

Reply to comment on essd-2022-324 (<https://doi.org/10.5194/essd-2022-324-RC4>) by an Anonymous Reviewer

The authors would like to thank the Anonymous Reviewer for the time on reading and reviewing the manuscript.

Our responses (plain text) to point-by-point comments are listed after Reviewer's comments (italic and bold font) followed by the main changes to the manuscript (italic font). Please, refer to the annotated manuscript and companion files.

Please, find the access to the new database files via a key provided by the editorial office.

Points of revision for the paper:

Line 121-122: "The whole database can be managed using R version 4.2.1 (R Core Team, 2022). It can be uploaded and managed with geographic information system software such as QGIS and ArcGIS after changing the table format from wide to long."

It would be great, if the database could be managed with geographic information system software such as ArcGIS. If I would have access to the database, I would have tested this myself. I think it would be really helpful, if the authors provide some more information on how to do this. For example, either the database could be uploaded in different file formats. At the moment the file that is uploaded at Pangaea seems to be an .xlsx file. This is of course great for standard users of .xls based software but makes it more complicated for other users (R for example), especially with such a big file. One solution could be to upload the database in different file versions. One text file in long format that could be directly imported into QGIS or ArcGIS, another one in wide format for other applications and the original .xlsx file. This would be very helpful for users but also would make updating the database in the future very tedious, since all the files would have to be exchanged. Another possibility would be to publish an R friendly text file that could easily be imported into .xls based software. If this file is in a wide format, it would be easy to provide an R script to convert it into a QGIS/ArcGIS friendly long format, either in the paper or in the supplement. It is also not directly clear to me, which columns would have to be merged to make the format GIS friendly. Here is an example for such a cookbook (using the gather() argument from the tidyr library): (...). These are of course only optional suggestions on how the accessibility of the manuscript might be improved. My other point is not really a change in the manuscript itself but more a question/suggestion about the future of this database.

Thanks for the Reviewer's suggestions and comments. Because it is our goal to reach all types of potential users interested in benthic foraminifera, regardless of their preferred method to handle large datasets; we have followed the Reviewer's suggestion and provided BENFEP (now BENFEP_v1) in two formats: short and long. Both versions are given in text format. For those who use more conventional software to read spreadsheet-type files, we keep the previously submitted format (BENFEP_v1_short). For those who prefer to access data using other types of software (e.g., geospatial software), we provide the database in long format (BENFEP_v1_long). The long format has been also suggested by Lukas Jonkers and another Anonymous Reviewer. Additionally, we provide general indications on how to manage the short version in R.

We have also implemented changes to the Supplement section to make clear to the potential users the taxonomic concepts provided by each author and to bring more transparency to the harmonization procedure. Some of the information included in the new Supplement files is incorporated in BENFEP_v1_long. Those changes are explained below.

Changes to the Supplement: Please, refer to the new Supplement section.

- The new Supplement File 1 incorporates more comprehensive taxonomic information (status, rank, order, family, class, AphiaID, etc) about the species harmonized in BENFEP (now BENFEP_v1) than the previously submitted file.
- The new Supplement File 2 includes the original authors' taxonomic concepts for each species harmonized in BENFEP_v1 and listed in new Supplement File 1.
- The new Supplement File 3 includes specific remarks on the harmonization procedure.
- The new Supplement File 5 provides general information on how to perform basic operations with BENFEP_v1_short in R

Changes to BENFEP (now BENFEP_v1): long and short formats.

The potential users of BENFEP_v1_short will have the original authors' specific assignments in a separate file (Supplement File 2). Conversely, the long-format version integrates all the information provided in the short format plus taxonomic information for each species (File 1) and each author's species assignments (File 2).

Please, refer to new section 2.5 of the revised version of the manuscript (which integrates former sections 2.4 and 2.6). There we explain the structure of the two formats of the database. Please, see the new Table C1 and Table C2, where we explain column names and column codes for BENFEP_v1_short and BENFEP_v1_long.

We would like to indicate that BENFEP_v1_short and BENFEP_v1_long can be opened and managed in GIS software, such as QGIS and ArcMap, among others. Both files can be opened as tables and plotted using the coordinate columns (long, lat) and the EPSG:4326 (WGS84). At this point, the user can transform the tables into a GIS format to facilitate the analysis and visualization. In the specific case of the short format database, the users might consider the possibility of managing it as a geodatabase or a geopackage file, because the number of fields is higher than 255 (limit of other GIS formats like shapefiles). In the case of the long format, there are no limitations in this regard.

Line 338-342: "This database is conceived as a springboard to store future quantitative data of benthic foraminifera in the East Pacific and make them available to the scientific community. It can be enlarged with new records as they are being generated or after the authors request, therefore providing an ongoing live resource. Any changes to add, correct, or update taxonomic categories to an existing version will be indicated in PANGAEA."

How do the authors think about the future of the database? Should there be an easy protocol or "cookbook", how to add data of new records as they are being generated? Are there any permanent members of the group that are able to sustain, clean up and update the database? I think for individual members this might be an impossible task but it should be important to avoid misuse of the feature to update the database. One suggestion would be to provide a review system like in WORMS: Uploaded datasets that are not reviewed, yet, will be marked as "not reviewed". Reviewed datasets might be marked as "accepted" or "unaccepted", if there are any issues considered by a reviewer. Reviewers could be volunteer foraminifera taxonomists. I think there might be a big support in the community for such an effort.

Unfortunately, we do not have at the moment the resources to migrate BENFEP_v1 to a system that enables the reception of external datasets, like WoRMS. We anticipate that the new entries to the database will be communicated via email to the corresponding author or any other of the current team members. After the reception of the dataset, the procedure would be very similar to the one described in the reviewed version of the manuscript, with entails taxonomic harmonization (2.4), curating metadata (section 2.6), and establishing a quality control on the data entered. In the reviewed version of the manuscript, we have now expanded on the taxonomic harmonization procedure.

Changes to the text: Text added (new section 2.4, Taxonomic harmonization, previously section 2.5).

In order to find the valid species name, we searched each author's original species assignment in the WoRMS research engine. This procedure enables to identify whether the original species name is accepted (valid species) or if it is a synonym of the valid species or taxa correspond to a variety or a subspecies. When the original species name was not currently in use, it was substituted by the valid species, subspecies, or variety name.

Finally, are there any plans to include other ocean basins into the database in the future? These are just some questions/suggestions about the future of the database and of course, this is an own project by itself and some things cannot be directly integrated into the paper and database. For example, the review system for future datasets would need an own platform or deeper collaboration with Pangaea. Though, I think it is worth it to think about the legacy of such a huge project and maybe to integrate some points about the discussion of the future of the database at the end of the paper about “Data availability and future plans”.

The reviewer is right, and we regret not having delved further into the process of updating the database. In order to facilitate updating, we have re-named BENFEP to BENFEP_v1 along the manuscript and in the PANGAEA repository. Besides, we have now clarified in the reviewed version of the manuscript how we plan to update BENFEP_v1 in the future.

Changes to the text: Text added and rephrased (new section 5. Data availability and future plans, previously section 6).

The BENFEP_v1 database can be accessed from <https://doi.org/10.1594/PANGAEA.947086> (Diz et al., 2022b). This database is conceived as a springboard to store future quantitative data of benthic foraminifera in the East Pacific and make them available to the scientific community. It will be open for any new quantitative data entry and thus, it welcomes any new data published or provided by any contributor. The database will be updated by the authors once a considerable number of new entries need to be incorporated or changes are required to update taxonomic categories to an existing version. New versions of BENFEP will be submitted and curated in PANGAEA. Collaborations with individual researchers and institutions are welcomed specially regarding potential expansion to other ocean basins.

The current reviewing processes and several interactions during conferences over the last couple of months have been insightful in understanding which the needs and the interest of the benthic foraminifera community are. Expanding BENFEP_v1 and extending the concept to other oceans requires external funding (e.g., for quality control and to host the database in a dedicated server and to migrate and upgrade it to PostGIS or similar), active community input and establishing/strengthening collaborations with other researchers, and open access repositories.

We will undertake steps in that direction, and it is in our mind to expand the scheme presented here (i.e., the improvements that might come) to other basins.