

Earth Syst. Sci. Data Discuss., referee comment RC3
<https://doi.org/10.5194/essd-2020-308-RC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2020-308**

Anonymous Referee #3

Referee comment on "Climate-Biogeochemistry Interactions in the Tropical Ocean: Data collection and legacy" by Gerd Krahnemann et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-308-RC3>, 2021

This manuscript describes a catalog of datasets obtained during a long-time research project. At this stage, I do not recommend to publish it for the following reasons:

The information about the long-time SFB project is not sufficient. The reader does not get a feeling why such a large project was done and what connects the different disciplines. Dissolved oxygen is mentioned and some textbook knowledge of oxygen distribution and their drivers are given. This is not enough to take the reader by the hand and show what great science has been done in the project.

The manuscript reads like a technical report.

The list of datasets is not complete, as the authors write. Therefore, this report comes too early. Some of the data appears in other data centers than Pangaea, which is also mentioned section 5. However, that summary does not say which data is concerned. The authors did not explain why some of the data were submitted elsewhere. Couldn't they have mirrored the data in Pangaea to really have all data from the SFB together?

The information about the data is found in section 4. However, in this section mainly the descriptions of the methods used are presented. This is not the same as a description of the data. What I expected to find here is a real description of the data and what has been done with that data: What kind of quality control was applied, how much data had to be discarded, which methods for data management were used, etc. A mere methods description is not useful in an ESSD paper.

Even when it is useful to have a list with all the cruises and datasets of this large project, this information can easily be provided in a technical report.

Some minor comments:

Figure 2 and 3: Please add some geographical names to the figures: countries, cities and in particular those names occurring in the text (e.g. ports)

L63-65 "The three 4-year long phases allowed for the development and adaptation of the observational and experimental program. Questions arising from the data already collected were incorporated into new sub-projects for the subsequent project phases." It would be nice to see some examples of this. As it is here, it is very abstract.

L88-89 "One of the first steps after the inception of the SFB 754 was the development and implementation of a common data policy (<https://oceanrep.geomar.de/47369>)." This is a sentence I expect in a technical report.

L91-92 "This data policy and its strict application is one of the reasons for the success of the SFB 754 with 421 peer reviewed publications at the time of writing." ditto

L97-98 "In the final step the data was published and made freely available at the World Data Center PANGAEA (<https://www.pangaea.de>) or at other more specific data centers." Why wasn't it part of the data management plan to have all data together in one data center? This is not in conducive to the FAIR data principles, in particular the findability.

L99-100 "the rules of the data policy were quite generic." I do not quite understand that. Rules should not be different for different data fields. Maybe the meaning behind "rules" as used here is not clear. Please modify the sentence to make clear what exactly you mean.

L102 "initial versions" of the data I suppose

L109-110 "Some of the data sets have been published elsewhere on more specialized databases. These are explicitly mentioned in the text below." It would be better to have all data sets here in this table. It would be easy in the table to discern the data sets not in Pangaea. Thus all data sets would be in overview together.