

# European Marine Omics Biodiversity Network (EMO BON)

*January 2025*

## *Updating EMO BON metadata spreadsheets*

In EMO BON, we are actively and continuously working on improving and standardizing the data management system, while at the same time ensuring adherence to the FAIR data principles. As often discussed, data management begins even before the sample collection. We have paid attention to the proper metadata collection from the beginning of EMO BON, and we have created the online spreadsheets to record all the metadata related to the samples.

From the beginning of EMO BON until now, the spreadsheets have changed and improved, and they now represent the core stepping stones for the whole EMO BON data management. You may have already noticed some changes in the EMO BON metadata spreadsheets. With this note we wish to summarize the changes and give you more details. As requested, you may also find a list of some commonly made errors. Please note that these changes are not yet applied in all online spreadsheets, but we are looking into having all spreadsheets fully updated in the coming months.

We believe that EMO BON metadata collection is among the most thorough and well-structured in the whole marine biology research landscape. It provides the basis not only for advancements in the genomic data management, but also for observation-based scientific questions.

We are thankful to all of you for keeping up with us during this procedure, for making the effort to fill the sometimes complex fields, for wanting to learn how to do this and for actively helping us improve the spreadsheets. We are grateful to the people who worked on this, put a great effort to improve the EMO BON metadata spreadsheets and keep working on our data management procedures: Christina Pavloudi (EMBRC-ERIC), Katrina Exter (VLIZ), Marc Portier (VLIZ), Cymon J. Cox (CCMAR), Bruno Louro (CCMAR), Stelios Ninidakis (HCMR), Melina Loulakaki (HCMR), Maria Luisa Chiusano (UNINA), Andrzej Tkacz (CCMAR), Salva Fernández Bejarano (VLIZ), Laurian Van Maldeghem (VLIZ), Melanthia Stavroulaki (HCMR).

As always, we are happy to hear from you (questions, issues, discussions), so feel free to contact us!



## Summary of updates on the EMO BOD metadata spreadsheets

1. The spreadsheets were migrated from the EMBRC secretariat account to an account dedicated to EMO BOD. Editing rights have been given to all relevant people. In case you wish to add more people from your observatories as editors, please contact us.
2. During the update process, some redundant sheets were created (Updated definitions, old-sampling). This was to ensure the passing of all information to the updated spreadsheets. Only necessary sheets will remain: READ\_THIS\_FIRST, observatory, sampling, measured, definitions.
3. We changed the entry format in the replicate column of the sampling sheet for the blank samples. An underscore was added and the new format is "blank\_1", "blank\_2" etc.
4. There is a new term to record environmental events taking place during sampling and may have affected the samples. We can take this information into account when analysing the data. If you want to add more than one value, use ; as the separator. The values should be from [ENVO](#) and some examples are:
  - o marine algal bloom [ENVO:02500036]
  - o coral bleaching process [ENVO:01000852]
  - o glacial abrasion [ENVO:01000682]
  - o storm [ENVO:01000876]
  - o drought [ENVO:1000745]
  - o rain [ENVO:01001564]
  - o snow [ENVO:01000406]
  - o hail [ENVO:03400011]
  - o oil spill [ENVO:00002061]
  - o increased temperature [PATO:0001305]
  - o environmental pollution [ENVO:02500036]
  - o biofouling [ENVO:06105023]

Spreadsheet: water, sediment, ARMS

Sheet name: sampling

### Term (column) name: noteworthy\_env\_cond

Definition: If there were any noteworthy/special environmental conditions during the sampling event, please note down the most appropriate term from the Environmental Ontology. Some suggestions for terms are: marine algal bloom [ENVO:02500036], coral bleaching process [ENVO:01000852], glacial abrasion [ENVO:01000682], storm [ENVO:01000876], drought [ENVO:1000745], rain [ENVO:01001564], snow [ENVO:01000406], hail [ENVO:03400011], oil spill [ENVO:00002061], increased temperature [PATO:0001305], environmental pollution [ENVO:02500036], biofouling [ENVO:06105023]. Additional appropriate terms may be selected from here: [https://ontobee.org/ontology/ENVO?iri=http://purl.obolibrary.org/obo/ENVO\\_01000203](https://ontobee.org/ontology/ENVO?iri=http://purl.obolibrary.org/obo/ENVO_01000203)

*Who is adding this information: sampling operators*

5. There are a few new terms related to data ownership.

Spreadsheet: water, sediment, ARMS



Sheet name: observatory

**Term (column) name: data\_owner**

Definition: the organization responsible for the curation, integrity, publication and use of generated data

*Who is adding this information: EMBRC*

**Term (column) name: data\_owner\_edmoid**

Definition: EDMO id of the data owner organization

*Who is adding this information: EMBRC*

**Term (column) name: data\_contact\_email**

Definition: data owner organization email

*Who is adding this information: EMBRC*

6. There is a new term to indicate in which batch each sample belongs to. This term is useful for managing the samples and instantly knowing where to find information related to the next steps

Spreadsheet: water, sediment, ARMS

Sheet name: sampling

**Term (column) name: seq\_batch**

Definition: batch of EMO BON samples processed together for DNA extraction, sequencing, and data acquisition. Example: "batch 1", "batch 2"

*Who is adding this information: EMBRC*

7. Terms related to the ENA submission of the data have been added.

Spreadsheet: water, sediment, ARMS

Sheet name: observatory

**Term (column) name: ENA\_accession\_number\_umbrella**

Definition: Umbrella study accession number is a unique identifier given to an umbrella study once the sample is registered to ENA (European Nucleotide Archive). It refers to a specific umbrella study and connected to all the projects, samples, sequencing runs and data included.

*Who is adding this information: EMBRC*

**Term (column) name: ENA\_accession\_number\_project**

Definition: Project accession number is a unique identifier given to a project once the project is registered to ENA (European Nucleotide Archive). It refers to a specific project and connected to all the samples, sequencing runs and data included.

*Who is adding this information: EMBRC*

Spreadsheet: water, sediment, ARMS

Sheet name: sampling

**Term (column) name: ENA\_accession\_number\_sample**

Definition: Sample accession number is a unique identifier given to a project once the sample is registered to ENA (European Nucleotide Archive). It refers to a specific sample and connected to all the sequencing runs and data included.

*Who is adding this information: EMBRC*

8. We found that the priorly used source\_material\_id was not easy to handle in later data management steps since there were cases of non-unique ids. This is why we simplified the source\_material\_id to include ascending numbers starting with the first sample collected (or added in the spreadsheet) in an observatory. The numbering is separate for each observatory. The new format is EMOBON\_[observatory name]\_ SampleType[Wa/So/Ha]\_iterator (1,2,3 etc), for example EMOBON\_BPNS\_Wa\_1. This is created via a google equation. There is no need for the sampling operators to fill in this column, as it will be filled by EMBRC. The intention is to lock this column and only allow EMO BON management team to edit it because this ID, once created, must NEVER be changed (so no moving the rows around!). The common sample name that is easily readable will remain in the spreadsheets under term source\_mat\_id\_orig and this is the term sampling operators should be filling in. The source\_mat\_id\_orig will always be the first column in each sheet to easily indicate which sample each row includes. Therefore, please add the source\_mat\_id\_orig for every sample in the sampling sheet according to the EMO BON Handbook and as written on your sample containers.

Spreadsheet: water, sediment, ARMS

Sheet name: sampling, measured

**Term (column) name: source\_material\_id**

Definition: Unique identifier to each row/sample: it consists of the terms 1) Project ("EMOBON") 2) Sampling Site ID; that is the Observatory ID (for example "SMN99") supplemented with the sampling site indicator ("Wa" for water column, "So" for soft substrates and "Ha" for hard substrates) and 3) an iterator (1,2,3 etc). This is created via a google equation. To add this identifier to a new row, simply drag the value from the cell above. This column can only be filled by EMBRC.

*Who is adding this information: EMBRC*

**Term (column) name: source\_material\_id\_orig**

Definition: A unique identifier assigned to a material sample according to EMO BON Handbook as added by the sampling operators and written on the sample containers. This identifier is not quality controlled but remains unchanged. This identifier consists of 5 terms: 1) Project (“EMOBON”) 2) Sampling Site ID; that is the Observatory ID (for example “SMN99”) supplemented with the sampling site indicator (“Wa” for water column, “So” for soft substrates and “Ha” for hard substrates) 3) Sampling Campaign Date formatted as YYMMDD. For example “220315” would be the campaign on 15 March 2022. Especially for Hard substrates, both the deployment and retrieval dates have to be recorded formatted as YYMMDD-YYMMDD 4) Size fraction (Wa: “3um” / “0.2um” / “20um” / “200um”) or organisms collected (So: “micro” for microorganisms, “meio” for meiobenthos and “macro” for macrobenthos) or ARMS fraction (Ha: “SF” for sessile fraction, “MF500” for motile fraction sieved through 0.5 mm, “MF100” for motile fraction sieved through 0.1 mm) 5) Replicate number (1-4). For labelling the negative control, this term will be replaced by the notation “blank”. All terms must be separated by “\_”.

*Who is adding this information: sampling operators*

9. The definitions of size\_frac\_low and size\_frac\_up were swapped so that the first is the lower filter value and the second is the upper filter value. This used to be a problem because the MlxS terms checklist had them reversed and we wanted to follow the checklist. We raised this with the Genomic Standards Consortium and the definitions will be corrected in the new MlxS version. This is already corrected in the water spreadsheets, and you don’t have to do anything for the samples already entered. Definitions are now:
  - a. size\_frac\_low: size-fraction lower threshold refers to the mesh/pore size used to retain the sample. Materials smaller than the size threshold are excluded from the sample
  - b. size\_frac\_up: size-fraction upper threshold refers to the mesh/pore size used to pre-filter/pre-sort the sample. Materials larger than the size threshold are excluded from the sample
10. All terms that are flagged as mandatory in sheet measured are coloured yellow.

**Frequently made errors**

1. Not double-checking the values added. Please always *double-check* after having filled in values. Please pay attention to not:
  - drag-drop values incorrectly (did you really want all numbers to increase by 1 for each row?)
  - copy-pasting from another row but forgetting to change e.g. the date
2. Make sure all mandatory cells are filled. What is required is
  - either write the actual value
  - or write “NA” if that cell will never have a value (e.g. a sensor broke, and no measurement was taken)
  - or write “expected later” or “expected MM-YY” if the value will come later
3. Using > or < in the size\_frac\_low and size\_frac\_up columns: the two values between them indicate a min to max range, so there is no need to use the upper or lower limits symbols.
4. Make sure you are using the indicated units for each measurement. The units are given in the definitions sheet, where terms are ordered alphabetically.



5. The following columns were always built to allow list entries, but please pay attention to how the lists are entered, see the example below

**Columns including list entries:**

biomass and biomass\_method  
chem\_administration  
n\_alkanes and n\_alkanes\_method  
organism\_count and organism\_count\_method  
particle\_class  
phaeopigments and phaeopigments\_method  
pigments and pigments\_method

**Example:**

*pigments*

B-caroten: 0.04;α-caroten 0.005

*pigments\_method*

HPLC (Alliance HPLC System) following the method described by Zapata et al. 2000; HPLC (Alliance HPLC System) following the method described by Zapata et al. 2000

*\*Please add as many methods as there are pigments. Even if the method is the same, please add it again)*

For the EMO BON management team,

**Ioulia Santi, PhD**

**Observation, Data and Service  
Development Officer**

**European Marine Biological  
Resource Centre (EMBRC-ERIC)**

Paris Headquarters

4 Place Jussieu,

Tour 46/00, 1er étage, bureau 101

75252 Paris Cedex 05 (FR)

[www.embrc.eu](http://www.embrc.eu)

Email: [ioulia.santi@embrc.eu](mailto:ioulia.santi@embrc.eu)

Phone: +30 2 81 03 37 727

Mobile: +30 6 94 51 78 545

-----  
[embrc.eu](http://embrc.eu) / [X](#) / [LinkedIn](#) / [YouTube](#)



**EMBRC**  
EUROPEAN  
MARINE  
BIOLOGICAL  
RESOURCE  
CENTRE