***Description of the headers of the ZooBase dataset (Benedetti et al., 2021, accepted in Nature Communications).***

1. "**occurrenceID**" : Unique occurrence identifier used to identify and remove any duplicated occurrence. Based on Latitude\_0.1 + Longitude\_0.1 + Month + Year + Depth\_10m + WoRMS\_ID.
2. "**Latitude\_0.1**": decimaLatitude rounded to the closest 0.1°. Created to define the occurrenceID.
3. "**Longitude\_0.1**" : decimaLongitude rounded to the closest 0.1°. Created to define the occurrenceID.
4. "**Depth\_10m**" : Depth rounded to the closest 10m. Created to define the occurrenceID.
5. "**eventDate**" : Date at which the species occurrence was observed. OBIS recommends using the extended ISO 8601 format with hyphens (e.g. ‘YYYY-MM-DD’).
6. "**Year**" : Year at which the species occurrence was observed. Used to define the occurrenceID.
7. "**Month**" : Month at which the species occurrence was observed. Used to define the occurrenceID.
8. "**Day**" : Day at which the species occurrence was observed.
9. "**decimalLatitude**" : Geographic Latitude in decimal degree, following the -180/+180 WGS84 Spatial Reference System (SRS).
10. "**decimalLongitude**" : Geographic Longitude in decimal degree, following the -180/+180 WGS84 Spatial Reference System (SRS).
11. "**Depth**" : Depth at which the occurrence was recorded (in meters below the local sea surface); > 0; = 0 is surface.
12. "**MinDepth**" : Minimum depth for depth-integrated quantities, in meters, > 0. E.g., the shallowest sampling depth of plankton net tows.
13. "**MaxDepth**" : Maximum depth for depth-integrated quantities, in meters, > 0. E.g., the deepest sampling depth of plankton net tows.
14. "**ParentEventID**" : Describes the parent event, which is composed of one or more sub-sampling (child) events (eventID below). (e.g. AMT29, Tara Oceans st.75, CPR tow XX etc.).
15. "**EventID**" : Labels the replicate samples (or sub-samples) from a ParenEventID (e.g., a sample number from a station if multiple samples were taken at same sampling station).
16. "**InstitutionCode**" : Custodian institution for the data record. Inherited from GBIF and/or OBIS.
17. "**SourceArchive**" : Describes the online archive where the data is stored (e.g., PANGAEA, OBIS, GBIF, DRYAD etc.).
18. "**OrigCollectionCode**" : Code given to the collection or the dataset within the SourceArchive. Inherited from GBIF and/or OBIS.
19. "**OrigCollectionID**" : occurrenceID given to the record/measurement in the SourceArchive. Inherited from GBIF and/or OBIS.
20. "**BiblioCitation**" : String indicating the bibliographic citation associated with the data (when possible; can be a dataset, a paper, a report).
21. "**CitationDOI**" : DOI of the associated bibliographic citation.
22. "**DateDataAccess**" : Date at which the data was downloaded from the SourceArchive.
23. "**DatasetKey**" : DatasetKey inherited from GBIF or OBIS. Can be used to identify the original dataset that was submitted to the corresponding SourceArchive.
24. "**OrigScientificName**" : Scientific name of the taxon recorded, as reported in the source dataset.
25. "**ScientificName**" : Manually corrected scientific name of the taxon. Used to retrieve the the WoRMS\_ID and WoRMS\_status from WoRMS using the ‘*worms’* R package (Holstein, 2018; <https://cran.r-project.org/web/packages/worms/worms.pdf>).
26. "**WoRMS**\_**ID**" : Species-level or taxon-level AphiaID code retrieved from WoRMS based on the ScientificName using the ‘*worms’* R package. Used to define the occurrenceID.
27. "**WoRMS**\_**status**" : Species-level or taxon-level status (marine, extant, fossil, freshwater etc.) retrieved from WoRMS based on the ScientificName using the ‘*worms’* R package.
28. "**TaxonRank**" : Taxonomic rank at which the species/taxon was identified and recorded (e.g. Species, Family, Order, etc.).
29. "**Kingdom**" : Kingdom the species/taxon recorded belongs to.
30. "**Phylum**" : Phylum the species/taxon recorded belongs to.
31. "**Class**" : Class the species/taxon recorded belongs to.
32. "**Order**" : Order the species/taxon recorded belongs to.
33. "**Family**" : Family the species/taxon recorded belongs to.
34. "**Genus**" : Genus the species/taxon recorded belongs to.
35. "**Species**" : Accepted species name recorded in WoRMS.
36. "**basisOfRecord**" : Nature of the data record (e.g. 'preserved specimen', 'fossil specimen', 'living specimen' etc.).
37. "**SamplingProtocol**" : String indicating the sampling protocol used to observe the species/taxon.
38. "**DeterminedBy**" : Name(s) of the people, groups or organizations who made the measurement, or identified the organism.
39. "**DeterminedDate**" : The date at which the occurrence was identified (can differ from the eventDate).
40. "**Note**" : Any note or comment on the measurement event.