



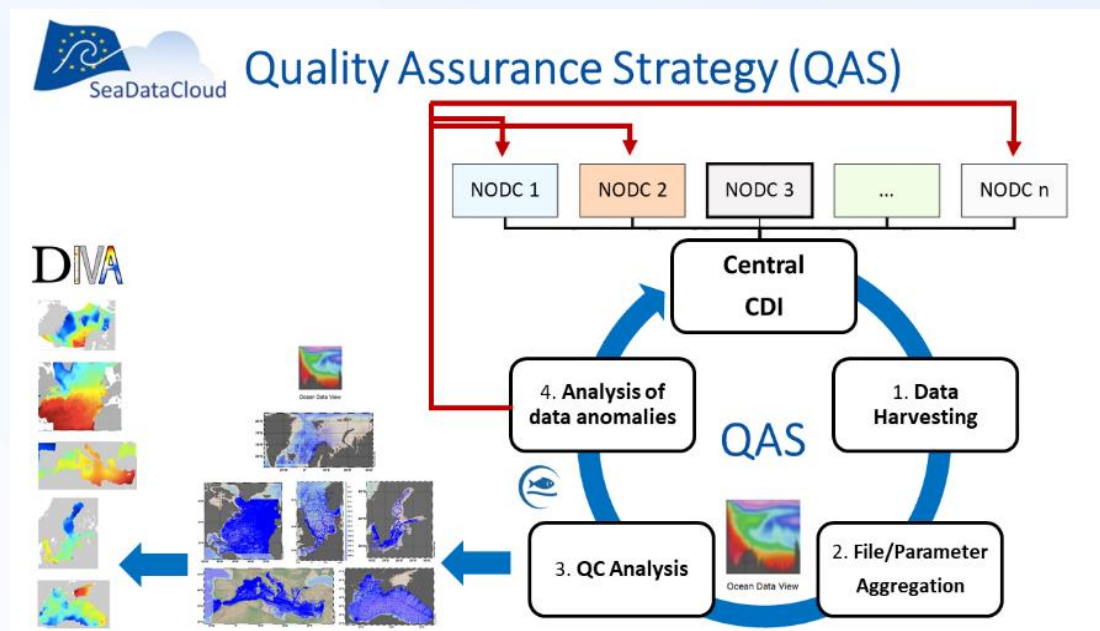
SeaDataCloud

WP11 – Aggregated Datasets

Regional leaders, presented by Christine Coatanoan (Ifremer)

Final Plenary Meeting, 30 October 2020
sdn-userdesk@seadatanet.org – www.seadatanet.org

Quality Assurance Strategy (QAS)



A **Quality Control Strategy (QCS)** was developed in SeaDataNet2 and continuously refined aiming at improving the quality of the data and creating the best data products.

The **QCS** iterative approach facilitates the upgrade of the data and it allows a versioning of data products (release of new data collections at the end of each loop and the generation of derived climatological products after a certain time lag dedicated to data processing).

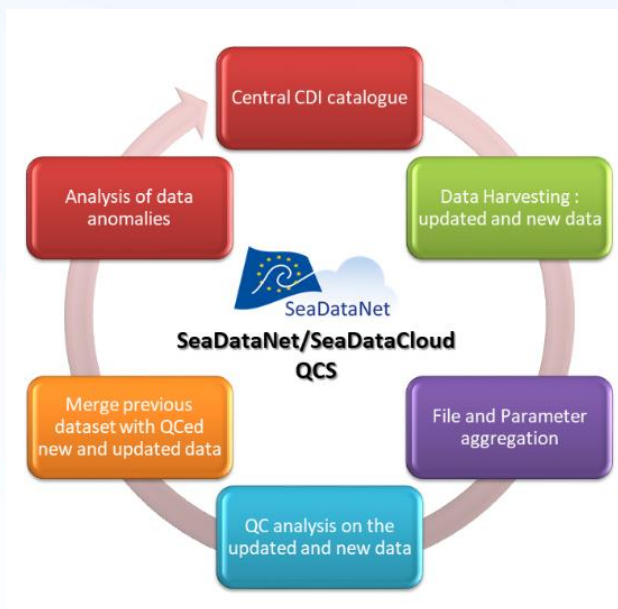
SeaDataCloud Aggregated dataset V1 & V2

- First SeaDataCloud aggregated dataset V1
harvest in 30th October 2017
- Second SeaDataCloud aggregated dataset V2
second semester of 2019

➔ Main difference : harvesting of new and updated data

- 1st harvest 31th July 2020 (=>BUG)
- 2^d harvest end of November 2020

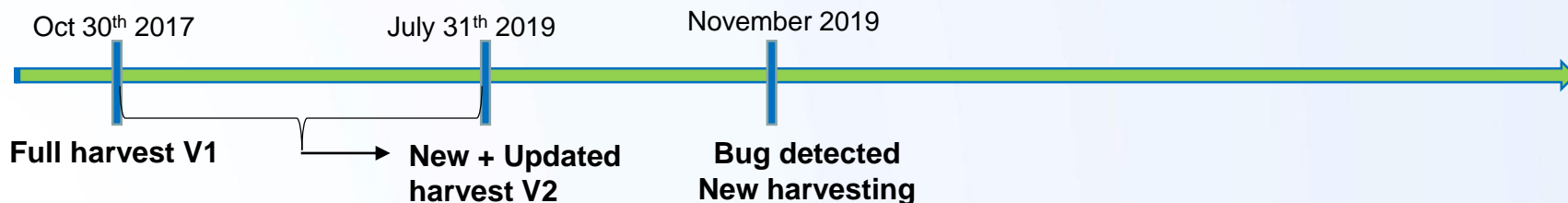
Harvesting of new and updated data



STEP 1 HARVESTING

Jul 31st 2019 ==> harvested 2 data collections (restricted and unrestricted) for T&S from the latest SeaDataNet CDI catalogue

- subset of complete data collection
ONLY for CDI-data sets that are New or Updated compared to the last harvest at Oct 30th 2017

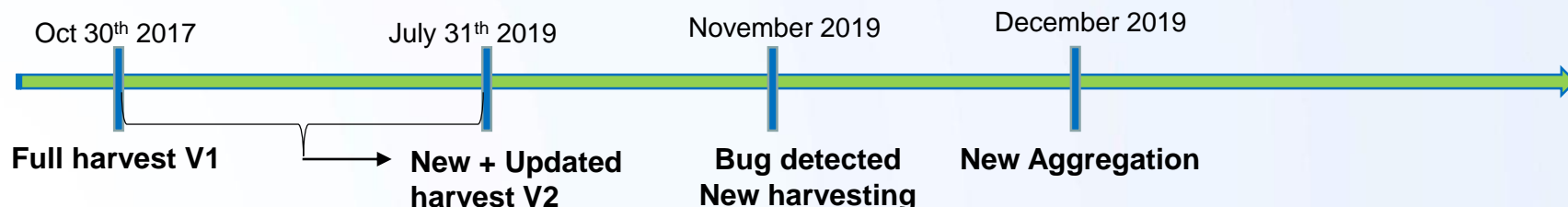
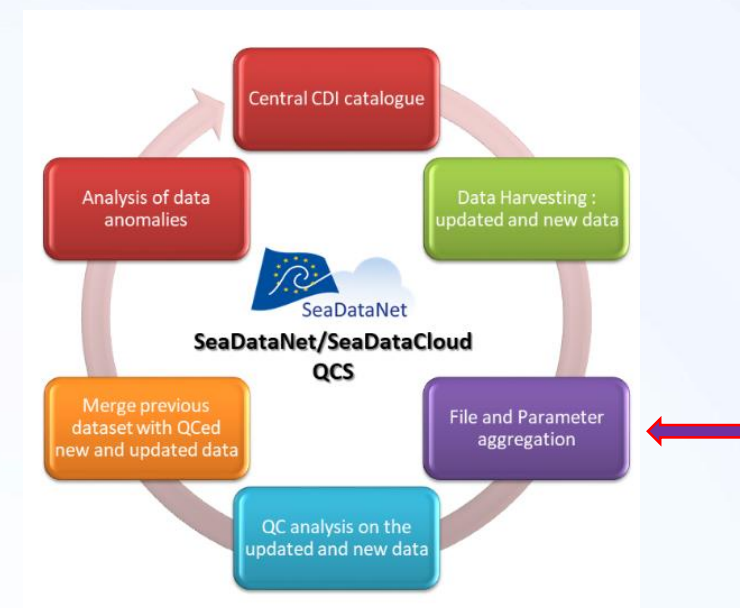


Harvesting of new and updated data

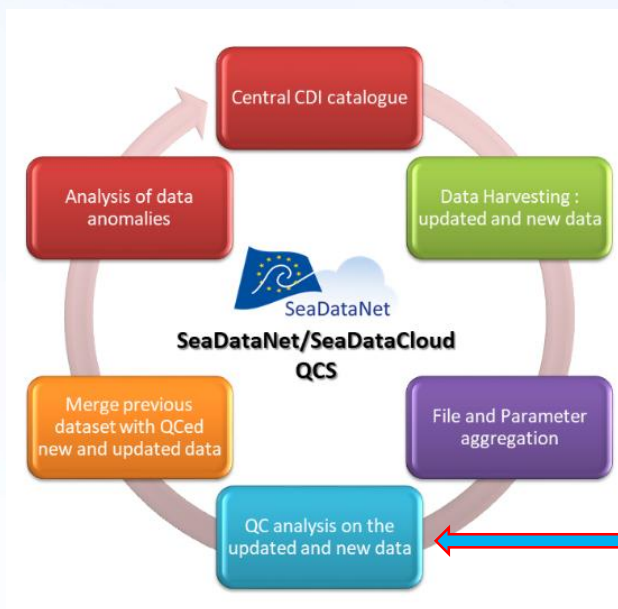
STEP 2 AGGREGATION

File and parameter aggregation of the "new_and_update_after_20171030" subset

- Split the results into regional collections [Merge the harvested regional data with the SDC V1 regional collections after QC V2]
- Analysis log files generated during import Error and warning messages → **there are still files with serious issues waiting to be corrected by the data centers**



QC analysis on updated and new data

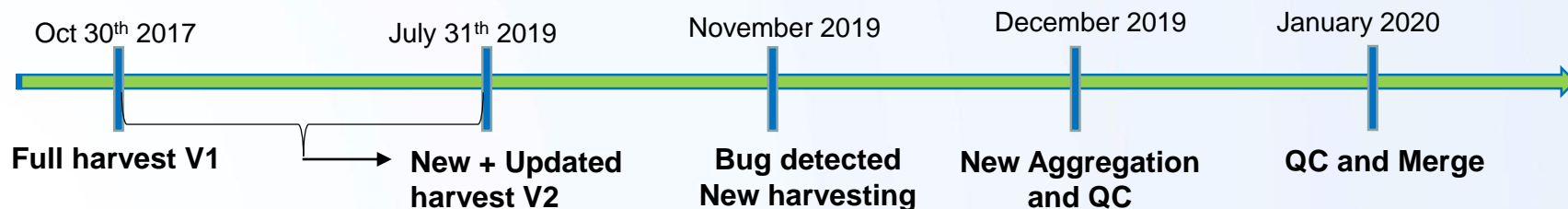


STEP 3 QC Analysis

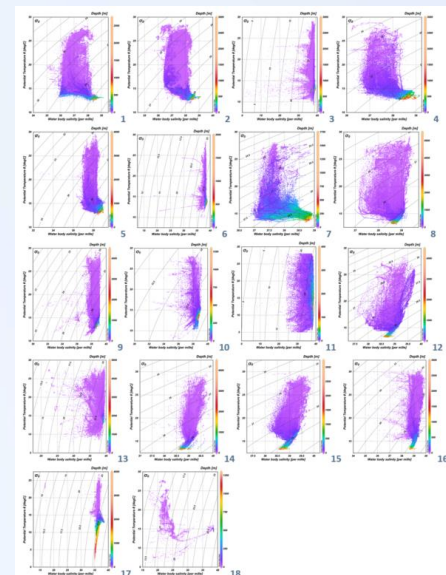
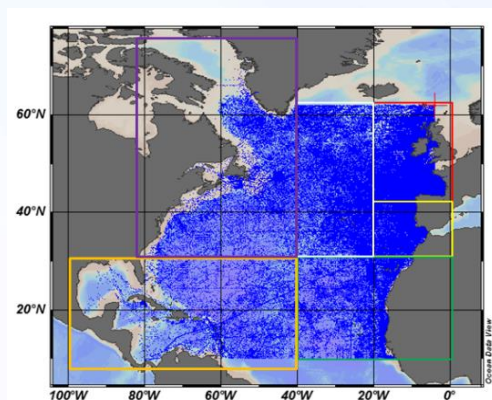
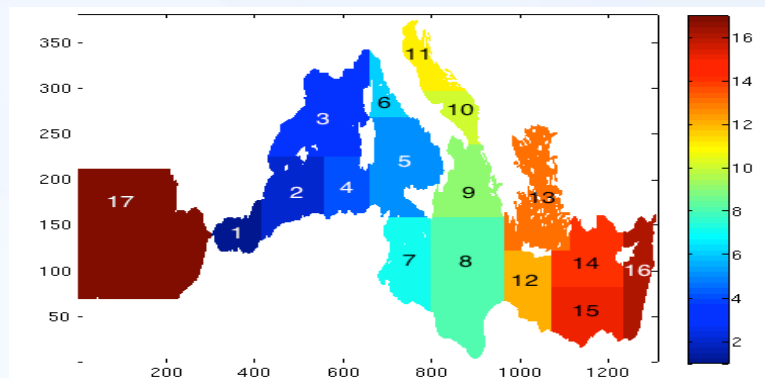
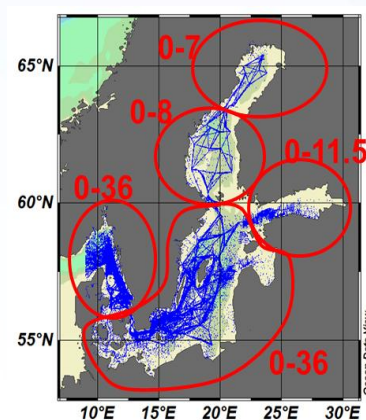
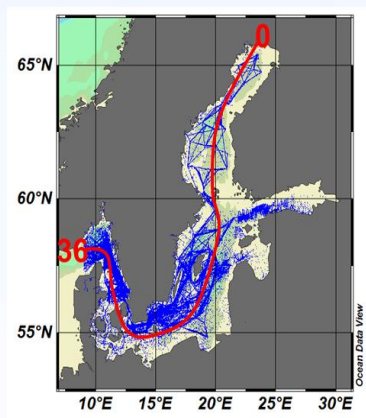
The QC analysis is conducted using ODV software and applying the guidelines defined for the previous aggregated datasets. (basic QC analysis steps, duplicates check).

The specific QC steps/procedures developed per each basins for the V2 data collection also have been applied.

The QC was applied both to non-restricted and restricted data but only non-restricted data were utilized in the products while restricted data are intended to be used internally for calculation of climatologies.

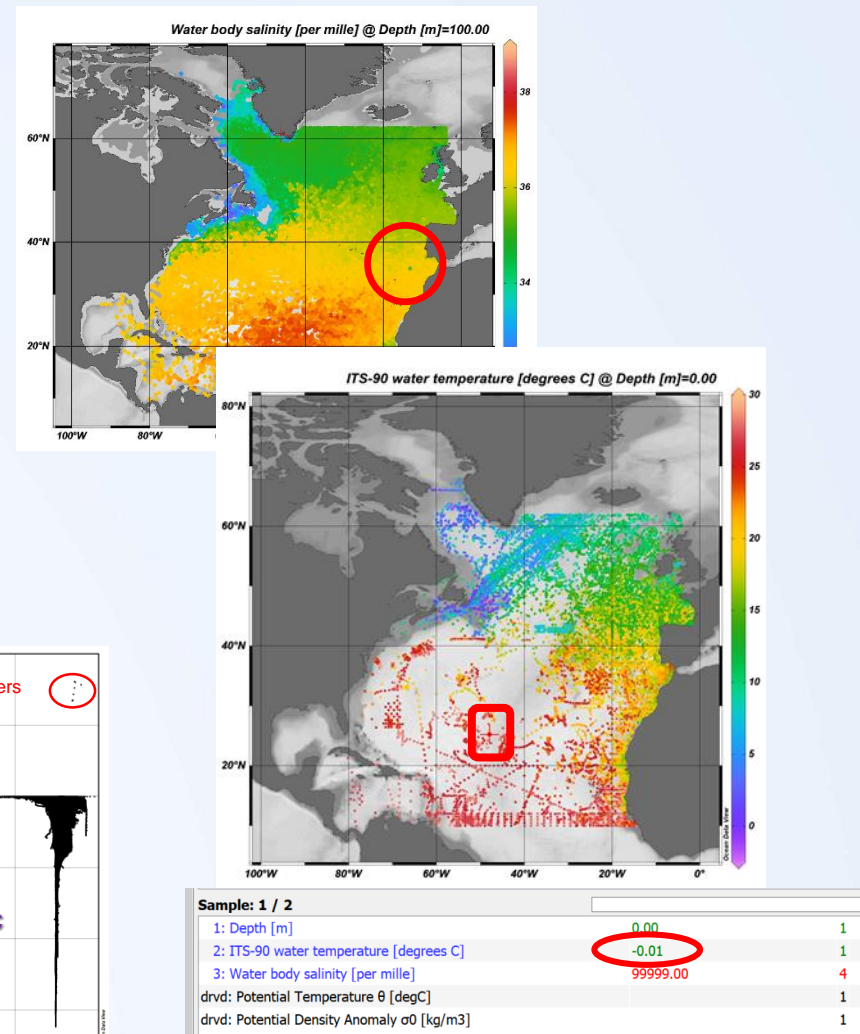
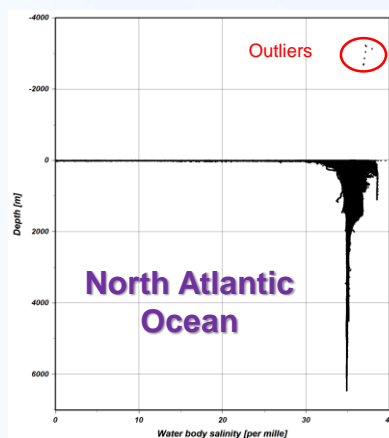
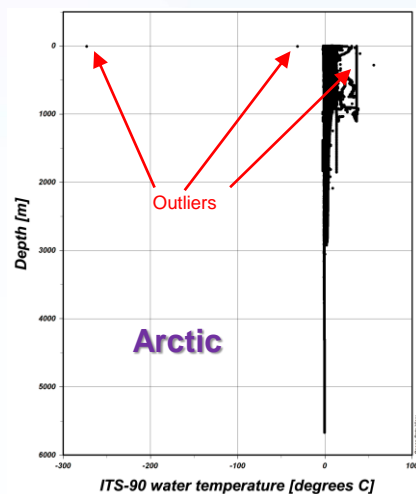
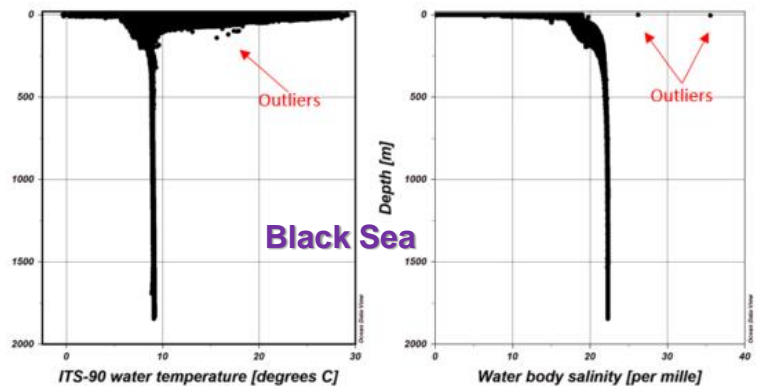


Examples of QC procedure (regional criteria)



- Specific area of studies
- Specific criteria for range check and specific to some areas (Baltic, Black Sea)
- Specific TS diagrams by sub-regions

Examples of anomalies



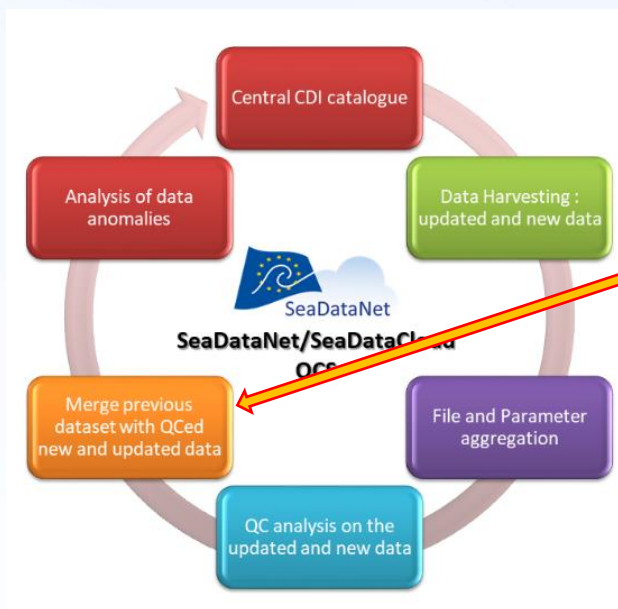
Duplicates and XBT issues (Med Sea)

Problem of duplicates highlighted by Med Sea region

- same tracks submitted by multiple data originators
- «duplicates» along the tracks have different metadata: project, instrument type (bt, xbt, ctd)
- «duplicates» along the tracks have different data : some of them are raw data, some have been processed (1m interp) but not following *Manzella et al. (2003 and 2007)*

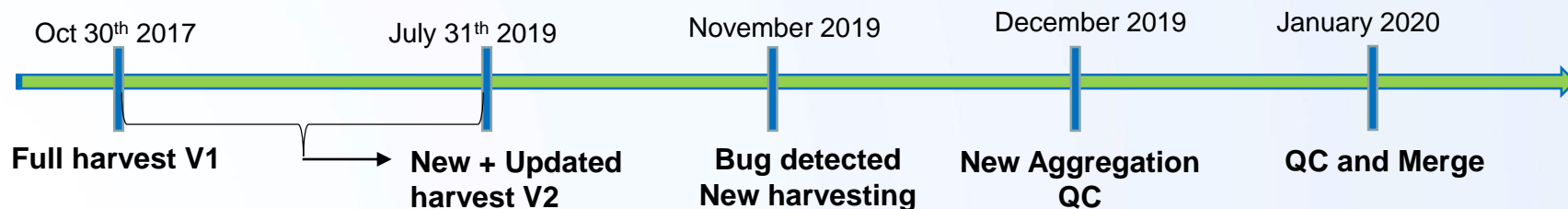
→ ENEA version has been kept even if not the best XBT metadata description but the data have been processed according the QC procedures defined by the Med obs and SOOP communities (btw in CORA and WOD the situation is even more confused)

QC analysis on updated and new data

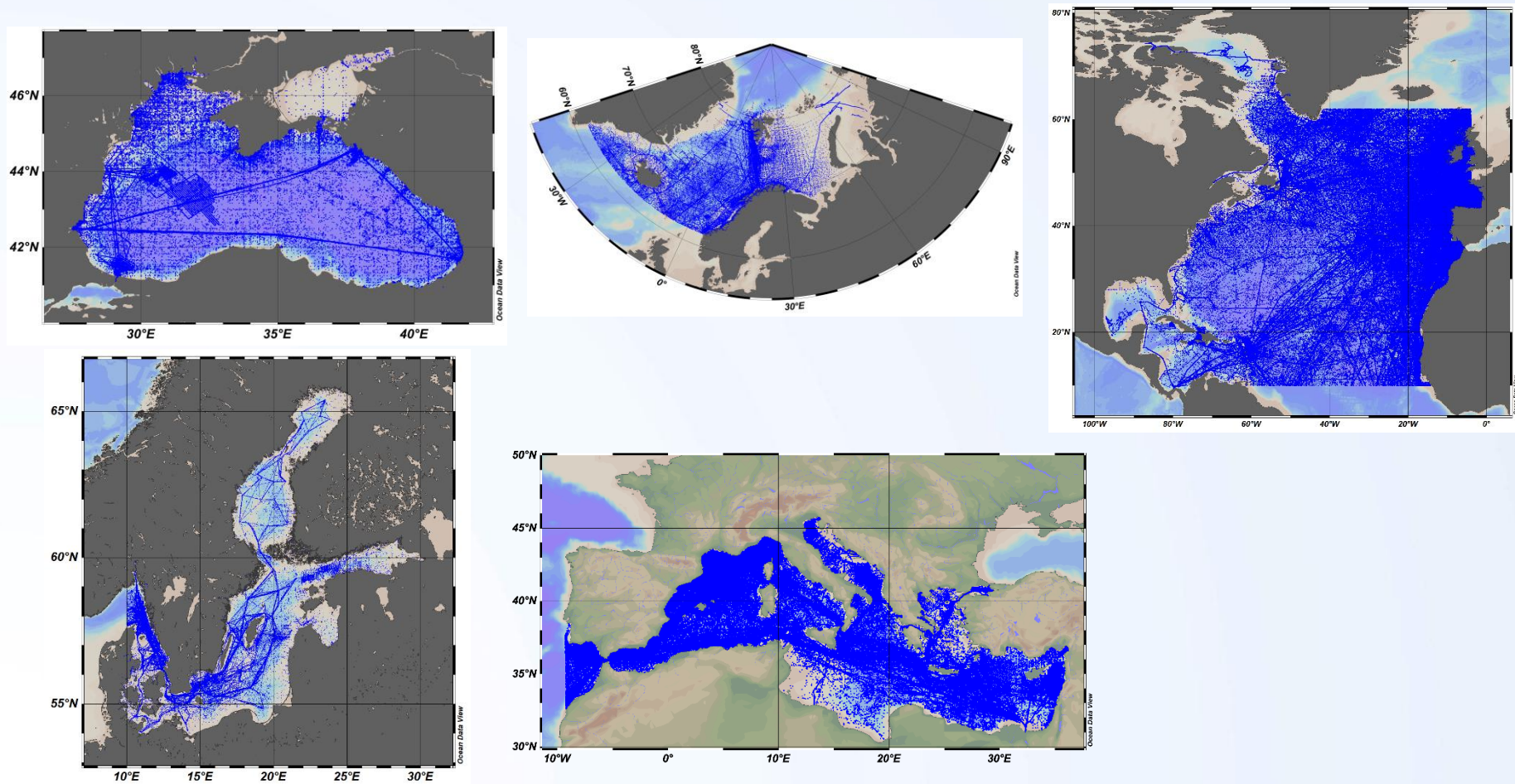


STEP 4 Merge new + updated V2 with V1

Taking care about duplicates based on `local_cdi_id` and `edmo_code`



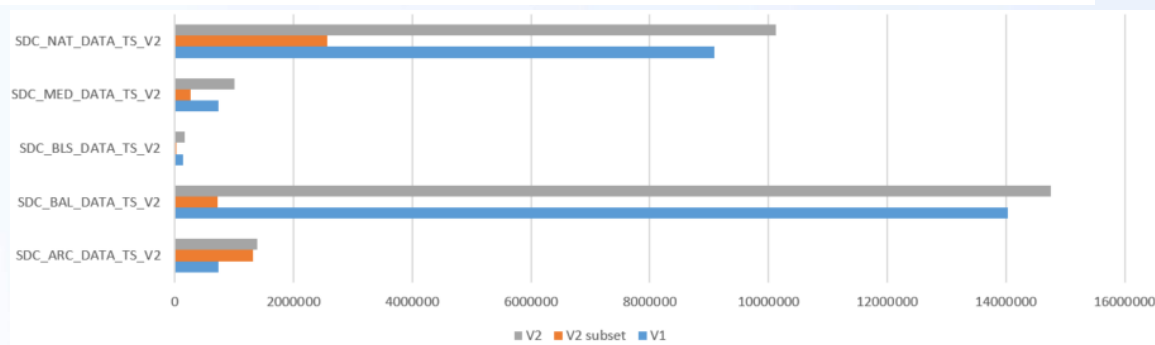
Aggregated dataset : Collection Map (V2)



Aggregated dataset : station number

Product	SDC_DATA_TS_V1	V2 subset	SDC_DATA_TS_V2	% increase V2/V1
SDC_ARC_DATA_TS_V2	731286	1315688	1392366	90
SDC_BAL_DATA_TS_V2	*14038820	*714222	*14753042	5
SDC_BLS_DATA_TS_V2	137723	24933	162656	18
SDC_MED_DATA_TS_V2	739784	263474	1003258	36
SDC_NAT_DATA_TS_V2	9091769	2572311	10119755	12
SDC_NS_DATA_TS_V2_DISCRETE	162452	109106		
SDC_NS_DATA_TS_V2_TRAJECTORIES	580376	260516		

* Samples for Baltic



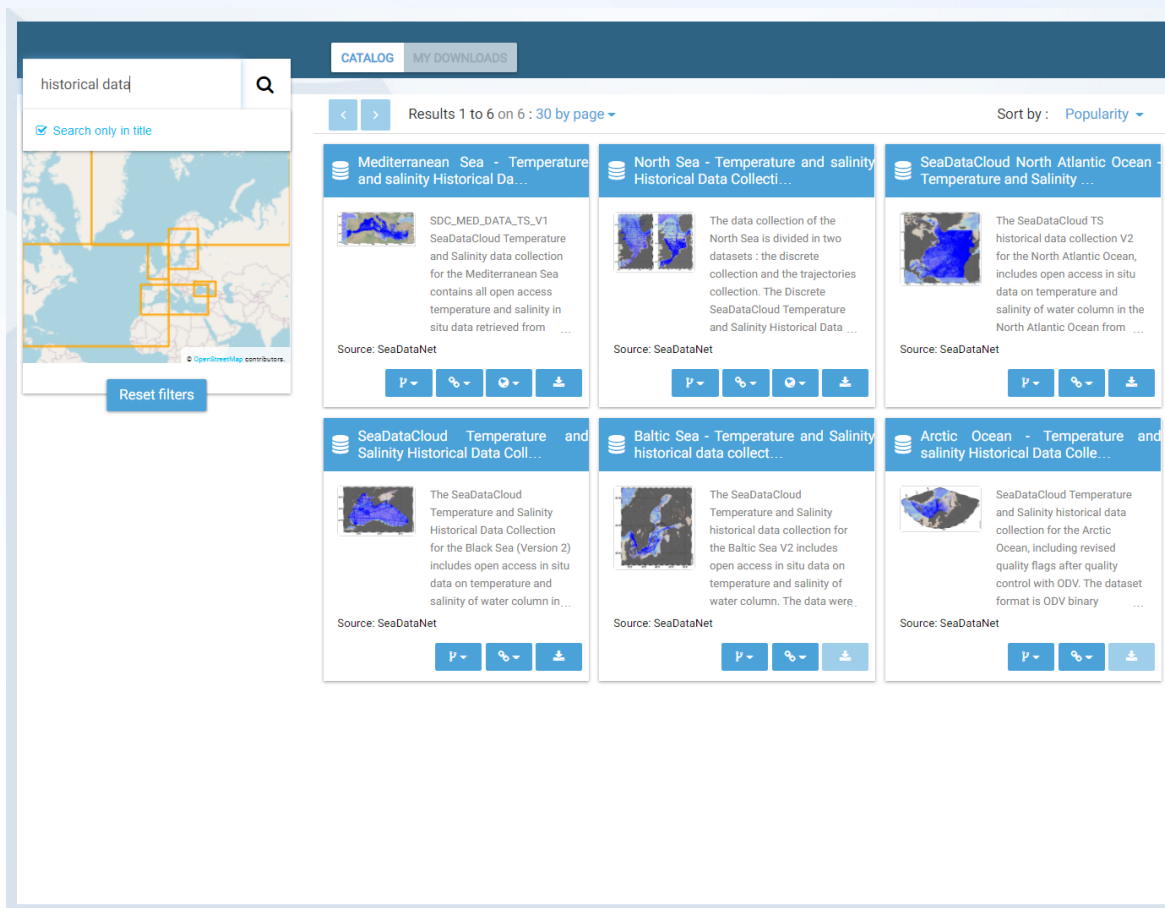
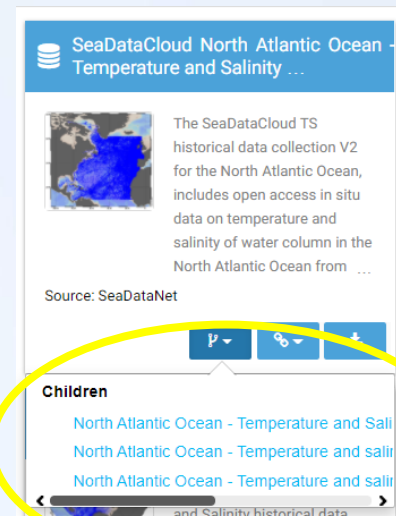
Aggregated dataset and external datasources

- Before working on climatology, one step was to integrate external datasets into the final SeaDataCloud aggregated dataset in order to improve data coverage for better climatology results.
- 2 sources used: WOD and CORA (one or the other depending on the region)
 - Extraction and conversion to ODV format
 - Duplicate check
 - Same quality control procedures on external data (with a few loops after the first tests with DIVAnd)
 - Need to send feedback about anomalies



Takes Time

Sextant catalogue

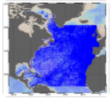
Previous versions

ERDDAP Vis



ERDDAP > [tabledap](#) > Make A Graph

SeaDataCloud North Atlantic Ocean - Temperature and Salinity H...



The SeaDataCloud TS historical data collection V2 for the North Atlantic Ocean, includes open access in situ data on temperature and salinity of water column in the North Atlantic Ocean from ...

Source: SeaDataNet

DOI of the product
ERDDAP link to aggregated dataset V2

Dataset Title: **SDC North Atlantic Aggregation V2** [✉](#) [RSS](#)

Institution: SeaDataNet (Dataset ID: SDC_NATL_AGG_V2)

Range: longitude = -97.1137 to -0.54916°E, latitude = 9.37043 to 76.45°N, depth = -3283.0476 to 6456.3m, time = 1893-08-10T04:45:00Z to 2019-06-04T08:10:41Z

Information: [Summary](#) | [License](#) | [FGDC](#) | [ISO 19115](#) | [Metadata](#) | [Background](#) | [Subset](#) | [Data Access Form](#)

Graph Type: **markers**

X Axis: **longitude**

Y Axis: **latitude**

Color: **temp**

Constraints

time >= 2019-01-01T00:00:00Z

depth >= 5

temp_qc = 1

depth_qc = 1

obs_date_qc = 1

Optional Constraint #1

Optional Constraint #2

Server-side Functions

distinct()

Graph Settings

Marker Type: **Square** Size: **3**

Color: **Continuity** Scale: **Ascending**

Color Bar: **Minimum** **Maximum** **N Sections**

Draw land mask: **Y Axis Minimum** **Maximum** **Ascending**

Redraw the Graph (Please be patient. It may take a while to get the data.)

Optional: Then set the File Type: **.htmlTable** (File Type information)

and Download the Data or an Image

or view the URL: http://www.ifremer.fr/erddap/tabledap/SDC_NATL_AGG_V2.htmlTable?lc

(Documentation / Bypass this form)

Click on the map to specify a new center point.

Zoom: **Out 8x** **Out 2x** **Out** **Data** **In** **In 2x** **In 8x**

Time range: **5** month(s)

Temperature

SDC North Atlantic Aggregation V2

time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z

depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet

Temperature

SDC North Atlantic Aggregation V2

time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z

depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet

Temperature

SDC North Atlantic Aggregation V2

time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z

depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet

Temperature

SDC North Atlantic Aggregation V2

time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z

depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet

Temperature

SDC North Atlantic Aggregation V2

time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z

depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet

Temperature

SDC North Atlantic Aggregation V2

time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z

depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet

Temperature

SDC North Atlantic Aggregation V2

time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z

depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet

Temperature


SDC North Atlantic Aggregation V2


time = 2019-01-01T00:00:00Z, time = 2019-06-05T00:00:00Z


depth = 5, temp_qc = 1, depth_qc = 1, obs_date_qc = 1

Data courtesy of SeaDataNet



ERDDAP Visualisation


ERDDAP
Easier access to scientific data


ERDDAP
Easier access to scientific data





ERDDAP
Easier access to scientific data


ERDDAP > tabledap > Make A Graph


Dataset Title: **SDC North Atlantic Aggregation V2**  


Institution: SeaDataNet (Dataset ID: SDC_NATL_AGG_V2)


Range: longitude = -97.1137 to -0.54916°E, latitude = 9.37043 to 76.45°N, depth =


Information: Summary  | License  | FGDC | ISO 19115 | Metadata | Background 

Graph Type: markers 


X Axis: longitude 


Y Axis: latitude 

Color: temp 



Constraints 




time	station_id	obs_date_qc	depth	temp_qc	position_qc	depth_qc	temp	obs_date_qc	psal_qc
>= 2018-01-01T00:00:00Z	<= 2019-06-05T00:00:00Z	<= 5	<= 1	<= 1	<= 1	<= 1	<= 1	<= 1	<= 1


Server-side Functions 


☐ distinct() 

Graph Settings




Marker Type: Square  Size: 3 


Color:  Continuity:  Scale: 

Draw land mask: 

Y Axis Minimum: 

Redraw the Graph (Please be patient. It may take a while to get the data.)

Optional: Then set the File Type:   (File Type information) and Download the Data or an Image 



or view the URL: http://www.ifremer.fr/erddap/tabledap/SDC_NATL_AGG_V2.htmlTable?c (Documentation / Bypass this form )


ERDDAP > tabledap > Make A Graph


Dataset Title: **SDC North Atlantic Agg**


Institution: SeaDataNet (Dataset ID: SDC_NATL_AGG_V2)


Range: longitude = -97.1137 to -0.54916°E, latitude = 9.37043 to 76.45°N, depth =


Information: Summary  | License  | FGDC


Graph Type: markers 

X Axis: longitude 


Y Axis: latitude 


Color: psal 

Constraints 



Optional Constraint #1 

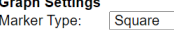


time	station_id	obs_date_qc	depth	temp_qc	position_qc	depth_qc	temp	obs_date_qc	psal_qc
>= 2018-01-01T00:00:00Z	<= 2019-06-05T00:00:00Z	<= 5	<= 1	<= 1	<= 1	<= 1	<= 1	<= 1	<= 1


Server-side Functions 


☐ distinct() 

Graph Settings




Marker Type: Square  Size: 3 


Color:  Continuity:  Scale: 

Draw land mask: 



Y Axis Minimum: 

Redraw the Graph (Please be patient. It may take a while to get the data.)

Optional: Then set the File Type:   (File Type information) and Download the Data or an Image 




or view the URL: http://www.ifremer.fr/erddap/tabledap/SDC_NATL_AGG_V2.htmlTable?c (Documentation / Bypass this form )


ERDDAP > tabledap > Make A Graph


Dataset Title: **SDC North Atlantic Aggregation V2**  


Institution: SeaDataNet (Dataset ID: SDC_NATL_AGG_V2)


Range: longitude = -97.1137 to -0.54916°E, latitude = 9.37043 to 76.45°N, depth = 3283.0476 to 6456.3m, time = 1893-08-10T04:45:00Z to 2019-06-04T08:10:41Z


Information: Summary  | License  | FGDC | ISO 19115 | Metadata | Background  | Subset | Data Access Form

Graph Type: markers 


X Axis: longitude 


Y Axis: latitude 

Color: psal 



Constraints 




time	station_id	obs_date_qc	depth	temp_qc	position_qc	depth_qc	temp	obs_date_qc	psal_qc
>= 2018-01-01T00:00:00Z	<= 2018-06-05T00:00:00Z	<= 5	<= 1	<= 1	<= 1	<= 1	<= 1	<= 1	<= 1


Server-side Functions 


☐ distinct() 

Graph Settings




Marker Type: Circle  Size: 3 


Color:  Continuity:  Scale: 


Draw land mask: 








Y Axis Minimum: 








Redraw the Graph (Please be patient. It may take a while to get the data.)

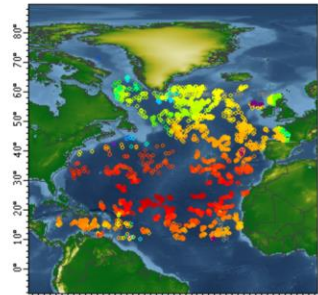
Optional: Then set the File Type:   (File Type information) and Download the Data or an Image 

or view the URL: http://www.ifremer.fr/erddap/tabledap/SDC_NATL_AGG_V2.htmlTable?c (Documentation / Bypass this form )

Click on the map to specify a new center point. 

Zoom:  Out 8x  Out 2x  Out  Data  In  In 2x  In 8x

Time range: 5  month(s)      



SDC North Atlantic Aggregation V2
time>=2018-01-01T00:00:00Z, time<=2018-06-05T00:00:00Z, depth=0, depth_qc=5, temp_qc=1, depth_qc=1, obs_date_qc=1
Data courtesy of SeaDataNet

ERDDAP Visualisation



The Dataset Attribute Structure (.das) for this Dataset

```

Attributes {
  s {
    time {
      Picking_CoordinateAsDateTime "Time",
      Picking_actual_range "1.470711e+9, 1.559635841e+9",
      String alias "T",
      String iose_category "Time",
      String long_name "Time",
      String source_name "Obs_data",
      String standard_name "Time",
      String time_string "701-JAN-1970 09:00:00",
      String time_precision "1970-01-01T00:00:00Z",
      String unit "seconds since 1970-01-01T00:00:00Z";
    }
    station_id {
      Int32 actual_range 1, 10119755;
      String long_name "Station ID";
    }
    platform {
      String long_name "Platform";
    }
  }
}

```

The Dataset Attribute Structure (.das) for this Dataset

```
Attributes {
  s {
    time {
      String _CoordinateAxisType "Time";
      Float64 actual_range -2.4107157e+9, 1.559635841e+9;
      String axis "T";
      String _axis_category "Time";
      String long_name "Obs Date";
      String source_name "Obs_date";
      String standard_name "time";
      String time_origin "01-JAN-1970 00:00:00";
      String time_precision "1970-01-01T00:00:00Z";
      String units "seconds since 1970-01-01T00:00:00Z";
    }
  }
  station_id {
    Int32 actual_range 1, 10119755;
    String long_name "Station Id";
  }
}
```

Minimum		Maximum		or a List of Values	
1893.08-10T04.45:00Z		2019-06-04T08:10:41Z			
1		10119755			
Q39					
A		B		C	
station_id,latitude,longitude,depth,temp,psal				D	
,degrees_north,degrees_east,m,				E	
1		9040941,32.8896,-28.8526,2.7886,19.915,36.585029999999999			
2		9040941,32.8896,-28.8526,3.67392,19.915,36.585029999999999			
3		9040941,32.8896,-28.8526,4.86545,19.914,36.585029999999999			
4		9040942,32.95486,-28.2233,2.78025,20.007,36.763290000000005			
5		9040942,32.95486,-28.2233,3.6739,20.005,36.76729			
6		9040942,32.95486,-28.2233,4.66683,20.005,36.763290000000005			
7		9040943,33.33241,-28.2263,2.97875,19.042,36.56309			
8		9040943,33.33241,-28.2263,3.87236,19.042,36.56309			
9		9040944,33.75041,-28.2203,3.67365,19.158,36.66383			
10		9040944,33.75041,-28.2203,4.666519999999999,19.156,36.66383			
11		9040945,32.82599,-28.7463,3.4753,20.28,36.776709999999994			
12		9040945,32.82599,-28.7463,4.68326,20.28,36.776709999999994			
13		9040946,32.75747,-28.3888,2.7803,19.382,36.572520000000004			
14		9040946,32.75747,-28.3888,3.673959999999999,19.382,36.572520000000004			
15		9040946,32.75747,-28.3888,4.66691,19.382,36.572520000000004			
16		9040947,34.11074,-27.864,2.97855,19.062,36.668			
17		9040947,34.11074,-27.864,3.871210999999999,19.062,36.668			
18		9040947,34.11074,-27.864,4.76566,19.063,36.668			
19		9040948,34.61031,-27.4841,3.17698,19.012999999999998,36.6068			
20		9040948,34.61031,-27.4841,3.971219999999999,19.011,36.6088			
21		9040948,34.61031,-27.4841,4.864730000000001,19.011,36.6078			
22		9040949,34.35789,-26.4964,2.77992,18.067999999999998,36.4508			
23		9040949,34.35789,-26.4964,3.77274,18.059,36.4478			
24		9040949,34.35789,-26.4964,4.76566,18.052,36.4458			
25		9040950,34.65624,-27.072,3.07769,18.12,36.4898			
26		9040950,34.65624,-27.072,4.07048,18.116,36.4918			
27		9040950,34.65624,-27.072,4.864719999999999,18.116,36.4918			

Conclusion

- Data aggregation was performed only on new and updated CDIs for the first time, and data merged with V1 → saving time and workflow optimization
- Improvement of harvesting and ODV from experience of this new procedure (from detected bug)
- Metadata need to be populated in order to improve products' quality and the relating scientific results (i.e. Originator, Instrument Type, ...) (highlighted by Med Sea [XBT] and Black Sea [cruises])
- A lot of data that are not processed according to the accepted QC DM best practices (highlighted by Med Sea, cases of XBT)
- Quality of the data has also to be taken into account in each data center to avoid wasting time at the regional product level (anomalies to be sent to concerned data center) : implementation of the QAS on the VRE to fully automate the process making it more efficient
- Deliverable D11.4 released in May 2020
- V2 PIDocs and assessment results have been updated from V1 → saving time for further checks
- PIDoc and aggregated datasets available in Sextant catalogue for most of the regions with relative DOIs (*see presentation Overview of the product catalogue*)

