

Interactive comment on “Palaeo sea-level and ice-sheet databases: problems, strategies and perspectives” by A. Düsterhus et al.

A. Düsterhus et al.

andre.duesterhus@uni-hamburg.de

Received and published: 20 November 2015

We would like to thank the reviewers for the very helpful comments and suggestions. We have thoroughly considered them and have decided to offer to restructure the manuscript to clarify its intent and make it more understandable for the audience. In this we will take up the suggestion to use the mentioned three phases (measurement, documentation, and interpretation) as orientation points and will reorganise the content accordingly.

We will refocus on the problems on how to create a database and how databases can be more successful for the future of the associated research fields. Additionally we will add at least one figure to illustrate the workflow of a successful database.

C2415

We will use a more direct language and will include concrete suggestions how database creators can address the raised problems. Furthermore, we will streamline the general discussion on databases, but still highlight the important points, as we think it is essential to demonstrate the communities that there is still a lot of work to be done (on the project based research side as well as on the policy based side) to get to a point to generate and facilitate appropriate databases. The manuscript currently addresses especially the need to bring new communities (data science, computer science etc.) into the currently very geological based community around these topics. Databases are a very interdisciplinary field and need a better incorporation of the different talents to gain the best result for everyone. As such we have submitted the paper as a review paper, to offer advice for scientists already in the field as well as giving scientists currently not involved an entrance point. Nevertheless, we acknowledge that we would incorporate with the reconstruction a considerable part, which includes genuine research. Therefore we do not think that it qualifies as a technical note. We are open to suggestions by the Editor on which category would fit best to the manuscript.

As the main points mentioned above are stressed by both reviewers, we will comment on the additional points made by reviewer 1 chronologically:

"P 2391, L 1: They are not essential, they are only helpful. If they were essential, then nothing that didn't use a database would be worth anything." Databases are more than just helpful tools. For applications like calibration, tuning or validation, they are really essential. We will clarify it and the remainder of the abstract according to the points of criticism in a rework of the manuscript.

"P 2392, L 5-21: This discussion seems out of place. This is a fine treatment of some uncertainties dealt with in the paleo community, but does not fit in the introduction. Such a long discussion makes it seem like the uncertainties are really the most important part of database development. Is this true? If it is not, then put this discussion elsewhere. If it IS the most important part, then the paper needs to be structured around a treatment of uncertainties..." Communication of uncertainties has the largest

C2416

room for improvement, which can be tackled with good databases as we describe them in the manuscript. However, it forms only part of the overall design of a database, and we will therefore reduce the explanations around this topic. As such we will in a rework not structure the whole manuscript around the treatment of uncertainties, but will focus on the essential implications to store them appropriately within a database. Nevertheless, we will continue to underline the importance of uncertainty documentation. Any scientific datum needs associated uncertainties. Without explicit and clear documentation of uncertainties, database users must make their own assumptions that will at best weaken the value of the data (due to conservative uncertainty estimates) or at worst introduce errors (due to uncertainty estimates that are too weak).

"P 2392, L 21-23: Starting with "Finally", this is a new topic. Is this introduction supposed to be a summary of the problems involved in a potential database for the paleo community? It is too disorganized for me to tell. One possible way to improve this would be to use the introduction to describe HOW the community does/could use databases. Alternatively, you could use the intro as a summary of problems, but you must then address more than just examples of uncertainties. Note that the ones listed in lines 5-21 are by no means exhaustive." We will address this in the rework and will restructure the shortened introduction around the reasons and history of databases in these fields.

"P 2393, L 16-17: What are the key components? It is not clear. This would be an area that could use a table." We will stress and structure this point much clearer either via bullet points or a table.

"P 2394, L 3-14: This seems out of place. Is this truly an aspect of the community structure? It seems more like suggested changes to funding structure, which might fit better toward the end of the paper. Alternatively, it could be incorporated into a section on paleo-community problems." We will restructure these topics and will point out the funding problems separately of the community structure.

"P 2394, L 15-17: Seems a bit bold. What other data are you comparing this to?"

C2417

The paleolimnologists, etc. might argue back." We will use an appropriate language to reflect this point.

"P 2394, L 24: Are the authors suggesting that in order to have a standardised database, researchers should have standardised field measurements?" At least for key components this is advised.

"P 2396, L 13-19: Do you suggest using one-to-one or relational databases? From your later discussion, it seems like one-to-one would not be sufficient. Go beyond the statement that there are different kinds of databases and propose a STRATEGY, as promised in the abstract, of how the PALSEA2 community thinks the database should be structured. At minimum, perhaps a short table of pros and cons of a one-to-one (or spreadsheet) versus a more complex database could be useful." We will focus on a strategy.

"P 2397, L 25-26: What are your suggested "consistent data types per column"? This is another example of a place to suggest strategies based on the PALSEA2 collective knowledge." We will clarify this point. It basically states that each column should only have one generic data type (no mix up of text and numbers).

"P 2398, L 15-17: Good point. What should be done to address this? Require citations of the database to include a "date last modified"? Preserve old versions of the database for reference?" Once database reach a point of maturity that it is sensible to cite them, a versioning with keeping all former versions or at least allow reproducibility of them is highly recommended (and in fact a requirement to gain a DOI).

"P 2398, L 9 to P 2399 L 9: How unique are these problems to the paleo community? Not very...? Have any other subfields addressed these problems? I agree that the problems raised are real and important, but it seems like this is a science-wide problem. This manuscript may not, therefore, be the place for a long-ish discussion of the problem. What IS important to address here is how the paleo community can develop a dataset that overcomes (or starts to overcome) the problem of citation. It is not help-

C2418

ful to list general problems without providing the PALSEA2 community's insight into potential solutions." Some valid points are raised here. We will shorten this discussion, and will emphasize that the problem for geology based science is unique in the sense that it is within our community still possible not to make data openly available once they are used in publications. We will also stress that when good databases should come into existence, the current citation rewarding system (h factors and the like) falls short while inadequate citations, and the failure to check the original sources, could mean that the quality of data could be called into question. Without addressing this it is highly questionable whether good future databases might come into existence, but overcoming all citation issues is perhaps not feasible. Palaeo might be the prime example why the current systems in data publishing might not be appropriate to address the need of the science at hand. Solutions will be given in the simplest form how database creators can help to keep the original data authors and sources visible. This will include to associate to every included data entry the original citation.

"P 2399, L 11-12: I am hoping that if the authors are putting in the effort to write a paper about a potential database, that it would indeed end up on the internet. That itself seems too obvious to include. However, if the authors wish to include such a statement, I would also make mention of access issues. I suspect that many young researchers, or researchers from outside of the country of origin (especially where language differences exist) do not know that of the existence of databases that are only acquired by contacting an author. I can also imagine instances when such researchers would hesitate to contact an intimidating bigwig scientist. Therefore, and especially when considered through the eyes of government funding agencies, I believe online access to be imperative." We will make a clear statement that databases should be openly accessible without restrictions.

"P 2399, L 20: How do you suggest using this "information as a basis for future databases"? Be specific." We will give a short and specific comment how a database might be build up, which fulfils the requirement.

C2419

"Table: I like the idea, but this needs to be expanded on. First, improve the consistency of detail between entries. Second, it would be useful to see what kind of data each database provides." This will be addressed.

Interactive comment on Clim. Past Discuss., 11, 2389, 2015.

C2420