effort to implement a stronger data stewardship policy.

General concerns

While we fully support the call for open data sharing and transparency in the palaeoclimate sciences, we do not agree that publication is always the ideal time to make data available. Specifically, we argue that data availability policies must include some flexibility so as to protect the interests and future career prospects of early career scientists and the longer-term viability of the research groups producing the palaeoproxy records.

The impact of rigid data policies formulated in a top-down manner by experienced researchers (often those involved in modelling or multi-proxy synthesis) with large teams will generally be negative on early-career researchers who are often working to schedules around their PhD study and cannot as rapidly produce the final products of their work as can a larger group. With a desire to succeed and contribute to the science, this leaves them vulnerable to ‘scientific exploitation’ and, in more serious cases, may compromise the successful completion of their postgraduate studies and future careers. It is these people in particular, that data policies should help protect. A data policy should also respect the fact that, in many cases, funding for the generation of new records comes from agencies (Government and NGOs) that expect a return on their investment from the funded group, not from a different group. Rather than a didactic top-down approach from data modellers or synthesizers who clearly require palaeoproxy data for their use, a data policy should be formulated with input from those who generate the records.

At the time of the Call for Papers for the CoP special issue, it was not apparent that all data would be required to be made publicly available upon publication of papers, nor that the policy for the special issue would differ from that for *Climate of the Past* generally. We are aware of at least one author who withdrew their paper due to this not being made clear at the outset. Certainly, if we had been aware of the strict ‘new policy’ and its heavy-handed application, we would not have included some data sets that were generously shared with us on the basis that they not be made publicly available yet. The group who generated these data sets was working towards their publication as part of an ongoing broader grant-funded project. This situation unnecessarily placed the PhD student writing this paper in an extremely difficult and stressful position.

Reference

Freund, M., B. J. Henley, D. J. Karoly, K. J. Allen and P. J. Baker, Multi-century cool and warm season rainfall reconstructions for Australia’s major climatic regions. *Clim. Past*, **13**, 1751–1770, 2017. <https://doi.org/10.5194/cp-13-1751-2017>

Appendix

Example 1

Extract from an email from a data provider in response to Kaufman:

“Releasing data from current projects that have yet to finish (or for unfunded projects) runs the real risk of being mined and so diluting my chances of future support. I have been burned before by releasing data. I’m sure this issue has been raised by others.

What particularly gets to me in all of this though, is that when Mandy, an early career scientist, asks for my data and I provide it, I am then the one being portrayed as the villain! I have asked for nothing in return for sharing the data with Mandy—neither co-authorship, nor restrictions on the use of the data. My only request is that a few of the chronologies that we provided be temporarily withheld from lodging with NOAA until after I have been able to publish something myself.”

Email from Darrell Kaufman

From: Darrell S Kaufman <Darrell.Kaufman@nau.edu>

Sent: Saturday, 29 July 2017 10:52 a.m.

Subject: PAGES /Freund data request

Dear XXXXX

I am writing in regard to [a manuscript](#) now under review at *Climate of the Past* by Mandy Freund, Benjamin Henley, David Karoly, Kathryn Allen and Patrick Baker, which is part of a special issue featuring the results of the PAGES 2k project. The special issue is also a contribution to the [PAGES Data Stewardship Integrative Activity](#). As such, all of the papers have been reviewed by the 2k special issue data review team in an effort to attain a high and consistent level of data stewardship across the volume. This involves enacting the [publisher's data policy](#) of including a 'data availability' section in each paper, which specifies where the essential data used in the study are located. If extenuating circumstances preclude the public release of data used in a paper, the reason must be clearly stated. In addition, all authors of the special issue are using [data citations](#) so the source of the data can be tracked and attributed to the data generators.

The rainfall reconstruction in Freund et al.'s paper was based on a large number of proxy records, including several of yours. My understanding is that the data that you provided the authors have not been submitted to a public repository and that you instructed the authors to not release the data as part of their larger synthesis.

The special issue data review team requests that you reconsider your position. Best practices dictate that all data used to formulate the major results of any study be made available along with the publication to assure reproducibility. As a signatory on the [International Accord on Open Data](#), PAGES is committed to promoting a high level of data stewardship across its activities. Progress within our community is advanced when data are made available for reuse.

The data you provided to Freund et al. have already been formatted by the authors in a way that can generate a NOAA-Paleoclimatology .txt file. We can facilitate the transfer of your data to NOAA or other long-term archive so that you can receive a persistent identifier, which can be used for a proper data citation in your name.

Can we proceed to help you transfer the dataset to a public repository so that it can be properly cited by Freund et al. and so this PAGES Data Stewardship Activity can achieve its goal?

Thank you.
Darrell

Darrell Kaufman, PAGES Executive Committee
On behalf of the PAGES 2k Special Issue Data Review Team (Belen Martrat, Scott St. George, Nerilie Abram, Raphael Neukom, Marie-France Loutre, Lucien von Gunten)

Example 2

Information from a data provider on Siple Dome shallow ice core B water isotopes:
The authors should cite Jones et al., 2014 (<https://www.clim-past.net/10/1253/2014/cp-10-1253-2014.pdf>). In that manuscript, the isotopic data for Siple Dome shallow ice cores B-H is provided in a supplemental file (<http://dx.doi.org/10.5194/cp-10-1253-2014-supplement>).

Response from Kaufman:

Climate of the Past discourages the use of supplemental files for long-term data archival and the publisher (Copernicus) is not an officially recognized data repository. Best practices would be to transfer the data to NOAA Paleo or similar and to obtain a persistent identifier for a proper data citation.