



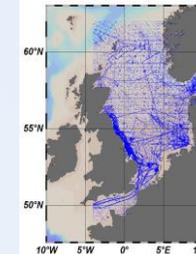
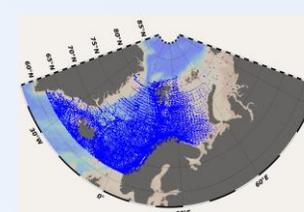
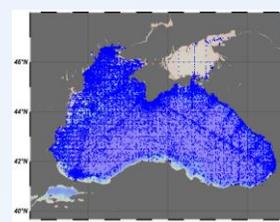
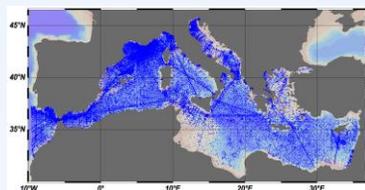
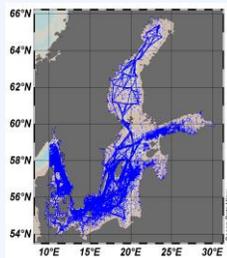
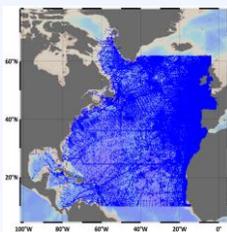
SeaDataCloud

WP11 achievements and the way forward

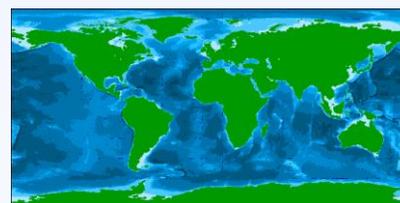
S. Simoncelli and WP11 team

- **Improve the quality of SDN infrastructure** content through systematic quality assessment (every 2 years)
- **Integrate external datasets** (CMEMS, WOD) to increase temporal and spatial coverage/resolution and further improve products' quality
- **Generate the best data products** to serve different user groups (operational oceanography, climate, marine environment, institutional, academia) adopting the most advanced methodologies
- **Increase user uptake providing reliable information** of the full product generation process and its quality
- test the Virtual Research Environment (VRE)

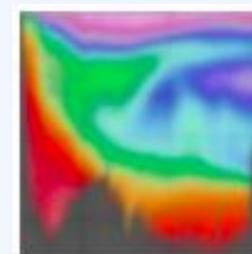
European marginal seas



Global Ocean



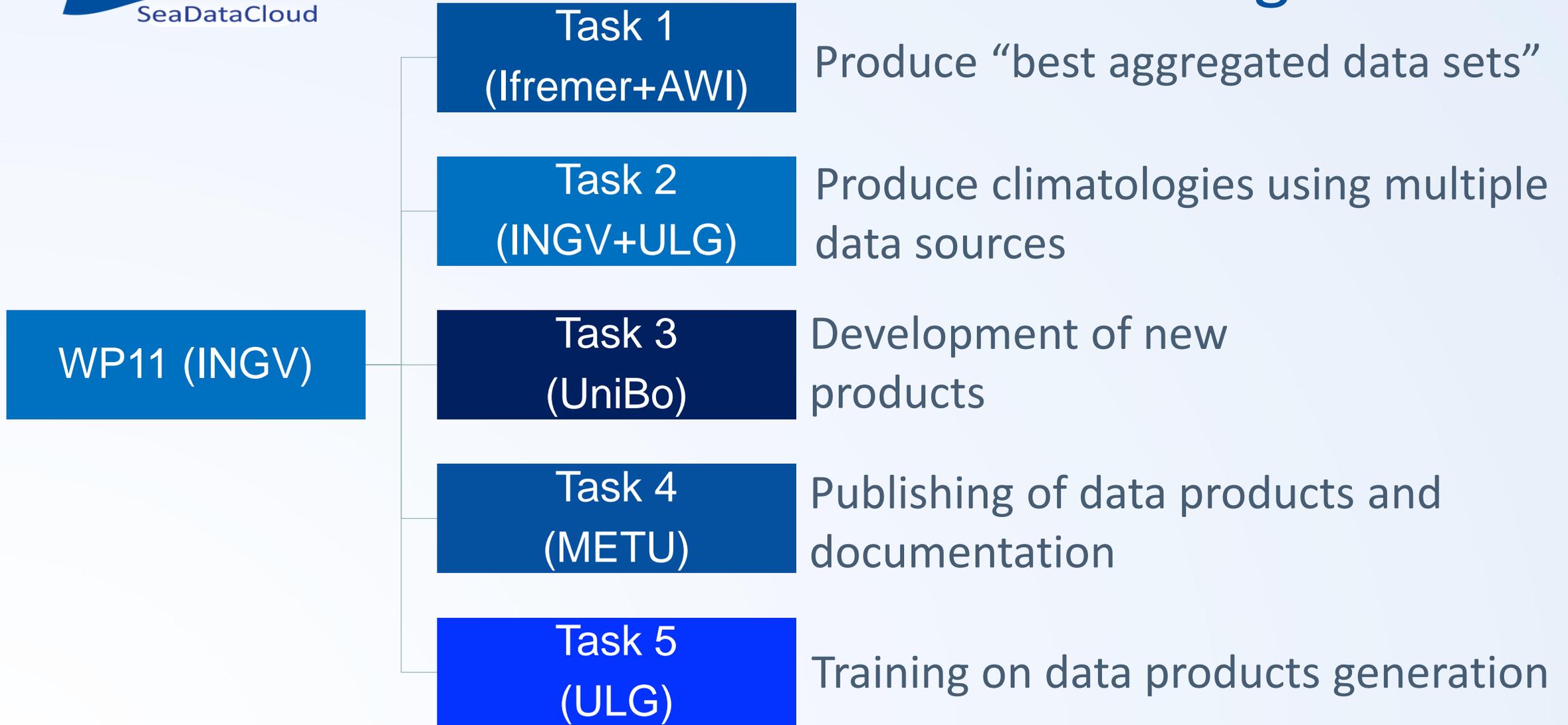
tools



Ocean Data View



Work Organization



QC and data sets (T1)

- release of V1 (Jun2018) and V2 (May2020) data sets
- Product Information Documents annexed to each product (data stats, methodology, QC results, tech specs, usability, improvement since previous version)
- QC procedure has been harmonized and improved, introducing check per time period, per depth layer/level, per sub-regions
- new metadata analysis thanks to ODV developments (i.e. data providers, instrument type) → identification of systematic errors, data and metadata omissions, duplicates
- feedback on data anomalies/duplicates to data providers

- production and release V1 and **V2** climatologies
- data **integration with external sources** CMEMS CORA and/or WOD (metadata track to provide a feedback on anomalies/duplicates)
- SDC climatologies have **higher vertical resolution** than SDN ones, include **decadal fields** and **different time resolutions** (monthly and seasonal)
- **global climatologies** have been generated with DIVAnd software using WOD data implementing a Non Linear Quality Control (**NLQC**) procedure
- each climatology, its methodology and validation described in **PIDocs** which passed through a double stage revision
- DIVAnd uptake and massive code testing made the code more robust
- **consistency analysis with WOA climatologies**

New products have been developed using DIVAnd for the first time to show the potential of the tools and SDN data, to enlarge our target audience and increase the scientific impact of SDN.

- research outcomes (GLO, BLS, ULG)
- beta versions (NAT, MED)
- improved statistics for data QC (BAL)

New data products (T3)

GLO	Seasonal Mixing Index at 1/4° (2003-2017)
GLO	AOU at 1/4° (2003-2017)
BAL	regional and sub-regional T and S monthly stats
NAT	Monthly climatology for MLD at 1/4°
MED	Monthly climatology for MLD at 1/8°
MED	OHC time series and trend (0-700m; 0-2000m)
BLS	Monthly climatology CIL at 1/8°
BLS	Decadal seasonal CIL cold content at 1/8°
BLS	sliding decades CIL cold content at 1/8°
ULG	Currents climatologies from HF radars

Products' publication (T4)

- each product available in the Sextant metadata catalogue with annexed documentation describing the methodology applied and informing about the product quality
- products and PIDocs have DOIs and acknowledge all data providers

Two DIVAnd training have been organized:

1. Apr2018 in Liege (ULG) → beginner users
 2. Jan2020 in Bologna (ULG-UNIBO) → both beginners and advanced users
- **training activities crucial for DIVAnd uptake (5/7 in V1 and 7/7 in V2) and support for products generation**
 - **success in terms of organisation and participation** → increasing interest of the scientific community
 - included additional seminars on adjacent topics (data quality control, data assimilation, consistency analysis for products):
 - how to use DIVAnd to get Ocean Monitoring Indicators, BGC indicators, coastal HR maps
 - how to harmonize the working flow
 - a posteriori QC to assure data representativeness (residuals, stats)

- collaboration with WP9 (recommendations) and WP10 (VRE)
- meeting with **NOAA** to discuss about the integration of WOD data with SDN for V2 climatologies and SDN acknowledgement (unique identifier)
- meeting with **CMEMS** to discuss about the integration of CMEMS data in V2 climatologies (unique identifier)
- collaboration with **IQuOD** team on **duplicates, uncertainties, XBTs** and use of Machine Learning in QC procedures
- "Feedback on found data problems in WOD" and the same for CORA

- production of V1 and V2 datasets and V1 and **V2** climatologies
- data **integration with external sources**: CMEMS CORA and WOD
- SDC climatologies have **higher vertical resolution** than SDN ones, include **decadal fields and different time resolutions** (monthly and seasonal)
- **new global products** have been generated with DIVAnd software using WOD data but also implementing a **NLQC**
- new data products: some are beta versions, others are new scientific advancements
- joint meetings and collaboration with CMEMS, WOD, IQuOD
- introduced products documentation, metadata analysis, detection of data omissions
- substantial contribution to the improvement of the analysis tools (ODV and DIVAnd)
- WP11 team was beta tester of webODV and the VRE, providing feedback and recommendations

Scientific Production (on-going papers)

- *Mieruch et al.* paper on ML application → **submitted**
- *Barth et al.* paper on surface currents mapping with DIVAnd → **in review**
- *Simoncelli et al.* book chapter on Data Products in Ocean Science Data → **in preparation**
- *Reseghetti et al.* (collaboration with ENEA on XBT reprocessing) → **in preparation**
- *Oliveri et al.* on SDC time series (aggregation, QC) and integration with CMEMS on the Med domain → **in preparation**
- *Shahzadi et al.* paper on GLO climatology → **in preparation**
- *Simoncelli et al.* paper on Med climatology → **in preparation**
- **data set paper on SDC data sets**
- **data set paper on SDC climatologies**

- data not processed or validated are still present in the CDI (raw CTDs, raw XBTs)
- there exist many duplicates and a precise metadata description is necessary in order to identify them
- data ownership is critical (e.g. Med XBT tracks have been duplicated from different data providers but metadata and data are different)
- there still exist many data omissions (existing data but not yet in SDN)

Duplicates

- how to define them?
- how to manage them during the data ingestion process?
- how to take decisions in case of duplicates

QAS

- how to implement/verify a more efficient/transparent/automated data set update process?
- how to implement an efficient feedback between producers and data providers in the VRE?
- how to keep updated and harmonized QC practices of data providers?

- collaboration and communication between data providers and producers should be enhanced → QAS loop on the VRE
- improve the QC with ML tools (ODV) or shared automatic procedures (Max/Min from CMEMS INS TAC) to be run on the VRE
- VRE should provide more computing resources (big data issue, e.g. NAT climatology) and a shared environment in order to improve the collaboration across regions or product applications
- time series data management, analysis and release
- underway data released as separate dataset at full resolution
- introduce quality on location and timing of data (profile flags)
- improve the metadata description of data
- synergy among the data providers to avoid duplicates and rescue the best metadata information available

- WP team is very collaborative and put a lot of efforts in all activities
- ODV developments help in optimizing the workflow and QC
- the continuous support of ULG and the DIVAnd feedbacks from producers allowed to solve many issues, making the code more robust
- the publishing of data products, dissemination activities needed a lot of dedication too
- deliverables were often postponed due to the many dependencies from other WPs activities, but we succeeded!