Reviewer 1

We thank the reviewer for taking the time to provide a useful and considered review of this manuscript, and the dataset detailed therein. We have reflected on the comments, made changes to the manuscript where appropriate, and provide below a point-by-point reply. The dataset itself is already published on PANGAEA, but all of the comments made here will be integrated into the development of v2.0, which will be uploaded to PANGAEA early next year. That being said, a live version of the dataset will be directly benefiting from your comments, and is available via a user interface that may be accessed via the WP1 website by clicking the 'Data access' tab.

We have addressed specific comments in line with the reviewers original text.

This is a very good step to provide an access to the coastal Arctic data.

It is worth pointing out here that the focus of the dataset is to provide data within seven specific Arctic fjords, not for coastal or open waters.

However some data are easy to obtained and self-explanatory (SST, ice cover, glaciers balance etc) some data are difficult to obtain, yet are simple (fisheries landings, jobs, tourists movements etc) and some data need to be carefully presented, as they have very broad meaning - in this work are data "species richness".

We agree with the reviewer that the data contained in this amalgamation range in difficulty and utility.

This for biologist means nothing. Species richnes in what taxon ? (there are no in the world complete species lists for the one site). It can be the species richness of mesozooplankton from specific water column (say 0-200 m) or microplankton species from euphotic waters (in spring ? in summer ?, all year ?) - all taxa of microplankton including Ciliata and Amoeba and other little known groups ? Same questions go for all groups of benthos (what habitat is presented ? rocky shore or soft bottom near the glaciers ? etc)

We agree that the 'species richness' field provided in this amalgamated dataset does not quite provide the intended information for which it was created. As detailed in the manuscript, the types of data provided in this dataset were determined via a published review paper on key drivers of change in EU Arctic fjords. Because species richness data tended to be lacking, we made a first attempt to fill this gap. v2.0 of the dataset will incorporate your feedback to provide a more specific accounting of the richness of species by different functional groups and/or taxa.

As to the data completeness there is much more to dig out - probably the best is to use the research institutions with long term observations (like authors have identified Norwegian Danish and German projects). There is also the Institute of Geophysics Polish Academy of Sciences research station in Hornsund, Svalbard with meteo data complete from 1957 to today, and glaciers balance last 40 years, oceanographic (SST) last 15 years.

The search for new open source datasets, that are without sharing restrictions, is an ongoing process that will continue for the length of the FACE-IT project, for which this amalgamation of data was commissioned. We thank the reviewer for their suggestions on these additional sources to investigate. Specifically though, Hornsund falls outside of the FACE-IT study sites, which is why no data were collected for it. As indicated in Section *2.3.5. (Additional sources of note)*, meteorological data also fall outside of the scope of this dataset

There is Institute of Oceanology PAS with r/v OCEANIA sailing every summer to the Svalbard coastal waters since 1989, with complete data set on hydrology, mesozooplankton and benthic life of Hornsund and Kongsfjorden.

We thank the reviewer for their suggestion on investigating the use of data from the r/v OCEANIA. While open ocean data is outside of the scope of this dataset, data within Kongsfjorden is very much something we want to include. Querying the dataset, we see that it does contain 8,869 OCEANIA hydrographic data points (temperature and salinity) from 1997 to 2003 (via PANGAEA) within both Kongsfjorden and Isfjorden. Investigating this issue more closely, we see that it also contains the zooplankton datasets as well. We will ensure in v2.0 that all available applicable data from the OCEANIA cruises are included.

There are Norwegian long time benthic-photo surveys for over 40 years in the rocky sublittoral of several Svalbard fjords (University of Tromso and AKVAPLAN NIVA) .

Photos surveys (unless converted to a spreadsheet of presence absence etc.) were determined to be outside of the scope of this dataset. The reason for this is that all of the data collected, from glaciology to national statistics, need to be able to be presented together in a single spreadsheet or text file. Since photographs cannot be included in such files, it has not been possible to include them in this dataset. The reasoning for this omission of photo survey data has been added to the introduction of the revised manuscript.

Lots of primary production data from Svalbard fjords have been published, need to be digged out from papers.

While primary production data are lacking in this amalgamated dataset, and are a top priority for v2.0, the extraction of data points from published literature is a high cost as well as a low accuracy and return activity that is avoided in favour of locating sources of FAIR data. It is for this reason that there is so much effort in the development of the FAIR data principles. In v1.0 of the dataset, many individual PAR data points for Kongsfjorden were harvested from published literature. It is from this experience that this method of collecting data is avoided.

I understand that authors used available data repositories, unfortunately these are still holding only a fraction of known information about coastal Arctic. The challenge is very difficult, and presented paper is a nice step, but only a minor step.

We agree with the reviewer that the amount of data not accounted for here is greater than the data that has been compiled. This is indeed a challenge in science, and one that must be addressed by the scientific community. Much work is being done in this regard, and many teams are also digitising pre-digital data in order to make them FAIR. It is regrettable that some data are sitting in drawers or in non-accessible databases and repositories. The work continues and the documentation for the assembly of this dataset is a step forward in that process. We are glad that you acknowledge that our efforts are a valuable step but beg to differ with the qualifier "minor". It has taken roughly one year to compile v1.0 of the dataset.