

# SeaDataCloud quality control of data collections

Christine Coatanoan, Institut Français de Recherche pour l'Exploitation de la Mer (France),  
christine.coatanoan@ifremer.fr

Simona Simoncelli, Istituto Nazionale di Geofisica e Vulcanologia (Italy), simona.simoncelli@ingv.it

Volodymyr Myroshnychenko, Middle East Technical University (Turkey), volodymyr@ims.metu.edu.tr

Örjan Bäck, Sveriges Meteorologiska och Hydrologiska Institut (Sweden), orjan.back@smhi.se

Helge Sagen, Institute of Marine Research (Norway), helge@hi.no

Serge Scory, Royal Belgian Institute of Natural Sciences (Belgium), sscory@naturalsciences.be

Reiner Schlitzer, Alfred Wegener Institute (Germany), reiner.schlitzer@awi.de

Michèle Fichaut, Institut Français de Recherche pour l'Exploitation de la Mer (France),  
michele.fichaut@ifremer.fr

Dick Schaap, Mariene Informatie Service 'MARIS' Bv (The Netherlands), dick@maris.nl

During the SeaDataNet II (SDN) EU-project, the Quality Control Strategy (QCS) has been implemented and continuously reviewed aiming at improving the quality of the global dataset and creating the best products. This QCS has also been used for the first aggregated dataset provided in SeaDataCloud (SDC). New regional temperature and salinity data collections covering the time period 1900-2017 have been released within the SeaDataCloud (SDC) project in 2018. A general description of these datasets, their data quality assessment procedure and results are presented.

The specific procedure implemented during SDN II allows assuring and certifying the best quality for the datasets (Fig. 1). After the data harvesting from the central CDI catalogue, QC has been performed at regional levels in a coordinated way, using the ODV software (5.0.0) as common and basic QC analysis tool. In SDC the additional checks have been performed per basin to consider the specific water masses characteristics, per instrument type to investigate data completeness and consistency, per data provider to better identify data anomalies.

This QCS allowed to highlight doubtful data and to organize the data anomalies in lists that have been sent to each concerned data originator together with guidelines explaining the expected corrections. The National Oceanographic Data Centers (NODC) have been asked, on the base of those lists, to check and eventually correct the original data and resubmit them in the SDC dataflow. The iterative procedure has been designed to facilitate the update and improvement of SDC database content.

A detailed description of each regional

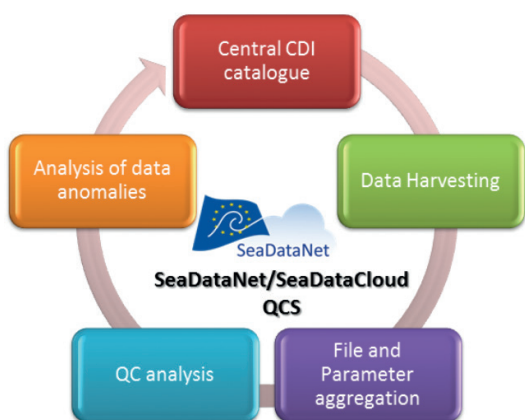


Fig. 1 - Quality Check Strategy implemented during SDN.SDC project.

dataset (Fig. 2) is contained in a Product Information Document (PIDoc): the general products' characteristics (space-time coverage, resolution, format), its quality (validation methodology results) together with experts' recommendations for its usability. ODV qualified dataset collections and PIDocs are available at <https://www.seadatanet.org/Products>.

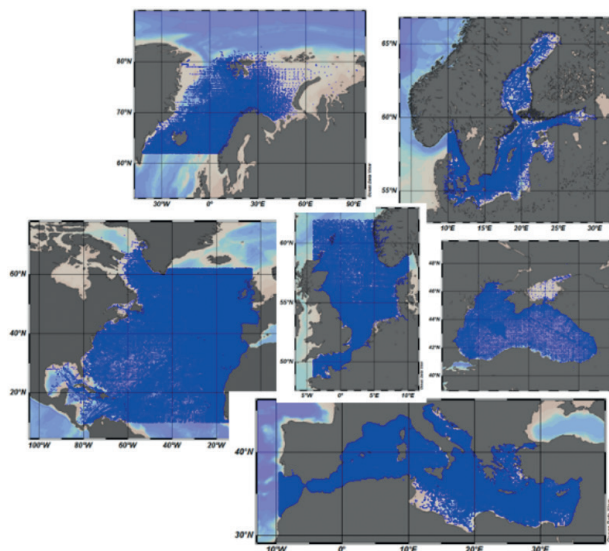


Fig. 2 - Regional data distribution maps of SDC temperature and salinity data collections.

Within SeaDataCloud, the implementation of a cloud environment (Virtual research environment, VRE in Figure 3) aims to optimize and automate the QCS at the central level assuring a continuous monitoring of the database content and its quality. The VRE gives the possibility of generating database snapshots on a regular basis, it facilitates data products versioning and it allows to combine data with subsets from external sources.

The VRE will offer to the users the opportunity to access SDC data and services in the cloud thus providing the possibility of generating their own temperature and salinity data products as well as products for other parameters.

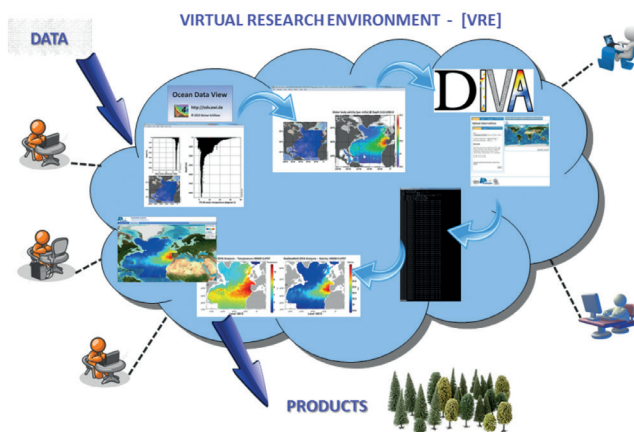


Fig. 3 - Virtual Research Environment.