Report #1

Submitted on 22 Aug 2023
Referee #1: Paul E. Renaud, pr@akvaplan.niva.no

Notes for the submission of interactive comments

Anonymous: Yes **No**

Formal manuscript rating and recommendation to the editor

(visible to authors and reviewers only)

1) Originality	**Excellent** Good Fair Poor
2) Significance	
Uniqueness	**Excellent** Good Fair Poor
Usefulness	**Excellent** Good Fair Poor
Completeness	Excellent **Good** Fair Poor
3) Presentation quality	Excellent **Good** Fair Poor
4) Data quality	Excellent **Good** Fair Poor

For final publication, the manuscript should be

accepted as is

accepted subject to **technical corrections**

accepted subject to minor revisions (review by editor)

reconsidered after **major revisions**

rejected

Please note that this rating only refers to this version of the manuscript!

Were a revised manuscript to be sent for another round of reviews:

I would be willing to review the revised manuscript.

I would not be willing to review the revised manuscript.

Referee Comment: 'Comment on essd-2023-263'

Report #1

With the recognition by researchers and funding agencies of the importance of archiving of research data we have seen a large increase in the availability of data for future comparative studies, use in other research activities (e.g. broad-scale modelling), and potentially a knowledge base for management. This will certainly increase the longevity and usefulness of the data, but this has also led to many (relatively) small data sets published in a wide variety of data outlets, many of which are not fully accessible and/or require registration, passwords, etc. This is counterproductive for the goals of requiring data publication. This ms describes a compilation of data from fragmented and previously unpublished data sources such that it useful. Benthic biota are important integrators of ecological conditions, and thus have the potential to track regional changes due to climate variability and other human impacts. Therefore, PANABIO, is an extremely valuable extension of the (not publicly available) ArcOD database.

GENERAL REPLY

First of all, we'd like to thank you for your positive evaluation, as well as for your constructive and valuable comments and suggestions, all of which we considered to revise and improve the manuscript. See below for our replies on your specific comments, including a short explanation how we addressed them in the revision (or a rebuttal in case we decided to not consider them).

Our most important revision is this: After consultation with the PANGAEA data curator, we decided to separately publish each of the 28 individual datasets that PANABIO encompasses. Within PANGAEA, PANABIO now also explicitly functions as a 'Data Collection' (with its own DOI), featuring 28 datasets, all of which also have their own DOI that are available under the overarching PANABIO data collection.

Consequently, we also revised our ESSD manuscript accordingly, by adding a new Table 2 with information about the 28 individual datasets within the data collection PANABIO.

This descriptive manuscript justifies and describes the database quite well. It gives a range of potential uses and ideas for upgrades that may be made in the future. The text is well written and the manuscript is complete. There are several inaccuracies/inconsistencies I would like to see remedied before final publication. **But please see the last comment below regarding access to the database itself (not the ms).**

1. In several locations the data are summarized at 'genus or species' level. A quick look at the data suggests that, contrary to the text on l 77, there are also data from fauna identified only to higher taxonomic levels (listed under 'Ranks' in Critterbase). Please clarify.

REPLY: Clarified and text appropriately modified ("… each record in the validated data collection represents a single taxonomic unit (mostly at species or genera, only in some cases at higher levels if a sound and reliable identification at species or genus level was not possible)".

2. On l 60 it is indicated that the domain of the database is the AMAP definition of the Arctic, but Fig. 2 and Table 1 indicate data from around 1500 stations from the Sea of Japan and Gulf of Ohkotsk. This area is not included in the AMAP definition so the text should be modified to both include these regions and explain why it is important to include them in a treatment of Arctic benthos.

   REPLY: Clarified and text appropriately modified, in both the Abstract ("… pan-Arctic realm, i.e., …, as well as some adjacent sub-Arctic regions, such as the Sea of Okhotsk and the Sea of Japan") and the text in section 2.1 ("… pan-Arctic study area, which also includes some adjacent sub-Arctic regions, such as the Sea of Okhotsk and the Sea of Japan (Figure 1; Table 1), since the distribution ranges of many species occurring in Arctic seas extend into the bordering areas").

3. On l 69-70 the authors indicate that more data will be added with a wider temporal range and spatial sampling density. This implies that some data sets/repositories have been identified. Perhaps a list of these would be useful. I am also curious how this will be accomplished. Is this an automated or semi-automated process or are we dependent on future funding (making this perhaps less likely to occur)?

   REPLY: The original text was obviously unclear in this regard and has now been appropriately modified, to clarify that further datasets to PANABIO will not be added from harvesting further repositories but through the contributions of further sets of data, either historical or novel, from PANABIO users, including ourselves ("We anticipate that additional sets of further historical or novel data will be added over time, by us and other data contributors using the Collector App of CRITTERBASE (Teschke et al; 2022), resulting in a steady growth of the data collection in number of records and samples, and thus also in spatial, temporal, taxonomic coverage and resolution").

4. Is it important that data sets on seafloor topography, chl a etc. (l 128) be included in a database on benthic fauna? Sure this could be helpful for someone who is looking for just these parameters to evaluate these data, but there is an entirely different set of files, data sources, and challenges involved. I suggest this database focus on faunal data.

   REPLY: We acknowledge your concerns in this regard. PANABIO, as CRITTERBASE (CB), is at its core an information system of faunal data – and

will remain so in the future. We also acknowledge that integrating environmental data directly into CB would be an ambitious goal that could not be achieved in due time (considering our current resources). Therefore, we toned down our outlook text to convey that we strive to create a workflow to ease linkage to environmental data in sources outside CB rather than integrating these data in CB itself ("... the development of an interface and a workflow to link the biotic data in PANABIO to ecological data layers from Arctic regions, such as, e.g., raster information on bottom topography, sea-ice and ocean dynamics, or Chlorophyll *a* distribution patterns, to support analysis and modelling work in day-to-day operations").

5. I am a little surprised that the numbers of stations and samples in the database are nearly identical. Benthic (grab) sampling is often conducted with triplicate (or higher) levels of replication. Were replicates pooled or omitted in some way?

   REPLY: No, replicate grab samples were not pooled or omitted during processing of PANABIO data input. Indeed, most historical data that account for the bulk of PANABIO'S current data are based on single samples taken at one station, for grab/corer and towed-gear data alike, and therefore overall station and sample numbers are close in the current PANABIO version. Please note, however, that PANABIO's data model includes the possibility of providing more than one sample at each station, for example if replicate grab samples were taken (as it is the case in most field studies nowadays).

6. **Note: I have gone into the database and it seems the front-end needs a bit of work to be user friendly. I was also not able to download any data, and it seems to hang up. Perhaps it is not quite ready for public use yet (there is a news item on the Critterbase site indicating major upgrades are being done). Shouldn't this all be in place when/before this ms is published/ officially accepted?**

   REPLY: Yes, we are aware that there are performance issues with the current version of CRITTERBASE. An improved and more versatile beta version is currently available under https://critterbase.awi.de/preview/. It will replace the old version by October 20, 2023.

Report #2

Submitted on 07 Sep 2023
Anonymous referee #2

Notes for the submission of interactive comments

Anonymous: Yes No

Formal manuscript rating and recommendation to the editor

(visible to authors and reviewers only)

1) Originality	**Excellent** Good Fair Poor
2) Significance	
Uniqueness	**Excellent** Good Fair Poor
Usefulness	Excellent **Good** Fair Poor
Completeness	Excellent **Good** Fair Poor
3) Presentation quality	**Excellent** Good Fair Poor
4) Data quality	**Excellent** Good Fair Poor

For final publication, the manuscript should be

accepted as is
accepted subject to **technical corrections**
accepted subject to minor revisions (review by editor)
reconsidered after **major revisions**
rejected

Please note that this rating only refers to this version of the manuscript!

Were a revised manuscript to be sent for another round of reviews:
I would be willing to review the revised manuscript.
I would not be willing to review the revised manuscript.

Referee Comment: 'Comment on essd-2023-263'

Report #2

The ms presents a PANABIO data collection that gathers records of benthos at species and/or genus level collected by various sampling gears. At the moment it holds over 124,000 records from samples taken at over 10,000 stations in years from 1800 to 2014. It is a very valuable initiative and such open databases are very needed for the Arctic scientific (and not only) community. I fully agree with Authors that open access databases, especially with quantitative ecological data, are essential for pan-Arctic comprehensive analyses and crucial in our work on understanding ecological processes in the Arctic Ocean, and I think Authors well justify the need for such data collection. They also provide good overview of what PANABIO offers in comparison to other available databases and good arguments why PANABIO is needed. In my opinion presented data set is significant and very useful, and of high quality, and should be published.

GENERAL REPLY

First of all, we'd like to thank you for your positive evaluation, as well as for your constructive and valuable comments and suggestions, all of which we considered to revise and improve the manuscript. See below for our replies on your specific questions and comments, including a short explanation how we addressed them in the revision (or a rebuttal in case we decided to not consider them).

Our most important revision is this: After consultation with the PANGAEA data curator, we decided to separately publish each of the 28 individual datasets that PANABIO encompasses. Within PANGAEA, PANABIO now also explicitly functions as a 'Data Collection' (with its own DOI), featuring 28 datasets, all of which also have their own DOI that are available under the overarching PANABIO data collection.

Consequently, we also revised our ESSD manuscript accordingly, by adding a new Table 2 with information about the 28 individual datasets within the data collection PANABIO.

I have however some minor questions that are listed below.

- Material and Methods, 2.1. Authors refer to Fig. 1 and Table 1 which shows the pan-Arctic study area. It does however look very broad with Table 1 listing also e.g. Sea of Japan in the Pacific and the Fig. 1 is probably bit misleading as it shows quite far south locations in both Pacific and Atlantic e.g. North Sea (which are however not listed in Table 1). Adding a line showing the most southern border of the data (stations) included would be helpful, and maybe a short justification for including non-Arctic locations can also be added.

  REPLY: Issue clarified and text appropriately modified, in both the Abstract ("… pan-Arctic realm, i.e., …, as well as some adjacent sub-Arctic regions,

such as the Sea of Okhotsk and the Sea of Japan") and the text in section 2.1 ("… pan-Arctic study area, which also includes some adjacent sub-Arctic regions, such as the Sea of Okhotsk and the Sea of Japan (Figure 1; Table 1), since the distribution ranges of many species occurring in Arctic seas extend into the bordering areas"). In the map in Figure 1, only the names of those sea regions are given, from which data are included in PANABIO.

- Material and Methods 2.3 – are only species and genus included or also higher taxonomic levels? If in original data set only higher taxonomic levels were present were these excluded from PANABIO?

  REPLY: Clarified and text appropriately modified ("… each record in the validated data collection represents a single taxonomic unit (mostly at species or genera, only in some cases at higher levels if a sound and reliable identification at species or genus level was not possible)".

- Data availability – a detailed information on the number of records and samples is provided. Later Authors say that the data set is expanding and new records are being added continuously. If this is the case it would be useful to add the exact date (instead of saying just "currently").

  REPLY: An exact date specifying the term "currently" is provided in the revised version ("5 October 2023").

- Outlook – Authors declare that there are plans to collect and include more information on the environmental settings that will supplement the existing data on benthos. That would be a very valuable add-on. Since plans to expand the benthic data are mentioned earlier a short information on that process here would be useful.

  REPLY: PANABIO, as CRITTERBASE (CB), is at its core an information system of faunal data – and will remain so in the future. We now realized that integrating environmental data directly into CB would be an ambitious goal that could not be achieved in due time (considering our current resources). Therefore, we toned down our outlook text in the revised version to convey that we strive to create a workflow to ease linkage to environmental data in sources outside CB rather than integrating these data in CB itself ("… the development of an interface and a workflow to link the biotic data in PANABIO to ecological data layers from Arctic regions, such as, e.g., raster information on bottom topography, sea-ice and ocean dynamics, or Chlorophyll *a* distribution patterns, to support analysis and modelling work in day-to-day operations").

- A general comment: I visited PANABIO and tried to download some data. It does however work very slow and I was not successful with neither

downloading nor showing data on the map. Is that because the work is still in progress?

REPLY: Yes, we are aware that there are performance issues with the current version of CRITTERBASE. An improved and more versatile beta version is currently available under https://critterbase.awi.de/preview/. It will replace the old version by October 20, 2023.

Report #3

Submitted on 19 Sep 2023
Anonymous referee #3

Notes for the submission of interactive comments

Anonymous: Yes No

Formal manuscript rating and recommendation to the editor

(visible to authors and reviewers only)

1) Originality	**Excellent** Good Fair Poor
2) Significance	
Uniqueness	**Excellent** Good Fair Poor
Usefulness	**Excellent** Good Fair Poor
Completeness	Excellent **Good** Fair Poor
3) Presentation quality	Excellent **Good** Fair Poor
4) Data quality	**Excellent** Good Fair Poor

For final publication, the manuscript should be

accepted as is
accepted subject to **technical corrections**
accepted subject to minor revisions (review by editor)
reconsidered after **major revisions**
rejected

Please note that this rating only refers to this version of the manuscript!

Were a revised manuscript to be sent for another round of reviews:
I would be willing to review the revised manuscript.
I would not be willing to review the revised manuscript.

Referee Comment: 'Comment on essd-2023-263'

Report #3

The overall goal of the manuscript is to describe the open access availability of benthic data (genus and species level, presence, abundance and/or biomass) obtained from various archives over the pan-Arctic scale. The ability to link data sets from various archives into one open access portal will provide a valuable and timely location for accessing data within a changing Arctic system. The plan for PANABIO is to connect both benthic faunal data sets as well as include associated environmental data. The strength of this effort is to have one open access portal for benthic faunal data that would be QA/QC'd for accuracy. It is important that there be doi numbers to the original data sets so that the user could go back to the original archived files for evaluating data quality, lead investigator interactions, and find associated detailed metadata that may not be translated directly to the PANABIO file. I do see some missing benthic files that would be appropriately linked to PABABIO, so there should be a clear mechanism for interested scientists to direct PANABIO managers to connect to those published data files. Overall, I rank the manuscript as well-written and publishable, with minor revisions as listed above and more specifics below.

GENERAL REPLY

First of all, we'd like to thank you for your positive evaluation, as well as for your constructive and valuable comments and suggestions, all of which we considered to revise and improve the manuscript. See below for our replies on your specific questions and comments, including a short explanation how we addressed them in the revision (or a rebuttal in case we decided to not consider them).

In response to your general comments above, our most important revision is this: After consultation with the PANGAEA data curator, we decided to separately publish each of the 28 individual datasets that PANABIO encompasses. Within PANGAEA, PANABIO now also explicitly functions as a 'Data Collection' (with its own DOI), featuring 28 datasets, all of which also have their own DOI that are available under the overarching PANABIO data collection.

Consequently, we also revised our ESSD manuscript accordingly, by adding a new Table 2 with information about the 28 individual datasets within the data collection PANABIO, including (if available) the DOIs of related scientific papers (based on the data available in the dataset) and data publication (where the data/metadata are, entirely or partly, also available).

Specific comments:

Line16-17: species or genus samples, but only the means of replicates will be provided. Files that have actual individual replicate data vs. means should also be identified in a table on the site.

REPLY: The original text was obviously unclear in this regard and has now been appropriately modified, to clarify that the actual replicate data (not means) are given in the datasets. In general, replicate grab samples were not pooled, averaged or omitted during processing of PANABIO data input. However, most historical data that account for the bulk of PANABIO'S current data are based on single samples taken at one station, for grab/corer and towed-gear data alike, and therefore overall station and sample numbers are close in the current PANABIO version. Please note, however, that PANABIO's data model includes the possibility of providing more than one sample at each station, for example if replicate grab samples were taken (as it is the case in most field studies nowadays).

Line 18-19: You mention the pan-Arctic collections, but your data files go further than the defined area. Since this is an Arctic-centric paper (central arctic basins and shelf systems), shouldn't you only list those species occurring in the pan-arctic region and not other areas, such as off Japan?

REPLY: Issue clarified and text appropriately modified, in both the Abstract ("… pan-Arctic realm, i.e., …, as well as some adjacent sub-Arctic regions, such as the Sea of Okhotsk and the Sea of Japan") and the text in section 2.1 ("… pan-Arctic study area, which also includes some adjacent sub-Arctic regions, such as the Sea of Okhotsk and the Sea of Japan (Figure 1; Table 1), since the distribution ranges of many species occurring in Arctic seas extend into the bordering areas"). In the map in Figure 1, only the names of those sea regions are given, from which data are included in PANABIO.

Line 24: I was able to look at the snapshot of data, but not download the file. Is this something that would be live by the time the paper is published?

REPLY: We are aware that there are performance issues with the current version of CRITTERBASE. An improved and more versatile beta version is currently available under https://critterbase.awi.de/preview/. It will replace the old version by October 20, 2023.

Line 35 and line 43: You cite the footprint paper for Wassmann et al 2011, but not the updated footprint paper by Brandt et al 2023. Should you not include the results from this updated footprint paper in your manuscript?

REPLY: Thank you for this suggestion. In the revised manuscript, the Brandt et al. paper from 2023 is now also cited.

Line 38: please correct the portion of sentence "…not only represents an own, specifically…" What is "an own" referring to?

REPLY: Text has been revised accordingly (by omitting the word "own").

Line 69-71: The data citation location is an example of an initial evaluation and the authors state that more data will be added as the project progresses. There should

be an easy and explicit procedure for the community to send doi numbers of already nationally archived species/genera benthic data to this open source project. How will you engage with the wider science community to solicit further benthic data that has already been submitted to individual country data archives to connect to the PANABIO project? Also, there needs to be a detailed record of where the raw data is located, the doi number in the national archive, and associated metafiles.

REPLY: Please note that CRITTERBASE (CB), the data warehouse that PANABIO is a regional component of, offers a special tool, the Collector App, that users can use for contributing their data to CB. The usage of this tool is described in more detail at CB's web-based frontend (https://critterbase.awi.de/preview/#app) and in Teschke et al. (2022) (https://doi.org/10.1038/s41597-022-01590-1). We modified the text of our manuscript to inform the reader (and potential PANABIO user and data contributor) about this way to provide data to PANABIO. This data-input workflow also guarantees that the "records of where the raw data is located" (such as "doi number in the national archive, and associated metafiles") are included in the PANABIO metadata (as evident in Table 2 of the revised manuscript for datasets currently included in PANABIO).

Line 127-129: The authors suggest adding raster information on bottom depth, sea-ice and ocean dynamics, or Chlorophyll a distribution patterns. However, this should be secondary activities or even just listing the doi for those type data available in archives to start. I suggest the ice cover parameter is not at a low enough scale, but some metric of the percent ice cover over a certain amount of days over the sampling site is more appropriate. Also, sediment parameters (TOC, grain size, sediment chlorophyll) would be valuable data for future modeling activities. Again, the authors should provide clear mechanisms for scientists interested in collaborating on PANABIO to connect known benthic data and ancillary environmental data doi numbers to the open access plans for this program.

REPLY: In response to the comments from the other reviews, we have reconsidered out plans regarding adding environmental data to PANABIO. PANABIO, as CRITTERBASE (CB), is at its core an information system of faunal data – and will remain so in the future. We now realized that integrating environmental data directly into CB would be a very ambitious goal that could not be achieved in due time (considering our current resources). Therefore, we toned down our outlook text in the revised version to convey that we strive to create a workflow to ease linkage to environmental data in sources outside CB rather than integrating these data in CB itself ("… the development of an interface and a workflow to link the biotic data in PANABIO to ecological data layers from Arctic regions, such as, e.g., raster information on bottom topography, sea-ice and ocean dynamics, or Chlorophyll *a* distribution patterns, to support analysis and modelling work in day-to-day operations"). Regarding "… clear mechanisms for scientists interested in collaborating on PANABIO to connect known benthic data and ancillary environmental data", see our reply above.

Table 1 and Figure 1 show data zones outside the pan-Arctic direction of PANABIO. The authors should explicitly explain why regions outside the pan-Arctic focus are being included in PANABIO.

REPLY: Please see above our response to your comment on Line 18-19.