Journal of Environmental Protection and Ecology 19, *No 1*, *300–308* (*2018*)

Environmental informatics

EMODnet CHEMISTRY – DATA AGGREGATION AND PRODUCT GENERATIONS IN THE BLACK SEA

L. BUGA[a], L. BOICENCO[a]*, A. GIORGETTI[b], G. SARBU[a], A.SPINU[a]

[a]*National Institute for Marine Research and Development 'Grigore Antipa', 300 Mamaia Blvd., Constanta, Romania*
E-mail: *lboicenco@alpha.rmri.ro*
[b]*OGS – Istituto Nazionale di Oceanografia e di Geofisica Sperimentale – Borgo Grotta Gigante, 42/C – 34010 Sgonico (TS), Italy*

Abstract. In order to unlock fragmented and hidden European marine data resources, to improve European marine data infrastructure, increase the availability of high quality data and assemble them under a common framework and to make these available to individuals and organisations (public and private), Directorate General – Maritime Affairs and Fisheries (DG-MARE) launched in 2009 a new initiative: European Marine Observation and Data Network (EMODnet) as proposed in the EU Green Paper on Future Maritime Policy. In present EMODnet (www.emodnet.eu) provides access to marine data and derived data products from eight thematic portals: bathymetry, geology, sea bed habitats, chemistry, biology, physics, human activity and coastal mapping. Through a stepwise approach, EMODnet Chemistry (www.emodnet-chemistry.eu/) aims to collect, standardise, check the quality of data developing new services to share and visualise information and products at the scale of regions and sub-regions defined by the Marine Strategy Framework Directive. Black Sea is one of the regional seas in EMODnet Chemistry. EMODnet Chemistry provides aggregated and validated Black Sea data collections for nutrients, dissolved gasses, chlorophyll, and contaminants, properly visualised with Open Geospatial Consortium Web Map Services and Web Processing Services (OGC WMS and WPS) viewing services. Black Sea concentration maps with 10-year moving window from 1960 to 2014, by season and for selected vertical layers, are computed and made available.

Keywords: EMODnet, Black Sea, data availability, nutrients, contaminants.

AIMS AND BACKGROUND

In the field of marine research, during the last decades, several oceanographic data management initiatives faced the challenges of data availability, interoperability, and resilience at pan-European level (EU MAST MTP II MATER 1996–1999, EU MAST-INCO MEDAR 1999–2001, FP5 Sea-Search 2002–2005, FP6/FP7 SeaDataNet 2006–2011, 2011–2015. Marine data management communities, developed in the framework of these European initiatives, had to solve the problem of providing access to the huge amount of already existing but fragmented and inaccessible data collected by EU oceanographic institutes.

* For correspondence.

Following the EU Marine Knowledge 2020 agenda[1] and the related roadmap, the European Marine Observation and Data Network (EMODnet) initiative was launched by DG MARE. EMODnet (www.emodnet.eu/) is a long-term marine data initiative developed through a stepwise approach aiming to ensure that European marine data will become easily accessible, interoperable and free of restrictions on use. Its main purpose is to unlock fragmented and hidden marine data resources, make them available to public and private organisations and to facilitate investment in sustainable coastal and offshore activities through improved access to quality-assured, standardised and harmonised marine data and data products which are interoperable and free of restrictions on use. The EMODnet lots with their infrastructure could play a central role specifically for countries where the Regional Sea Conventions are less mature to support the need of qualified and standard information at national, regional, and bigger scales.

Supported by the European Commission, EMODnet Chemistry (www.emodnet-chemistry.eu/) started in 2009 to fulfil the Marine Strategy Framework Directive (MSFD) requirements for the assessment of eutrophication and contaminants[2], following INSPIRE Directive rules[3].

The EMODnet Chemistry 1, pilot project, started in 2009 and was followed by the EMODnet Chemistry 2 project in mid-2013. The consortium for both projects largely consisted of SeaDataNet partners. SeaDataNet (www.seadatanet.org) is the leading infrastructure in Europe for managing, indexing and providing access to ocean and marine data sets and data products, acquired from research cruises and other observational activities in European coastal marine waters, regional seas and the global ocean. The third phase of EMODnet Chemistry started in the beginning of 2017 and involves 45 institutes from 27 countries (20 EU member states) and 3 international organisations (ICES, Black Sea Commission, UNEP/MAP).

The overall objective of the EMODnet Chemistry 2 project was to provide access to marine chemistry data sets and derived data products concerning eutrophication and contaminants for 5 major European sea regions: Baltic, N.E. Atlantic (Celtic Seas, Iberian coast and Bay of Biscay and Macaronesia), Greater North Sea, Mediterranean Sea and Black Sea (Fig. 1). These data products are specifically relevant for Marine Strategy Framework Directive Descriptors 5 (eutrophication), 8 (chemical pollution) and 9 (contaminants in seafood), based on the guidance of the MSFD Common Implementation Strategy.

Fig. 1. Geographic focus of EMODnet Chemistry 2

Black Sea data collection and products on nutrients, chlorophyll, oxygen as well as contaminants, as the result of EMODnet Chemistry 2, will be presented, highlighting the availability of historic and present time data.

EXPERIMENTAL

EMODnet Chemistry is focused on the collection and management of data on some chemical parameters relevant for the MSDF (contaminants and fertilisers) in three matrices (sediment, seawater, and biota). Data managed by EMODnet Chemistry 2 include also silicates, chlorophyll, partial pressures of dissolved gases (oxygen and carbon dioxide), plastics (polyethylene, polypropylene) and acidity (pH, pCO_2, total inorganic carbon, alkalinity).

For each of the 5 regions in EMODnet Chemistry 2 a regional coordinator has been tasked with preparing regional data products: Greater North Sea (including Norwegian Sea and Celtic Sea) (AU-DCE, Denmark), Atlantic Sea (including Atlantic Coast and Macaronesia) (Ifremer, France), Baltic Sea (SMHI, Sweden), Black Sea (NIMRD, Romania), Mediterranean Sea (HCMR/HNODC, Greece).

One of the major challenges for EMODnet Chemistry has been the management of the heterogeneity and complexity of parameters addressed. To illustrate the situation:

- 3 matrices (water, sediment, biota) for 14 groups of variables (such as fertilisers, heavy metals, polychlorinated biphenyls, and others) each having multiple parameters, measurement methods, laboratory methods, instruments used, etc.;

- different data distributions in time and space;
- different organisations leading environmental and research data in the different countries;
- heterogeneous data policy.

Data collection and access. The technical set-up is based on the principle of adopting and adapting the SeaDataNet infrastructure for ocean and marine data which are managed by National Oceanographic Data Centres (NODCs). SeaDataNet is actively involved in the development of standards that follow the INSPIRE implementing rules to ensure interoperability.

Data harvesting is a fundamental activity of EMODnet and it is carried out by the network of NODCs which maintain regular contact with data originators collecting and enriching data with the best set of relevant metadata to ensure the reliability of the information. Within EMODnet Chemistry relevant data sets are managed by tens of data centres. For efficiency, a robot harvesting system that retrieves from connected data centres automatically all data sets within a configured query filter has been developed. The harvested data sets are split over regions and transferred to the regional coordinators as 'raw' input for further processing. However, these data sets are retrieved from tens of data centres and despite common approaches for formats, vocabularies and quality flags there is still a lot of heterogeneity which requires further Quality Control (QC) and harmonisation on a regional scale.

Data quality and validation. The quality of the data is a key issue when merging heterogeneous data coming from different sources, periods, and geographic areas.

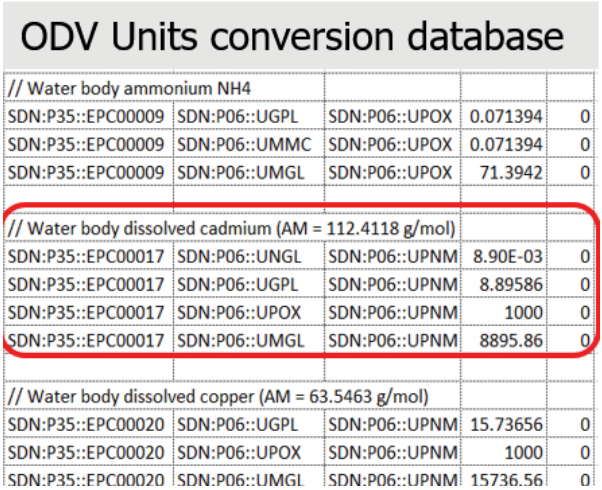
Within EMODnet Chemistry community, commonly agreed and standardised data QC protocols have been defined[4–6] and a data validation loop on format error files and qualification of the data has been implemented (Fig. 2).

As a first step, the data are checked and completed by collators with a standard set of metadata that provide the basic information. The second step, which consists of regional quality control, is performed at regional scale on aggregated datasets. The QC ensure that position and time of data are realistic and compare measurements with broad ranges and specific regional ranges. Whenever available, data are also compared with climatology.

The results of the regional QC are sent to the data collators (NODCs) to correct errors or anomalies in the original copy of the data available in the EMODnet infrastructure. This feedback loop guarantees data quality upgrade (Fig. 2).

For aggregation and harmonisation of parameters and associated units a new controlled vocabulary was initiated in order to aggregates multiple parameters (P01 terms) under unified P35 terms (Fig. 3). The Ocean Data View (ODV) software has developed a built-in aggregation module applying a number of business rules like possible units' conversions[7] (Fig. 2).

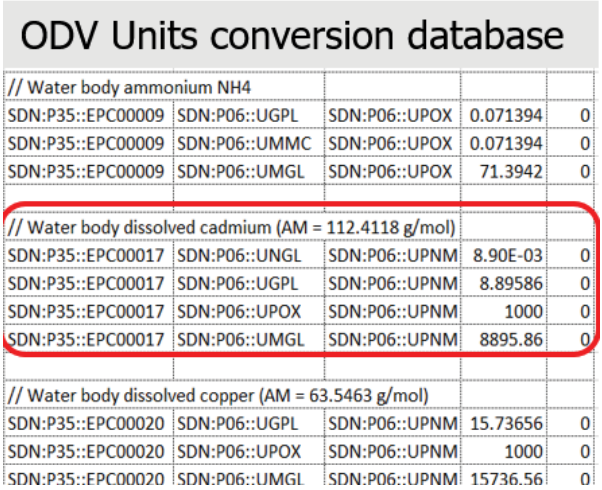
Besides this, the consortium started the collection of quality information 'ex-ante', related to the source laboratories analysis (based on ISO/IEC 17025/2005) (Ref. 6).



ODV Units conversion database

// Water body ammonium NH4				
SDN:P35::EPC00009	SDN:P06::UGPL	SDN:P06::UPOX	0.071394	0
SDN:P35::EPC00009	SDN:P06::UMMC	SDN:P06::UPOX	0.071394	0
SDN:P35::EPC00009	SDN:P06::UMGL	SDN:P06::UPOX	71.3942	0
// Water body dissolved cadmium (AM = 112.4118 g/mol)				
SDN:P35::EPC00017	SDN:P06::UNGL	SDN:P06::UPNM	8.90E-03	0
SDN:P35::EPC00017	SDN:P06::UGPL	SDN:P06::UPNM	8.89586	0
SDN:P35::EPC00017	SDN:P06::UPOX	SDN:P06::UPNM	1000	0
SDN:P35::EPC00017	SDN:P06::UMGL	SDN:P06::UPNM	8895.86	0
// Water body dissolved copper (AM = 63.5463 g/mol)				
SDN:P35::EPC00020	SDN:P06::UGPL	SDN:P06::UPNM	15.73656	0
SDN:P35::EPC00020	SDN:P06::UPOX	SDN:P06::UPNM	1000	0
SDN:P35::EPC00020	SDN:P06::UMGL	SDN:P06::UPNM	15736.56	0

Fig. 2. Data quality and validation loop



ODV Units conversion database

// Water body ammonium NH4				
SDN:P35::EPC00009	SDN:P06::UGPL	SDN:P06::UPOX	0.071394	0
SDN:P35::EPC00009	SDN:P06::UMMC	SDN:P06::UPOX	0.071394	0
SDN:P35::EPC00009	SDN:P06::UMGL	SDN:P06::UPOX	71.3942	0
// Water body dissolved cadmium (AM = 112.4118 g/mol)				
SDN:P35::EPC00017	SDN:P06::UNGL	SDN:P06::UPNM	8.90E-03	0
SDN:P35::EPC00017	SDN:P06::UGPL	SDN:P06::UPNM	8.89586	0
SDN:P35::EPC00017	SDN:P06::UPOX	SDN:P06::UPNM	1000	0
SDN:P35::EPC00017	SDN:P06::UMGL	SDN:P06::UPNM	8895.86	0
// Water body dissolved copper (AM = 63.5463 g/mol)				
SDN:P35::EPC00020	SDN:P06::UGPL	SDN:P06::UPNM	15.73656	0
SDN:P35::EPC00020	SDN:P06::UPOX	SDN:P06::UPNM	1000	0
SDN:P35::EPC00020	SDN:P06::UMGL	SDN:P06::UPNM	15736.56	0

Fig. 3. Business rules as used by ODV for converting units for parameters

Data products. EMODnet Chemistry developed products suitable to visualise the time evolution of a selected group of measurements and to calculate spatially distributed data products specifically relevant for MSFD descriptor 5 (eutrophication), 8 (chemical pollution), and 9 (contaminants in seafood).

Two types of data products are prepared:

- Interpolated maps of specific parameters in time and depth per sea region;
- Graphical time series of specific parameters at point locations

The interpolated maps have been produced with the Variational Inverse Method (VIM) using the software DIVA (Data-Interpolating Variational Analysis)[8]. DIVA is an appropriate numerical implementation of VIM suitable for oceanographic data spatial analysis as it is designed to obtain a gridded field from the availability of non-uniformly distributed observations[9,10] and it gives major benefits above standard interpolations. During the projects several upgraded releases of DIVA have been prepared and delivered, following project requirements.

Interpolated maps are generated, mainly for nutrients, with 10-year overlapping moving window (every map is representing the year of the middle 10-year period) in order to find a balance between the duration of the environmental evaluation cycle for member states (to provide maps with a time frame near to the 6-year process of the member states evaluation) and the number of years that guarantee a sufficient data coverage, and at different standard depths, depending on data availability.

Regional Quality controlled and validated data has been entered into the geodatabase, and consequently being instantly available from the OceanBrowser EMODnet portal (http://ec.oceanbrowser.net/emodnet/). Furthermore, an application that provides additional visualisation services for the aggregated and validated data collections has been developed. The goal is the generation of server-side plots of time series, profiles, time-profiles and DIVA generated maps of selected parameters from data sets of selected stations. The DIVA maps can not only be viewed but also downloaded as Network Common Data Form (Climate and Forecast) (NetCDF (CF)) data files.

RESULTS AND DISCUSSION

National Institute for Marine Research and Development 'Grigore Antipa' – NIMRD, as Regional leader for the Black Sea, received 3 harvested data collections as 'raw' data for further QC and validation and data products: (1) nutrients (included silicates); (2) oxygen, chlorophyll *a*, acidity, and (3) contaminants parameters (hydrocarbons, heavy metals, polychlorinated biphenyls, pesticides and biocides, radionuclides) in water, sediment and biota data collection.

NIMRD applied established standard quality controls to the data which consists in: data format checks, units conversions, broad range control checks to exclude erroneous high values, negatives, handling of LOD, duplication eliminations and comparison of interpolated data with spatially averaged profiles. The parameters aggregation was done automatically by the new enhanced ODV built-in aggregation module. The QC was done following the methodology that was already defined and documented.

All processing steps as well as the methodology followed for the preparation of the aggregated data sets, were documented and sent to originators for confirmation and updating of the local data sets. The errors/problems occurred during the aggregation were corrected (using ODV software, Excel and in house developed scripts) in collaboration with partners.

DIVA analysis and parameters optimisation was done following the project specifications described in the 'Methodology for data QA/QC and DIVA products' document[4]. Based on the temporal and vertical distributions of data, the following 10-years overlapping running horizontal maps have been produced (Table 1).

Table 1. Number of QC controlled and validated data profiles and DIVA products currently available for the Black Sea

Parameter	No of profiles	Years	Depth interval
Phosphate	22728	1960–2013	0–300
Nitrates	8274	1975–2013	0–150
Nitrites	11619	na	na
Nitrate + nitrite	9354	1976–2013	0–150
Ammonium	6935	1976–2013	0–150
Total nitrogen	5393	na	na
Total phosphorus	14563	na	na
Silicate	19024	1960–2013	0–200
Oxygen	23102	1970–2014	0–250
Chlorophyll *a*	1658	1995–2014	0–20
pH	20270	na	na

DIVA analyses are made for ten year moving average. Available depths are IODE standard depths within the interval given in the table (na = no DIVA products produced).

In total, 28 seasonal DIVA products (concentration maps), in SeaDataNet NetCDF format, were prepared and made available for visualisation and downloading at OceanBrowser and uploaded to the OceanBrowser Viewing Service EMODnet Chemistry – Map Server (http://oceanbrowser.net/emodnet/). An example of a DIVA horizontal map for oxygen parameter in Black Sea is given in Fig. 4.

In the case of contaminants, heterogeneity of data is particularly high, as EMODnet Chemistry manages data measured in different matrices (water, sediment, biota), in different sediment size classes, in different phases (dissolved/particulate), in different marine species and target tissues/organs, with different sampling, analytical and normalisation protocols.

A large number of P01 terms for contaminants were present in the Black Sea data collection (more than 220 P01 terms). After aggregation to P35 Vocabulary it was found that, except for antifoulants data, all priority contaminants parameters were present in the Black Sea data collection in water and sediment matrices and in biota (Mytilus) only Mercury data were found (Table 2).

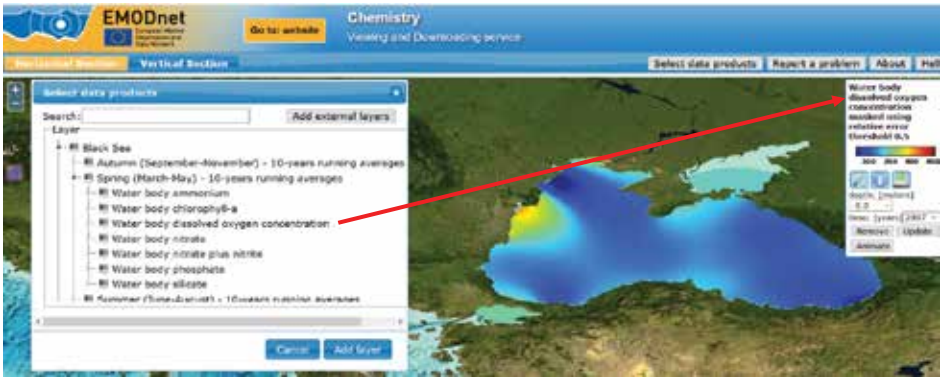


Fig. 4. DIVA horizontal map for oxygen in the Black Sea in spring, in the water surface, decade 2002–2011, centred on year 2007

Table 2. Total number of vertical profiles (CDIs) (after data quality control) for contaminants in the Black Sea

Parameter group	Water	Sediment	Biota
Hydrocarbons	308	654	no data
Heavy metals	3528	4177	138
PCBs	1982	1990	no data
Pesticides	10968	3448	no data
Radionuclides	1517	56	no data

All validated and QC controlled data collections were sent to Central Buffer and used for EMODnet Chemistry – Dynamic Plots (http://ec.oceanbrowser.net/emodnet/).

CONCLUSIONS

The development of the Quality Assurance – Quality Control (QA-QC) procedures and of the unified P35 vocabulary to aggregate variables has been essential to the project.

Work has been spent finding and flagging bad and questionable values, but in general the quality of the Black Sea data has been fairly good. The main issues, and the most time-consuming work, has been to handle format errors in the ODV files. Most of these errors have been obvious and actions have been taken to correct them. This shows that the feedback validation loop is one of the most important steps for improving data quality towards a better harmonised workflow. DIVA maps can also be used as a tool/step in the data qualification.

With the start of EMODnet phase II, DG MARE and DG ENV started a coordination table to agree on a joint process and to identify how EMODnet can best contribute in practical terms to the MSFD. The situation is not homogeneous in EU sea basins. While much of the chemistry and contaminant data are well organised

within OSPAR Commission and Helsinki Convention (HELCOM) (North Sea and Baltic Sea), EMODnet Chemistry has a more useful role in the Black Sea where these outputs are less well organised. Two major Black Sea projects, MISIS and EMBLAS have been using EMODnet Chemistry outputs in their strategic papers[12,13]. A Memorandum of Understanding with the Commission on the Protection of the Black Sea against Pollution (Bucharest Convention) is under preparation to formalise the cooperation in terms of providing dedicated access to EMODnet Chemistry regional products for supporting management of MSFD indicators as well as increasing participation in the advisory groups meetings.

Acknowledgements. This work was supported by DG MARE, Call for Tender 2012/10. Authors acknowledge the contribution of the EMODnet Chemistry partners involved in the technical infrastructure development, in marine chemical data gathering, harmonisation, standardisation and quality control, and in data products preparation, as well as all Black Sea data providers.

REFERENCES

1. European Commission: Green Paper – Marine Knowledge 2020 –from Seabed Mapping to Ocean Forecasting. COM (2012) 473 final.
2. European Commission: Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008, Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Framework Directive). 2008. 22 p.
3. European Commission: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). 2007. 14 p.
4. A. BARTH, L. BUGA, L. FRYBERG et al.: EMODnet Thematic Lot n° 4 – Chemistry -Methodology for Data QA/QC and DIVA Products. 2015.
5. M. LIPIZER, L. BUGA, M. CHATZINAKI et al.: EMODnet Chemistry Rules for Aggregation and Visualization of Data on Contaminants. 2016 (www.emodnet-chemistry.eu/default/products/documents/EMODnet_Chemistry_AggregationRules.pdf).
6. MATTEO VINCI et al.: EMODnet Chemistry 2 Quality Control inventory. 2015.
7. R. LOWRY, A. LEADBETTER, M. VINCI: Semantically-enhanced Aggregation of SeaDataNet Data. In: Proc. of IMDIS 2013 International Conference on Marine Data and Information Systems, 23–25 September 2013, Lucca, Italy, 2013, 49–51.
8. C. TROUPIN et al.: High-resolution Climatology of the North-East Atlantic using Data-interpolating Variational Analysis (Diva). J Geophys Res, **115**, C08005 (2010).
9. A. BARTH, A. ALVERA-AZCÁRATE, C. TROUPIN, M. OUBERDOUS, J.-M. BECKERS: A Web Interface for Griding Arbitrarily Distributed *in situ* Data Based on Data-interpolating Variational Analysis (DIVA). Adv Geosci, **28**, 29 (2010).
10. C. TROUPIN et al.: Generation of Analysis and Consistent Error Fields Using the Data Interpolating Variational Analysis (DIVA). Ocean Model, **52–53**, 90 (2012).
11. M. GOLUMBEANU, A. OROS, M. NENCIU, M. ZAVATARELLI, A. DRAGO: Contribution of Environmental Indices in Meeting the Objectives and Principles of the Marine Strategy Framework Directive (MSFD). J Environ Prot Ecol, **15** (3), 1130 (2014).
12. L. BOICENCO et al: MISIS Deliverable. Black Sea Monitoring Compliance to Marine Strategy Framework Directive (MSFD) – Position Paper. 2014. 40 p.
13. V. MEDINETS et al.: Diagnostic Report II (EMBLAS): Guiding Improvements in the Black Sea Monitoring System. EMBLAS EC-UNDP Project Publications. 2013. http://www.emblasproject.org.

Revised 2 November 2017
Revised 12 December 2017