SeaDataCloud Data Products for the European marginal seas and the Global Ocean

Simona Simoncelli (1), Christine Coatanoan (2), Volodymyr Myroshnychenko (3), Örjan Bäck (4), Helge Sagen (5), Serge Scory (6), Paolo Oliveri (1), Kanwal Shahzadi (7), Nadia Pinardi (7), Alexander Barth (8), Charles Troupin (8), Reiner Schlitzer (9) Michèle Fichaut (2) and Dick Schaap (10)

(1) Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, via Franceschini 31, 40128 Bologna, Italy, simona.simoncelli@ingv.it

(2) Ifremer Centre de Bretagne, Plouzané, Brest, France

(3) METU, Turkey

(4) SMHI, Sweden

(5) IMR, Norway

(6) RBINS, Belgium

(7) University of Bologna, Bologna, Italy

(8) University of Liege, Liege, Belgium

(9) Alfred Wegener Institute, Bremerhaven, Germany

(10) MARIS, The Netherlands

Abstract: Data products, based on in situ temperature and salinity observations from SeaDataNet infrastructure, have been released within the framework of SeaDataCloud (SDC) project. The data from different data providers are integrated and harmonized thanks to standardized quality assurance and quality control methodologies conducted at various stages of the data value chain. The data ingested within SeaDataNet are earlier validated by data providers who assign corresponding quality flags, but a Quality Assurance Strategy has been implemented and progressively refined to guarantee the consistency of the database content and high quality derived products. Two versions of aggregated datasets for the European marginal seas have been published and used to compute regional high resolution climatologies. External datasets, the World Ocean Database from NOAA and the CORA dataset from the Copernicus Marine Service in situ Thematic Assembly Center, have been integrated with SDC data collections to maximize data coverage and minimize the mapping error. The products are available through the SDC catalogue accompanied by Product Information Documents containing the specifications about product's generation, characteristics and usability. Digital Object Identifiers are assigned to products and relative documentation to foster transparency of the production chain, acknowledging all actors involved from data providers to information producers.

Keywords: data value chain, aggregated datasets, climatologies, quality assurance, co-production

1. INTRODUCTION

Data products, based on in situ temperature and salinity observations distributed by the SeaDataNet infrastructure (SDN https://www.seadatanet.org), have been released within the framework of SeaDataCloud (SDC) project (2016-2020). SDN is a distributed Marine Data Infrastructure that connects more than a hundred professional data centers in Europe. The nodes of this Pan-European network are interoperable and provide on-line integrated databases of standardized quality, thanks to the adoption of common standards and the use of common technologies developed within the SeaDataCloud project and its precursors. The data and their full associated metadata description are integrated and harmonized thanks to standardized Quality Assurance and Quality Control (QA/QC) methodologies conducted at various phases of the data value chain.

The data ingested within SDN have been prior quality checked by data providers who submit the data with their corresponding Quality Flags (QF). During the ingestion process all formats and standards are harmonized and checked. The data can thus be accessed by users through the SDN data access service, selected and downloaded for further use, but in order to guarantee the consistency of the database content, the internal Quality Assurance Strategy (QAS) has been implemented. The QAS is an iterative procedure (Simoncelli et al., 2019), which involves many experts at various stages that work in a collaborative framework and, it permits to enhance the overall quality of the database content at each iteration. The QAS starts harvesting all temperature and salinity data and full metadata description contained in SDN system. The files and parameters are then aggregated using Ocean Data View software (ODV, Schlitzer 2002, https://odv.awi.de/) and the obtained ODV collection is then split into regional collections, one per each EU marginal sea. The regional collections are then quality checked by regional experts to verify their completeness and consistency. The experts identify data and metadata anomalies, defined as data flagged as good but that present undetected quality issues, and change the corresponding QFs thanks to ODV functionality. The QAS loop ends with the reporting of the detected anomalies to the corresponding data provider, thanks to the ODV logs which record all adjustments by unique station identifier and the provider's EDMO code (https://www.seadatanet.org/Metadata/EDMO-Organisations) which identify the data originator. The data center inspects the list of data anomalies and applies the proposed corrections if approved, updating the data and metadata in the database.

The purpose of the QAS is also to deliver validated data products, whose quality increases at each QA/QC loop. The SDC team of regional experts has in fact the twofold task of conducting data QC and generating derived data products. Temperature and salinity regional aggregated datasets for the EU sea basins are the first level of data products released from SDN infrastructure. The regional datasets are then used to produce climatologies through DIVAnd mapping tool (Barth et al., 2014) and further develop new data products to serve a diverse user community.

Section 2 presents the aggregated datasets, Section 3 the climatologies and Section 4 introduces the new SDC data products. Products' documentation and access is described in Section 5, together with the main conclusions.

2. Aggregated Datasets for the European Marginal Seas

Two versions of SDC aggregated datasets of temperature and salinity have been released in 2018 and 2020 for the North Atlantic Ocean (NAT), North Sea (NS), Baltic Sea (BAL), Arctic (ARC), Mediterranean Sea (MED) and Black Sea (BLS). The aim was to provide to the users a delay mode quality checked data collections enriched with extensive metadata and characterized by high data quality.

Basic QC steps are applied by visual inspection to: analyze spatial and temporal data distribution and coverage; inspect temperature and salinity data distributions through scatter plots (spikes, outliers); identify stations falling on land or wrong/missing data; and compute statistics about Quality Flags (QF). The checks are conducted per specific areas having similar hydrodynamics, layers (surface, intermediate, bottom), time periods, according to the specific characteristics of each basin. QF assigned by the data centers are modified by the regional products' leaders when/if a data anomaly is detected. Many ODV functionalities have been exploited to further inspect temperature and salinity data, such as the spatial distribution at specific depths or isosurfaces (i.e. potential density anomaly) or to filter data according to the many different metadata. Analysis by instrument type or by data providers are examples that allowed to identify omission (data existing in literature but not publicly shared) within the infrastructures and systematic errors (format, flagging) at the data center level. All the validation results are included in a Product Information Document (PIDoc) annexed to each dataset, which provides also important usability instructions by the experts and acknowledges all data originators.

Table I presents statistics from the two SDC datasets versions. The number of stations and samples increased from version 1 (V1) to version 2 (V2) in all sea regions. The largest percentages of station increase (i.e. ARC, MED, BAL) are due to the availability of underway data, characterized by one station per measurement along the track. The sample statistics in these cases provide the best indicator of SDN database population.

Table I - Summary of stations in the SeaDataCloud regional data collections from V1 to V2 version and the percentage of data increase, in terms of stations or samples. In the NS region only V1 was released.

Product	V1	V2	% increase (stations)	% increase (samples)
ARC	731286	1392366	+90%	+4%
BAL	14038820	14753042		+5%
BLS	137723	162656	+18%	+21%
MED	739784	1003258	+36%	+8%
NAT	9091769	10119755	+11%	
NS	742828			

3. Climatologies for the European Marginal Seas and the Global Ocean

Two versions of SDC climatologies have been released for the EU marginal seas and the global ocean. The first release was designed with a harmonized approach to cover the time period 1955-2017, adopting the World Ocean Atlas (WOA, Garcia et al., 2019) vertical discretization and decadal fields definition ans using WOA for the final validation/consistency analysis. Two major achievements were (1) the adoption of DIVAnd mapping tool (5 over 7 products) and, (2) the integration of external sources of data, such as World Ocean Database (Boyer et al., 2019) from NOAA and Coriolis Ocean Dataset for ReAnalysis (CORA, Szekely et al., 2019) distributed from the CMEMS in situ TAC. The production of the second version aimed at improving the workflow making it more efficient, in particular: to ameliorate QC during the data integration process, tracking external data through unique station identifier in order to report anomalies and duplicates. Efforts have been made to improve the duplicates detection/removal, to optimize DIVAnd parameters, to improve the consistency analysis versus WOA.

A SDC climatology for Global Ocean has been created for the first time and improved with two different time coverages (see Tab. II) using data from the WOD, since the spatial coverage of SDN data at the global scale is still too sparse, but in the future all data sources should be integrated as done in the other regions. A Non-linear Quality Control has been developed and implemented in the global domain (Shahzadi et al. 2021) eliminating less than 15% of input data per month, mainly outliers and non-representative data, not suitable to estimate the large scale climatology. This procedure could be extended and adapted to all regions in the next production cycle.

All V2 climatologies (see details in Tab. II) have been produced with DIVAnd and cover approximately the time period 1955-2018 on monthly basis and also provide seasonal decadal fields.

Table II - Summary of main characteristics of SDC climatologies.

	Horizontal	Time	Season	Month	External

	resolution	coverage	al	ly	data sets
GLO_1	1/4°	1900-2017		x	WOD18
GLO _2	1/4°	2003-2017		x	WOD18
ARC _1	1x1/2°	decades	x	x	WOD18
ARC _2	1x1/2°	1955-2019	x	x	WOD18
BAL_1	1/16x1/32°	decades	x	x	CORA5.2
BAL_2	1/16x1/32°	1955-2018	x	x	CORA5.2
NS_1 (V1)	1/8°	1955-2014		x	WOD18
NS_2 (V1)	1/8°	decades	x		WOD18
NAT _1	1/2°	decades	x	x	CORA5.3
NAT_2 (V1)	1/4°	1955-2017	x	x	CORA5.1
NAT_3 (V1)	1/4°	decades	x	x	CORA5.1
MED_1	1/8°	1955-2018	x	x	CORA5.2
MED_2	1/8°	1955-1984	x	x	CORA5.2
MED_3	1/8°	1985-2018	x	x	CORA5.2
MED_4	1/8°	decades	x		CORA5.2
BLS_1	1/8°	1955-1994	x	x	WOD18 & CORA5.2
BLS_2	1/8°	1995-2019	x	x	WOD18 &CORA5.2
BLS_3	1/8°	1955-2019	x	x	WOD18 & CORA5.2
BLS_4	1/8°	decades	x		WOD18 & CORA5.2

Figure 1 shows the percentage of data from SDC and the additional data integrated from external sources into the climatology input data sets per sea basin. In all sea regions SDC represents the main contributor with the percentage of external data that range from 5% in the NAT region to 37% in the BAL. In fact, the SDC version of the detected duplicate casts have been retained. The external datasets integrated per each region are specified in Tab. II.

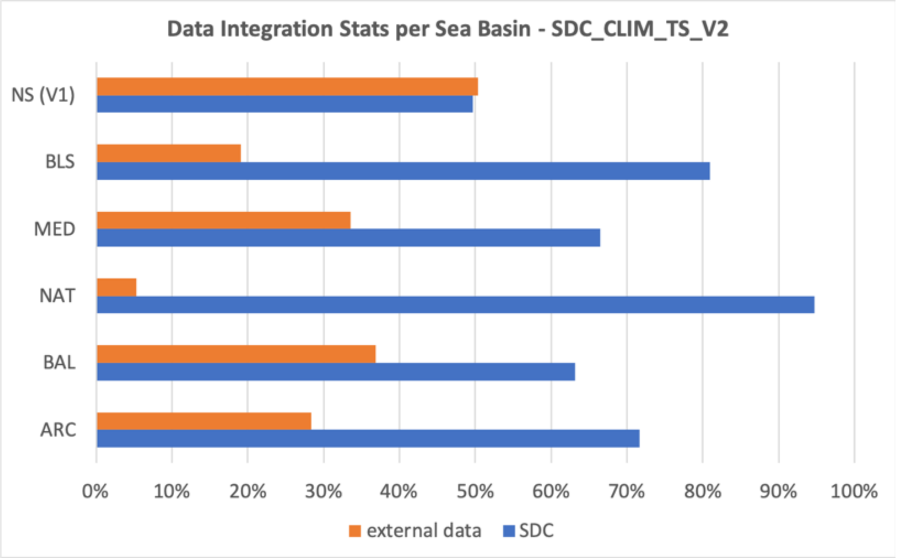


Fig. 1. Percentage of data from SDN infrastructure and external sources in the climatological input datasets.

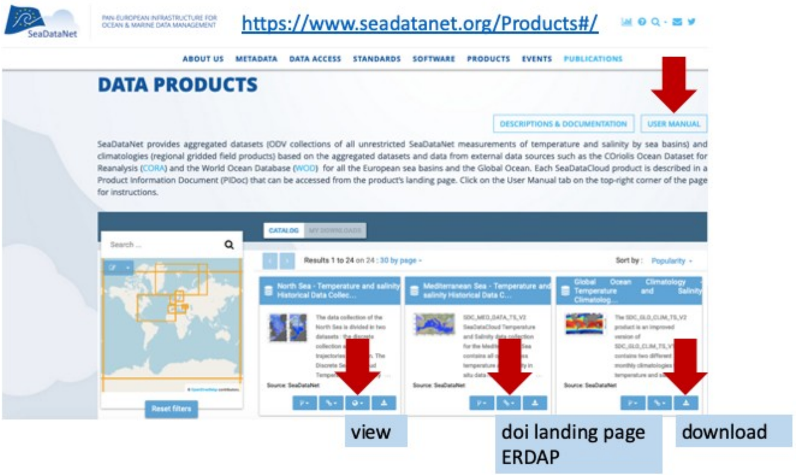
4. New Products

The experts team explored the feasibility of new data products and the capability of SDN infrastructure to release systematically advanced products to monitor the ocean state in view of the Ocean Decade and in line with other initiatives like EMODnet and CMEMS. Eleven new products have been released: three for the Global ocean; three for the Black Sea; two for the Mediterranean Sea, one for the Baltic; one for the North Atlantic and one coastal product. These products mainly apply DIVAnd software to generate advanced products, such as Mixed Layer Depth climatology, Ocean Heat Content estimate, Apparent Oxygen Utilization climatology, coastal currents maps from HF radars (Barth et al., 2021), etc. The Baltic Sea product instead provides temperature and salinity statistics, that could be applied for data QC purposes. All products have been published in SDC catalogue (https://www.seadatanet.org/Products#/) with the relative PIDoc.

5. Conclusions

The SDN data value chain ends with the generation of data products, whose quality reflects the coordination capacity in managing multidisciplinary in situ data but also developing and adopting software/tools through continuous feedback. The analysis of the regional data collections showed a progressive increase of the available data and quality. A novel metadata analysis allowed to monitor the EU data sharing landscape, to detect systematic (format, flagging) errors and data/metadata omissions. SDC climatologies were designed with a harmonized approach to integrate for the first time SDC aggregated datasets with external sources. A SDC global climatology has been created for the first time too.

All products have been published in Sextant catalogue (Figure 2) and have an annexed unified documentation (Product Information Document, PIDoc) describing the methodology applied, the product quality, the usability, acknowledging the data sources and the tools used. All products and PIDocs obtain a persistent Digital Object Identifier - DOIs - for their citation in scientific publications. The products entries in the catalogue (https://www.seadatanet.org/Products#/) are supplied with the web links to the product data files, documentation and visualization tools, as displayed in Figure 2.


SeaDataNet
PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT
https://www.seadatanet.org/Products#/
ABOUT US
METADATA
DATA ACCESS
STANDARDS
SOFTWARE
PRODUCTS
EVENTS
PUBLICATIONS
DATA PRODUCTS
DESCRIPTIONS & DOCUMENTATION
USER MANUAL
CATALOG
MY DOWNLOADS
Search ...
Results 1 to 24 on 24 : 30 by page
Sort by : Popularity
North Sea - Temperature and salinity Historical Data Collec...
Mediterranean Sea - Temperature and salinity Historical Data C...
Global Ocean Climatology Temperature and Salinity Climatolog...
Source: SeaDataNet
Reset filters
view
doi landing page
ERDAP
download

SeaDataNet provides aggregated datasets (ODV collections of all unrestricted SeaDataNet measurements of temperature and salinity by sea basins) and climatologies (regional gridded field products) based on the aggregated datasets and data from external data sources such as the COriolis Ocean Dataset for Reanalysis (CORA) and the World Ocean Database (WOD) for all the European sea basins and the Global Ocean. Each SeaDataCloud product is described in a Product Information Document (PIDoc) that can be accessed from the product's landing page. Click on the User Manual tab on the top-right corner of the page for instructions.

Fig. 2. Screenshot of the SDN products' catalogue with principal indications on how to access the available information.

Acknowledgements

This project has received funding from the European Union Horizon 2020 and Seventh Framework Programmes under grant agreement number 730960 SeaDataCloud.

REFERENCES

Barth, A., Beckers, J.M., Troupin, C., Alvera-Azcárate, A., Vandenbulcke, L., (2014). divand-1.0: n-dimensional variational data analysis for ocean observations. Geoscientific Model Development 7, 225–241. doi:10.5194/gmd-7-225-2014.

Barth, A., Troupin, C., Reyes, E. et al. Variational interpolation of high-frequency radar surface currents using DIVAnd. Ocean Dynamics 71, 293–308 (2021). https://doi.org/10.1007/s10236-020-01432-x

Boyer, T.P., O.K. Baranova, C. Coleman, H.E. Garcia, A. Grodsky, R.A. Locarnini, A.V. Mishonov, C.R. Paver, J.R. Reagan, D. Seidov, I.V. Smolyar, K. Weathers, M.M. Zweng, (2018): World Ocean Database 2018.

A.V. Mishonov, Technical Ed., NOAA Atlas NESDIS 87. https://www.ncei.noaa.gov/sites/default/files/2020-04/wod_intro_0.pdf

Garcia H.E., T.P. Boyer, O.K. Baranova, R.A. Locarnini, A.V. Mishonov, A. Grodsky, C.R. Paver, K.W. Weathers, I.V. Smolyar, J.R. Reagan, D. Seidov, M.M. Zweng (2019). World Ocean Atlas 2018: Product Documentation. A. Mishonov, Technical Editor.

Schlitzer, R. (2002). Interactive analysis and visualization of geoscience data with Ocean Data View. Computers & Geosciences, 28(10), 1211-1218

Shahzadi, K., Pinardi, N. and Lyubartsev, V., 2021, A Non-linear Quality Control Procedure for Representativeness errors in Ocean Historical Datasets, International Conference on Marine Data and Information Systems (IMDIS) 2021. https://imdis.seadatanet.org/files/IMDIS2021_119_abstract.pdf and https://imdis.seadatanet.org/files/IMDIS2021_poster_119.pdf

Simoncelli, S., Fichaut, M., Schaap, D., Schlitzer, R., Barth, A., and Fratianni, C. (2019). “Marine Open Data: a way to stimulate ocean science through EMODnet and SeaDataNet initiatives,” In: INGV Workshop on Marine Environment, Vol. 51, eds L. Sagnotti, L. Beranzoli, C. Caruso, S. Guardato, and S. Simoncelli (Rome), 1126. https://doi.org/10.13127/misc/51

Szekely, T., Gourrion, J., Pouliquen, S., and Reverdin, G.: The CORA 5.2 dataset for global in situ temperature and salinity measurements: data description and validation, *Ocean Science*, 15, 1601–1614, https://doi.org/10.5194/os-15-1601-2019, 2019.