

SeaDataCloud quality control of data collections

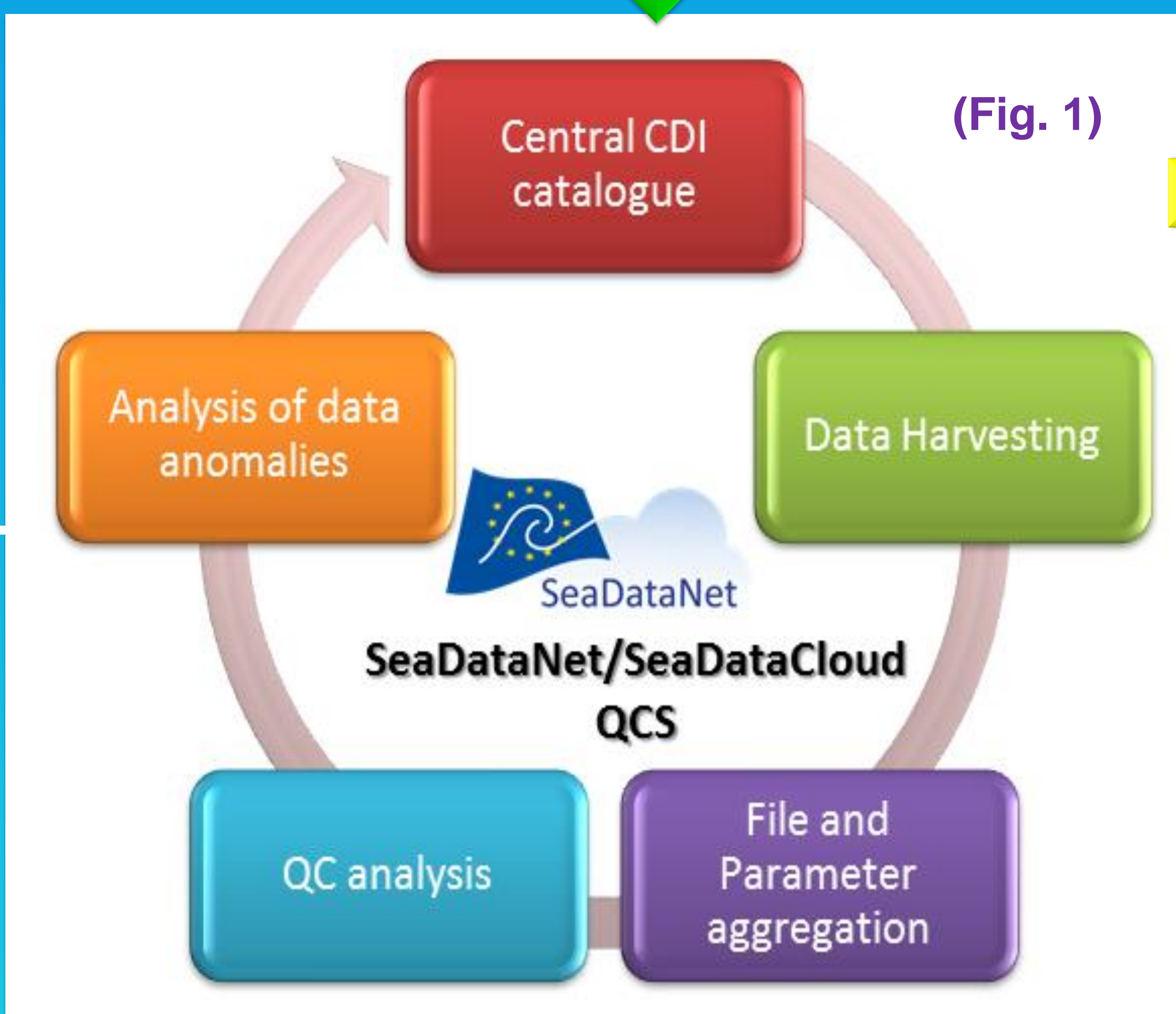
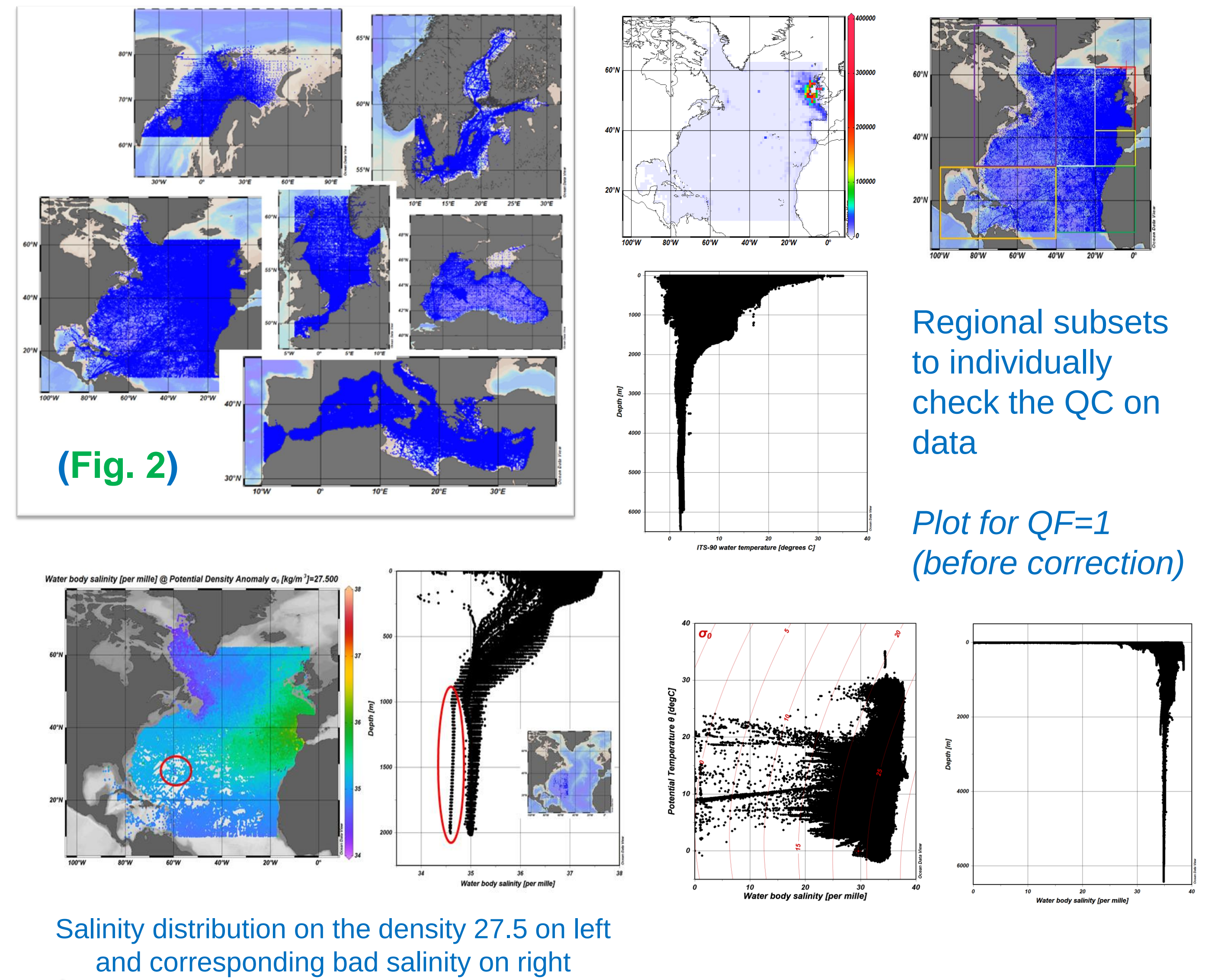
C. Coatanoan¹, S. Simoncelli², V. Myroshnychenko³, O. Bäck⁴, H. Sagen⁵, S. Scory⁶, R. Schlitzer⁷, M. Fichaut¹ and D. Schaap⁸
(1) IFREMER, (2) INGV, (3) METU, (4) SMHI, (5) IMR, (6) RBINS, (7) AWI, (8) MARIS

During the SeaDataNet II (SDN) EU-project, 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.

Quality check strategy (QCS)

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) 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.

Regional datasets (ex of QC for North Atlantic Ocean)



Quality control analysis

The quality control work followed the best practices that were defined during the project SeaDataNet 2:

- Checks of the data coverage, by sub-region when necessary (distribution for T, S, TS couples), by time periods, by layers (distinction between surface, intermediate and bottom layers);
- TS scatter plots of the entire dataset: T versus Z, S versus Z, θS diagram with isopycnal levels for all the QF<3 (check the outliers and change the QF to 4); sometimes the outliers were the missing data values with not appropriate QF;
- By sub-region, scatter plot of observations with QF=1 (good) with a secondary plot showing the density; scatter plot of observations with QF=2 (probably good) with a secondary plot showing the density;
- Scatter plot observations with QF=0 (no quality check): only change the bad data with QF4;
- Identification of stations falling on land; having unreal depth (depth values<0);

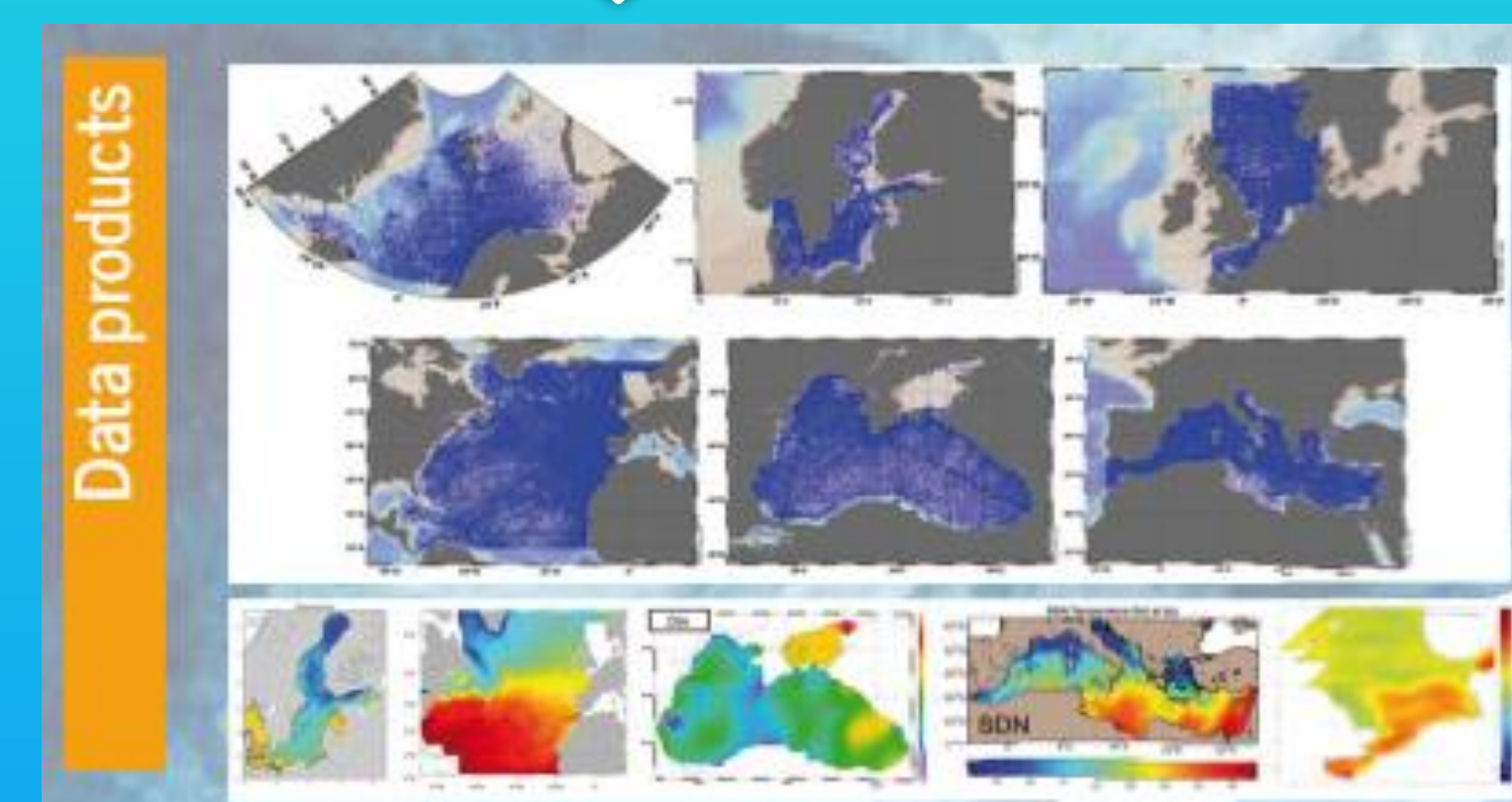
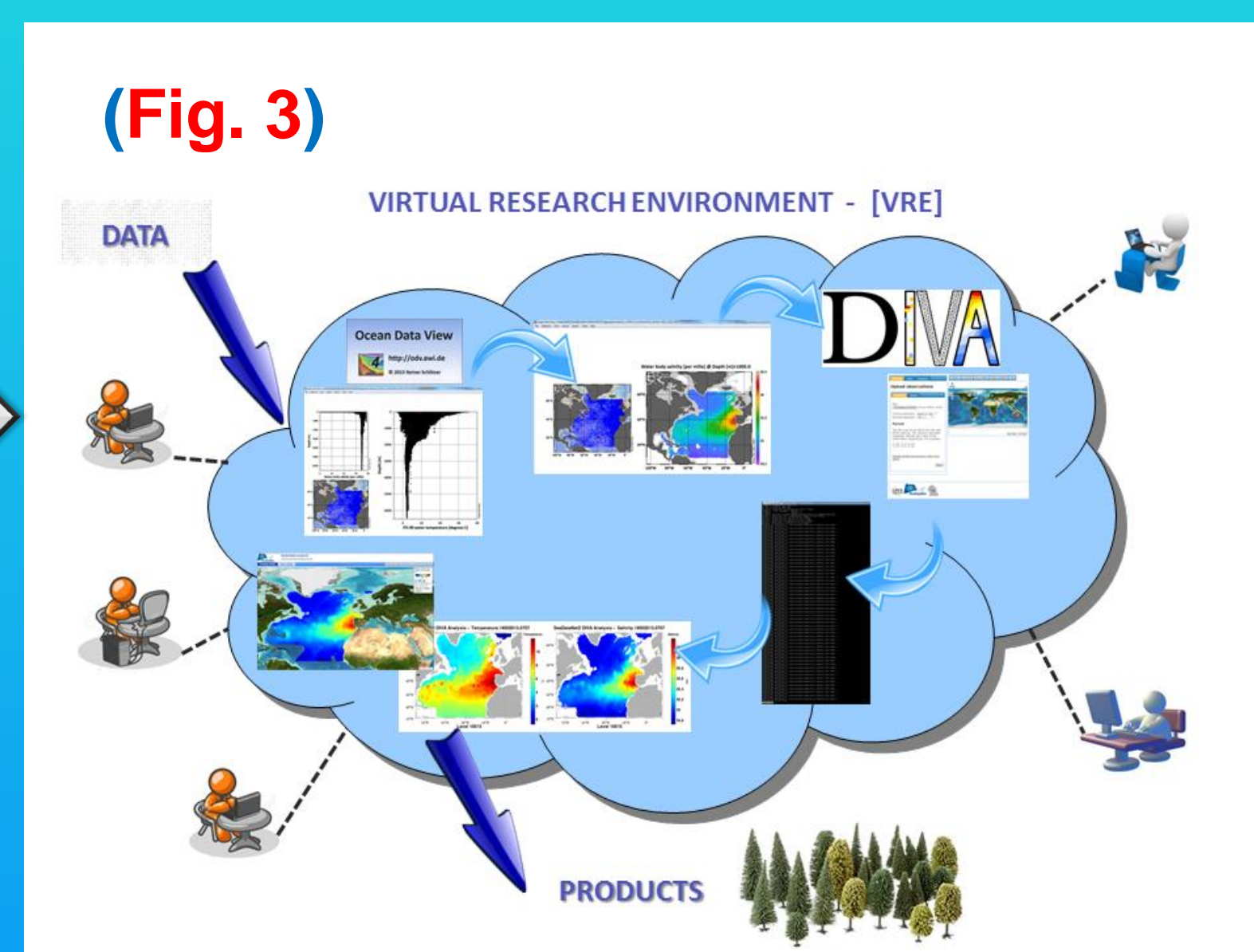
The most useful and powerful quality control used was visual inspection of subsets of data in ODV to discover spikes, outliers, unstable profiles and stations on land. (Fig 2)

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

SeaDataCloud Innovation : Virtual Research Environment

Within SeaDataCloud, the implementation of a cloud environment (Virtual research environment, VRE in Fig. 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.



<https://www.seadatanet.org/Products/>

<https://www.seadatanet.org/>