

Interactive comment on “Data compilation on the biological response to ocean acidification: an update” by Y. Yang et al.

Y. Yang et al.

gattuso@obs-vlfr.fr

Received and published: 6 January 2016

We thank Stéphane Pesant (referee #2) for his constructive comments and suggestions which helped considerably improve the data set associated with this manuscript. Our replies, in italics, are below.

- Overall quality of the discussion paper

This is an important data compilation. The authors clearly explain the interest in this data compilation and provide convincing evidence that the first version has been used in the past and that an updated publication is well justified. The data compilation has grown from 81 data sets to 581 data sets since it was first described by Nisumaa et al. (2010).

C469

The general methods of the data compilation were previously described in a first publication (Nisumaa et al. 2010), and additional methods are clearly explained in the Data Section.

Generally, however, I was surprised that data set [doi:10.1594/PANGAEA.149999](https://doi.org/10.1594/PANGAEA.149999) includes only a reference list of source papers and their related data sets archived at PANGAEA. I would expect this data set to include contextual data that help navigate through the data collection (581 individual data sets) and facilitate data re-use. I provide here several specific comments below to improve this shortcoming. Improvements of the data set merit additional discussion in the Data Section.

The supplement is useful and deserves a citation of its own.

Technical corrections will be provided in a separate comment.

We appreciate the positive comments on the OA-ICC data compilation. We agree that it is useful to add contextual data in the data set [doi:10.1594/PANGAEA.149999](https://doi.org/10.1594/PANGAEA.149999), which will be revised as suggested. Concerning the supplement, we were actually invited by a scientific journal to turn it into a paper but we declined because we do not think that it can stand on its own. Adding it as a supplement to this manuscript submitted to Earth System Science Data is the option we chose to make it publicly available.

- Specific comments for the authors about the data set:

I am not convinced of the usefulness of breaking down the citations of the source papers into six parameters in [doi:10.1594/PANGAEA.149999](https://doi.org/10.1594/PANGAEA.149999), i.e. Author, Title, Persistent Identifier, Journal/report title, Journal volume, and Page(s). Instead, I recommend including all source papers in the “Related to:” reference list of the data set (see my “Editorial comments for PANGAEA”). A single parameter with the persistent ID (DOI) of the source paper is enough to properly cite the data

source in the data set. Nevertheless, for text mining purposes, it would indeed be useful to also include the title of the source paper as a second parameter.

The authors explain in the Data Section (S2;L22-26) that some of the metadata were extracted directly from the original papers in which data were obtained, i.e. taxonomy, processes measured, experimental manipulations, and country of first author. These are indeed interesting metadata that will guide readers and facilitate data re-use. However, these metadata are currently anonymised in the discussion paper by showing only results in the form of figures (Fig. 3-6). I strongly recommend that these metadata be included in doi:10.1594/PANGAEA.149999. This was done to some extent in the first data compilation published in 2010. I see several advantages to doing this:

1. Making data behind each figure available (here figures 3-6);
2. Consistently describing “data sets” in the text would ease the reading of Section 2, instead of describing sometimes papers and sometimes data sets;
3. Providing additional contextual metadata in doi:10.1594/PANGAEA.149999 would guide readers in sorting through the list of data sets and identifying which ones they wish to re-use for their analyses;
4. Allowing users to programmatically query doi:10.1594/PANGAEA.149999 using for example <https://ws.pangaea.de/dds-fdp/>

Similarly to the previous comment, I feel it is essential to include in doi:10.1594/PANGAEA.149999 the contextual data presented in Figures 2, 7 & 8 so that readers can sort through the list of data sets and identify which ones they wish to re-use for their analyses. These parameters are the latitude and longitude of the sampling/experimental sites (fig. 2), variables/scale of the carbonate system that are reported (fig. 7) and the environmental context of the sampling site or experimental setup (fig. 8). (see my detailed recommendations about doi:10.1594/PANGAEA.149999). Additionally, parameters such as date/time of

C471

sampling/experiment should be included in doi:10.1594/PANGAEA.149999.

I also recommend including metadata about the nb of observations in each dataset (this can be provided by PANGAEA), geographic location names such as Marine biome (Longhurst 2007), Ocean and sea region (IHO General Sea Areas 1953) and Biogeographical province (Longhurst 2007). The geographic names can be obtained for each location at www.marineregions.org.

Finally, following the “Guidelines for reporting ocean acidification data in scientific journals” provided by the authors as supplement, it is recommended to include the following information in doi:10.1594/PANGAEA.149999: hydrostatic pressure (if the measurements were not performed at atmospheric pressure), total dissolved inorganic phosphorus whenever possible, total dissolved inorganic silicon (in $\mu\text{mol kg}^{-1}$) whenever possible, the carbonate system computation flag, and the nb. of replicate measurements to address reproducibility.

Following the previous comments, I recommend reorganising and augmenting the current data set into two complementary data sets:

Ocean Acidification, International Coordination Centre: Data compilation on the biological response to ocean acidification: environmental and experimental context of data sets and related literature.

doi:10.1594/PANGAEA.149999, 2015.

1. persistent identifier of the data set
2. persistent identifier of the source paper
3. title of the source paper
4. country of affiliation of the first author
5. # of observations
6. marine biome (Longhurst 2007)

C472

7. ocean and sea region (IHO General Sea Areas 1953)
8. biogeographical province (Longhurst 2007)
9. latitude
10. longitude
11. date/time
12. taxonomic groups targeted by the data
13. biological processes targeted in the data
14. experimental treatment(s)

Ocean Acidification, International Coordination Centre: Data compilation on the biological response to ocean acidification: measurements and calculations of the carbonate system.

doi:10.1594/PANGAEA.xxxxxx, 2015.

1. persistent identifier of the data set
2. variables of the seawater carbonate system reported in the data set
3. temperature
4. salinity
5. hydrostatic pressure (if the measurements were not performed at atmospheric pressure)
6. total dissolved inorganic phosphorus whenever possible
7. total dissolved inorganic silicon (in $\mu\text{mol kg}^{-1}$) whenever possible
8. carbonate system computation flag
9. reproducibility (nb of replicate measurements)
10. pH calculated using seacarb after Nisumaa et al. (2010)

C473

11. pH scale
12. carbon dioxide calculated using seacarb after Nisumaa et al. (2010)
13. fugacity of carbon dioxide (water) at sea surface temperature (wet air)
14. fCO₂water_SST_wet calculated using seacarb after Nisumaa et al. (2010)
15. bicarbonate ion calculated using seacarb after Nisumaa et al. (2010)
16. carbon, inorganic, dissolved calculated using seacarb after Nisumaa et al. (2010)
17. aragonite saturation state calculated using seacarb after Nisumaa et al. (2010)
18. calcite saturation state calculated using seacarb after Nisumaa et al. (2010)
19. carbonate ion calculated using seacarb after Nisumaa et al. (2010)

In the first paragraph of Section 2, it is indicated that in some cases, data were downloaded from other repositories such as BODC, BCO-DMO and AADC. Please make sure that in these cases the new data sets archived at PANGAEA includes a reference or an URI pointing to the original data. These references or URI should appear as "other version" in the reference list.

Excellent suggestion! We will include the contextual metadata in doi:10.1594/PANGAEA.149999. In page 2 line 51-53, S. Pesant suggests "Consistently describing 'data sets' in the text would ease the reading of Section 2, instead of describing sometimes papers and sometimes data set". But in section 2.1-2.5, we cannot describe the statistics of the geographical region, type of organism and biological process studied, other factors manipulated, as well as the country of affiliation of the first author by "data sets", because these information were extracted from papers which sometimes led to several data sets. For example, data from the same paper which studied a coral as well as a mollusk are archived in data sets doi:10.1594/PANGAEA.777720 and

C474

doi:10.1594/PANGAEA.777719. Data set doi:10.1594/PANGAEA.777720 only archived the data of coral, while the other one only included the data of mollusk. When we do statistical analysis, we do not check data sets one by one, we only know there are two data sets supplement to the same publication. Hence, if we describe by "data sets", we will say there are two data sets archived data on coral and mollusk, but actually it should be one on coral and one on mollusk.

We will also include the data of Figure 2, 7 & 8 in

doi:10.1594/PANGAEA.149999. But we are unable to include the marine biome (Longhurst 2007), Ocean and sea region (IHO General Sea Areas 1953), and Biogeographical province (Longhurst 2007), because this information cannot be extracted from the bibliographic database. Also, the number of replicate measurements cannot be added because it has not been archived in the compilation before. We will try to add the other parameters suggested in doi:10.1594/PANGAEA.149999 but need to check whether it is possible with the support staff at Pangaea.

We will set doi:10.1594/PANGAEA.149999 as parent data set to which three children data sets supplement for: (1) the metadata (figures 3-6), (2) date/time, latitude and longitude (figure 2) and (3) for the original (measured) and calculated parameters of the carbonate system (figures 7 and 8).

In all the data sets which archived data already archived in other repositories, a reference or a link to the original data has been added under "other version".

- Editorial Comments for PANGAEA:

1. Please correct "Accidifiacion" to "Acidification" in the data set title.
2. Please change the author of the data set in the editorial system of PANGAEA to last name="Ocean Acidification" and first name="International Coordination Centre", so to better conform to conventions about naming authors. This will au-

C475

tomatically appear in citations as "Ocean Acidification, I.C.C." or "Ocean Acidification, International Coordination Centre".

3. Please ask the authors for an abstract.
4. Please include all source papers in the "Related to:" reference list; Please ask the IT team to implement expandable/retractable lists of references when there are more than three references, as with Events.
5. Please include all 581 data sets in the "Source data set:" reference list; Please ask the IT to implement expandable/retractable lists of references when there are more than three references, as with Events.
6. Please change doi:10.5194/essdd-8-889-2015 from "Related to:" to "Supplement to:"
7. Please include doi:10.1594/PANGAEA.735138 in the "Other version:" reference list
8. Please include doi:10.5194/essd-2-167-2010 as first reference in the "Related to:" reference list

Excellent suggestions! We will verify the data set following all these suggestions and will change Ocean Acidification International Coordination Centre" to "Ocean Acidification, I.C.C." or "Ocean Acidification, International Coordination Centre" in the citation of this data set.

- Editorial comments for ESSD:

It would be useful to provide quick references to "data presented in the discussion paper", and to "additional data resources". I recommend inserting these quick references between the Abstract and the Introduction Sections, in the way suggested here.

C476

Access to data presented in this paper:

Summary of data sets and related literature: doi:10.1594/PANGAEA.149999

Experimental and environmental context of data sets:

doi:10.1594/PANGAEA.xxxxxx

Access to additional data resources:

Growing collection of data sets: <http://www.pangaea.de/search?q=Project:OA-ICC>

Ocean Acidification (OA-ICC) bibliography: <https://www.mendeley.com/groups/4333941/ocean-acidification-oa-icc/papers/>

I also recommend including in the Reference Section the full citation of the two data sets presented in the discussion paper, in the way suggested below. These citations could be included among the paper citations or ideally could appear in a sub-section dedicated to data citations.

Ocean Acidification, International Coordination Centre: Data compilation on the biological response to ocean acidification: environmental and experimental context of data sets and related literature. doi:10.1594/PANGAEA.149999, 2015.

Ocean Acidification, International Coordination Centre: Data compilation on the biological response to ocean acidification: measurements and calculations of the carbonate system. doi:10.1594/PANGAEA.xxxxxx, 2015.

We will add following information in the abstract and the introduction. Data presented in this paper can be accessed at doi:10.1594/PANGAEA.149999.

Access to additional data resources: (1) growing collection of data sets: <http://www.pangaea.de/search?q=Project:OA-ICC> (2) Ocean Acidification (OA-ICC) bibliographic database:

<https://www.mendeley.com/groups/4333941/ocean-acidification-oa-icc/papers/>

Interactive comment on Earth Syst. Sci. Data Discuss., 8, 889, 2015.

C477