SeaDataCloud
Temperature and Salinity data collections

S. Simoncelli (INGV)
simon@ingv.it

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

**Support Activities**
- Coordination and Communication
- System Monitoring
- Technology Infrastructure
- Service Development

**Value Chain Steps**
- Discover
- Collect
- Organize
- Integrate
- Analyze
- Validate & Report
- Visualize & Disseminate
- Use

**Support Activities**
- FAIR guiding data management principle and linking data approach
- Science support and supervision → Up to date QC procedures
- User driven and provider/producer appreciative

**Outputs**
- Create Knowledge, Derive Products, Develop Applications and Services, Drive Predictive Models, Rapid Environmental Assessment

**Innovation**
- Effective & Transparent Decision Support
- Blue Sustainable Growth
SeaDataCloud Products

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

1. **Aggregated data sets EU marginal seas** → historical temperature and salinity data harvested from the central CDI and validated by regional leaders

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 (e.g. gliders, Argo, ships, drifters, fixed platforms) and remote sensed observations, Ocean Monitoring Indicators for tracking ocean mechanisms and/or climate modes and trends
Quality Control Strategy (QCS)

**GOAL**: to improve the quality of SeaDataNet database content and create the best data products

→ **iterative approach** to facilitate the **upgrade** of the database and **versioning** of data products
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

<table>
<thead>
<tr>
<th>par</th>
<th># stations</th>
<th>%</th>
<th># samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>739784</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>737102</td>
<td>99,6</td>
<td>41232938</td>
</tr>
<tr>
<td>S</td>
<td>667232</td>
<td>90,2</td>
<td>28518744</td>
</tr>
<tr>
<td>TS</td>
<td>665388</td>
<td>89,9</td>
<td>28119926</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>%</th>
<th>QF0</th>
<th>QF1-2</th>
<th>Q3-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>dpt</td>
<td>3.0→0</td>
<td>96.9→99.8</td>
<td>0.1</td>
</tr>
<tr>
<td>T</td>
<td>2.7→0</td>
<td>97.0→99.8</td>
<td>0.3</td>
</tr>
<tr>
<td>S</td>
<td>4.5→0</td>
<td>94.6→99.2</td>
<td>0.9</td>
</tr>
<tr>
<td>dpt&amp;T&amp;S</td>
<td>3.0</td>
<td>94.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>
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)
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

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

!!!Actions needed to complete crucial metadata information!!!

**MED XBTs 1950-2000**

**MEDATLAS II**

SDC

Data Omission

!!!Actions needed to ingest missing XBTs!!!
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

<table>
<thead>
<tr>
<th>Instrument Info</th>
<th>Probe</th>
<th># stations</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>26162</td>
<td>46</td>
</tr>
<tr>
<td>'SDN:P01::ADEPZZ01 SDN:L22::TOOL0262 SDN:L33::011</td>
<td>SDN:P01::TEMPET01 SDN:L22::TOOL0262'</td>
<td>T-5 XBT</td>
<td>1239</td>
</tr>
<tr>
<td>'SDN:P01::ADEPZZ01 SDN:L22::TOOL0263 SDN:L33::041</td>
<td>SDN:P01::TEMPET01 SDN:L22::TOOL0263'</td>
<td>T-7 XBT</td>
<td>9995</td>
</tr>
<tr>
<td>'SDN:P01::ADEPZZ01 SDN:L22::TOOL0263 SDN:L33::041</td>
<td>SDN:P01::TEMPPR01 SDN:L22::TOOL0263'</td>
<td>T-7 XBT</td>
<td>14</td>
</tr>
<tr>
<td>'SDN:P01::ADEPZZ01 SDN:L22::TOOL0435 SDN:L33::001</td>
<td>SDN:P01::TEMPET01 SDN:L22::TOOL0435'</td>
<td>T-4 XBT</td>
<td>16732</td>
</tr>
<tr>
<td>'SDN:P01::ADEPZZ01 SDN:L22::TOOL0592 SDN:L33::710</td>
<td>SDN:P01::TEMPET01 SDN:L22::TOOL0592'</td>
<td>XCTD-2</td>
<td>6</td>
</tr>
<tr>
<td>'SDN:P01::ADEPZZ01 SDN:L22::TOOL0718 SDN:L33::061</td>
<td>SDN:P01::TEMPET01 SDN:L22::TOOL0718'</td>
<td>T-10 XBT</td>
<td>2126</td>
</tr>
</tbody>
</table>

!!!Actions needed to complete crucial metadata information!!!
ENEA, OGS and Ifremer have been informed:

- ENEA will insert all the probe info available and missing metadata
- Ifremer checked and added additional available metadata and ~91k French XBTs
- OGS will insert missing probe and instrument type info

<table>
<thead>
<tr>
<th>EDMO</th>
<th># stations</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>OGS</td>
<td>120</td>
<td>491</td>
</tr>
<tr>
<td>ENEA</td>
<td>136</td>
<td>10488</td>
</tr>
<tr>
<td>HCMR</td>
<td>269</td>
<td>869</td>
</tr>
<tr>
<td>Ifremer</td>
<td>486</td>
<td>13549</td>
</tr>
<tr>
<td>SHOM</td>
<td>540</td>
<td>30876</td>
</tr>
<tr>
<td>MHI</td>
<td>727</td>
<td>1</td>
</tr>
</tbody>
</table>

**Filling data and metadata omissions** highly improves the quality of the infrastructure content and increases users’ confidence.

Reprocessing at the data center level allows:

→ to secure and preserve crucial historical info (international approach of preserving the original data)
→ to apply the latest XBT bias correction (ODV development)
New French XBTs

- **NAT**: +23331 STATIONS
- **MED**: +5355 STATIONS
## Results: database progresses

<table>
<thead>
<tr>
<th>Cruises</th>
<th>Stations</th>
<th>Samples (data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDN2_V2</td>
<td>SDN2_V2</td>
<td>SDN2_V2</td>
</tr>
<tr>
<td>SDC_V1</td>
<td>SDC_V1</td>
<td>SDC_V1</td>
</tr>
<tr>
<td>±%</td>
<td>±%</td>
<td>±%</td>
</tr>
</tbody>
</table>

### 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

|       | 115596 | 162452 | +41%   | 6670529 | 7817193 | +17% |
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**
Results

• 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** highlights omissions and suggests the need of systematic check and data reprocessing at the data centers level
SDC_DATA_TS_V1 release (June 2018)
Products
Catalogue
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

Visualization
Restricted data analysis

- PIDocs are available
- Metadata stats to monitor continuously the amount of restricted data, the providers

GOAL unlock old data and avoid stagnation

### Restricted/Unrestricted ratio

<table>
<thead>
<tr>
<th>STATIONS</th>
<th>unrestricted</th>
<th>restricted</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAT</td>
<td>9091773</td>
<td>23217</td>
<td>0,3</td>
</tr>
<tr>
<td>BAL</td>
<td>10221</td>
<td>10221</td>
<td>1,0</td>
</tr>
<tr>
<td>MED</td>
<td>734957</td>
<td>33022</td>
<td>4,5</td>
</tr>
<tr>
<td>BLS</td>
<td>137723</td>
<td>10528</td>
<td>7,6</td>
</tr>
<tr>
<td>ARC</td>
<td>731286</td>
<td>374</td>
<td>0,1</td>
</tr>
<tr>
<td>NS</td>
<td>1385289</td>
<td>12011</td>
<td>0,9</td>
</tr>
</tbody>
</table>
Unlock restricted data before 2010? 2014?

Black Sea Restricted Stations:
- **BLS**: 7541 (72%), 10067 (96%)
- **ARC**: 374 (100%)
- **BAL**: 8721 (85%), 10033 (98%)
- **MED**: 20640 (62%), 22032 (67%)
- **NAT**: 16995 (73%), 22306 (96%)
- **NS**: 8629 (72%), 12011 (100%)

10522 stations from INSTM with missing depth (QF9)
<table>
<thead>
<tr>
<th>Region</th>
<th>EDMO</th>
<th>Originator</th>
<th>% st</th>
</tr>
</thead>
<tbody>
<tr>
<td>MED</td>
<td>1232</td>
<td>Institut National des Sciences et Technologies de la Mer (INSTM)</td>
<td>51</td>
</tr>
<tr>
<td>BLS</td>
<td>723</td>
<td>State Oceanographic Institute (SOI)</td>
<td>56</td>
</tr>
<tr>
<td>ARC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAL</td>
<td>193</td>
<td>Institute of Meteorology and Water Management National Research Institute, Maritime Branch in Gdynia (IMWM MB)</td>
<td>81</td>
</tr>
<tr>
<td>NAT</td>
<td></td>
<td>IEO</td>
<td>66</td>
</tr>
<tr>
<td>NS</td>
<td>2135</td>
<td>Marine Scotland Science</td>
<td>37</td>
</tr>
</tbody>
</table>

Coming soon → monitoring of restricted/unrestricted rationper data provider to assure that the balance is reasonable
TRAINING and DISSEMINATION

DIVA training 3-6 April 2018 ➔ wide participation, all RC were present
Very good and efficient course

➔ Importance Quality Control SDC WP11 Introduction
Presentation to the 1st SDC training workshop (Serge, Simona, Christine)

WP11 promoted and presented SDC in many workshop and conferences (WP4 presentation)
Work Plan and Timeline

Year 2
- SDC_V1 aggregated datasets
  - Apr 18

Publication of aggregated datasets and PIDocs

Year 3
- SDC_V1 climatologies
  - Feb 19
- 5th QCS loop
  - Jan 19

Year 4
- SDC_V2 aggregated datasets
  - Oct 19
- New Products
  - Apr 20
- SDC_V2 climatologies
  - Sept 20

!!! Correct and ingest data and metadata ASAP !!!
Estimates from metadata: 60% duplicates 80000 stations (+50%)

<table>
<thead>
<tr>
<th></th>
<th>unrestricted</th>
<th>restricted</th>
<th>WOD2013</th>
<th>CORA 5.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>stations</td>
<td>137723</td>
<td>10528</td>
<td>120845</td>
<td>103721</td>
</tr>
<tr>
<td></td>
<td>148251</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ongoing Activities: collaboration with CMEMS

Ins TAC

AGENDA

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
Advanced estimates of GLOBAL climatologies
K. Shahzadi, N. Pinardi, M. Zavatarelli (UNIBO), S. Lyubartsev (CMCC), S. Simoncelli (INGV)

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

Need to develop a new NONLINEAR QC (NQC) to reduce the representativeness error
Climatologies SDC_CLIM_TS_V1

- 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
Conclusions

- Introduction of **PI Docs** represented a very good progress
- SDC\_DATA\_TS\_V1 publication → *dataset 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 (upcoming milestone)

- **WP11** met regularly online to monitor and harmonize the activities
- High participation/great collaboration/very good progresses
- Deliverables and reports were all submitted in time
Open Issues

- **underway data** → data have been subsampled in SDC_DATA_TS_V1 release (1 over 7) → wish to manage underway data (FerryBox and TSG) separately and provide the full resolution (ongoing discussion)

- **restricted data** → need to reduce its percentage (unlock data older than 2010)

- **Time series** → fragmented and difficult to manage for QC and data products → how do we want to proceed? WP11 wish a reprocessing at the data center level

- Meeting with Ins-TAC CMEMS ASAP
Thanks to
Christine Coatanooan, IFREMER
Volodymyr Myroshnychenko, METU
Örjan Bäck, SMHI
Helge Sagen, IMR
Serge Scory, RBINS
Reiner Schlitzer, AWI
Michèle Fichaut, IFREMER
Dick Schaap, MARIS