EMODnet Chemistry

Guidelines for QC of contaminants

M. Lipizer, M.E. Molina Jack and A. Giorgetti (OGS), Luminita Buga (NIMRD), Lotta Fyberg and Karin Wesslander (SMHI), Ann Kristin (IMR), Sissy Athanasia Iona (HCMR), Martin Larsen (AU), Reiner Schlitzer (AWI)

Date: 12/05/2020

Content:

1. Introduction

EMODnet Chemistry aims to provide access to marine chemistry data sets and derived data products concerning eutrophication, ocean acidification, contaminants and marine litter. The chemicals chosen are relevant for the Marine Strategy Framework Directive (MSFD). Parameter names are based on P01, BODC Parameter Usage Vocabulary, which is available at: http://seadatanet.maris2.nl/bandit/browse_step.php. Each measurement value has a quality flag indicator, based on the standard SeaDataNet scheme (L20 SEADATANET MEASURAND QUALIFIER FLAGS[1]). This document describes the steps of data and metadata validation needed to achieve the standardised, harmonised and validated datasets concerning **contaminants in the marine environment**, in water, sediment and biota matrices. EMODnet Chemistry is focused on the following list of parameter groups:

P36 Code	P36 Parameter groups	P02 Code	P02 Parameter discovery vocabulary
ANTIFL	Antifoulants	OMBI	Organometallic species concentration parameters in biota
		OMPW	Organometallic and organometalloid species concentration parameters in sediments
		OMWC	Organometallic and organometalloid species concentration parameters in water bodies
HCARBS	Hydrocarbons	BCAH	Concentration of polycyclic aromatic hydrocarbons (PAHs) in biota
		BCOC	Concentration of other organic contaminants in biota
		OHWC	Concentration of other hydrocarbons in the water column
		PCAH	Concentration of polycyclic aromatic hydrocarbons (PAHs) in suspended particulate material
		PCHW	Concentration of polycyclic aromatic hydrocarbons (PAHs) in the water column
		PCOC	Concentration of other organic contaminants in suspended particulate material
		SALK	Concentration of aliphatic hydrocarbons in sediment samples

---

[1]

http://seadatanet.maris2.nl/v_bodc_vocab_v2/browse.asp?order=conceptid&formname=search&screen=0&lib=l20&v0_0=&v1_0=conceptid%2Cpreflabel%2Caltlabel%2Cdefinition%2Cmodified&v2_0=0&v0_1=&v1_1=conceptid&v2_1=3&v0_2=&v1_2=preflabel&v2_2=3&v0_3=&v1_3=altlabel&v2_3=3&v0_4=&v1_4=modified&v2_4=9&v0_5=&v1_5=modified&v2_5=10&x=19&y=22&v1_6=&v2_6=&v1_7=&v2_7=

		SCAH	Concentration of polycyclic aromatic hydrocarbons (PAHs) in sediment samples
		SCOC	Concentration of other organic contaminants in sediment samples
		WCOC	Concentration of other organic contaminants in the water column
HVYMTL	Heavy metals	BCMT	Metal concentrations in biota
		MDBO	Trace metalloid concentrations in biota
		MTPW	Metal concentrations in sediment pore waters
		MTWC	Colloidal metal concentrations in the water column
		MTWD	Dissolved metal concentrations in the water column
		MTWP	Particulate metal concentrations in the water column
		MTWT	Total metal concentrations in water bodies
		RMIN	Inorganic chemical composition of sediment or rocks
PCBSXX	Polychlorinated biphenyls	BCPB	Concentration of polychlorobiphenyls (PCBs) in biota
		PPCB	Concentration of polychlorobiphenyls (PCBs) in suspended particulate material
		PPWC	Concentration of polychlorobiphenyls (PCBs) in the water column
		SPCB	Concentration of polychlorobiphenyls (PCBs) in sediment samples
PESTB	Pesticides and biocides	PEBI	Pesticide concentrations in biota
		PESD	Pesticide concentrations in sediment
		PEWB	Pesticide concentrations in water bodies

2. Validation steps

A step-wise approach is needed to obtain validated regional aggregated datasets, due to the large heterogeneity in matrices (seawater total, filtered, unfiltered samples; different size classes of sediment; different biota taxa, target organs, size,..), in parameters measured by the different laboratories all over Europe, in analytical protocols used, in measurement units as well as in monitoring frequency and purposes. The multiple combinations of the above mentioned variables results in a huge list of parameters included in the datasets, often with limited temporal – spatial distribution for single parameters. Additionally, there is not enough knowledge on concentration ranges for the specific variables in the different areas. As a consequence, the applied validation procedure involves[2]:

☐ harmonization of unit and parameter naming

☐ metadata completeness and dataset format control

☐ check of quality flagging of data and metadata

☐ checks for clearly impossible data ranges (ex. Different orders of magnitude)

☐ checks for inconsistent measurement units

The results of validation are reported to data originators to evaluate, revise and eventually correct encountered issues. Part of the validation process can be carried out using ODV standard interface, while part requires the use of the harmonized, transposed and decomposed matrix and the handling outside ODV. The use of ODV transposed and decomposed matrix enables to filter per substance, matrix characteristics, analytical method, …

---

[2] These Guidelines derive in part to activities carried out in the framework of the project HarmoNIA (Harmonization and Networking for contaminant assessment in the Ionian and Adriatic Seas) financed by Interreg ADRION (2018-2020).

3. Harmonization of unit and parameter naming

In order to harmonize measurement units, a bibliographic analysis of most widely used measurement units for classes of contaminants in the different matrices has been performed. EU directives (2013/39/UE; Comm. Dec. EU 2017/848) have been taken into account to define short list of «preferred units» and to establish, when possible, conversion factors.

Ocean Data View (ODV) allows to harmonize (i.e. performs unit conversion, merges the same P01 expressed by different local names and keeps track of "harmonized" variables) and to allow easier retrieval of variable names from P01 standard codes.

All aggregated datasets are prepared and validated using ODV according to the following steps:

- ❖ File format checks during import in ODV

- ❖ Check of consistency of primary variables, especially in sediment and biota datasets

- ❖ Separation of data related to different matrices into different dataset collections

- ❖ Measurement units and parameter harmonization with ODV harmonization tool

- ❖ Inspection of not-harmonized variables to verify correctness of measurement units as well as correctness of P01 terms used

- ❖ Check for consistency between measured variables (P01) and measurement units

- ❖ Visual checks of concentration ranges to identify clearly impossible concentrations

4. Metadata completeness and dataset format control

Metadata availability:

In order to be correctly used, data of contaminants need to be accompanied by relevant metadata associated to the datasets:

➤ For water: Depth (if time series),

➤ For sediment: Depth below seabed (if time series); additionally, proportion of sizes of particles, Parameters related to granularity, water content, organic matter content, thickness of sediment properties and sedimentation rates. Due to heterogeneity in grain size, missing information on grain size, as well as lack of indication of station depth and of sample thickness strongly affect QC of contaminants in the sediment matrix. Organic carbon and grain size are relevant supplementary data required for QC and for the application of normalization procedures.

➤ For biota: parameters related to the sample (biota sizes, sample ids, sex, life stage…), water and lipid contents, wet weight, depth…

Furthermore, temporal and spatial (position, sampling depth, station bottom depth) information must be available otherwise data cannot be used.

Dataset format check:

EMODnet Chemistry regional aggregated datasets are available as "extended-ODV" spreadsheet file, that contains data plus almost all metadata associated. Up to seven types of aggregated datasets can be expected from each region:

• Time-series for each matrix (water, sediment and biota)
• Profiles for each matrix (water, sediment and biota)
• Biota datasets with the sample identification as the primary data variable

During the import phase (CDI+ODV files) with ODV software, different types of collections are generated:

• Ocean depth profiles: Primary variable will be depth (DEPHPR01 or ADEPZZ01), so the collection may be primarily for water. It is necessary to be careful because if there are profiles of biota, this collection would contain them. If there are parameters related to biota

in the collection, the stations that contain biota can be filtered with the station filter described below in step 2. Additionally, if there are parameters related to sediment, this would be an indication of profiles related to sediment with the wrong primary variable attribution; in that case, stations have to be filtered, primary variable changed to COREDIST and merge with the collection related to sediment profiles.

- Sediment depth profiles: Primary variable will be depth below seabed (COREDIST), so the collection would be related to sediment.
- Time Series: Primary variable will be time. This collection may contain parameters related to water, sediment and biota.
- Sample-ID primary variable collection: If data originators have used sample -ID as the primary variable for biota datasets, an additional collection would be created.

It might happen that during the import phase more than one collection of sediment or water is created, due to a different primary variables and or units. Examples:

- Sediment profiles, with COREDIST in as primary variables but given in different units (metres or centimetres). In this case, it is necessary to harmonise units for depth, the preferred unit is metres.
- Ocean profiles, with different primary variables, DEPHPR01 and ADEPZZ01, that can be considered equivalent, so DEPHPR01 can be edited into ADEPZZ01 for the whole collection.

There is a functionality inside ODV to aggregate them performing previously the transformation needed. It is necessary to add the data variables from the second collection to the main collection. This can be done through the properties of the collection→ data variables → new→ add from another collection. Then you start with the import dialog in the option "add/replace station data". For water, depth variables can be associated without any transformation (as they are in the same units and considered equivalent parameters). For sediment, it is necessary to convert the primary variable during the import to have the same units in the three collections.

For timeseries, parameters related to the three matrices must be splitted, using the filter for the metadata field "variables measured" (using the station filter described in step 2). Export stations related to biota, sediment and water. Open each of the new created collections, delete the empty variables and export the collection using the option to obtain a SDN harmonized ODV collection. Some parameters of the collection may belong to other matrix if the station samples multiple matrices.

Once the 6 or 7 collections are ready, the next step is to harmonise the collection in terms of units and P01 parameters. To carry out this step, ODV software contains a functionality that aggregates the same P01s present in the whole collection and harmonizes the units when possible:
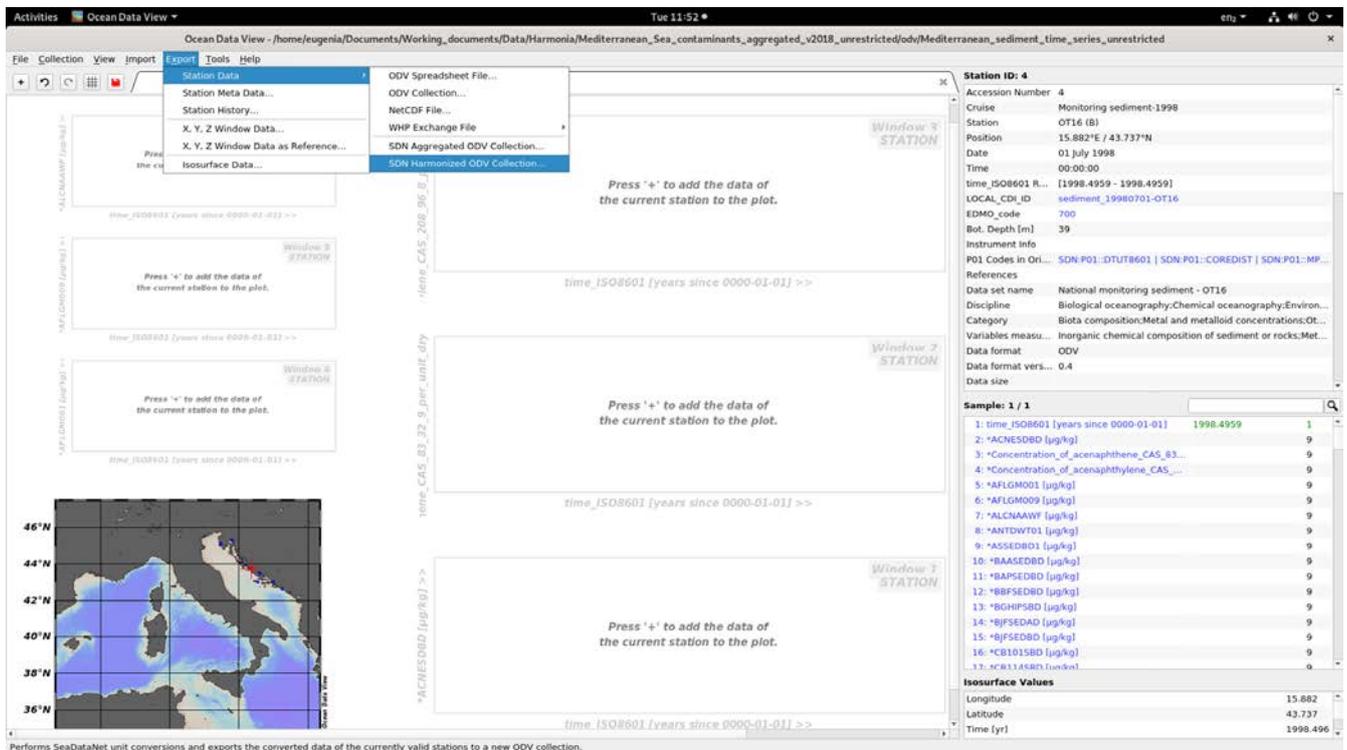
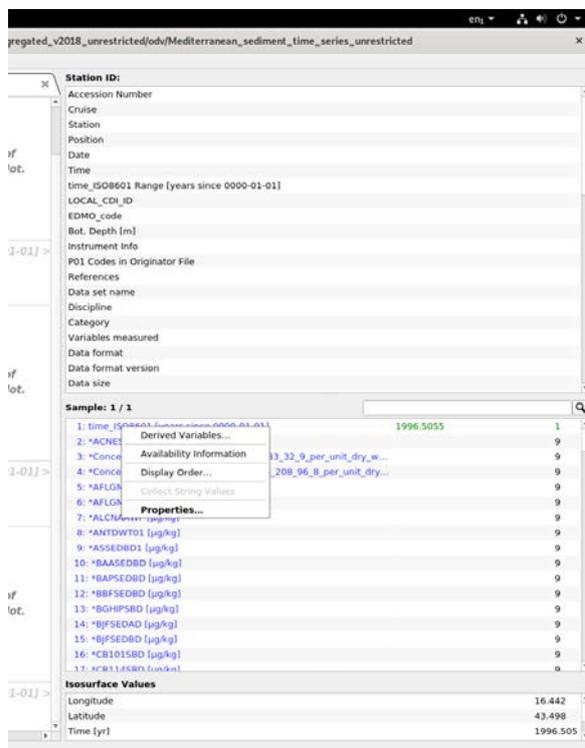After the harmonisation of the collection, some format and content controls can be performed.
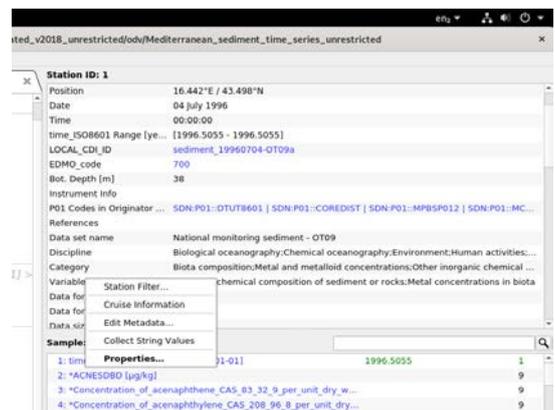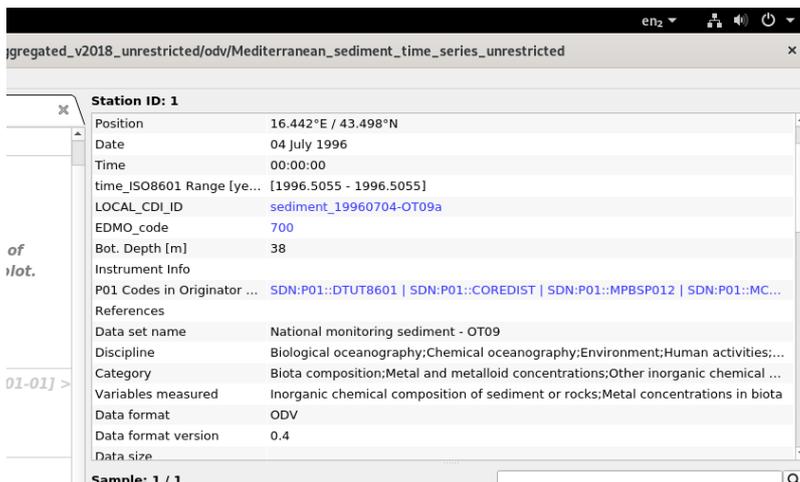
1. Check P02 present in the collection: Right-click in metadata field "variables measured", select collect string values. Copy and paste in a text or spreadsheet file. Check if all P02s are related to the collection
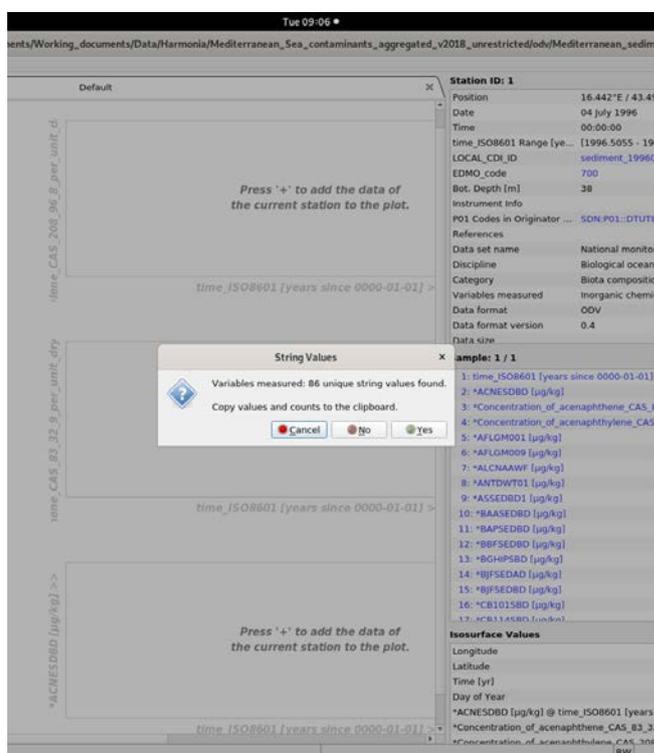
2. If any P02 is not related to the matrix/collection that is being prepared or to split collections that belong to different matrices: Right-click in metadata field "variables measured", select "station filter". Inside station filter window select "meta data" sheet and within text meta variables select "variables measured". Write the keyword related to your collection: water, sediment or biota with * before and after the key word. If there is more than one keyword separate them with || as ODV general guidelines establish.
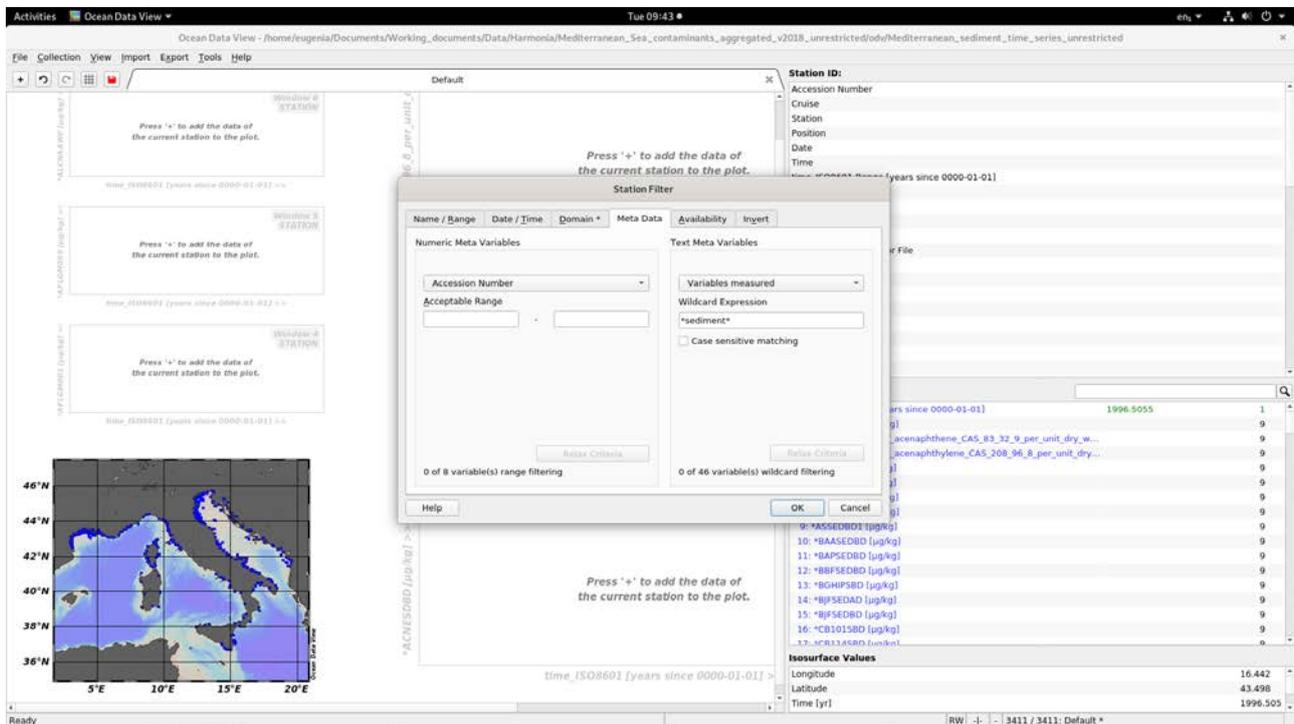
3. Export the stations and variables related to the previous filter. An additional "inverse selection" is needed to avoid parameters related to * **pore waters** * in the water collection. To do this, inside station filter window select "meta data" sheet and within text meta variables select "variables measured". Write * **pore waters** * and check "invert selection" on the last sheet. This step cannot be performed for sediment because stations where sediment is measured and stations where sediment pore waters are measured are the same, but variables related to pore waters are measured in units /L, so this can be addressed in a further step.

4. Check if the primary variable is present in the whole collection. Right-click in the primary variable field, select availability information. If the percentage is not 100% for the primary variable, check if there is another P01 that is the primary variable (sample_id; depth instead of coredist for sediment profiles…).

5. If primary variable is not 100% present, you can filter which stations don't contain the parameter associated to the primary variable. Right-click in metadata field "variables measured", select "station filter". Inside station filter window select "availability" and choose the primary variable as required variable. Select invert selection in the invert sheet and stations without the correct primary variable will be displayed.
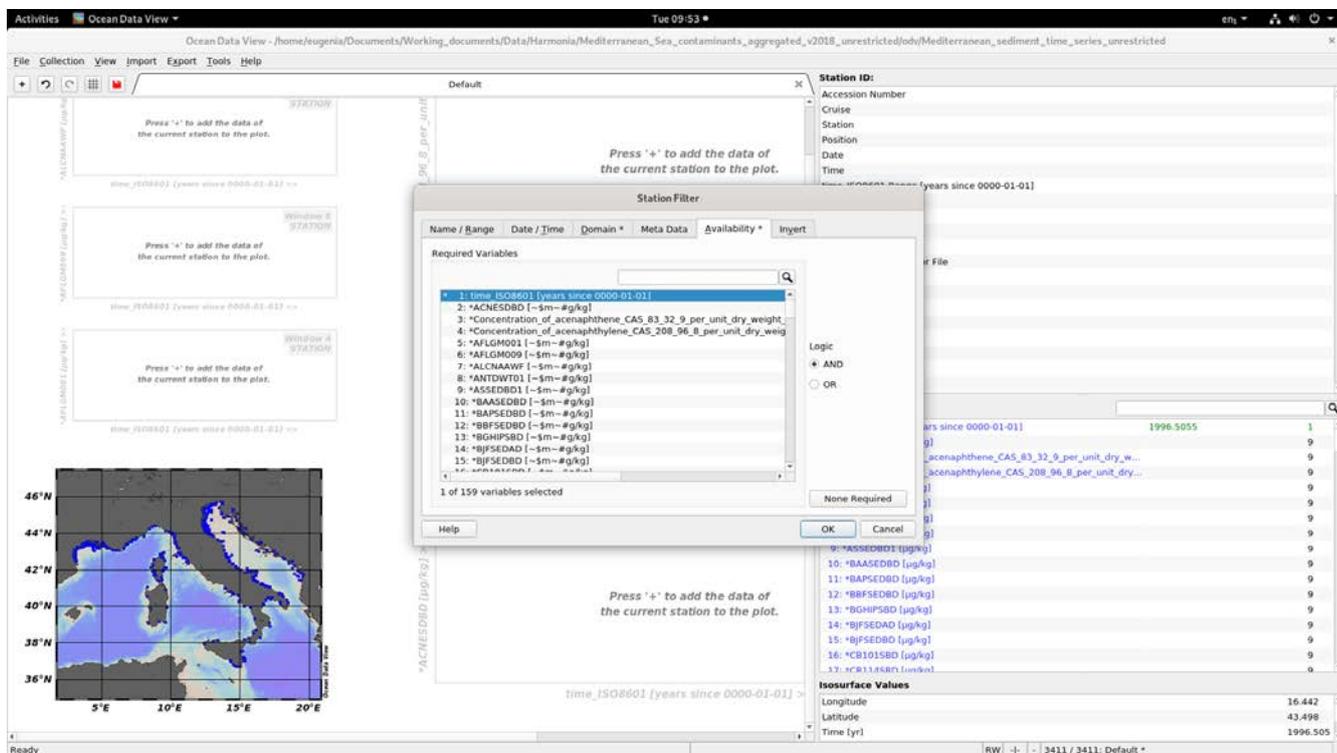
6. Check if there are empty variables in the whole collection. In the window collection, inside properties, data variables, select empty variables. If there' s any empty variable for the whole collection, it can be deleted.
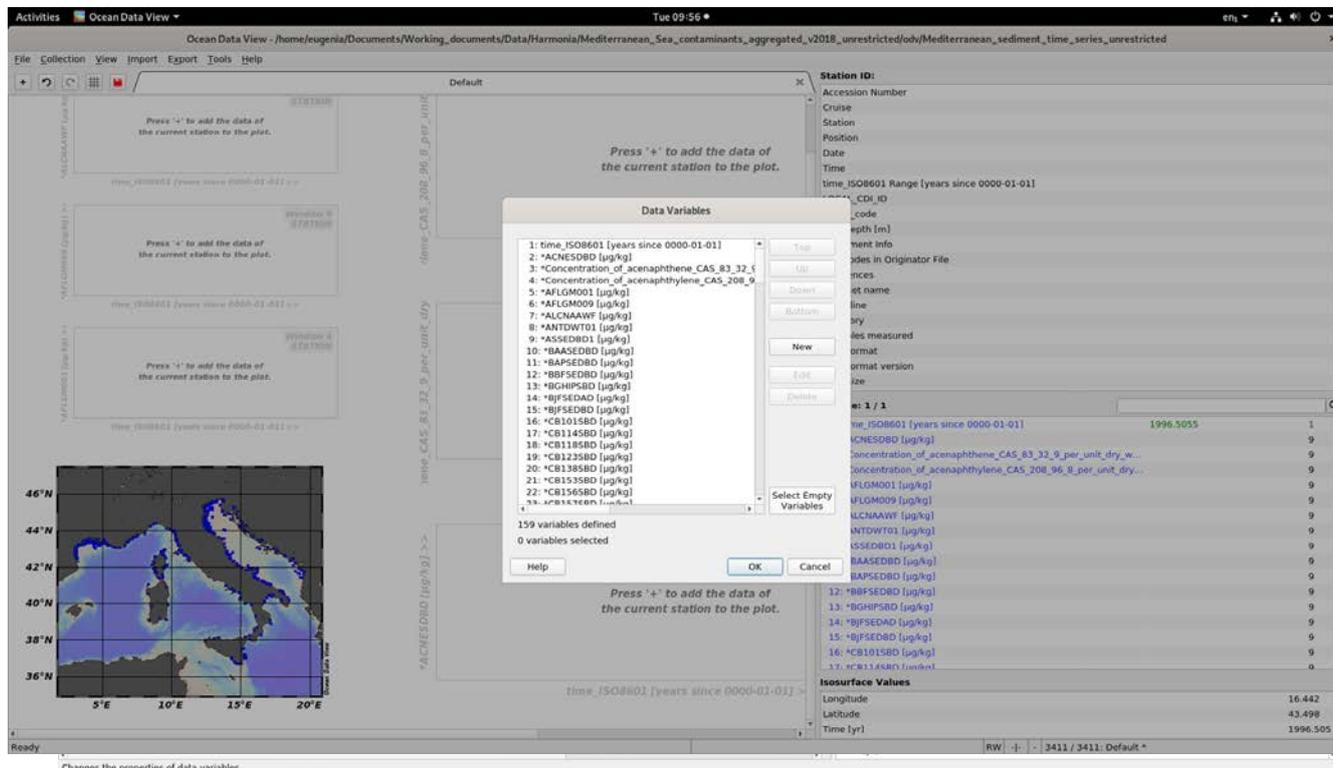
7. Check if there are variables that do not correspond to the matrix (i.e. depth in sediment instead of depth below seabed). See the parameters list on the right part of ODV; one by one with the cursor over each variable the name can be visualized, taking note of which of them are not suitable in the collection. Parameters related to water are usually in units /L (but not always). Parameters related to sediment or biota are usually in units /kg. It is necessary to

be careful with the end of the parameters list that contain parameters related to the sample and not to contaminants measurements. Try to avoid deleting this kind of parameters.

8. Check at the end of the variable list if there is any non-harmonized variable that is not linked with contaminants data (i.e. chlorophyll data…). It is better to avoid exporting these parameters. This step needs to be carefully performed because there are a set of parameters at the end of the list that are related to the sample and not to contaminants measurements. Try to avoid deleting this kind of parameters:

- For water: Depth (if time series), pH, Salinity, Transparency Depth, Dissolved Oxygen

- For sediment: Depth below seabed (if time series), Proportion of sizes of particles, Parameters   related to granularity, water content, organic matter content, thickness of sediment properties   and sedimentation taxes...

- For biota: parameters related to the sample (biota sizes, sample ids, sex, life stage…), water and lipid contents, wet weight, depth…

Regarding time-series collections, sometimes there is more than one parameter related to the sampling depth (with equivalent P01s or same P01 in different units). In this case the parameters should be merged.

9. Once the parameters collection is "clean" and all parameters are supposed to belong to the matrix, a final check can be done. Open again the station filter, and in the availability sheet, choose "OR" and select only related to contaminants, not to the sample (i.e. depth, sample_id, composition of the sample….). It is possible to have stations in the collection that

contain only parameters related to the sampling and not contaminants and these stations have to be avoided. Export the final clean selection of stations. A double check before the export can be performed inverting the selection of the stations and making sure that the "discarded stations" don't contain parameters related to contaminants.

10. If there are still stations that contain potential wrong variables, contact the originators. In the station filter, choosing the sheet "availability", choose the variable to check.

It is very important for sediment samples to have the sampled depth (COREDIST: depth below seabed)

11. If there are variables in different units, there is a functionality inside ODV to aggregate them performing previously the transformation needed.
12. For contaminant in sediment only measurements till 10 cm depth (COREDIST=0.1 m) are useful. During the final spreadsheet export, this can be done exporting values with the data filter as function of the range, choosing values from 0 – 0.10 for COREDIST parameters

13. Check manually inside ODV or outside using a script that variables related to contaminants with value=0, have a QF=6 (under LOD)

5. Check of quality flagging of data and metadata

Check values of quality flags on data and metadata and modifications according to SDN standards:

- In case of empty cells → QF=9
- In case of values = 0 for contaminant concentrations → QF=6

6. Checks for inconsistent measurement units

When using the harmonized, transposed and decomposed matrix outside ODV, it is possible to select all parameters related to "concentration" variables and check for inconsistent measurement units (eg. Kg, meters, °C for concentrations) and attribute the proper QF.

7. Checks for clearly impossible data ranges (ex. Different orders of magnitude)

Due to the multiple combinations of matrices, chemical substances and analytical methods, the resulting list of variables included in the datasets is huge. There is not enough information about the concentration ranges for the specific variables in the different areas. With this in mind, visual inspection of data is carried out in order to check for clearly impossible data ranges (ex. different orders of magnitude). Suspect data are flagged according to SDN standards and data originators must be contacted.