

Authors reply to: Aleksandra Lewandowska (Referee): This is an important and much needed tool with a great potential for further development.

Dear Professor Lewandowska,

Thank you for the positive thoughts and suggestions you have provided to improve our manuscript, we greatly appreciate you taking the time to review it. Below we address each of your points, including your text in black font, and our response in blue font.

Referee comment on "The Plankton Lifeform Extraction Tool: A digital tool to increase the discoverability and usability of plankton time-series data" by Clare Ostle et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-171-RC2>

The manuscript is very well written, all functions of the database are clearly described and justified. I can only hope that this tool finds its way to the stakeholders and the policy makers in Europe.

*I especially admire the functional groups **look-up table and the confidence rating**. It is easy to expand and continuously improve when more data are added. The spatial coverage of the database can be rapidly expanded, especially if SMHI extends access to their time-series from the Baltic Sea. If this happens, it would make sense to expand the lifeforms table by filamentous cyanobacteria to track their blooms in the Baltic Sea. Such information would be highly relevant to policy makers in the Baltic Sea region and some other coastal areas in Europe.*

This is a great suggestion for future versions of PLET and one that we would hope to implement if the datasets were expanded to this area. Thank you for this insight.

*I also appreciate that data from different sources are not aggregated. This gives a lot of **freedom for the users**, who can apply and develop their own statistical techniques to make generalisations.*

*Figures 5-8 are **wonderful examples how to use the PLET** and what kind of information can be extracted. I do not expect that the database developers will offer such visualisation tool, but this content of the manuscript is a great source of inspiration for the users. **Figures 2-3 on the sampling effort are extremely important** from the point of view of statistical diagnostics. It would be great if such figures could be included in the metafile description on the website, so that the user can easily see where are the potential gaps in each dataset and what are the limitations.*

We really like this suggestion, however some of the dataset's gaps will change depending on the regions selected for outputting lifeforms. We are hoping that we can provide a link to this manuscript and figures within when it is published, to act as a resource for the use of PLET and the current source datasets.

Although the manuscript and the database are impressive, below are my suggestions for some improvement.

Regarding the manuscript:

*It might be a good idea to highlight the **advantage of PLET over satellite derived information in the introduction**. There is a short sentence in the discussion about the limitation of bulk indices, such as total chlorophyll a concentration, but I think it would be good to have it earlier in text.*

Yes, we agree that this is an important point that we have overlooked. We have accordingly inserted a short paragraph at line 75 (i.e. making it the new third introductory paragraph) to add these Earth Observation data strengths and weaknesses.

“To map changes in ocean colour, Earth Observation (EO) satellite tools provide unparalleled spatial coverage, and now offer the prospect of 20 years of ocean colour data, with increasingly resolved information, for example on trends of specific size-fractions of Chlorophyll a (Schmidt et al. 2020). However, the EO techniques are still not yet sufficiently developed to obtain information on changes in abundance of the key component planktonic functional groups, particularly for the zooplankton. Additionally, some taxonomic datasets now have up to 90 years of data which provide a critical perspective in assessing long-term change and which is unparalleled by satellites. We therefore need to maintain direct monitoring approaches for a holistic view of the plankton, and northwest European waters are particularly well-blessed with these time-series.”

*There is no mention of **current development of plankton trait databases**, such as nutrient utilisation traits database (Edwards et al. 2015 - Ecological Archives), Baltic Sea phytoplankton traits database (Klais et al. 2017 - Functional Ecology), French phytoplankton traits database (Treyture et al. 2020 - Scientific Data). It would be good to place the PLET in their context. Maybe it would be worth adding the links to such trait databases in the future, if they exist for individual datasets, e.g. in the metafile description. This would be especially valued in those cases where taxonomic lists are made available.*

We agree these are important contributions that should be mentioned, we have added the following text has been added at line 129 to place PLET in the context of these datasets:

There are a number of plankton trait datasets and plankton compilation efforts that are complementary to the PLET with the potential to feed into future versions of the tool, such as the nutrient utilisation traits dataset (Edwards et al., 2015), the Baltic Sea phytoplankton traits dataset (Klais et al., 2017), and the French lakes phytoplankton traits database (Laplace-Treyture et al., 2021). While these are highly valuable resources, the authors are not aware of a platform to bring such information together and disseminate it in a consistent format. The design of PLET allows for this lifeform extraction and dissemination, with the aim to incorporate further plankton trait datasets in future versions.

References added to manuscript:

Edwards, K. F., Klausmeier, C. A. and Litchman, E.: Nutrient utilization traits of

phytoplankton, *Ecology*, 96(8), 2311–2311, doi:10.1890/14-2252.1, 2015.

Klais, R., Norros, V., Lehtinen, S., Tamminen, T. and Olli, K.: Community assembly and drivers of phytoplankton functional structure, *Funct. Ecol.*, 31(3), 760–767, doi:10.1111/1365-2435.12784, 2017.

Laplace-Treyture, C., Derot, J., Prévost, E., Le Mat, A. and Jamoneau, A.: Phytoplankton morpho-functional trait dataset from French water-bodies, *Sci. Data*, 8(1), 1–9, doi:10.1038/s41597-021-00814-0, 2021.

Please add a short information how the PLET is **dealing with synonyms and updates in plankton taxonomy**. Is there an automatic check applied (e.g. with WORMS or AlgaeBase) or does it need to be made manually by data providers? In general, is there a systematic data quality check performed upon submission of the time-series? How often such quality check should be performed? I wonder how to ensure the consistent data quality among PLET database, if the data quality check is the responsibility of data providers. I am sure this is not a problem at the moment, but how to guarantee it in the future when the tool expands?

Thank you for this point, it is something that we have worked hard to ensure consistent data quality, and as you can imagine it is an evolving process, we have added the following text has been added at line 160 to describe current procedure:

When new datasets are submitted to the PLET the data providers supply aphia IDs of all of the taxa within their dataset. Following the pre-processing of the data by the data providers, the data manager of PLET and the manager of the Plankton Lifeform Traits Master List does a check of the submitted aphia IDs to highlight any missing taxa. Any taxa that are not included within the Plankton Lifeform Traits Master List are checked for compatibility with the lifeforms and their traits are added in discussion with an expert group and the data providers.

Please add a **link to the SMHI portal** in the chapter 3.1.8 similarly as you did for the other time series (<https://www.smhi.se/en/services/open-data/national-archive-for-oceanographic-data/download-data-1.153150>).

Thank you, the link has been added to the manuscript at line 351.

Regarding the PLET:

*The **website performance** needs significant optimisation. I believe the problem is not in PLET, but rather in the host server, but this should be fixed before the tool expands. I tried different browsers and different computers, but the problem persists and the service website jams easily, even when I'm trying to limit my search and download data in small pieces. If this causes problems now, it will grow in the future.*

Thank you for your suggestion, we agree that we need to develop PLET further as the database increases in size in future versions of PLET. We have developed a caching system whereby if a user has already requested an area/period than that data is cached and outputted to speed up processing, but there are still queries that have not been cached and it can hang on the larger data requests. We are looking at

aggregating some of the largest datasets (such as CPR data) into subsets such as geographic average boxes of 1 x 1 degree. We have also written some scripts using curl/wget to run through iterations of data queries to pull out lifeform outputs in bulk, and for the CPR data these outputs can be grouped by degree boxes. We hope to publish these scripts on the PLET site once they have been tested and finalized.

*The short **description of sampling methodologies** (chapter 3.1) is excellent and could be added to the metafile together with the information on sampling effort (see my comment to Fig 2-3). This would make the service more user friendly.*

We really like this suggestion, and have added each method paragraph to the individual data providers metadata page within the tool.

*As the tool is meant for biodiversity assessment, it might be good to add some basic **information on changes in taxonomic resolution**. For example, species accumulation curves for each time series could give a clue on significant change in resolution, which can affect interpretation of the outcomes. As many diversity indices, including the most popular species richness, are sensitive to changes in sample size, this is an important information on data quality. Depending on the visualisation, such curves could have annotations with information on changes in methodology or instrumentation, which correspond to observed inconsistency.*

Thank you for this suggestion. The reason we have used lifeforms as indicators is because this is a robust grouping that many different datasets can contribute to, despite differences or changes in taxonomic resolution. For example, even if the taxonomic resolution during the times series improves, such as Decapoda larvae becoming identified to infraorder level (Brachyura), you would still have the same abundance values in the meroplankton and crustacea lifeform outputs from PLET. We stress in the first paragraph of section 3.1 that all the source time series data were screened such that they were internally consistent (i.e. taxa were recorded at the highest taxonomic level in which they were consistently recorded in, through each time series, lumping where necessary). For this reason, the issue of changes in taxonomic resolution within each time series is not an issue. However, each time series was analysed differently, with different levels of taxonomic resolution, so the comparison of diversity indices across multiple time series is not recommended. We have added text on these limitations on lines 570 with an added sentence (italicised below):

“..... coverage does not bias the combined results. For example, due to the differing taxonomic resolution and sampling methods across the various component time series, we do not recommend simple comparisons of indices of species richness or diversity. There is also the flexibility to.....”

Thank you again for your insightful and thorough review, we hope you find our responses satisfactory.

Dr. Clare Ostle and co-authors.