



SeaDataCloud

SeaDataNet Products, Improvement of QC loop, and feedback to data centers

S. Simoncelli (INGV) and the WP11 TEAM

C. Coatanoan (Ifremer), V. Myroshnychenko (METU), Ö. Bäck (SMHI), H. Sagen (IMR), S. Scory (RBINS), N. Pinardi (UniBO), R. Schlitzer (AWI), A. Barth (ULG)

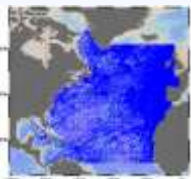
sdn-userdesk@seadatanet.org – www.seadatanet.org

SeaDataCloud Products

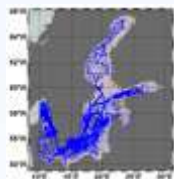
GOAL: to provide the best **data products** from SDN at **regional and global scale** and serve diverse user communities (op.oceanography, climate, marine environment, institutional, academia)

- 1. Aggregated data sets** → historical T and S data harvested from the central CDI and validated by regional leaders

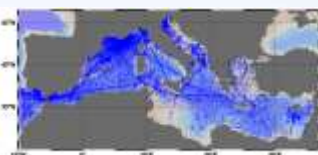
NAT



BAL



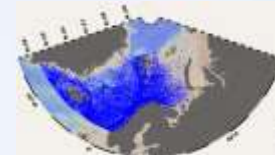
MED



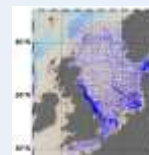
BLS



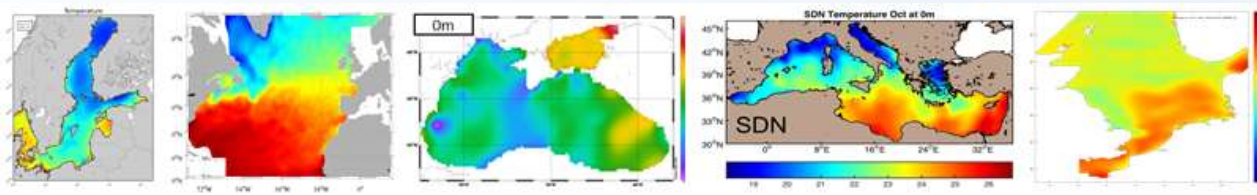
ARC



NS

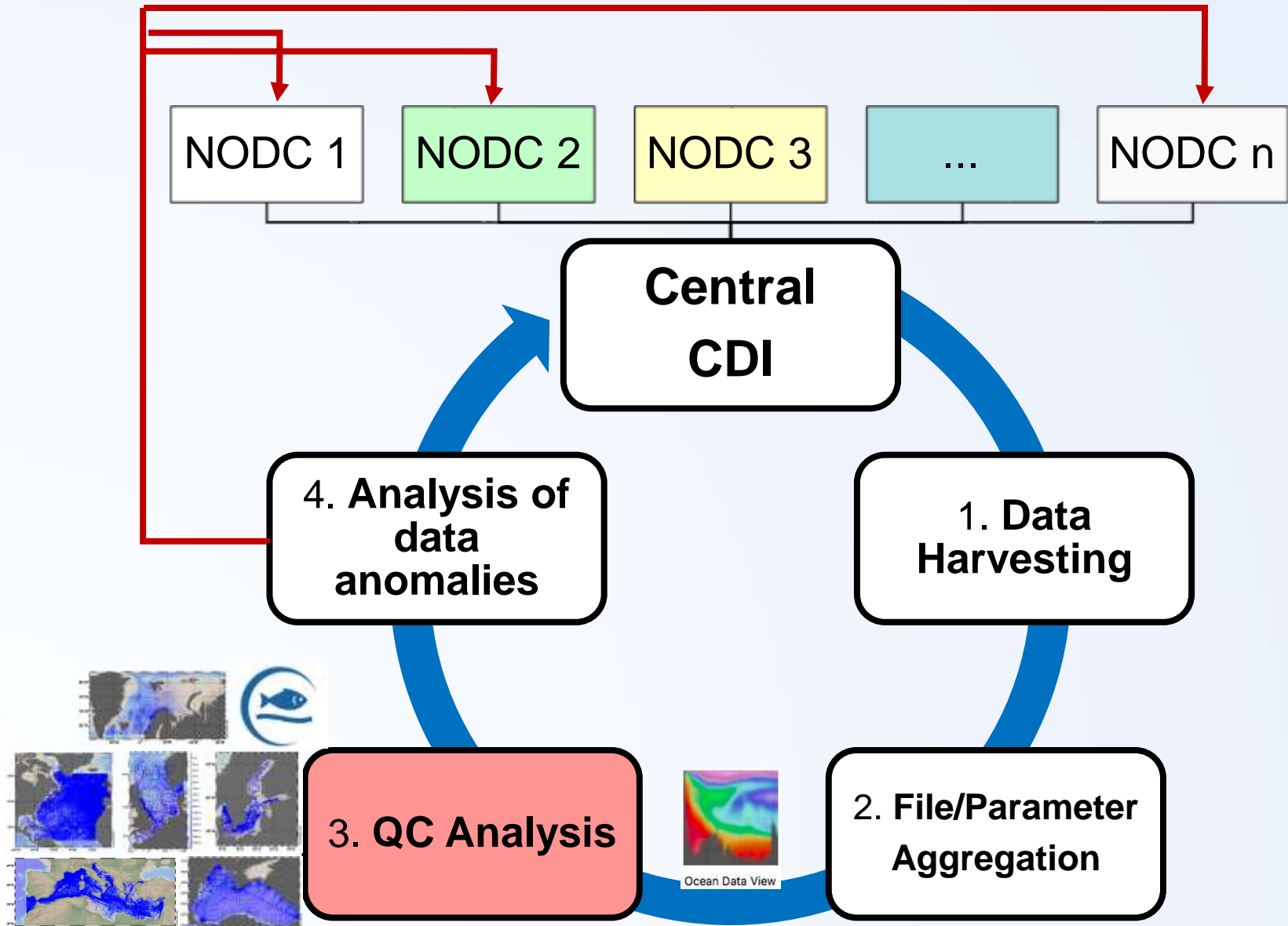


- 2. Climatologies** → gridded fields obtained through DIVA mapping tool and representing the climate of the ocean at **regional and global scale**



- 3. New data products** → multi-platform and multi-disciplinary approach combining in situ and remote sensed observations, Ocean Monitoring Indicators for tracking ocean mechanisms and/or climate modes and trends

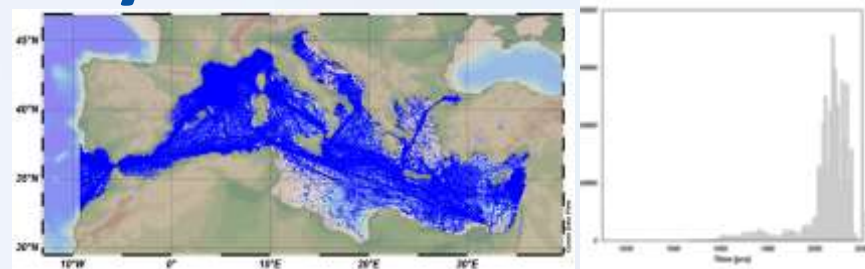
Quality Control Strategy (QCS)



Quality Control Analysis

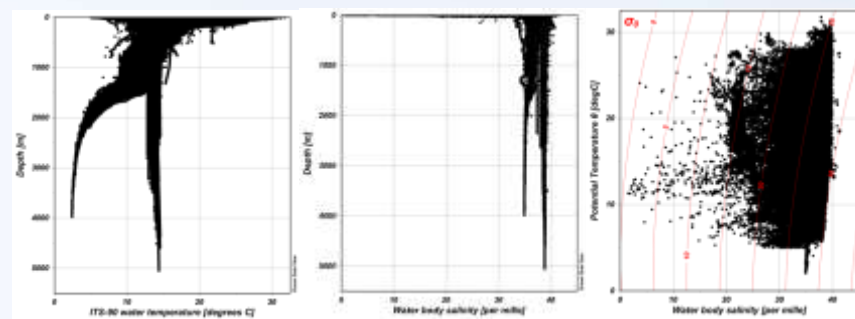
Common guidelines

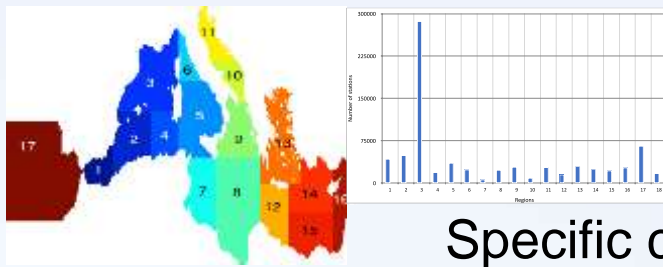
1. Spatial data distribution
2. Temporal data distribution (annual, seasonal and monthly)
3. Quality Flags statistics
4. Scatter plots of good/probably good (QF1/QF2) observations
5. Gross range check
6. Analysis of QF0 data (not checked) to disclose good data
7. Visual check to identify wrong profiles (spikes, outliers)
8. Identification of stations on land
9. Identification of wrong/missing data (time, measurements)
10. Stability check



par	# stations	%	# samples
total	739784		
T	737102	99,6	41223938
S	667232	90,2	28518744
TS	665388	89,9	28119926

%	QF0	QF1-2	Q3-4
dpt	3.0→0	96.9→99.8	0.1
T	2.7→0	97.0→99.8	0.3
S	4.5→0	94.6→99.2	0.9
dpt&T&S	3.0	94,4	0.3

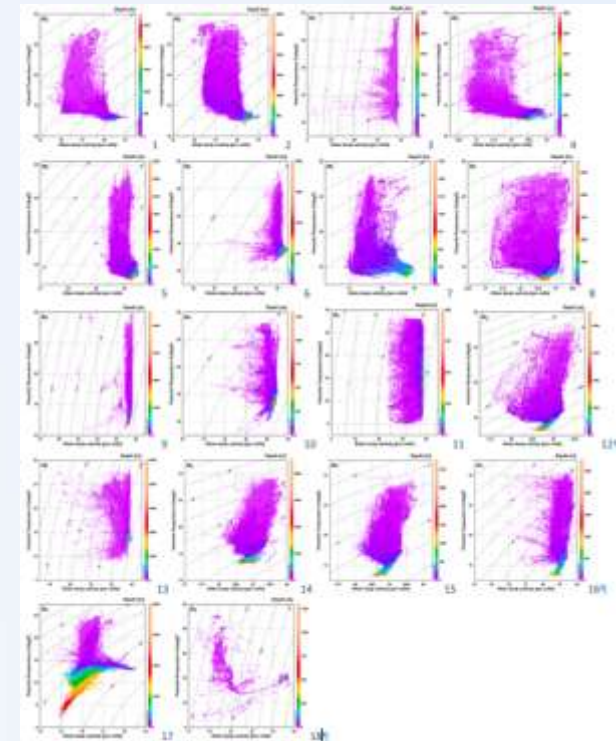
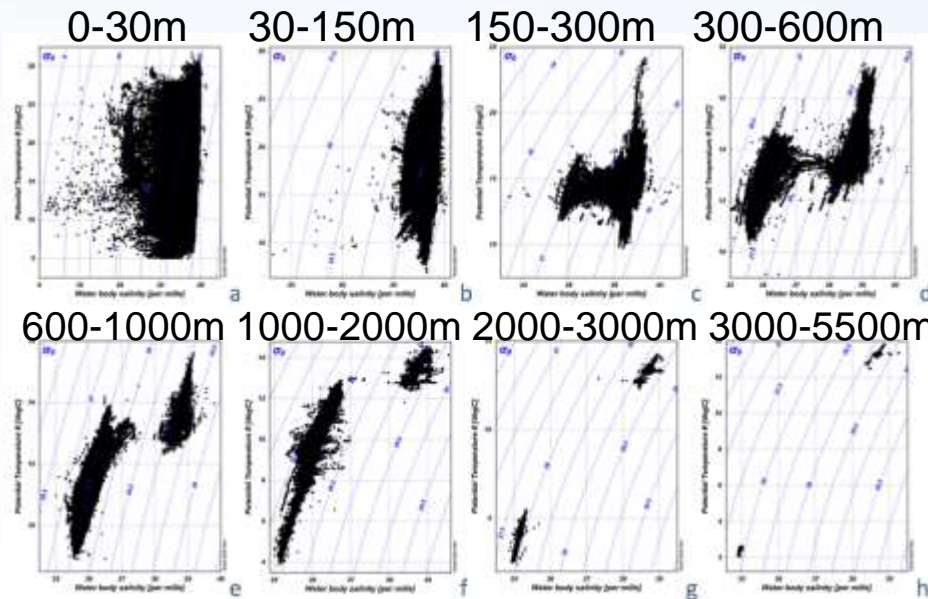
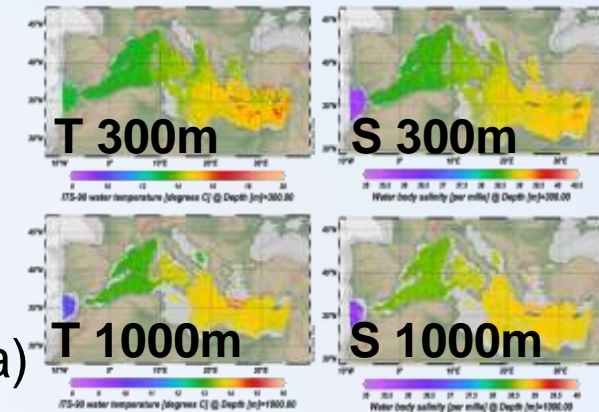




Quality Control Analysis

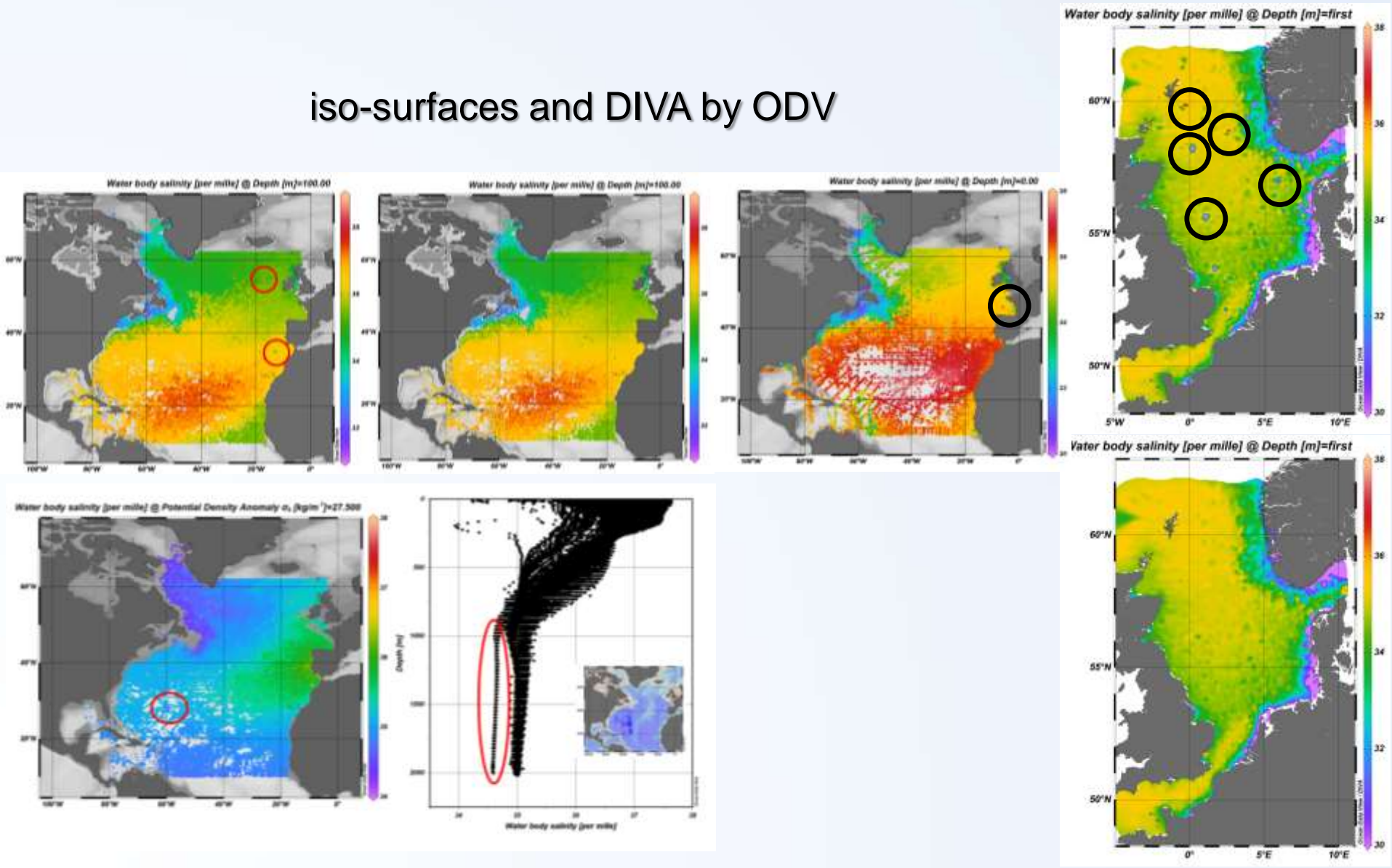
Specific checks per

- **areas** (similar hydrodynamic characteristics)
- **layers** (surface, intermediate, bottom)
- **time periods** (decades or specific periods)
- **Instrument type** (consistency issue of historical data)



Quality Control Analysis

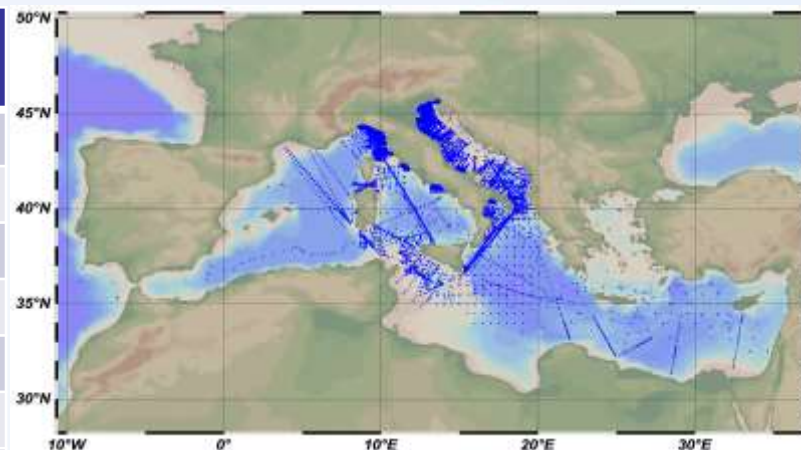
iso-surfaces and DIVA by ODV



Metadata Analysis

- New **data distributors/originators statistics** → QC filtering by EDMO code → detection of systematic (format, flagging) errors
- New **instrument type statistics** → analysis of monitoring space-time coverage → detection of data omissions

Instrument/Gear Type	# stations	%
CTD	52031	7
bathythermograph	56558	8
discrete water sampler	32258	4
thermosalinograph	555269	75
thermistor chains	22	0
continuous water sampler	1577	0
salinity sensor; water temperature sensor	19852	3
salinometers	100	0
salinity sensor	143	0
water temperature sensor	1	0
none info	21973	3

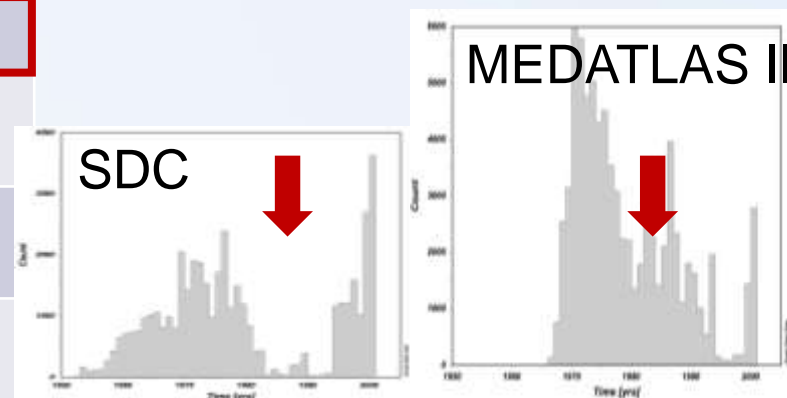


Metadata Analysis

- New **data distributors/originators statistics** → QC filtering by EDMO code → detection of systematic (format, flagging) errors
- New **instrument type statistics** → analysis of monitoring space-time coverage → detection of data omissions

Instrument Info	Probe	# st	%
"		26162	46
'SDN:P01::ADEPZZ01 SDN:L22::TOOL0262 SDN:L33::011 SDN:P01::TEMPET01 SDN:L22::TOOL0262'	T-5 XBT	1239	2
'SDN:P01::ADEPZZ01 SDN:L22::TOOL0263 SDN:L33::041 SDN:P01::TEMPET01 SDN:L22::TOOL0263'	T-7 XBT	9995	18
'SDN:P01::ADEPZZ01 SDN:L22::TOOL0263 SDN:L33::041 SDN:P01::TEMPPR01 SDN:L22::TOOL0263'	T-7 XBT	14	0
'SDN:P01::ADEPZZ01 SDN:L22::TOOL0435 SDN:L33::001 SDN:P01::TEMPET01 SDN:L22::TOOL0435'	T-4 XBT	16732	30
'SDN:P01::ADEPZZ01 SDN:L22::TOOL0592 SDN:L33::710 SDN:P01::TEMPET01 SDN:L22::TOOL0592'	XCTD-2	6	0
'SDN:P01::ADEPZZ01 SDN:L22::TOOL0718 SDN:L33::061 SDN:P01::TEMPET01 SDN:L22::TOOL0718'	T-10 XBT	2126	4

MED XBTs 1950-2000



Data Omission
**Actions to ingest missing
 XBTs**

**Actions to complete
 metadata information**

Data and Metadata Omissions

Filling data and metadata omissions highly improves the quality of the infrastructure content and increases users' confidence

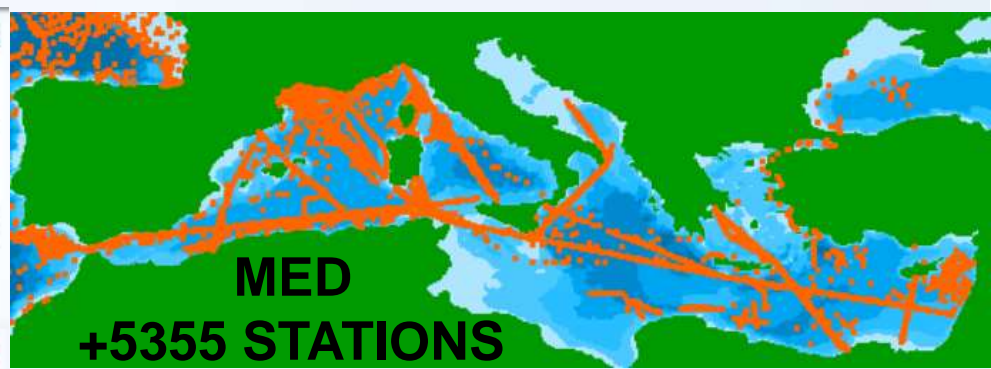
Reprocessing at the data center level

→ to secure and preserve crucial historical info (international approach of preserving the original data)

→ to apply the latest XBT bias correction and further use the data consistently with other data types

Implemented Actions

- Ifremer checked and added additional available metadata and ~91k French XBTs
- XBT data providers were asked to ingest missing metadata (new terms in L05 to distinguish XBT/MBT, probe type L22, fall rate eq L33)



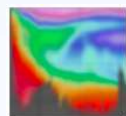
Management of Data Anomalies

Data centers:

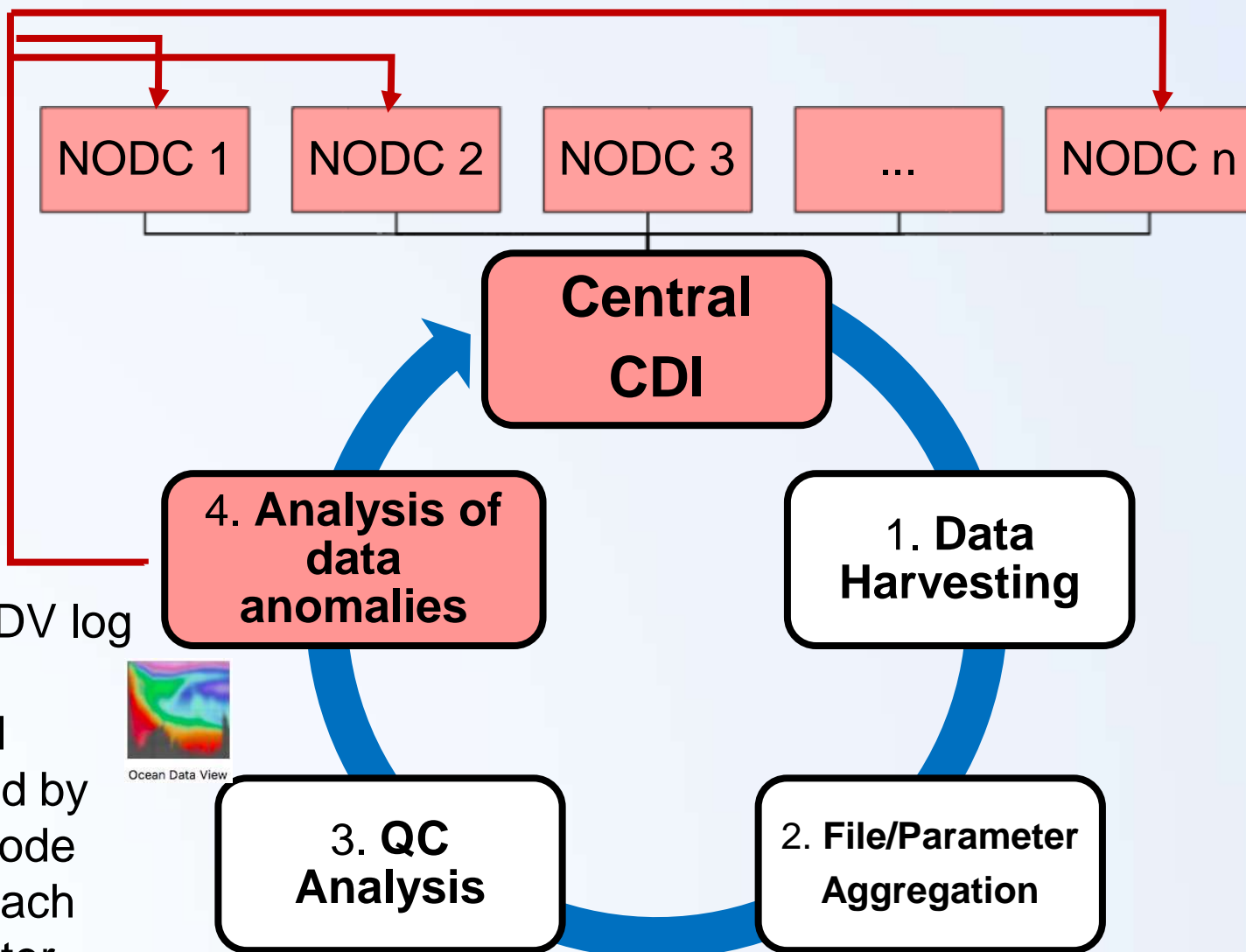
- check the anomalies
- can apply corrections or reject
- update the CDI
- report

Regional ODV log files are:

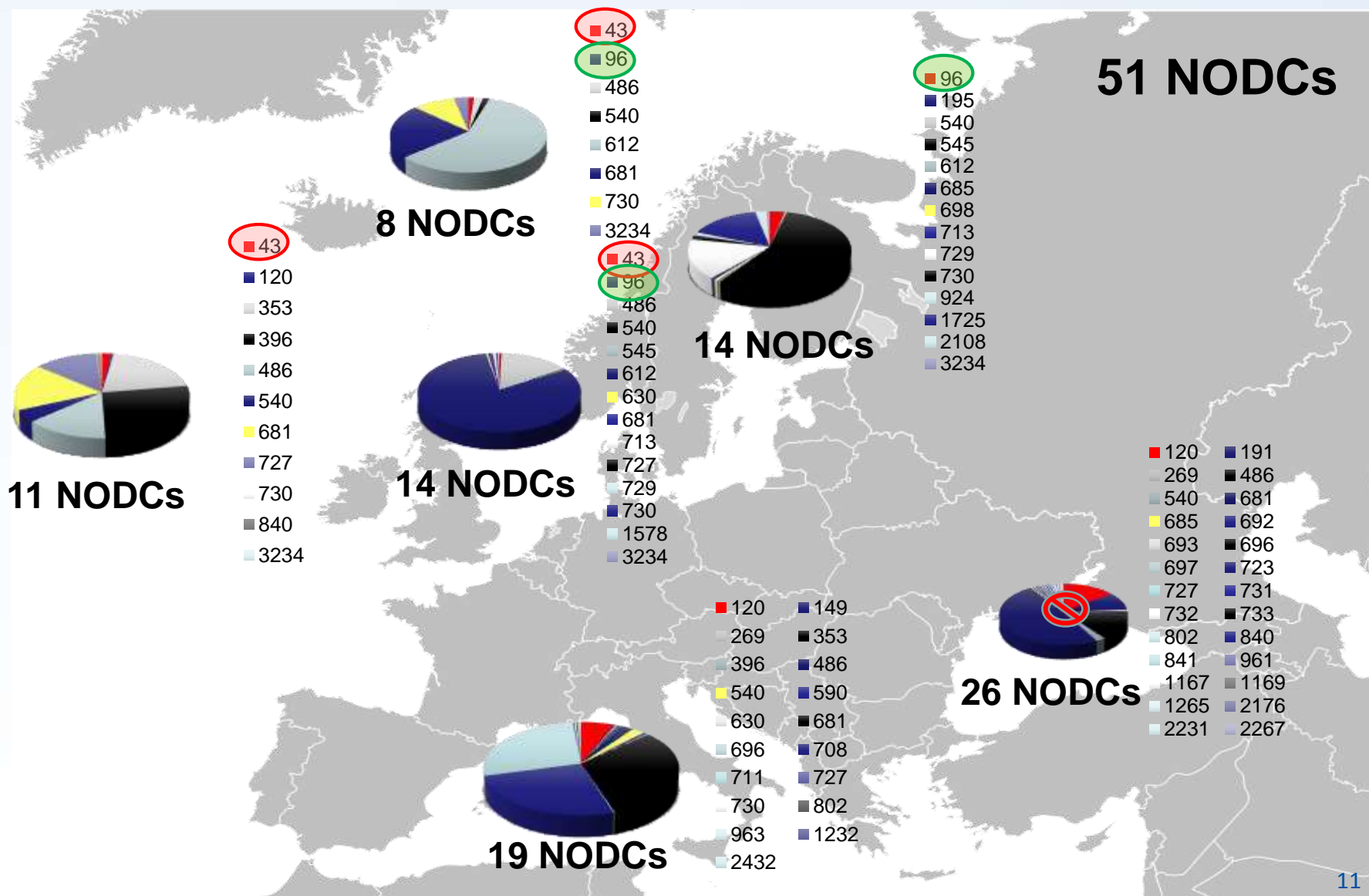
- collected
- organized by EDMO code
- sent to each data center



Ocean Data View



Analysis of Data Anomalies



Analysis of Data Anomalies

First message: Jul 10 – 1st Reminder: Oct 10 – 2nd Reminder: Oct 31 + Stress at the Annual Meeting!

List of anomalies with LOCAL_CDI_ID, EDMO_CODE, PARAMETER_LEVEL, and OLD_QF NEW_QF + Guidelines (instructions about what to do and what is expected)

What is expected? → Deadline end of November

- NODC comments about anomalies and send to MARIS the updated CDI XML
- A report with informations (list of parameters in errors, details on why corrections have not been taken into consideration, etc.....)

Feedback

70% of answers with ~39% of work already processed (on answers)

Results: database progresses

Cruises			Stations			Samples (data)		
SDN2_V2	SDC_V1	±%	SDN2_V2	SDC_V1	±%	SDN2_V2	SDC_V1	±%
Mediterranean Sea								
			212887	734957	+245,%	26625173	42294299	+59%
Black Sea								
1723	2284	+32.6%	96487	137723	+43%	2696215	4240346	+57%
Arctic Sea								
1075	1956	+82%	266291	731286	+175%	19681474	24203161	+23%
Baltic Sea								
						11100238	13780801	+24%
North Atlantic								
			1807266	9091773	+403%			
North Sea								
			115 596	162 452	+41%	6670529	7817193	+17%



Product Information Documents (PIDoc)

PIDocs contain all specifications and descriptions of:

- Product's characteristics (format, space-time coverage, resolution)
- Quality (validation methodology and results)
- Product's usability
- **data distributors and data originators list** (add statistics)
- instrument type statistics

PIDocs have DOI and are available through the product landing page

→ **Big effort to produce/revise and publish PIDocs**

→ **Major improvement to increase user confidence and products uptake**

- Data population statistics per sea basin show a progressive increase of available data
- Data quality also improved thanks to the introduction of additional checks by regional experts (sub-regions, depth layers, iso-surfaces)
- QF statistics after QC present very high percentages of good data (QF1,2) ~99% MED; 98-99% BLS; ~99% ARC; ~99% BAL; 98-99% NS 96(S)-99% NAT
- **metadata statistics** about data distributors/originators highlights systematic (format, flagging) errors and allows the monitoring the EU data sharing landscape but also fair acknowledgment to providers
- **instrument type statistics** highlight omissions and suggests the need of systematic check and data reprocessing at the data centers level

SDC_DATA_TS_V1 release

Datasets released in June but the publication on the web catalogue has been finalized last week

ABOUT US

METADATA

DATA ACCESS

STANDARDS

SOFTWARE

PRODUCTS

EVENTS

PUBLICATIONS

PRODUCTS

SeaDataNet provides aggregated datasets (ODV collections of all SeaDataNet measurements of temperature and salinity by sea basins) and climatologies (regional gridded field products based on the aggregated datasets) for all the European sea basins.

[Read more](#)

Aggregated datasets

Climatologies

Documentation

**SDC_DATA_TS_V1
EU basins (DOI)
and
Product Information
Documents
(DOI)**

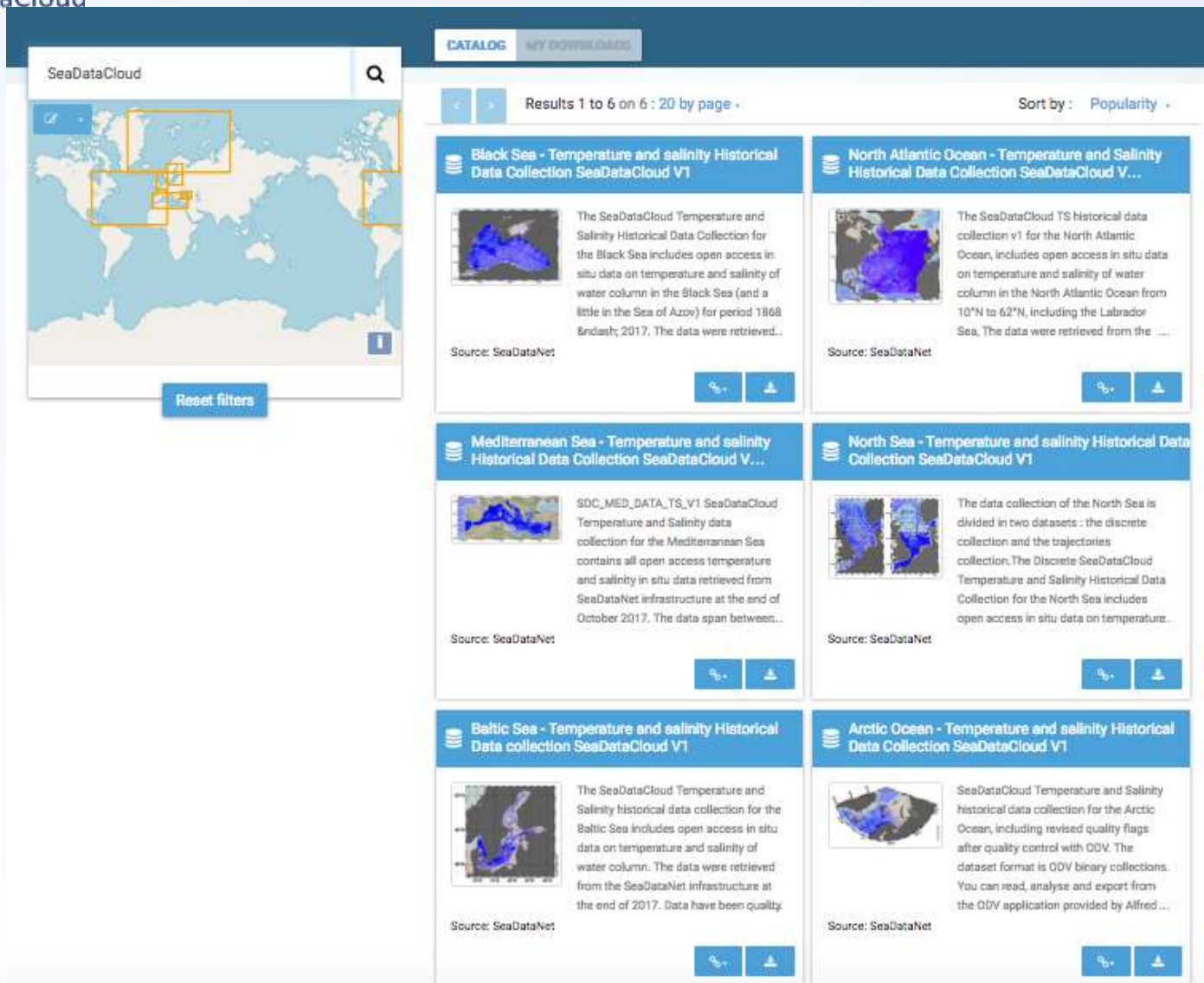
**gridded
climatologies
(DOI)**

First aggregated dataset V1.1
Regional climatologies V1.1
Second aggregated dataset V2

catalogue

sdn-userdesk@seadatanet.org – www.seadatanet.org

Products Catalogue



The screenshot displays the SeaDataCloud Products Catalogue interface. On the left, a map of the world shows the location of the data collection area. The main content area is divided into two columns, each displaying a list of data products. The products are sorted by popularity, and the results are shown on page 1 of 6, with 20 items per page.

Products Catalogue

Results 1 to 6 on 6 : 20 by page

Sort by : Popularity

Black Sea - Temperature and salinity Historical Data Collection SeaDataCloud V1

The SeaDataCloud Temperature and Salinity Historical Data Collection for the Black Sea includes open access in situ data on temperature and salinity of water column in the Black Sea (and a little in the Sea of Azov) for period 1868 – 2017. The data were retrieved...

Source: SeaDataNet

North Atlantic Ocean - Temperature and Salinity Historical Data Collection SeaDataCloud V...

The SeaDataCloud TS historical data collection v1 for the North Atlantic Ocean, includes open access in situ data on temperature and salinity of water column in the North Atlantic Ocean from 10°N to 62°N, including the Labrador Sea. The data were retrieved from the ...

Source: SeaDataNet

Mediterranean Sea - Temperature and salinity Historical Data Collection SeaDataCloud V...

SDC_MED_DATA_TS_V1 SeaDataCloud Temperature and Salinity data collection for the Mediterranean Sea contains all open access temperature and salinity in situ data retrieved from SeaDataNet infrastructure at the end of October 2017. The data span between...

Source: SeaDataNet

North Sea - Temperature and salinity Historical Data Collection SeaDataCloud V1

The data collection of the North Sea is divided in two datasets: the discrete collection and the trajectories collection. The Discrete SeaDataCloud Temperature and Salinity Historical Data Collection for the North Sea includes open access in situ data on temperature...

Source: SeaDataNet

Baltic Sea - Temperature and salinity Historical Data collection SeaDataCloud V1


The SeaDataCloud Temperature and Salinity historical data collection for the Baltic Sea includes open access in situ data on temperature and salinity of water column. The data were retrieved from the SeaDataNet infrastructure at the end of 2017. Data have been quality...


Source: SeaDataNet

Arctic Ocean - Temperature and salinity Historical Data Collection SeaDataCloud V1

SeaDataCloud Temperature and Salinity historical data collection for the Arctic Ocean, including revised quality flags after quality control with ODV. The dataset format is ODV binary collections. You can read, analyse and export from the ODV application provided by Alfred...

Source: SeaDataNet


ARCHIMER
French Institutional Repository


FR

SeaDataCloud Temperature and Salinity Historical Data Collection for the Mediterranean Sea (Version 1)

[View](#)
[Download this PDF](#)
[PDF](#)

[Download metadata](#)
[DOI: 10.5281/zenodo.102102](#)

[Project: GEMMUS-00000](#)
[Dataset: Local](#)

[Related resources](#)
[Available in the French National Information System for the GEMMUS-00000](#)
[Archival report: Sea Temperature and Salinity Historical Data Collection \(Version 1\) - 10/04/2017](#)

[Share](#)
[M](#)
[T](#)
[W](#)
[F](#)
[D](#)

[Full text](#)

File	Pages	Size	Access
Author's official version	42	1.1 MB	Download

[How to cite](#)

[Benoît Stévenin, MyrtheGénérique Hingray, Gauthier Deshayes \(2017\) 'SeaDataCloud Temperature and Salinity Historical Data Collection for the Mediterranean Sea \(Version 1\)' Individual Information Document \(IIDM\), Institut Français de Recherche pour l'Exploitation de la Mer](#)

How To Cite

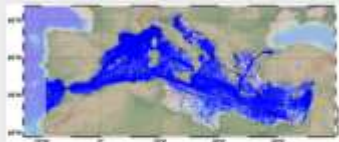
Landing Page

PIDoc

PIDoc

How To Cite





Mediterranean Sea - Temperature and salinity Historical Data Collection SeaDataCloud V1



SDC_MED_DATA_TS_V1 SeaDataCloud Temperature and Salinity data collection for the Mediterranean Sea contains all open access temperature and salinity in situ data retrieved from SeaDataNet infrastructure at the end of October 2017. The data span between -9.25 and 37 degrees of longitude, thus including an Atlantic box and the Marmara Sea. It covers the time period 1900-2017. Data have been quality checked using ODV 5.0 software. Quality Flags of anomalous data have been revised using basic QC procedures. The dataset format is ODV binary collections...

Source: SeaDataNet



-  SDC_MED_DATA_TS_V1
-  Water_body_salinity
-  ITS-90_water_temperature
-  Add all 3 layers to the map



HORIZONTAL SECTION VERTICAL SECTION

CONTACT ABOUT HELP



Visualization

Restricted data analysis

PIDocs are available
Metadata stats
→to monitor continuously
restricted data and their
providers



GOAL
unlock old data and
avoid stagnation

Restricted/Unrestricted ratio

	STATIONS		
	unrestricted	restricted	%
NAT	9091773	23217	0,3
BAL		10221	1,0
MED	734957	33022	4,5
BLS	137723	10528	7,6
ARC	731286	374	0,1
NS	1385289	12011	0,9

Unlock restricted data before **2010**? **2014**?

BLS

Black Sea Restricted Stations

7541 (72%)
10067 (96%)

10522
stations
from INSTM
with missing
depth (QF9)

MED

20640 (62%)
22032 (67%)

BAL

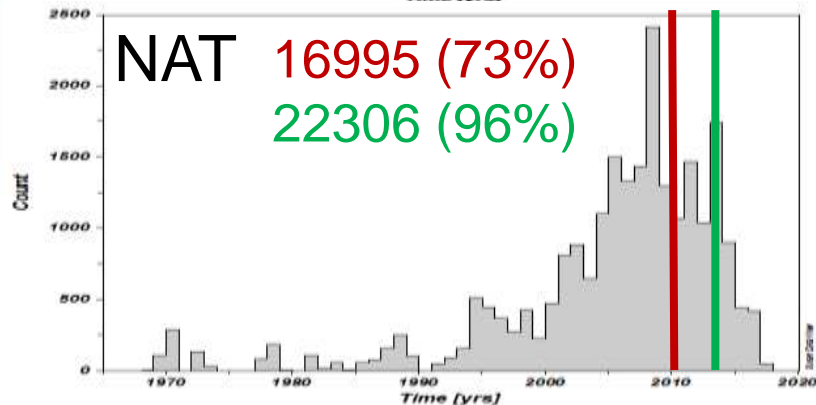
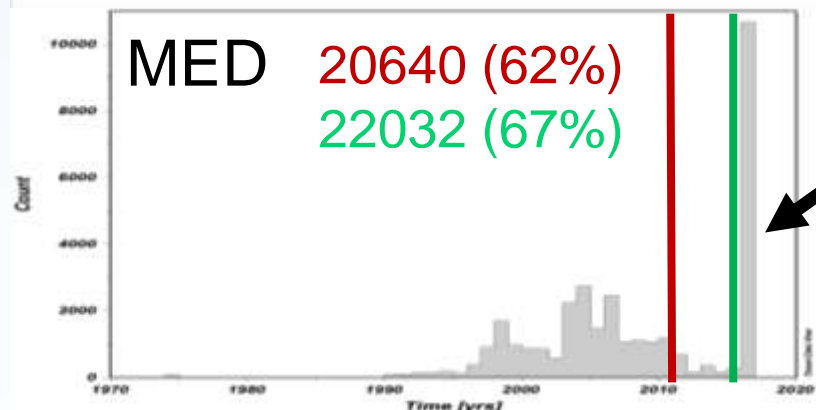
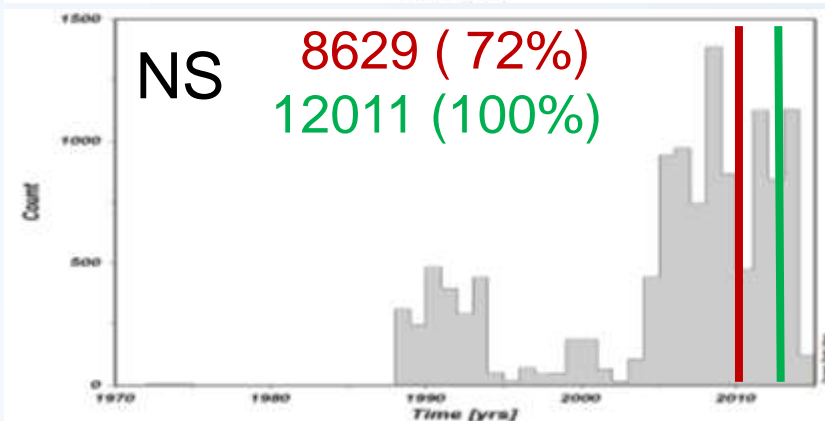
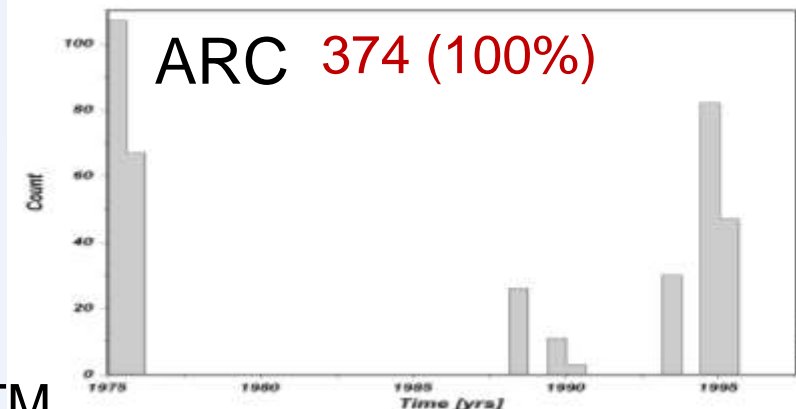
8721 (85,%)
10033 (98%)

NAT

16995 (73%)
22306 (96%)

NS

8629 (72%)
12011 (100%)



TRAINING and DISSEMINATION

WPWP11 promoted and presented SDC in many workshop and conferences (WP4 presentation)

DIVA training 3-6 April 2018 → wide participation, all RC were

 [SDC_WP11_D11.14_StrategyForTrainingActivities.pdf](#)

Modified on: 22 December, 2017 By: Michele FICHAUT

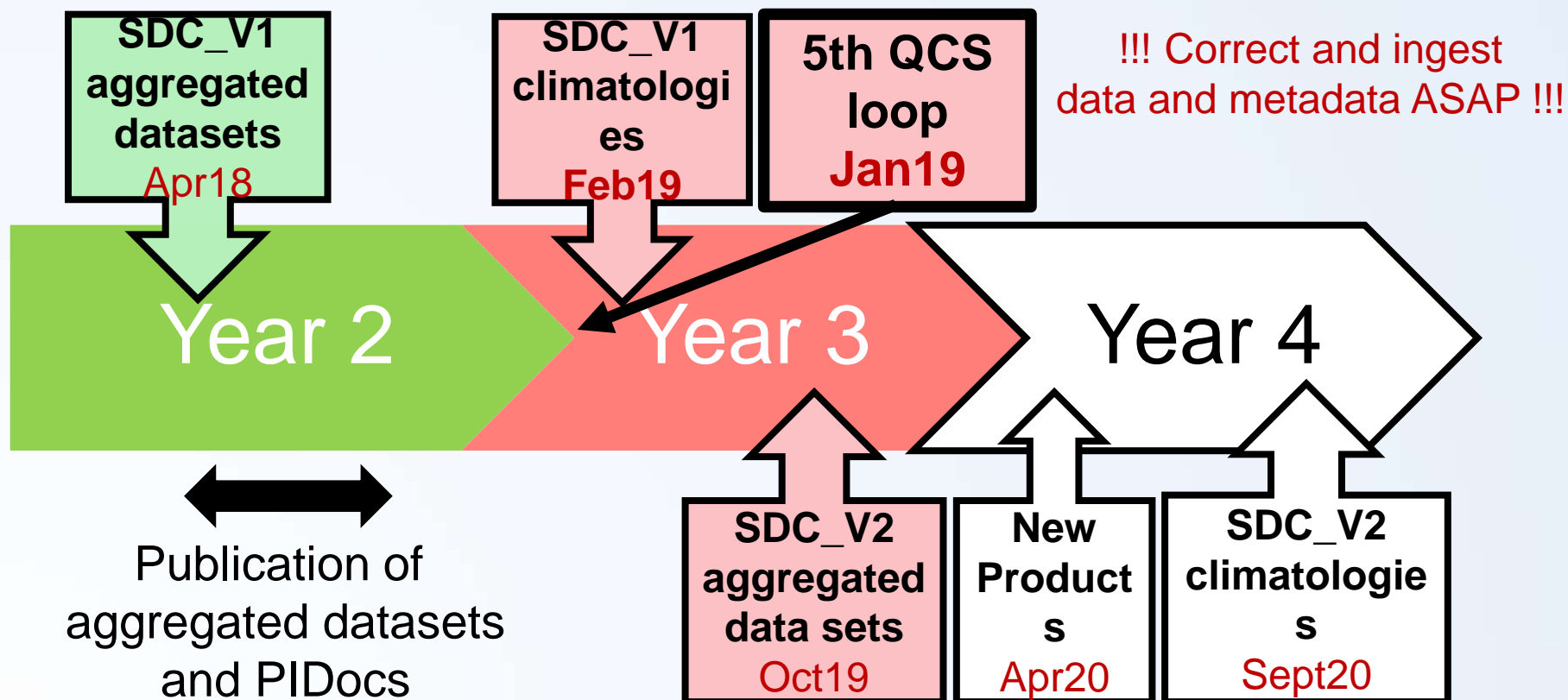
 [SDC_WP11_D11.15_OutcomeOfTrainingActivities.pdf](#)

Modified on: 20 April, 2018 By: Michele FICHAUT

1st SDC training workshop

→ Importance Quality Control SDC WP11 Introduction(WP11 team)

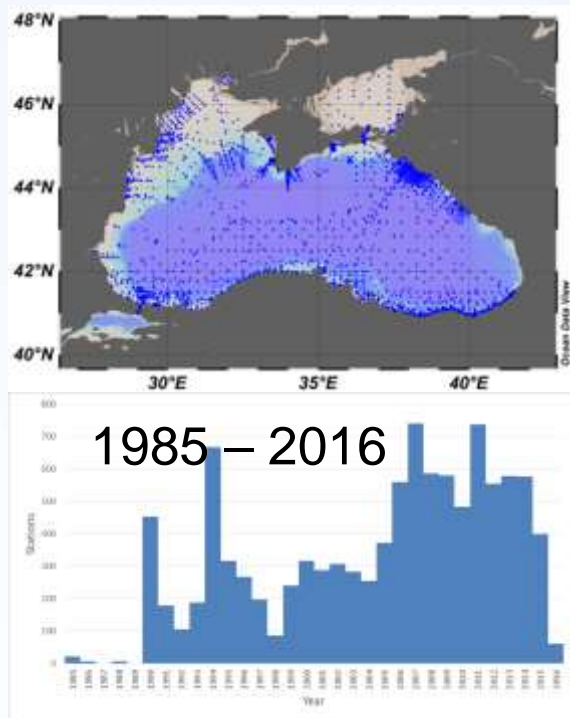
Work Plan and Timeline



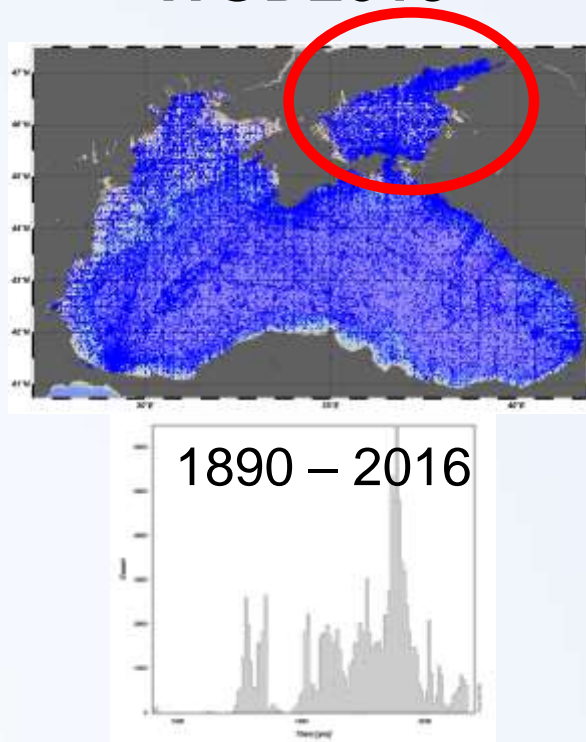
Ongoing Activities: data integration

Black Sea

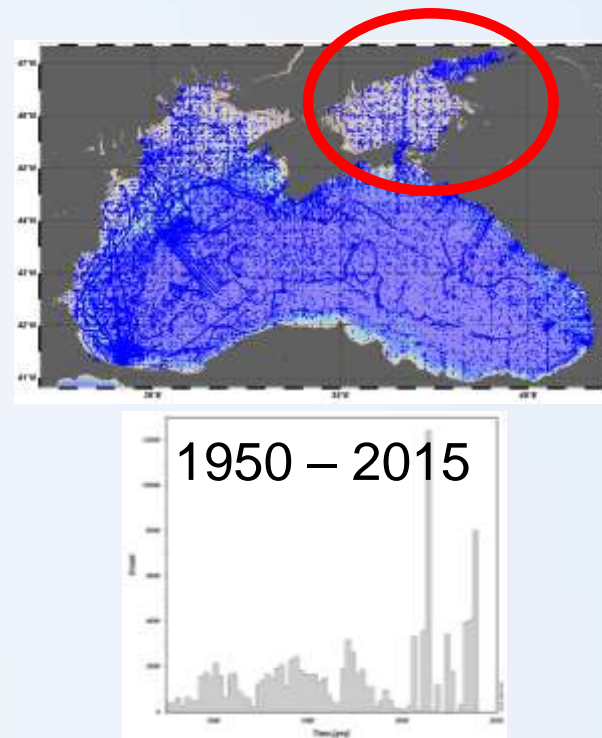
Restricted Dataset



WOD2013



CORA 5.1



Estimates from metadata:
60% duplicates
80000 stations (+50%)

sdn-userdesk@seadatanet.org – www.seadatanet.org

	SDC_BLS_D ATA_TS_V1	Restricted Dataset	WOD2013	CORA 5.1
stations	137723	10528	120845	103721
	148251			

Ongoing Activities: collaboration with CMEMS Ins TAC

1st Joint Meeting July 12th 2018



AGENDA

General introduction

1. From SeaDataNet to SeaDataCloud: new data products and innovation
2. From phase I to phase II: CMEMS in situ TAC developments

DISCUSSION

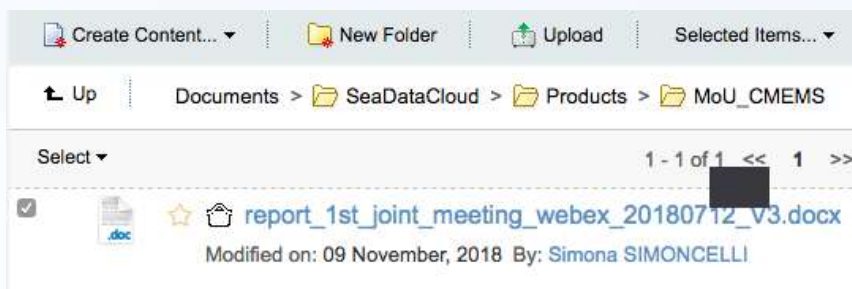
Review of Quality Check procedures

3. SDN Quality Check Strategy: from visual inspection to automatic approach
4. CMEMS quality check procedures

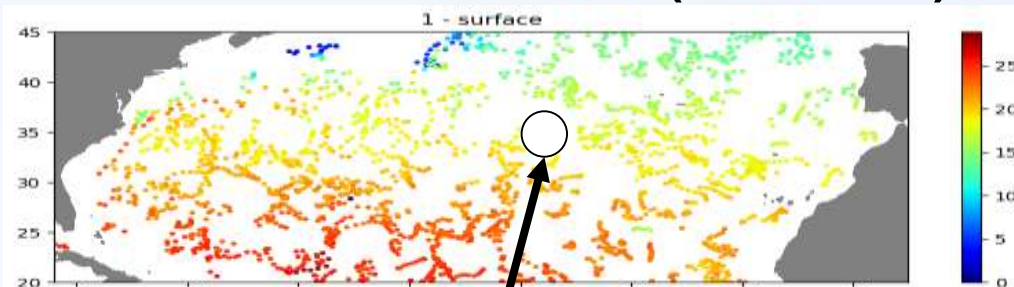
DISCUSSION → find a common strategy

5. First release of SDC_XXX_DATA_TS_V1 and Product Information Documents (PIDocs)
6. CMEMS new data types?

DISCUSSION --> data type approach/timelines/mutual requirements

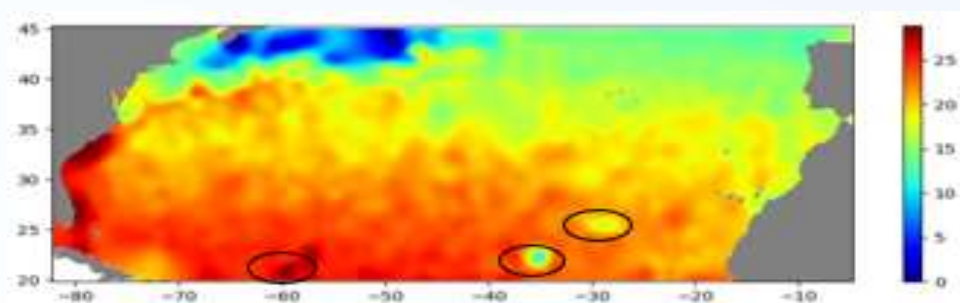


World Ocean Database (WOD2013)



Temperature from ARGO, Spring (2012-2013)

What's the **CLIMATOLOGICAL** value at this location?



DIVA estimate of 2012-2013 observations

Preliminary results from DIVA

→ anomalous features

→ **better quality control to eliminate outliers**

Errors in observations could be:

Instrumental Error (limited precision or bias of the sensor)

Representativeness Error

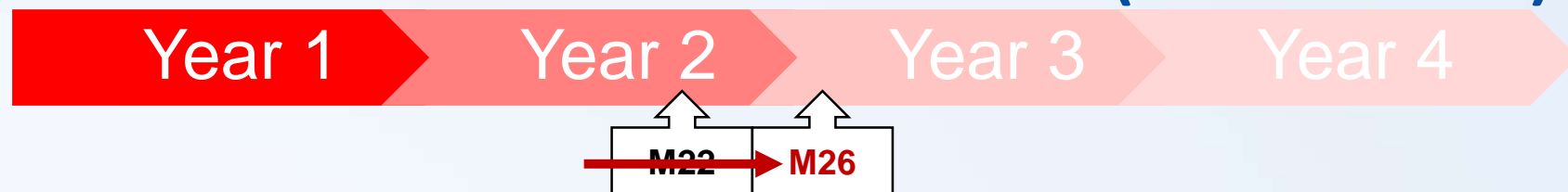
Synoptic Error (time)

Gross Error:
Human error, Instrumental failure, Incorrect communication, calibration error

Quality Control



Need to develop a new **NONLINEAR QC (NQC)** to reduce the representativeness error

Climatologies SDC_CLIM_TS_V1 (due Feb2019)



- T and S monthly and seasonal climatologies with **increased horizontal and vertical resolution** (WOA standard depth) covering the time period 1955-2017
 - **Integration** of SeaDataCloud data collections **with external data sets** to increase data coverage (CMEMS, WOD2018, ICES)
 - Analysis of (1) space/time data distribution; (2) data types consistency; (3) long term variability to compute climatologies on a decadal basis (sliding decades when possible)
 - **Product validation** → consistency analysis with WOA and CMEMS products (satellite reprocessed data sets and reanalysis)
- PIDoc will contain all this information
- SDC_CLIM_TS_V1 will be also accessible through EMODnet Physics portal

Conclusions

- Introduction of **PIDocs** represented a very good progress
- SDC_DATA_TS_V1 publication (link data approach) → *data set paper* and submission of SDN QCS as Ocean Best Practice
- **Metadata analysis** will be extended to all the regional seas in the next QCS loop
- Hunting of **data omissions** will be intensified  
- **QC analysis per data type** will continue to assure data consistency
- **Integration with external datasets** (WOD2018, CMEMS) for climatology production (ongoing activity)
- **WP11 met regularly online to monitor and harmonize the activities**
- **High participation/great collaboration/very good progresses**