

SeaDataCloud quality control of data collections

C. Coatanoan¹, S. Simoncelli², V. Myroshnychenko³, O. Bäck⁴, H. Sagen⁵, S. Scory⁶, R. Schlitzer⁷, M. Fichaut¹ and D. Schaap⁸
(1) IFREMER, (2) INGV, (3) METU, (4) SMHI, (5) IMR, (6) RBINS, (7) AWI, (8) MARIS

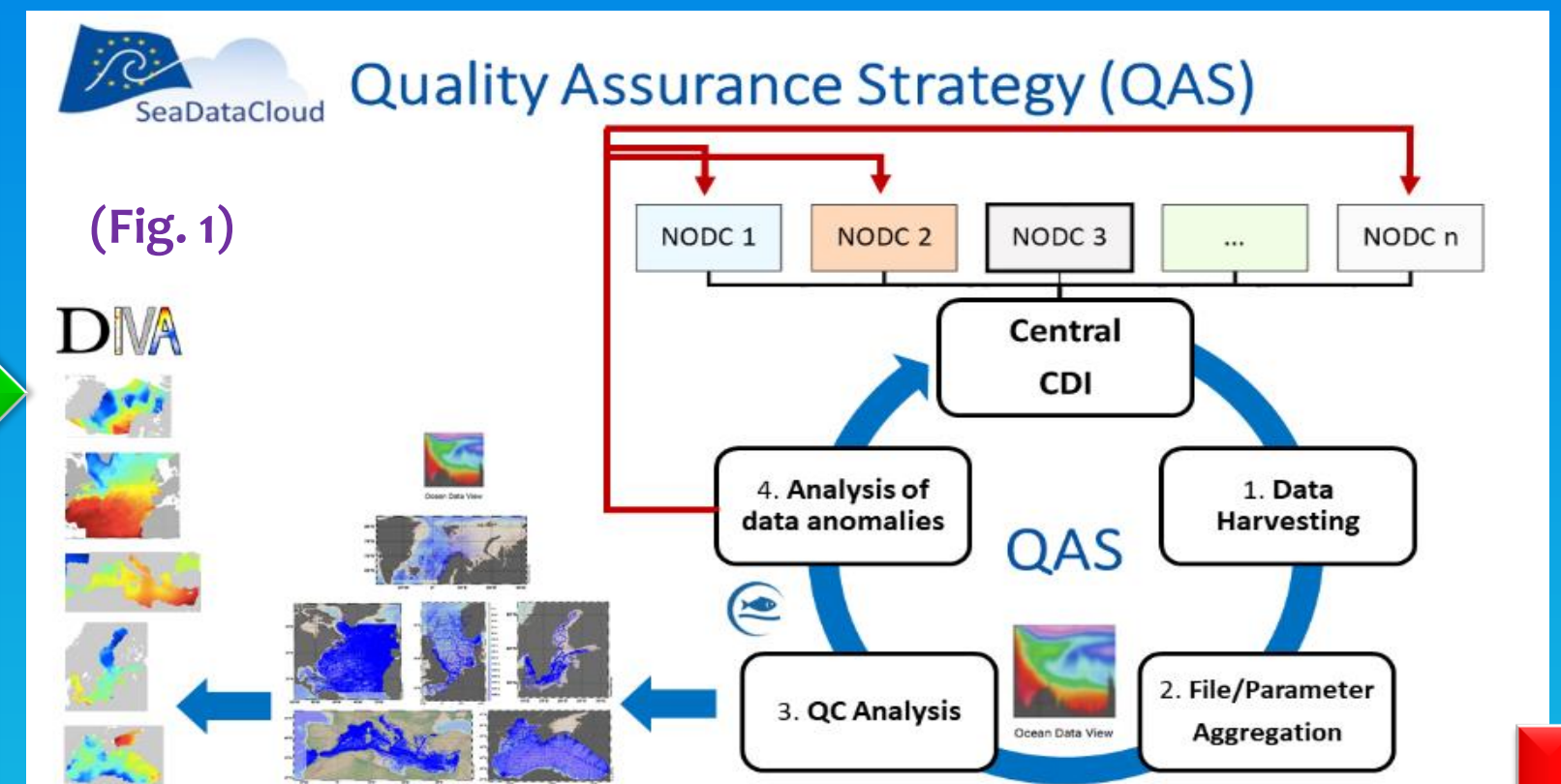
During the SeaDataNet II (SDN) EU-project, a Quality Assurance Strategy (QAS) was implemented, aiming at improving the data quality and creating the best products. The QAS was applied to produce the first aggregated datasets released in SeaDataCloud (SDC) in 2018. It has been continuously reviewed and was upgraded regarding the step for harvesting data for the second version of the aggregated datasets. These last regional temperature and salinity data collections for 6 European sea regions (Arctic Ocean, Baltic Sea, Black Sea, Mediterranean Sea, North Atlantic, and North Sea) covering the time period 1900-2019 (SDN infrastructure content at July 2019) have been released within the SeaDataCloud (SDC) project in 2020.

Quality Assurance Strategy (QAS)

During the SeaDataNet2 activities, the Quality Assurance strategy QAS (Fig. 1) has been developed and continuously refined by WP11 Regional Coordinators in order to improve the SeaDataNet database content and to create the best product deriving from SeaDataNet data (phase 1). The QAS is made by four main phases:

- data harvesting from the central CDI
- file and parameter aggregation
- quality check analysis at regional level
- and analysis and correction of data anomalies.

The QC strategy involved NODCs and data providers that, on the base of data quality assessment outcome, checked and eventually corrected anomalies in the original data. The QC procedure has been designed to be iterative and facilitate the update of SDN database content. The Quality Assurance Strategy was implemented in collaboration with CMEMS in order to implement a true synergy at regional level, create the best historical datasets to serve operational oceanography and climate change communities.



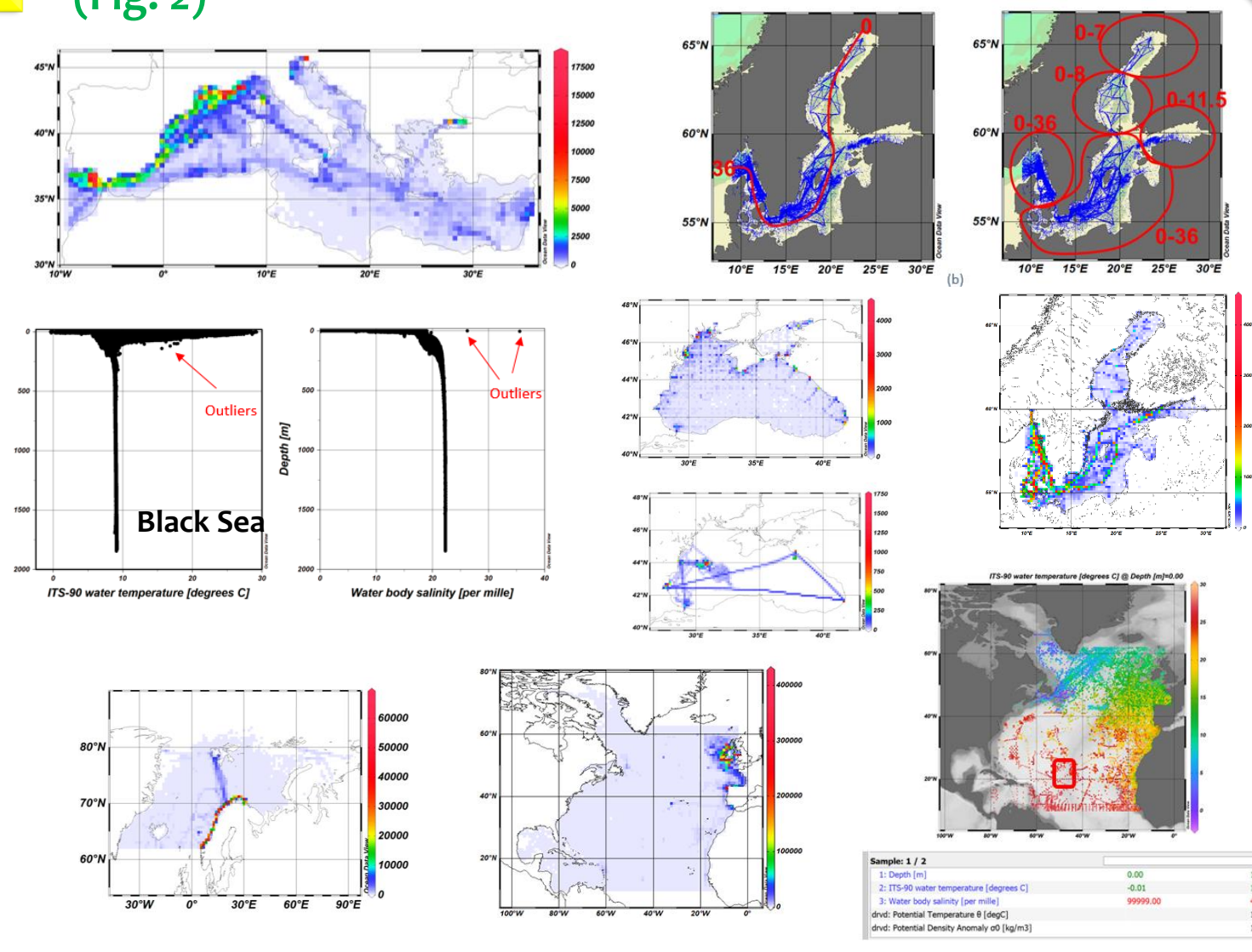
Guidelines and dissemination

The QAS describes how to highlight doubtful data and to organize the data anomalies in lists that are sent to each concerned SDC partners together with guidelines explaining the expected corrections.

A detailed description of each regional dataset is published in a **Product Information Document (PIDoc)**. It includes general product's characteristics (space-time coverage, resolution, format), its quality (description of QC applied, validation methodology results), experts' recommendations for its usability and main changes from previous version.

ODV qualified dataset collections and PIDocs are available at: <https://www.seadatanet.org/Products>

Regional datasets (data density and some examples of quality criteria) (Fig. 2)



Data harvesting for new and updated data and QC

The specific procedure implemented during SDN II allows assuring and certifying high quality of the datasets. Comparing to the first version, the new procedure has been put in place for the step corresponding to data harvesting from the central CDI catalogue. **Only new and updated data have been harvested to focus on this subset of data for the quality control (QC).** This adjustment allowed to skip the quality check of data already validated in the previous QA (Quality Assurance) loop and contained in the preceding data product version. The QC is performed at regional levels in a coordinated way, using the ODV software (5.2.1) 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 (Fig. 2). After performing the quality control on this specific dataset, the second version of the regional data collections consist of a merge between the previous version (delivered in 2018) and the new subset.

