Overall, I greatly appreciate this manuscript, including the inclusion of extensive metadata, representing a large amount of work; I think that this manuscript and its product (BENFEP) provide a great service to the community and should be published. In general, the figures read well and are illustrative. I have some remarks, including some suggesting improvements, some of which are rather minor and nitpicky, but will improve easy of reading and database use by the community.

Remark 1: In my opinion research on paleoceanographic use of benthic foraminifera is indeed (as stated by the authors of this manuscript) greatly hampered by the lack of integration of (the many) published data, which at least in part is caused by the problems on integrating taxonomy. In this manuscript the authors 'unify' the taxonomy to the WORMS database, which is a good first goal. However, the authors do not evaluate the taxonomic concepts of different authors, and do not aim to do so. In fact, that task might be difficult to impossible, because none of the compilated papers provides good descriptions and/or figures of all taxa mentioned. This database does not include descriptions and/or figures of the taxa, thus the user remains uncertain about the uniformity of the species concepts included in this study. I would think that definitely some of the more 'difficult taxa', e.g., various species names of *Gyroidina/ Gyroidinoides*, of *Lagena*, of *Lenticulina* could well be used differently by different authors, a problem not (fully) resolved by the synonymies given in WORMS. One also cannot exclude authors making mistaken identifications. As an example, *Nuttallinella florealis* White has been extinct since the end Paleocene benthic foraminiferal extinction, thus is unlikely to have been found in surface samples, and was probably misidentified by Liu, 2001. *Obesopleurostomella brevis* and many other uniserial taxa with a complex aperture became extinct in the mid-Pleistocene benthic foraminiferal extinction event (Hayward et al., 2012, Cushman Foundation Foram Res vol 43), and have been marked as 'fossil only' in WORMS. Taxa with such a relatively recent extinction date have been found in surface samples by e.g., Brady in the Challenger Report: see Hayward and Kawahata, 2005, Extinct foraminifera in Brady's *Challenger Report*, J Micropalont. 24, 171-175, who list *O. brevis* as such, as well as *P. alternans*, also listed in the Supplement. This of course might be caused by the age of the surface-sediment sample not being not well controlled (see below, Remark 2). The authors should in my opinion say something about the possibility of mistaken identification, possibly where they are discussing taxa mentioned by one author only (lines 244-on), or where they are discussing 'taxonomic concepts (section 3.6.1). In my opinion, this lack of detailed checking of the taxonomy definitely does not invalidate the present manuscript. The goal of comparing data from different authors (thus taxonomic 'schools of thought'), as well as different collection methods is a very good and worthwhile first step: after all, the Paleobiology Database is extremely useful for the community without all taxonomic confusion having been solved. In fact, this database effort will make taxonomic evaluation in the future easier to do since so many disparate datasets are here collated. However, I think that the authors should mention the potential problems due to taxonomic misidentification and non-realized synonymies.

Remark 2: The apparent occurrence of various extinct species in the database: I suggest that the authors mark it in the Supplement (Species list) if the species is given as 'fossil only' in WORMS, in order to warn the reader/user. In such a case, the explanation could be one or

more of the following: the species may be misidentified (would be interesting to know whether it was tagged as 'living'), the species may not be extinct (in disagreement with WORMS), or the surface sample may in fact not be recent (or even Holocene) in age. Many of the sampling methods described will not recover 'undisturbed' sediment, i.e., the sediment/water interface will not be recovered in most samplers except multicorers. Could the authors say something about this, either in the text or in the description of sample methods? They mention this in line 296, but could they have added a bit more information, e.g., as described above?
It would e.g., be good to know if the authors of the cited article did radiocarbon dating to determine a Holocene age. It would be specifically of interest to know if species described as 'fossil only' in WORMS had been stained and recognized as 'living'.

Remark 3: I am not sure whether I am convinced by the authors that the eastern equatorial Pacific should get priority for such a compilation (maybe there are more data for the Atlantic?). However, the question whether this is the highest priority region on Earth for such a compilation - which could be made because we are looking at a large part of the largest ocean on Earth - is unimportant. After all, we must start somewhere with a global database, and exactly where we start does not really matter - as long as the database is open for extension in the future (see below).

Remark 4: The authors mention that BENFEP is indeed open to such expansion (e.g., lines 23-24), and my earlier remarks indicate that it would indeed be important for the community if BENFEP can be expanded. However, the potential pathway(s) to such extension, e.g., into different oceans, adding additional sources (e.g., articles by Saidova published in Russian), is not very clearly spelled out. How do the authors envisage to keep BENFEP going and growing? Are they proposing to supervise and/or impose quality control on the database? will there be an opportunity for community input? Which entity (university? museum? association?) will; hoist the database in the long term? Potential cooperation which groups hosting databases for other microfossils?

Notes by line number:
23-24: the authors say ' *We complement BENFEP with an additional database integrating metadata and stations geolocation of benthic foraminiferal studies dearth of quantitative data*'. I do not understand the English in this sentence. I see in lines 93-94 and section 4 (lines 305-on) that this is a database containing qualitative record, but this sentence does not explain that. Maybe the authors want to say that they add these because there are so few quantitative datasets? What data does this database contain: presence -absence, or such data as 'rare, few, common, abundant'?

36-37: for the eastern equatorial Pacific, I think that the statement ..*'ongoing deoxygenation..induced by coastal eutrophication*' is incorrect or at least incomplete. After all, the eastern equatorial Pacific is the location of the largest open-ocean Oxygen Minimum Zone not linked to coastal eutrophication (see e.g. Breitburg et al., 2018, Declining oxygen in the global ocean and coastal waters. Science 359, eeam7240). Comparison of figure 1 and the

figure in Breitburg et al shows clear overlap of the studied sites with that open ocean OMZ, and the authors refer to deep-water deoxygenation in line 53.

164-165: The authors say: *'see Supplement for full species description*'. Either I am missing something, or the supplement contains a full list of species, but there is no description of species (i.e., description of their morphology). Please change this text to explain what actually is in the Supplement.

181-190: Text and Caption figure 2. why use words such as 'epi**pelagic**', mesopelagic, bathypelagic, etc.? I think that the words 'xxx-pelagic' are defined for planktonic and nektonic organisms, not for benthos. For example, in general (and in oceanographic textbooks) benthic organisms are described as bathyal or hadal, planktic organisms as bathypelagic or hadal pelagic. The cited paper by Costello & Breyer does not use these xxxpelagic terms for benthos, they use them for planktonic organisms, and for benthics they give the not very useful word 'deep-sea' for benthos below 2000 m. Benthic foram users commonly follow van Morkhoven et al., 1986 depth zones:

< 200 m depth: **Neritic**

- coastal: 0–30 m,
- inner neritic: 30–50 m,
- middle neritic: 50–100 m
- outer neritic; 100–200 m

200-2000 m depth: bathyal

- 200-600: upper bathyal
- 600-1000: middle bathyal
- 1000-2000: lower bathyal

>2000 m depth: abyssal

- 2000-3000: upper abyssal
- >3000 m: lower abyssal

194: delete 'being'.

213-217: maybe add a remark that nowadays Rose Bengal staining is not thought to be very reliable as indicator of actually living specimens at the time of collection?? e.g. Bernhard, J., 2006, Comparison of two methods to identify live benthic foraminifera: A test between Rose Bengal and CellTracker Green with implications for stable isotope paleoreconstructions, Paleoceanography 21, PA4210 states: *'On average, less than half the Rose Bengal–stained foraminifera were actually living when collected*'. In my opinion the authors thus should add some information on the concept of 'stained' vs. 'living.

222-225: The definition of studied grain sizes is confusing. The text states: *For example, 62.5% of the samples were analysed in the >61-74 µm size fraction, 6% in the > 88-106 µm size fraction and 10.5% in the > 125-150 µm size fraction*. I think this must mean that the lower boundary of the studied grainsize is in the stated interval, e.g. ' *the >61-74 µm size fraction'* means that the

full range of grainsizes >63 μm was investigated (as listed in Table A1). I think this must be so, because that is what most people do (e.g. Loubere 1994). However, the text appears to read either as if only specimens in between 61 and 74 μm were studied, or as if specimens >71 μm were studied, either of which is incorrect. Similarly, people study either the 125 μm or >150 μm size fraction, but definitely do not limit themselves to specimens between 125 and 150 μm. Then the caption of Figure 5 shows different numbers again, i.e. instead of >63 μm or >61-74 μm the piechart shows (60, 80) (which I think is the same size fraction). Please standardize this to make sure that the reader understands what is listed unequivocally.

234: 'valid taxa' - I assume this includes subspecies, varieties and such? could the authors add the number of species? below-species level groups may be included in the species by other researchers, thus making species diversity lower. Again, note that some of the species in this database are listed as extinct, and might have been misidentified. The authors mention how many of the species in their list are 'rare' according to Murray 2013. Could they say a bit more about the total number of species - how many of the more common ones are overlapping with Murray?

235: please provide reference for these Classes- who defined these?

248: are these all (90) at the species level or including below species level taxa?

251: typo in Nuttallides (spelled with two letters t)

Figure 7: figurer 7D says 'Biodiversity' but plots number of species. the Number of species is NOT how diversity is defined, it is 'species richness'.

256: what is a 'heatmap'?

261: 'made' or 'drawn' rather than 'elaborated'.

286: rounding could lead to both >100 as well as < 100%?

316-on (section 5): here the authors should have mentioned that presently many journals require making data available, and will not publish papers without such data, which must be provided not just in the journal, but in an accepted database such as Pangaea. They are escribing some of the broadly accepted FAIR data practices: Findable, Accessible, Interoperable, and Reusable. Their text here should have pointed this out: in many cases 'authors should not be encouraged' - they will be required to do so by reputable journals. - https://www.nature.com/articles/sdata201618

Table A1: for studies describing 'living + dead', it is possibly to provide the information how many of these species were observed as 'living', how many as 'dead' (with overlap, some species could be present both living and dead)? it would be good information to know whether species were present both as living and as dead.

Table B1:
I think there are mistakes in N200, N300,which are both described as ' It indicates whether sample counts are equal to or higher than 100 individuals'. I think this should be 200 and 300, respectively.

Note to supplement:
There are two entries for *Oridorsalis tener* (681, 682). meant or mistake?