

EMODnet Chemistry

Quality Control steps and dataset formats for EMODnet Chemistry Contaminant aggregated datasets

M. Lipizer, M.E. Molina Jack and A. Giorgetti (OGS), L. Buga (NIMRD), K. Wesslander (SMHI), J. Gatti (IFREMER), A. K. Østrem (IMR), A. Iona and M. Tsompanou (HCMR), M. M. Larsen (AU), R. Schlitzer (AWI)

Date: 08/04/2021

<u>**Citation:**</u> M. Lipizer, M.E. Molina Jack, A. Giorgetti, L. Buga, K. Wesslander, J. Gatti, A. K. Østrem, A. Iona, M. Tsompanou, M. M. Larsen, R. Schlitzer, 2021, Quality Control steps and dataset formats for EMODnet Chemistry Contaminant aggregated datasets, 08.04.2021, 20 pp., DOI 10.6092/8b52e8d7-dc92-4305-9337-7634a5cae3f4



Content:

1.	Introduction	2
2.	Validation steps	4
3.	Metadata completeness and dataset format control	5
F	format and content controls roadmap:	7
4.	Harmonization of unit and parameter naming	12
5.	Check of quality flagging of data and metadata	16
6.	Checks for inconsistent measurement units	16
7.	Checks for clearly impossible data ranges (ex. Different orders of magnitude)	16
8.	Dataset output formats	16
S	eaDataNet ODV csv spreadsheet format	17
Т	ransposed and decomposed csv spreadsheet format	17



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 Vocabulary, available Usage which is 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 FLAGS¹). 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 parameter groups listed in the table below. The links between P36 (Terms based on the EU MSFD used by the EMODNET chemistry lot to provide coarse granularity groupings of the chemical parameters it covers) and PO2 (terms describing fine-grained related groups of measurement phenomena designed to be used in dataset discovery interfaces) are in the process of revision as more experience is gained in the field. The updated links are available here: http://vocab.nerc.ac.uk/collection/P36/current/

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	AQYN	Concentration of alkenes and alkynes in the water column
		BCAH	Concentration of polycyclic aromatic hydrocarbons (PAHs) in biota
		DALK	Concentration of alkanes in the water column
		OHWC	Concentration of other hydrocarbons in the water column

1

 $[\]label{linear} 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=\\ \end{tabular}$



EMODnet Chemistry Quality Control steps for EMODnet Chemistry Contaminant aggregated datasets

		РСАН	Concentration of polycyclic aromatic hydrocarbons (PAHs) in suspended particulate material
		PCHW	Concentration of polycyclic aromatic hydrocarbons (PAHs) in the water column
		SALK	Concentration of aliphatic hydrocarbons in sediment samples
		SCAH	Concentration of polycyclic aromatic hydrocarbons (PAHs) in sediment samples
HVYMTL	Heavy metals	BCMT	Metal concentrations in biota
	-	MDBO	Trace metalloid concentrations in biota
		MDPW	Trace metalloid concentrations in sediment pore water
		MDWD	Dissolved trace metalloid and inorganic selenium concentrations in water bodies
		MDWP	Particulate trace metalloid and inorganic selenium concentrations in water bodies
		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
PCBSXX	Polychlorinated biphenyls	ВСРВ	Concentration of polychlorobiphenyls (PCBs) in biota
		РРСВ	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
RADNUC	Radionuclides	BRAD	Radioactivity in biota
		SRAD	Geological sample radioactivity
		WRAD	Radioactivity in water bodies
OTHERS	Other contaminants	BCOC	Concentration of other organic contaminants in biota
		PCOC	Concentration of other organic contaminants in suspended particulate material
		SCOC	Concentration of other organic contaminants in sediment samples
			Concentration of other organic contaminants in the



2. Validation steps

A step-wise approach is carried out to obtain validated regional aggregated datasets, taking into account 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 involves²:

- metadata completeness and dataset format control
 - check of consistency of primary variables, especially in sediment and biota datasets
 - o separation of data related to different matrices into different dataset collections
- □ harmonization of unit and parameter naming
- □ check of quality flagging of data and metadata
- □ checks for inconsistent measurement units
- □ checks for clearly impossible data ranges (ex. Different orders of magnitude)

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 This approach derives 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, <u>https://harmonia.adrioninterreg.eu/</u>) financed by Interreg ADRION (2018-2020).



3. 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, some automatic file format checks are performed and 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.

Additional checks needed:

An additional check could be necessary to investigate if biota data should rather be in the time series collection. This check consists in looking at the number of depths, and if it's a profile with only one depth, check how other biota data from this EDMO code have been submitted. Contact with data originator is needed to evaluate to check the dataset type.

It might happen that during the import phase more than one collection of sediment or water is created, due to a different primary variable 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 \rightarrow data variables \rightarrow new \rightarrow 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 SDN harmonized ODV collection. Some parameters of the collection may belong to other matrix if the station samples multiple matrices.



Format and content controls roadmap:

1. Check PO2 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 PO2s are related to the collection being analysed.

	Station ID: 1			×	Station I					
- × \		10 44205 / 42 4000			Position		6.442°E / 43.498°N			
-	Position	16.442°E / 43.498°N		- <u> </u>	Date		14 July 1996 10:00:00			
	Date	04 July 1996		_			1996.5055 - 1996.5	0551		
	Time	00:00:00		LOCAL_CO		ediment_19960704				
	time_ISO8601 Range [ye			EDMO_co	de 7	00				
	LOCAL_CDI_ID	sediment_19960704-OT09	a		Bot. Dept		18			
	EDMO_code	700		_	Instrumer			SDN:P01::COREDIST SDN		
	Bot. Depth [m]	38		Reference		DN:P01::D1018601	SDN:P01::COREDIST SDN	:P01::MP85P012 [:	SDN:P01::MC	
	Instrument Info			Data set r		lational monitoring	sediment - OT09			
	P01 Codes in Originator	SDN:P01::DTUT8601 SDN	I:P01::COREDIST SDN:P01::MPBSP012 SDN:P01::MC		Discipline			phy:Chemical oceanography	:Environment;Hum	nan activities;
	References				Category	E		etal and metalloid concentra		
	Data set name	National monitoring sedim	ent - OT09	1.4	Variable	Station Filter	chemical c	omposition of sediment or ro	cks;Metal concent	rations in biota
	Discipline	Biological oceanography;C	hemical oceanography;Environment;Human activities;		Data for	Cruise Information	n			
1] >	Category	Biota composition;Metal a	nd metalloid concentrations;Other inorganic chemical		Data siz	Edit Metadata				
17	Variables measured	Inorganic chemical compo	sition of sediment or rocks;Metal concentrations in biota		Sample:	Collect String Va	lues			
	Data format	ODV			1: time	Properties	01-011	1996	5055	1
	Data format version	0.4				ESDBD [µg/kg]				9
	Data size			Ŧ	3: *Con	centration_of_acena		2_9_per_unit_dry_w		9
	Sample: 1 / 1			2	4: *Con	centration_of_acena	phthylene_CAS_208	96_8_per_unit_dry		9
		9		Date Time		04 July 1996 00:00:00				
		Default	×/	Station ID: Position		16.442°E / 43.	49			
			•	Date						
		P								
		an		time_ISO86		e [1996.5055 - 1				
		5		EDMO code		sediment_199 700	00			
		5	Press '+' to add the data of	Bot. Depth		38				
		36	the current station to the plot.	Instrument						
		80				SDN:P01-DTU	TB			
				References Data set na		National moni				
		CAS			MTRE					
				Discipline						
		<u>p</u> '		Discipline Category		Biological oce Biota composi				
		lene_	time_IS08601 {years since 0000-01-01] >	Discipline Category Variables n	neasured	Biota composi Inorganic cher	tio			
		tiene	time_IS08601 [years since 0000-01-01] >	Category Variables n Data forma	ıt	Biota composi Inorganic cher ODV	tio			
		ilette_	time_ISO8601 [years since 0000-01-01] >	Category Variables n Data forma Data forma	ıt	Biota composi Inorganic cher	tio			
		dry iteme_		Category Variables n Data forma Data forma Data size	it it version	Biota composi Inorganic cher ODV	tio			
		nit_dry itene_	String Values ×	Category Variables n Data forma Data forma Data size ample: 1	it it version / 1	Biota composi Inorganic cher ODV 0.4	tion			
		r_unit_dry ilene_		Category Variables n Data forma Data forma Data size ample: 1 1: time_!!	it it version / 1	Biota composi Inorganic cher ODV 0.4	tion			
		per_unit_dry dene_	String Values ×	Category Variables n Data forma Data forma Data size ample: 1 1: time_II 2: *ACNE	t version / 1 508601 (year SDBD (µg/kg	Biota composi Inorganic cher ODV 0.4	nic 1]			
		2.9 per unt. dy lane	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma Data size ample: 1 1: time_!! 2: *ACNE 3: *Conce 4: *Conce	it version /1 SO8601 (year SDBD (µg/kg entration_of_ entration_of_	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0] accenaphthene_CA5 accenaphthylene_C	tion mic			
		.32.9 per unit, day dane	String Values × Variables measured: 86 unique string values found	Category Variables in Data forma Data forma Data size ample: 1 1: time_0 2: *ACNE 3: *Conci 4: *Conci 5: *AFLG	it version /1 SOB601 (year SDBD (µg/kg entration_of_ M001 (µg/kg)	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0] acenaphthene_CAS acenaphthylene_C	tion mic			
		.83.32.9. per, unit, dity	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma Data size ample: 1 1: time_II 2: *ACNE 3: *Conce 4: *Conce 5: *AFLG 6: *AFLG	it version /1 SOB601 (year SDBD (µg/kg entration_of_ M001 (µg/kg) M009 (µg/kg)	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0 1 acenaphthene_CAS acenaphthylene_Ca	tion mic			
		45,83,32,9,per,untt,dry ritene.	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma Data size ample: 1 1: time_it 2: *ACNE 3: *Conce 4: *Conce 5: *AFLGI 6: *AFLGI 7: *ALCN	t t version /1 SOB601 (yea SDBD (µg/kg entration_of_ M001 [µg/kg] M009 [µg/kg] AAWF [µg/kg	Biota composi Inorganic cher ODV 0.4 rs aince 0000-01-0 1 acenaphthene_CA acenaphthene_CA	tion mic			
		, CAS, 32, 32, 9, per, unit, dry viene.	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard. Copy values	Category Variables in Data forma Data forma Data size ample: 1 1: time_It 2: *ACNE 3: *Conce 5: *AFLGi 6: *AFLGi 7: *ALCN 8: *ANTD	t t version / 1 508601 (year SDBD [µg/kg entration_of_ M001 [µg/kg] AAWF [µg/kg WT01 [µg/kg	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0 1 accenaphthene_CAS accenaphthylene_Ca	tion mic			
		one_CAS_83_32_9 per unit_ary none_	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma Data sizie ample: 1 1: time_II 2: *ACNE 3: *Conci 5: *AFLGI 6: *AFLGI 7: *ALCN 8: *ANTD 9: *ASSEI	t t version /1 508601 [year 5080 [µg/kg entration_of_ Mo09 [µg/kg] AAWF [µg/kg WT01 [µg/kg] DBD1 [µg/kg	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0 l acenaphthene_CAS acenaphthene_CAS	tion mic			
		have CAS. 03.32.9.per. unit. dry dame.	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard. Copy values	Category Variables in Data forma Data forma Data size amplet 1 1: time_R 2: *ACHE 3: *Conct 5: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 8: *ANTD 9: *ASSE 10: *BAAS	t t version / 1 508601 (year SDBD [µg/kg entration_of_ M001 [µg/kg] AAWF [µg/kg WT01 [µg/kg	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0 l acenaphthene_CAS accenaphthene_CAS accenaphthylene_CAS accenaphthylene_CAS accenaphthylene_CAS accenaphthylene_CAS accenaphthylene_CAS accenaphthylene_CAS	tion mic			
		1010 _ CAS, 82, 32, 9, per, unit, dry 1600_	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard. Copy values	Category Variables n Data forma Data forma Data size ample: 1 1: time_!: 2: *ACNE 3: *Conci 3: *Conci 5: *AFLGi 6: *AFLGi 7: *ALCN 8: *ANTD 9: *ASSEI 10: *BAPS 11: *BAPS 12: *BEFS	it version / 1 SOB601 (year SOB0 (up/kg instration_of_ instration_of_ instration_of_ MOD1 (up/kg) BD1 (up/kg) SEDD0 (up/kg) SEDD0 (up/kg) SEDD0 (up/kg) SEDD0 (up/kg)	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0 1 accenaphthene_CAS accenaphthene_CAS accenaphthylene_Ca 1 1 1 1 1 1 1 1 1	tion mic			
		tone_cAS_43_3_9_per, unit_dryriene_	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard. Copy values	Category Variables n Data forma Data forma Data size ample: 1 1: time_ti 2: *ACNE 3: *Conci 4: *Conci 5: *AFLGI 6: *AFLGI 6: *AFLGI 7: *ALCN 8: *ANTD 9: *ASSEI 10: *BAAS 11: *BAPS 12: *BBFS	tt version / 1 SOB601 (yea SDB0 (µg/kg entration_of_ intration_of_ Mo09 (µg/kg) MO09 (µg/kg) MO09 (µg/kg) SEDB0 (µg/kg SEDB0 (µg/kg) SEDB0 (µg/kg) SEDB0 (µg/kg) SEDB0 (µg/kg)	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0 accenaphthene_CA3 accenaphthene_CA3 accenaphthylene_CA 1 1 1	tion mic			
		>> 100, CAS, 83, 32, 9, per, unit, dry 16me.	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard. Copy values	Category Variables n Data forma Data forma Data forma nata size. ample: 1 1: fime_li 2: *ACNE 4: *Conce 5: *AFLG 6: *AFLG 6: *AFLG 7: *ALCN 8: *ANTD 9: *ASSE! 10: *BASS 11: *BASS 11: *BASS 11: *BASS 11: *BASS 11: *BASS 11: *BASS 11: *BASS	tt version / 1 S08601 (yeas S0801 (yeas S0801 (yeas S0801 (ye/kg) M001 (µg/kg) M001 (µg/kg) M001 (µg/kg) S0801 (µg/kg	Biota composa inorganic cher ODV 0.4 Its since 0000-01-0 I accenaphthene_CA accenaphthene_CA I I I I I I I I I I I I I I I I I I I	tion mic			
		0] >> tone_CAS,83,32,9,per,unit,dry tione_	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard. Copy values	Category Variables n Data forma Data forma Data forma nata size. ample: 1 1: time,it 2: *ACNE 3: *Conci 4: *Conci 5: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 10: *BASE 10: *BASE 10: *BASE 12: *BGH 14: *BGH 14: *BGH	tt version /1 SOB601 (yea SDB0 (yg/kg entration, of_ entration, of_ moot (µg/kg) MOOF (µg/kg) AAWF (µg/kg) AAWF (µg/kg) EDB0 (µg/kg) EDB0 (µg/kg) EDB0 (µg/kg) EDB0 [µg/kg)	Biota composi Inorganic cher ODV 0.4 rs since 0000-01-0 1 acenaphthene_CAS conspitthylene_C 1 1 1 1 1 1 1 1 1	tion mic			
		b/b] >> tene_cAS_d3_32_9_per,unt_dry nene_	String Values × Variables measured: 86 unique string values found. Copy values and counts to the clipboard. Copy values	Category Variables in Data forma Data forma Data size ample: 1 1: time_!: 2: *ACNE 3: *Conce 3: *Conce 4: *Conce 5: *AFLGI 6: *AFLGI 6: *AFLGI 6: *AFLGI 7: *ALCN 8: *ANTD 9: *ASSE 10: *BASS 11: *B	tt version / 1 S08601 (yea S0800 (yea/kg entration_of_ entration_of_ entration_of_ entration_of_ entration_of_ entration_of_ M001 (µg/kg) K0001 (µg/kg) K0001 (µg/kg) E000 (µg/kg) E000 (µg/kg) E000 (µg/kg) E000 (µg/kg) E000 (µg/kg) E000 (µg/kg) E000 (µg/kg)	Biota composi- Inorganic cher OOV 0.4 Is since 0000-01-0 I accraphthylene_CA caraphthylene_CI I I I I I I I I I I I I I I I I I I	tion mic			
		[19/kg] >> tene_CAS_83_32_9_per_unit_dry risme_	String Values X Variables measured: 66 unique string values found Copy values and counts to the clipboard. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma Data size ample: 1 1: time_!: 2: *ACNE 3: *Conce 3: *Conce 4: *Conce 5: *AFLGI 6: *AFLGI 6: *AFLGI 6: *AFLGI 7: *ALCN 8: *ANTD 9: *ASSE 10: *BASS 11: *B	tt version /1 SoB601 (yea SOB601 (yea SOB601 (yea SOB0 (yea MO01 (yea/kg) MO01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB0 (yea/k	Biota composi- Inorganic cher OOV 0.4 Is since 0000-01-0 I accraphthylene_CA caraphthylene_CI I I I I I I I I I I I I I I I I I I	tion mic			
		8D[juj/kg] >> tene_CAS_,83_,32_9_per_unit_dry itene_	String Values × Watiables measured: 86 unique string values found. Copy values and counts to the clipboard. © gancel @ go @ res time_ISO3601 [years since 0000-01-01] >	Category Variables in Data forma Data forma Data size ample: 1 1 time, it 2: *ACNE 3: *Conci 4: *Conci 4: *Conci 5: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 9: *ASSE 10: *BAAG 11: *BAAG 12: *BASS 11: *BASS	tt version /1 SoB601 (yea SOB601 (yea SOB601 (yea SOB0 (yea MO01 (yea/kg) MO01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB0 (yea/k	Biota composi- Inorganic cher OOV 0.4 Is since 0000-01-0 I accraphthylene_CA caraphthylene_CI I I I I I I I I I I I I I I I I I I	tion mic			
		SDBD [Jug/kg] >> tene_CAS_83_32_9_per_unit_dry riene_	String Values X Variables measured: 66 unique string values found Copy values and counts to the clipboard. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma Data size ample: 1 1: time_ti 2: *ACNE 2: *ACNE 3: *Conci 4: *Conci 4: *AFLGI 6: *AFLGI 6: *AFLGI 7: *ALCN 8: *ANTD 9: *ASSE 10: *BASS 10: *BASS 11: *CB1C 11: *CB	tt version /1 SoB601 (yea SOB601 (yea SOB601 (yea SOB0 (yea MO01 (yea/kg) MO01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB01 (yea/kg) SOB0 (yea/k	Biota composi- Inorganic cher OOV 0.4 Is since 0000-01-0 I accraphthylene_CA caraphthylene_CI I I I I I I I I I I I I I I I I I I	tion mic			
		NESDBD [jig/kg] >> tene_CAS.83_32,9_per_unit_dry risene_	String Values X Variables measured: 66 unique string values found Copy values and counts to the clipboard. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma	tt version /1 SOBEO Lignks SOBEO Lignks Instration_ef_ mitration_ef_ more Lignks More Lignks More Lignks EDBD Lig	Biota composi- Inorganic cher OOV 0.4 Is since 0000-01-0 I accraphthylene_CA caraphthylene_CI I I I I I I I I I I I I I I I I I I	tion mic			
		ACKESDBD [jig/kg] >> teine_CAS_43_32_9_per_unit_dry riene_	String Values X Variables measured: 66 unique string values found Copy values and counts to the clipboard. Copy values and counts to the clipboard.	Category Variables in Data forma Data forma Data Size ample: 1 1 : time,ti 2 : 4ACNE 3 : *Conce 5 : 4AE(6) 6 : *AFLGI 6 : *AFLGI 7 : *ALCN 9 : *ASSEI 10 : *BAAG 11 : *BAAG 11 : *BAAG 13 : *BGHI 13 : *BGHI 15 :	tt version /1 SoB601 (yaks SoB00 (yaks SoB01 (yaks SoB01 (yaks Mool (yaks) Mool (yaks) Mool (yaks) Mool (yaks) SoB01 (yaks) SoB01 (yaks) SoB01 (yaks) SoB01 (yaks) SoB0 (yaks	Biota composi Interganic chen ODV 0.4 Instance 0000-01-0 Instance 0000-01-0 Instance 0000-01-0 Instance 0000-01-0 Instance 0000-01-0 Instance 0000-01-0 Instance 000-01-0 Instance 000-00-0 Instance 000-00-00-0 Instance 000-00-00-0 Instance 000-00-00-00-00-00-00-00-00-00-00-00-00	tion 11 11 12 13			
		*ACNESDBD [Jug/kg] >> tene_CAS_B3_32_9_per_unit_dry ritene_	String Values X Variables measured: 66 unique string values found Copy values and counts to the clipboard. Copy values and counts to the clipboard.	Category Variables in Data forma Data Size ample: 1 1: time, t 2: *ACNE 3: *Conce 5: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 6: *AFLG 10: *BASS 11: *BASS 1	tt version / 1 SOBGO I jvea SDBO [lug/kg entration_of_ moral lug/kg MOD I lug/kg MOD I lug/kg MOD I lug/kg BOD I lug/kg EDBD lug/kg EDBD lug/kg EDBD lug/kg SDBD lug/kg FSBD I lug/kg TSBD I lug/kg AGED Lug/kg FSBD I lug/kg AGED Lug/kg FSBD I lug/kg AGED I l	Biota composition of the second secon	tion 11 5 5 75 1			

2. If any PO2 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. For water, if the filter is "*water col* || *water bod*", it will avoid getting parameters related to sediment pore waters.

Activities 🔛 Ocean Data View 👻	Tue 09:4	43 •			en1 👻 📇 🕸	0+
Ocean Data View - /home/eugenia/Documents/V	Norking_documents/Data/Harmonia/Mediterranean_Sea_co	ntaminants_aggregated_	v2018_unrestricted/odv	//Mediterranean_sediment_time_series_unrest	ricted	×
File Collection View Import Export Tools Help						
+) C # • /	Default	×	Station ID:			
	U. I	~	Accession Number			-
Window 6 STATION	it.		Cruise			
Press '+' to add the data of	5		Station			
the current station to the plot.	Jer.		Position			
NA.4	Press '+' to ad	dd the data of	Date			
0.14	% the current sta	tion to the plot.	Time	years since 0000-01-01]		
* time_ISO8601 [years since 0000-01-01] >>	Station Filt	er		years since 0000-01-01]		
Window 5	Name / Range Date / Time Domain * Meta Data	Availability Invert				
STATION						
Press '+' to add the data of the current station to the plot.	Numeric Meta Variables	Text Meta Variables		or File		
o o o o o o o o o o o o o o o o o o o						
L T C	Accession Number *	Variables measured	٣			
time_ISO8601 [years since 0000-01-01] >>	Acceptable Range	Wildcard Expression				
	· · · · · · · · · · · · · · · · · · ·	*sediment*				
Mindow 4 STATION		Case sensitive mate	hina			
6						
Press '+' to add the data of the current station to the plot.						
						٩
				ars since 0000-01-01]	1996.5055	1 4
time_IS08601 [years since 0000-01-01] >>				al	1990.3033	9
				acenaphthene_CAS_83_32_9_per_unit_dry_w		9
	Relax Criteria			acenaphthylene_CAS_208_96_8_per_unit_dry.		9
	0 of 8 variable(s) range filtering	0 of 46 variable(s) wild	card filtering	91 1		9
46°N			-	1]		9
	Help		OK Cancel	9]		9
44°N	Thep		9: *ASSEDBD1 (µg/	91		9
			10: *BAASEDBD [µg			9
42°N	[6]		11: *BAPSEDBD [µg			9
	/b		12: *BBFSEDBD [µg			9
40°N	Press '+' to ad		13: *BGHIPSBD [µg	[kg]		9
	the current star	tion to the plot.	14: *BJFSEDAD [µg/			9
38°N	S		15: *BJFSEDBD [µg/			9
30 11	ACN		16: *CB101SBD [µg			9
	V*		Isosurface Values			u i
36°N	time ISO8601 [years	since 0000-01-011	Longitude		16	442 *
5°E 10°E 15°E 20°E	unie_iso8001 [years	ames 0000-01-01] >	Latitude			498
4			* *****			96.505
Ready				RW - 3411 / 3411: Default	t *	

- 3. Export the stations and variables related to the previous filter.
- 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 PO1 that is the primary variable (sample_id; depth instead of coredist for sediment profiles...).

EMODnet Chemistry



Quality Control steps for EMODnet Chemistry Contaminant aggregated datasets

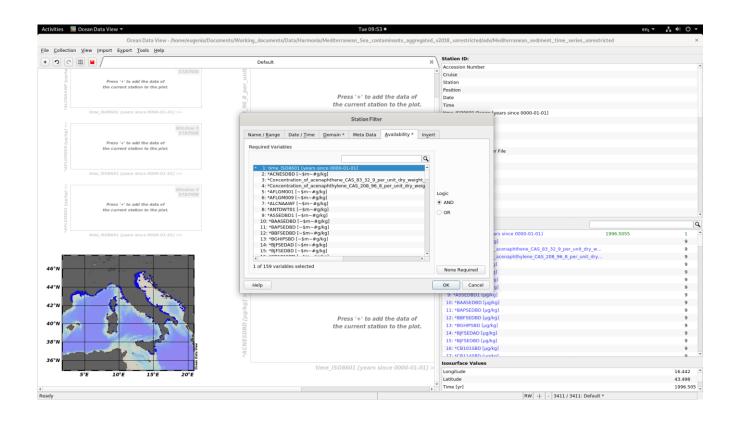
	Chatles ID.				De	fault				×	Station ID:	
× \	Station ID:					ioun				~ \	Accession N	lumber
	Accession Nu	mber		-	in l					-	Cruise	
	Cruise										Station	
	Station				De						Position	
	Date				co		Press '+' to ad				Date	
	Time				6		the current stati		the pl	ot.	Time	
		Range (years since 0000-01-0	1		508		Data Availabili	ty [%]				Range (years
	LOCAL CDI IE		(1)								0	
	EDMO code	,			AS						٩	<u>1</u>
	Bot. Depth [m	1					Variable	_	ct Set	Cruise S		2 0
	Instrument In				6116	1	time_ISO8601 [years since 0000-01-01]	94	94	100	100	o
1] >		Originator File			1	2	*ACNESDBD [µg/kg]	0	0		-	Driginator File
	References	originator rine				3	*Concentration_of_acenaphthene_CAS_83	. 7	7			
	Data set nam	e				4	*Concentration_of_acenaphthylene_CAS	7	7			5
	Discipline				C g	5	*AFLGM001 [µg/kg]	9	9			
	Category				±1	6	*AFLGM009 [µg/kg]	9	9			
	Variables me	asured			- In In	7	*ALCNAAWF [µg/kg]	58	58			isured
	Data format				L D	8	*ANTDWT01 [µg/kg]	0	0			
	Data format v	rersion			9	9	*ASSEDBD1 [µg/kg]	0	0			ersion
	Data size				32_9.	10	*BAASEDBD [µg/kg]	0	0			
						10	*BAPSEDBD (µg/kg)	0	0			
	Sample: 1 / 1			٩	8	11	*BBFSEDBD [µg/kg]	0	0			3601 (years si
	1: time_ISO	Derived Variables	1996.5055	1 *	CAS		*BGHIPSBD [µg/kg]	0	0			BD [µg/kg]
	2: *ACNES			9	G	13	15.5.					ration_of_acen
	3: *Conce	Availability Information	33_32_9_per_unit_dry_w	9	10	14	*BJFSEDAD [µg/kg]	65	65			ration of acen
1] >	4: *Conce	Display Order	_208_96_8_per_unit_dry	9	10	15	*BJFSEDBD [µg/kg]	0	0			01 [µg/kg]
	5: *AFLGN	Collect String Values		9		16	*CB101SBD [µg/kg]	0	0			09 [µg/kg]
	6: *AFLGN	Properties		9		17	*CB114SBD [µg/kg]	2	2			WF [µg/kg]
	7: *ALCNA		1	9		18	*CB118SBD [µg/kg]	0	0			01 [µg/kg]
	8: *ANTDW			9		19	*CB123SBD [µg/kg]	2	2			D1 [µg/kg]
	9: *ASSEDB			9	[<i>b</i>]	20	*CB138SBD [µg/kg]	0	0			BD [µg/kg]
	10: *BAASED			-	g/k	21	*CB153SBD [ua/ka]	0	0		Ψ	BD [µg/kg]
	11: *BAPSED			9	LP.				Clipt	oard Copy	Close	BD [µg/kg]
	12: *BBFSED 13: *BGHIPS			9	DBD							BD [µg/kg]
	13: *BGHIPS 14: *BJFSED			9	SD							EDAD [µg/kg]
	15: *BJFSED			9	NE							EDBD [µg/kg]
	16: *CB1015			9	ACI							1SBD [µg/kg]
	17: +CB1145			9 *	*							ASRD Juarkal
	Isosurface V										Isosurface	Values
1] >	Longitude			16.442 *							Longitude	
	Latitude			43.498							Latitude	
	Time [yr]			1996.505							Time [yr]	

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.

EMODnet

EMODnet Chemistry

Quality Control steps for EMODnet Chemistry Contaminant aggregated datasets

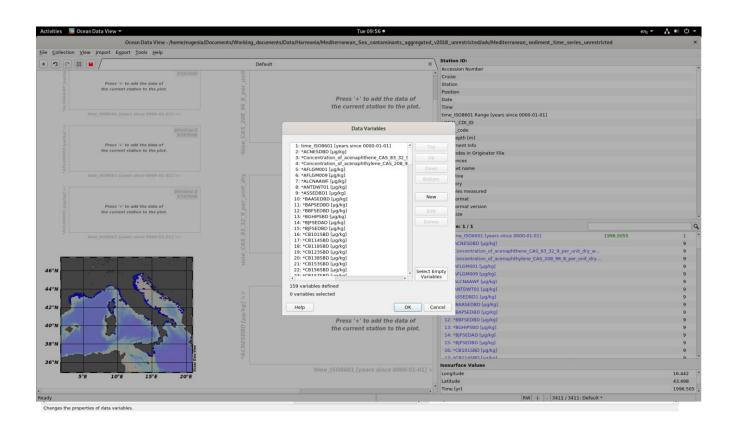


- 6. Check if there are empty variables in the whole collection. In the window collection, inside properties, data variables, select empty variables. If there's any empty variable for the whole collection, they should be inspected in order to verify any possible errors deriving from import into ODV. Only if it is confirmed that these are empty, they 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.

EMODnet Chemistry



Quality Control steps for EMODnet Chemistry Contaminant aggregated datasets



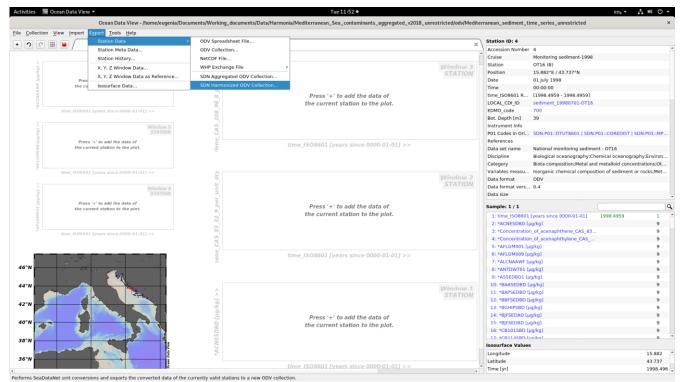
Activities 🔚 Ocean Data View 🕶	Tue 10:10 •			en ₂ -	• • • •
	Vorking_documents/Data/Harmonia/Mediterranean_Sea_contaminants_aggregated_v2018_unrest	ricted/odv/Medite	rranean_sediment_time_series_unrestrict	ed	×
File Collection View Import Export Tools Help	Default	×	Station ID: 849		
STATION		STATION	Accession Number 849 Cruise Contaminants_Sedim	ent	
Press '+' to add the data of the current station to the plot.	o Press '+' to add the data of	Station ITHAZSED0428 (B) Position 12.855°E / 45.521°N Date 15 April 2009			
time_IS08601 [years since 0000-01-01] >>	6 the current station to the plot.		Time 00:00:00 time_ISO8601 R [2009.2849 - 2009.24] LOCAL_CDI_ID Contaminants_ITHAZ		
Window 5 STATION Press '+' to add the data of			EDMO_code 3009 Bot. Depth [m] Instrument Info		
the current station to the plot.	time_ISO8601 [years since 0000-01-01] >>		P01 Codes in Ori SDN:P01::DTUT8601 References Data set name CONTAMINANTS IN S		/N:P01::A
time ISOB401 (years since 0000-01-01) >> Window & STATION Press 's' to add the data of the current station to the plot.	Lip Yuun - 190 Press '+' to add the data of the current station to the plot.	Window 2 STATION	Category Hydrocarbons:Metal	ohy:Environment:Human ac and metalloid concentratio er organic contaminants in	ns:Other
time_t508601 (years since 0000-01-01) >>			Sample: 1 / 1 1+0 KW 35E02 / (µg/kg) 144: *RWSSED28 [µg/kg]	0.12) 1
46'N	0 time_ISO8601 [years since 0000-01-01] >>		145: *RWSSED32 [µg/kg] 146: *RWSSED33 [µg/kg] 147: *Concentration_of_silver_Ag_per_ur 148: *Concentration_of_tributyltin_expr 149: *TBTXSD63 [µg/kg]		1 1 9 9
44°N 42'N	< fay	Window 1 STATION	150: Depth below seabed [m] 151: Depth [m] 152: Max_depth [m] 153: Proportion_dry_weight_of_particle 154: Concentration of total organic mal		9 9 9 9
40°N	Press '+' to add the data of the current station to the plot.		155: CLAYZZZZ [mg/kg] 156: ONCNCIWF [mg/kg] 157: SANDZZZZ [mg/kg] 158: SILTZZZZ [mg/kg]	51.30 24.70 51.30	1 9 1
	*AC		159: STTMUBS [mg/kg]		9
36'N 5'E 10'E 15'E 20'E	time_ISO8601 [years since 0000-01-01] >>		Kosurface Values *Concentration_of_acenaphthene_CAS_83 *Concentration_of_acenaphthylene_CAS_ *Concentration_of_acenaphthylene_CAS_	208_96_8_per_unit_dry_w	
4 Ready		b.	*AFLGM001 [µg/kg] @ time_ISO8601 [yea RW - 3411 / 3411: Default *	n's since 0000-01-01]=first	10.00



4. Harmonization of unit and parameter naming

The next step involves harmonising the collection in terms of units and PO1 parameters. 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 considered 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.



1. 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 additional parameters related to other themes (e.g. eutrophication). However, this step needs to be carefully performed because there may be a set of parameters at the end of the list that are related to the sample (e.g. sediment size and porosity, water content, ...) and not to contaminants measurements, which need however to be maintained in the collection. Try to avoid deleting this kind of parameters:



- For water: Depth (if time series), pH, Salinity, Temperature, 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 PO1s or same PO1 in different units). In this case the parameters should be merged.

- 2. 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.
- 3. 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.



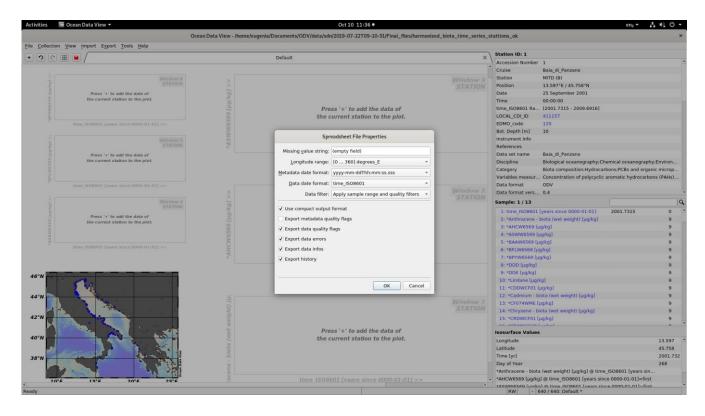
It is very important for sediment samples to have the sampled depth (COREDIST: depth below seabed)

			ents/Working_documents/Data/Harmonia/Mediterranean_Sea_contaminants_aggregated	_v2018_un	restricted/odv/N	lediter	rranean_sediment_time_series_	unrestricted	×		
<u>File</u> <u>Collecti</u>	ion ⊻iew	Import Export Tools Help				_	Station ID: 849				
+) (• # 😐		Default			×					
2		STATION			STATIO		Accession Number 849 Cruise Contaminat	nts_Sediment			
0.0		Press '+' to add the data of	un .		STATIO	10	Station ITHAZSED0				
10		the current station to the plot.	or				Position 12.855°E /				
VV.			Press '+' to add the data of				Date 15 April 20				
10			the current station to the plot.				Time 00:00:00	55			
14			o and current station to the prot					- 2009.2849]			
			Station Filter					nts ITHAZSED0428 H09			
		Window 5					EDMO code 3009	113_111423200426_103			
102		STATION	Name / Range Date / Time Domain * Meta Data Availability * Inve	t			Bot. Depth [m]				
0.0		Press '+' to add the data of		20			Instrument Info				
0.0		the current station to the plot.	Required Variables					TUT8601 SDN:P01::AFLGM001 S	DN-P01-A		
140			9				References				
174			145: *RWSSED32 [~\$m~#g/kg]					ANTS IN SEDIMENT			
2			145: *RW55ED32 [~5m~#g/kg] 146: *RW55ED33 [~\$m~#g/kg]					ceanography;Environment;Human	activities:M		
			147: *Concentration_of_silver_Ag_per_unit_dry_weight_of_sediment [~\$m			2		ins:Metal and metalloid concentrati			
		Window 4	148: *Concentration_of_tributyltin_expressed_as_tinTBTasSn_per_un		8			ion of other organic contaminants i			
(ka)		STATION	149: *TBTXSD63 [~\$m~#g/kg] 150: Depth below seabed [m]	Logic			Data format ODV				
Unit		Press '+' to add the data of	* 151: Depth [m]	AND			Data format vers 0.4				
101		the current station to the plot.	152: Max_depth [m]	O OR			Data size				
CIMIC			153: Proportion_dry_weight_of_particles63um_in_the_sediment_by_ 154: Concentration_of_total_organic_matter_per_unit_dry_weight_of_sediment_bry_backstering								
AFL			155: CLAYZZZZ [mg/kg]				Sample: 1 / 1		Q		
	time	1508601 (years since 0000-01-01) >>	156: ONCNCIWF [mg/kg]				144: *RWSSED28 [µg/kg]	0.12			
			157: SANDZZZZ [mg/kg]				145: *RWSSED32 [µg/kg]	33000.00	1		
			159: STTMLIBS [mg/kg]				146: *RWSSED33 [µg/kg]	7700.00	1		
			4 P				147: *Concentration of silver				
46°N			1 of 159 variables selected	Nora Z	Required		148: *Concentration_of_tributy		9		
40 /4				None	required		149: *TBTXSD63 [µg/kg]		9		
200000			Help	OK	Cancel		150: Depth below seabed [m]		9		
44°N			Heip	UK	Cancer	7	151: Depth [m]		9		
					STATIO	N	152: Max_depth [m]		9		
42°N			(6)				153: Proportion dry weight	of particles	9		
			6				154: Concentration of total or		9		
40°N	-		Press '+' to add the data of				155: CLAYZZZZ [mg/kg]	51.30	1		
40 14			the current station to the plot.				156: ONCNCIWF [mg/kg]		9		
		and the second of the second s	D S				157: SANDZZZZ [mg/kg]	24.70	1		
38°N	_		NE				158: SILTZZZZ [mg/kg]	51.30	1		
	-		AC				159: STTMUBS [mg/kg]		9		
36°N							Isosurface Values				
	1 10 100 V	Š	time_ISO8601 [years since 0000-01-01] >	>				ne CAS 83 32 9 per unit dry weig	a		
	5°E	10°E 15°E 20°E						ene_CAS_208_96_8_per_unit_dry_w			
								8601 [years since 0000-01-01]=firs			

- 4. If there are variables in different units, there is a functionality inside ODV to aggregate them performing previously the transformation needed.
- 5. 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.



6. 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 \rightarrow QF=9
- In case of values = 0 for contaminant concentrations → QF=6

Negative values should be carefully inspected:

- in case of negative values (not related to P01 that can have negative values, and different from missing values: -9999, -999, ...) --> QF=4
- in cases of P01 admitting negative values, originator Quality Flagging should be maintained, unless scientific evidence suggests modification

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 (e.g. 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.

8. Dataset output formats

Data collections for marine contaminants are available in two formats:



- as SeaDataNet ODV csv spreadsheet format, which can be easily handled and visualised with ODV Software (More information can be found at <u>www.seadatanet.org/Software/ODV</u>);
- □ as transposed and decomposed csv spreadsheet format.

SeaDataNet ODV csv spreadsheet format

The fundamental data model underlying the format is the spreadsheet: i.e. a collection of rows each having the same fixed number of columns. There are three different types of columns:

- Metadata columns
- • Primary variable data columns (one column for the value plus one for the qualifying flag)
- Data columns

Further details on standard Data Transport Formats are provided here (<u>https://archimer.ifremer.fr/doc/00454/56547/</u>)

Transposed and decomposed csv spreadsheet format

Due to high heterogeneity and complexity of data on marine contaminants, to assist data processing and selection of specific information such as chemical substance, matrix characteristics, biological entity, basis of determination, etc., connected to each single variable, EMODnet Chemistry provides as additional dataset format the "transposed and decomposed" csv spreadsheet format. This dataset format, as the standard ODV csv format, contains metadata and data together.

The differences consist in:

- data on concentrations of chemical substances are transposed, which results in one column containing the parameter names (i.e. P01 codes) and one column with the concentration value
- P01 terms are decomposed into the different subcomponents (namely parameter entity name, parameter statistic, substance name, matrix characteristics, biological entity names, analytical method, ...), when available

The transposed and decomposed format allows users to independently filter and aggregate data according to their specific needs. As some information (eg. station name, bottom depth, water depth, ...) is present both in the metadata and data, some information can be either present twice or available only as data or as metadata. The list of column names with description and example is available in the table below. All vocabularies are available here:

https://vocab.seadatanet.org/search.



EMODnet Chemistry

Quality Control steps for EMODnet Chemistry Contaminant aggregated datasets

Column names	Description	Example
Cruise	Name of the	148
Station*	cruise Code of monitoring station	TZ-8-S
yyyy-mm- ddThh:mm:ss.sss	time	2014-09-27T11:00:00
Longitude [degrees_east]	Longitude	18.741
Latitude [degrees_north]	Latitude	42.474
LOCAL_CDI_ID	LOCAL_CDI_ID	SED148_3/v1
EDMO_code	Originator code	2432
Bot. Depth [m]**	Bottom depth of the station	34
Instrument Info	Information on sampling device	sediment grabs
References	References	https://cdi.seadatanet.org/report/edmo/2432/SED148_3/v1
Datum	Geodetic Datum	World Geodetic System 84 (4326)
Water depth [m]**	Water column depth	33.8
Depth reference	Depth reference	sea level (D08)
Minimum	Minimum	
instrument	instrument	
depth [m]	depth [m]	
Maximum	Maximum	
instrument	instrument	
depth [m]	depth [m]	
Instrument / gear type	Sampling and analytical instrument	sediment grabs (50) inductively-coupled plasma mass spectrometers (LAB15) rulers (LAB29)
Station name*	Station name	TZ-8-S
Originator	Originator	Institute of Marine Biology (IMBK) (2432)
Project name	Project name	Monitoring program of the state of coastal sea ecosystem of Montenegro (12207)
EDMED	EDMED	Monitoring of contaminants in Montenegrin coastal areas 2009-
references	references	2015 (6820)
Access restriction	Access restriction	SeaDataNet licence (LS)
CDI-record id	CDI-record id	2676153



Quality Control steps for EMODnet Chemistry Contaminant aggregated datasets

Sample identifier	Sample	500
C. have a la	identifier	500.4
Subsample identifier	Subsample identifier	500.1
ODV internal	ODV internal	1
		1
sample number	sample number	
Depth below	Depth below	0
seabed [m]	seabed [m]	1
QV:SEADATANET	Quality Value	1
	(according to SEADATANET)	
Value	Value of	270
value		270
	parameter indicated by S06	
QV:SEADATANET	Quality Value	0
QV.SEADATAINET	(according to	0
	SEADATANET)	
Units	Measerurement	ug/kg
Onits	Units	ug/ kg
P01_conceptid	P01 code	CD04ICP3
P01_preflabel	BODC Parameter	Concentration standard deviation of cadmium {Cd CAS 7440-43-
	usage	9} per unit mass of the water body [dissolved plus reactive
	vocabulary	particulate <0.4/0.45um phase] by filtration, acidification and
		inductively-coupled plasma mass spectrometry
S06_preflabel	Parameter entity	Concentration
	names	
S07_preflabel	Parameter	standard deviation
	statistic	
S27_preflabel	BODC	cadmium
	substances	
S27_altlabel	short name	Cd
CAS no	Chemical	7440-43-9
	Abstracts	
	Service number	
S02_preflabel	Where/what	per unit mass of the
	relationships	
S26_preflabel	BODC matrices	water body [dissolved plus reactive particulate <0.4/0.45um
		phase]
S25_preflabel	Biological entity	Dinophysis caudata (ITIS: 9939: WoRMS 109612)
	names	
S03_preflabel	Sample	filtration, acidification
	preparation	
S04_preflabel	Analytical	inductively-coupled plasma mass spectrometry
	•	



S05_preflabel	Data processing	volume computation from mass and mean density
S21_preflabel	Sphere names	water body

* duplicate information, however sometimes not exactly matching

**duplicate information, however sometimes not exactly matching (due to decimal rounding)