SeaDataCloud

WP9 progress and results

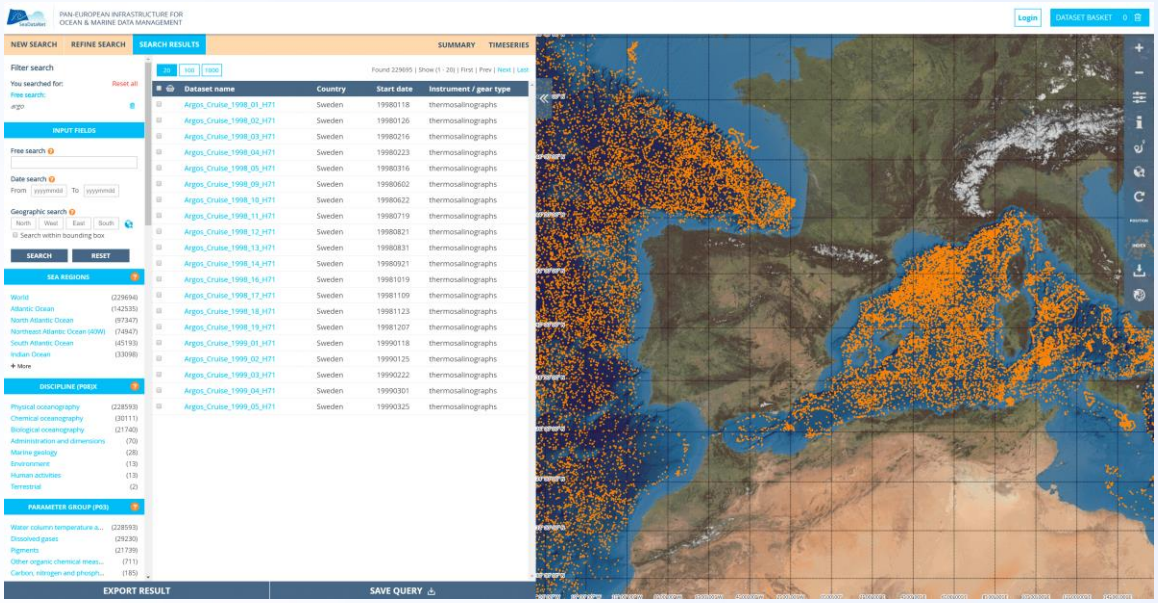
Dick MA Schaap – MARIS

SeaDataCloud Plenary Group meeting – Web conference, 29 October 2020

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

SeaDataNet CDI service

- **CDI Data Discovery and Access service** giving facilities for searching and retrieving marine data sets
- Upgraded CDI service deployed in October 2019



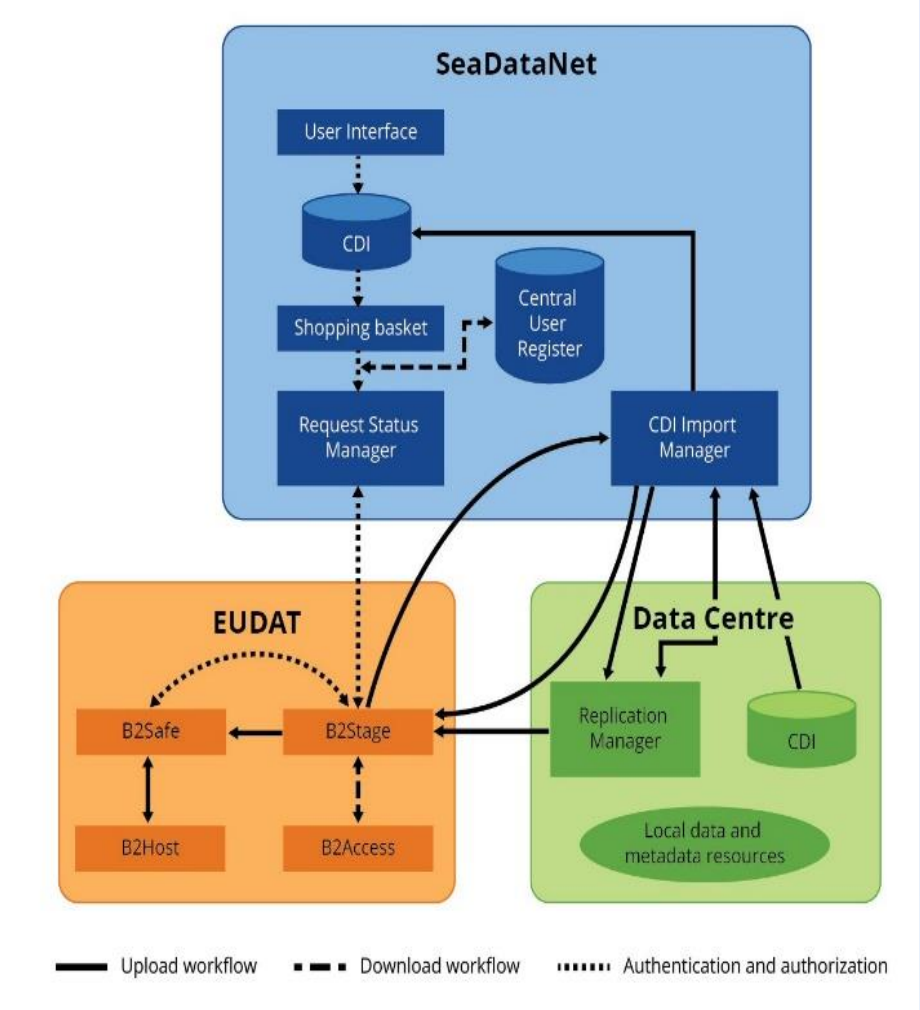
- https://cdi.seadatanet.org/search

Features of GUI

- Combination of search criteria from pull down lists + search by facets + powerfull full text search over all contents
- Full screen mapping
- IT Components:
  - GeoServer for mapping
  - SQL server for database
  - Elastic Search for free text search (very fast!)
- MySeaDataNet for customized services and integration:
  - Marine-ID registration
  - Accepting SDN licence
  - Shopping basket
  - RSM for users to follow data requests and to download
  - RSM for CDI data centres to oversee data requests and to take decisions on requests for restricted data
  - Import Manager for CDI data centres to update and add new CDI – Data entries

Upgraded CDI service set-up

- **Local software tools** at data centres to prepare ingestions
- **Replication Manager** (RM) at data centres for exchanging to Import Manager and EUDAT cloud
- **EUDAT cloud** with adapted EUDAT services
- Upgraded **CDI User Interface**, ordering and downloading facility



Activities undertaken

- Finalising technical developments for a robust and operational system with versions for development, testing, import, and production (like current CDI system)
- Validating and working up the central CDI database and associated unrestricted data collection, including completing different formats.
- Performing various actions for further improving the consistency between the CDI metadata and data sets
- Integrating all components for public launch
- Amending selected website pages
- Connecting all SeaDataNet nodes to the upgraded CDI service

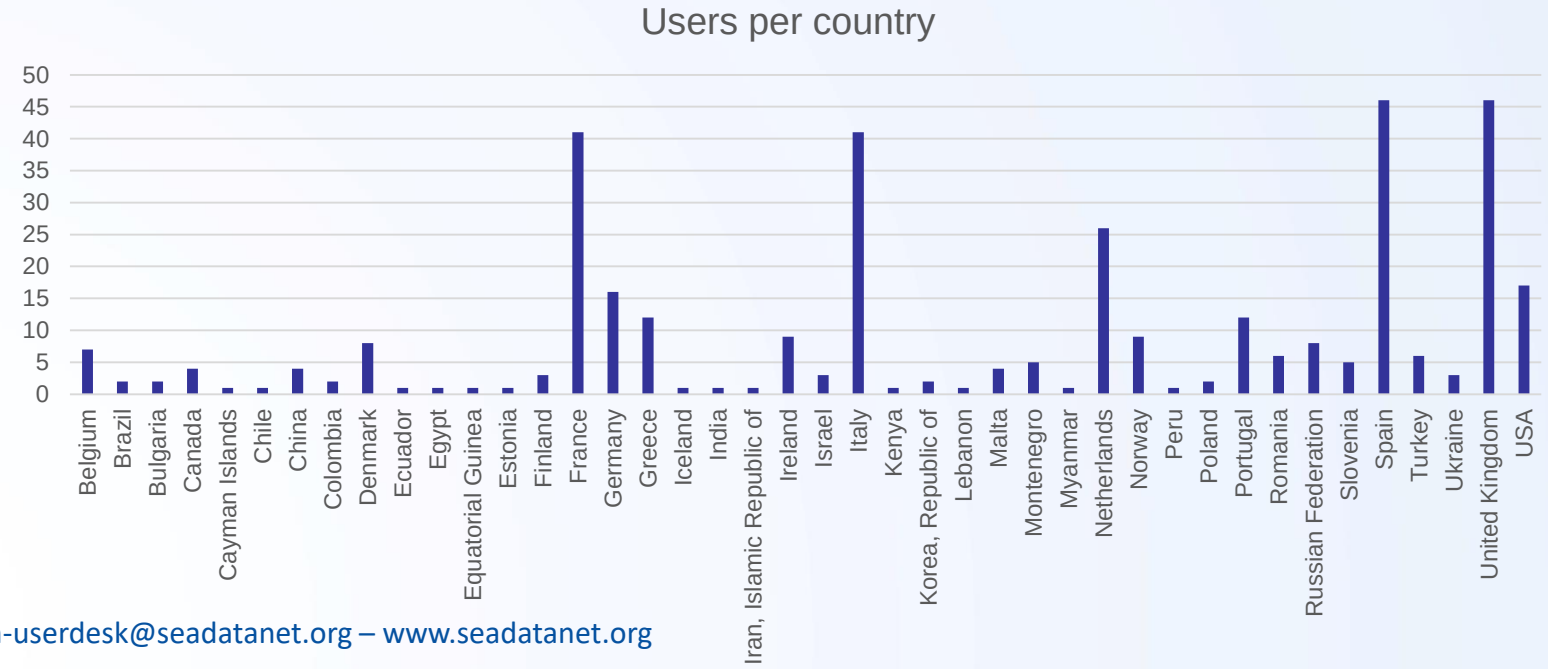
CDI service operation since October 2019

- Import service works fine and lot of activity; we had some hick-up's but solved by EUDAT helpdesk and upgrades of RM

month	batches	partners	new_cdis	updated_cdis
2019-10	8	6	13823	46
2019-11	13	8	1377	5
2019-12	35	21	5741	3268
2020-1	60	16	13740	32466
2020-2	60	18	23310	70911
2020-3	72	17	23632	44210
2020-4	39	13	15815	7981
2020-5	62	16	19391	82473
2020-6	50	12	81013	2490
2020-7	60	17	13735	14439
2020-8	30	13	27352	3190
2020-9	102	26	15263	25051
2020-10	96	26	71900	27235
Total	687		326092	313765

CDI service user activity since launch of upgraded service

- In total **374** unique users in one year
- In total **1715** transactions for **4.079.526** CDI – data sets
- Extra for robot harvesting: **2590** transactions for **10.448.584** CDI – data sets for use in SeaDataCloud products and EMODnet products
- Users from **44** Countries



Adding extra metadata to CDI interface

- Adding more metadata tags to CDI query and details, using Linked Data principles, to be developed in synergy with ENVRI-FAIR project and EMODnet
- Using CSR for Ship name (C17) and Platform type (L06)
- Using EDMERP for project type (monitoring / research)
- Using geographic coordinates for tagging Sea regions (C19), ICES areas, Helcom areas, OSPAR areas, MSFD regions
- Using EMODnet Bathymetry REST service to complete empty water depths
- Extracting P01 from buffers to ensure consistency between P01 in data and P02 in CDIs

Connecting SeaDataNet CDI nodes

Connecting SeaDataNet CDI nodes

- **2 models:**
  - **Replication Manager (V1.0.47)**: self managing population of updates and new entries to CDI catalogue via RM - Import Manager + delivery of restricted data via RSM, if requests accepted
  - **Interim Solution**: population of updates and new entries to CDI catalogue facilitated through CDI support desk + delivery of restricted data on bilateral basis, if requests accepted
- **78 CDI nodes** have agreed to install the Replication Manager (RM)
  - **71 nodes** in production mode of which **24** with latest RM
  - **7 nodes** not yet in production mode (MHI (UA), NOA (GR), EPA (LT), AU (TR), DHMO (UA), RBINS (BE), MHI-RAS (RU)
- **46 CDI nodes** function with Interim solution
- **All** SDC partners with RM, except for RBINS (BE)

ACTION FOR DATA CENTRES: Upgrade to latest RM for best performance

Status of SeaDataNet tools

- **MIKADO** release 3.6.1 on 14/05/2020
- MIKADO release 3.6.2 on 28/05/2020
- Next MIKADO release 3.6.3 ➔ November 2020

- **NEMO** current version is 1.7.0, released on 16/04/2020
- Plan for new NEMO development in near future

- **OCTOPUS** current version is 1.5.3, released on 16/04/2020
- Next Octopus 1**.6.0 version** ➔ November 2020
- Includes: Checker of HF-Radar netCDF data files
- Early 2021 – new version planned for handling Glider data formats

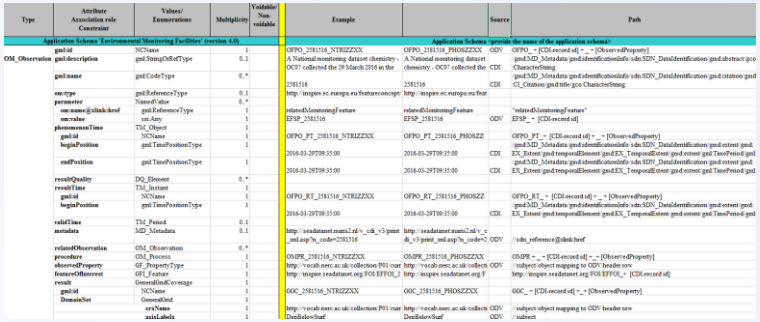
- Latest **Replication Manager** release is V1.0.47, released 18/08/2020

SdnToInspire Transformation service

- Target: service to transform SeaDataNet formats into INSPIRE formats (metadata and data)
- Earlier analysis in D8.6. This recommends the use of data models from the INSPIRE data themes:
  - Environmental Monitoring Facilities (EF)
  - Oceanographic Geographical Features (OF)
  - Observations and Measurements (O&M)
- Two use cases:
  - research vessel with subsurface mooring and lowered unmanned
  - one for water sampling for analyses of nutrients
- Mappings prepared for both cases from SeaDataNet to INSPIRE

Cruise: D324
CSR: 20097458

	CDI	ODV	GML Encoding	Matching Tab
Platform			SDN_EF_Vessel_EFVS_74E3	EF Vessel
Cruise			SDN_CruiseActivity_EFAC_74E3_D278	Activity
Time Series Data			Temperature values, half daily	
• **Sampling Point**	2075842	b0686762	SDN_EF_SamplingPoint_EFSP_D278_TEMPPR01	EF SamplingPoint
• **FeatureOfInterest**	2075842	b0686762	SDN_FOI_EFFOI_D278_TEMPPR01	FoI
• **Process**	2075842	b0686762	SDN_Process_OMPR_TEMPPR01	Process
• **Observation**	2075842	b0686762	SDN_PointTimeSeriesObservation_OFTS_D278_TEMPPR01	TimeSeriesObservation
Profile Data			Chlorophyl at pressure depths	
• **Sampling Point**	1597207	b1061981	SDN_EF_SamplingPoint_EFSP_D278_CPHLPM01	EF SamplingPoint
• **FeatureOfInterest**	1597207	b1061981	SDN_FOI_EFFOI_D278_CPHLPM01	FoI
• **Process**	1597207	b1061981	SDN_Process_OMPR_CPHLPM01	Process
• **Observation**	1597207	b1061981	SDN_ProfileObservation_OFPO_D278_CPHLPM01	ProfileObservation
Trajectory Data			Sea-floor depth {bathymetric depth}	
• **Sampling Point**	2034903	b1051624	SDN_EF_SamplingPoint_EFSP_D278_MBANCT01	EF SamplingPoint
• **FeatureOfInterest**	2034903	b1051624	SDN_FOI_EFFOI_D278_MBANCT01	FoI
• **Process**	2034903	b1051624	SDN_Process_OMPR_MBANCT01	Process
• **Observation**	2034903	b1051624	SDN_ProfileObservation_OFTO_D278_MBANCT01	TrajectoryObservation



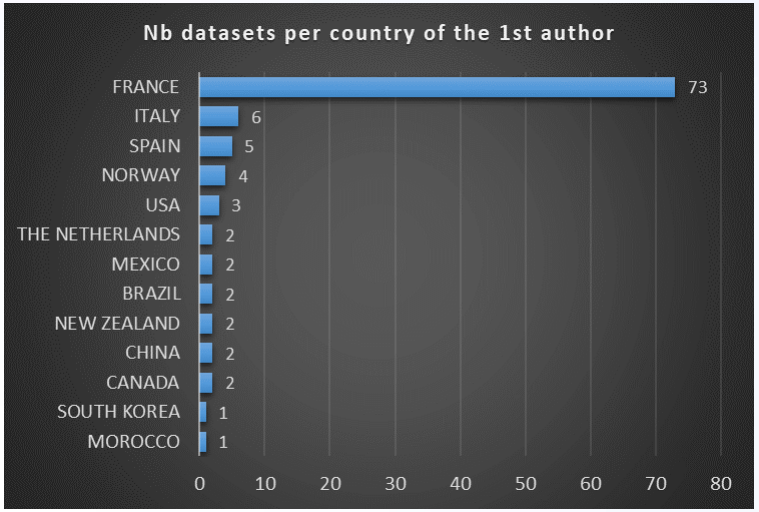
SdnToInspire Transformation service

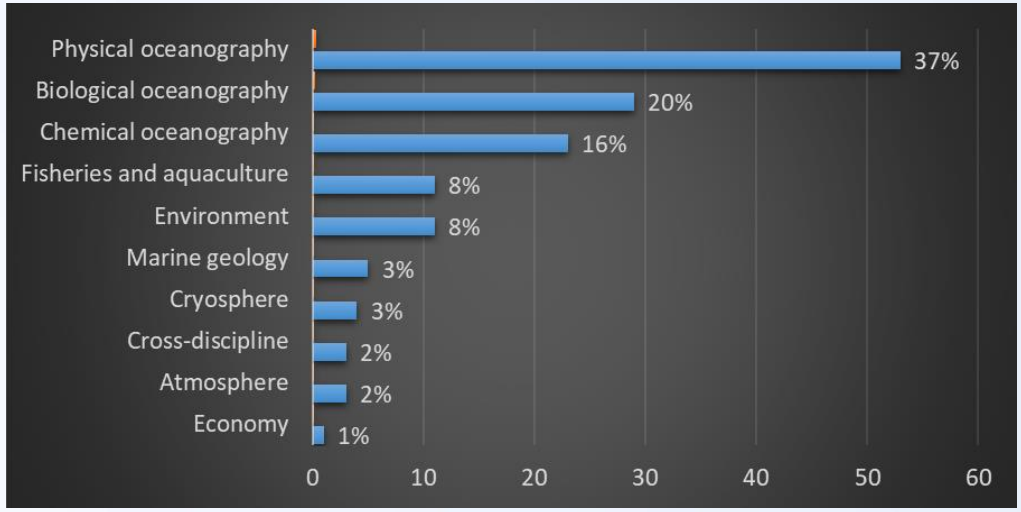
- **INSPIRE Conversion Tool** developed by IFREMER
- A batch process that takes as input a list of files in SeaDataNet formats (CDI, CSR, ODV)
- It implements the SDN-INSPIRE mapping of the BODC SeaDataNet and OGS Nutriment use cases
- It requires access to BODC vocabularies
- This java software runs on a command line
- It generates INSPIRE compliant **xml** files:
  - **Vessel, Activity, SamplingPoint, FeatureOfInterest, Process**

- Validation performed against INSPIRE validator with support of CNR
- This gives good validation results, but also a few items which need resolving on short term
- Will be finalised and documented in **D9.8** as Proof of Concept
- Deployment as online service is best way forward

SEANOE DOI publishing service

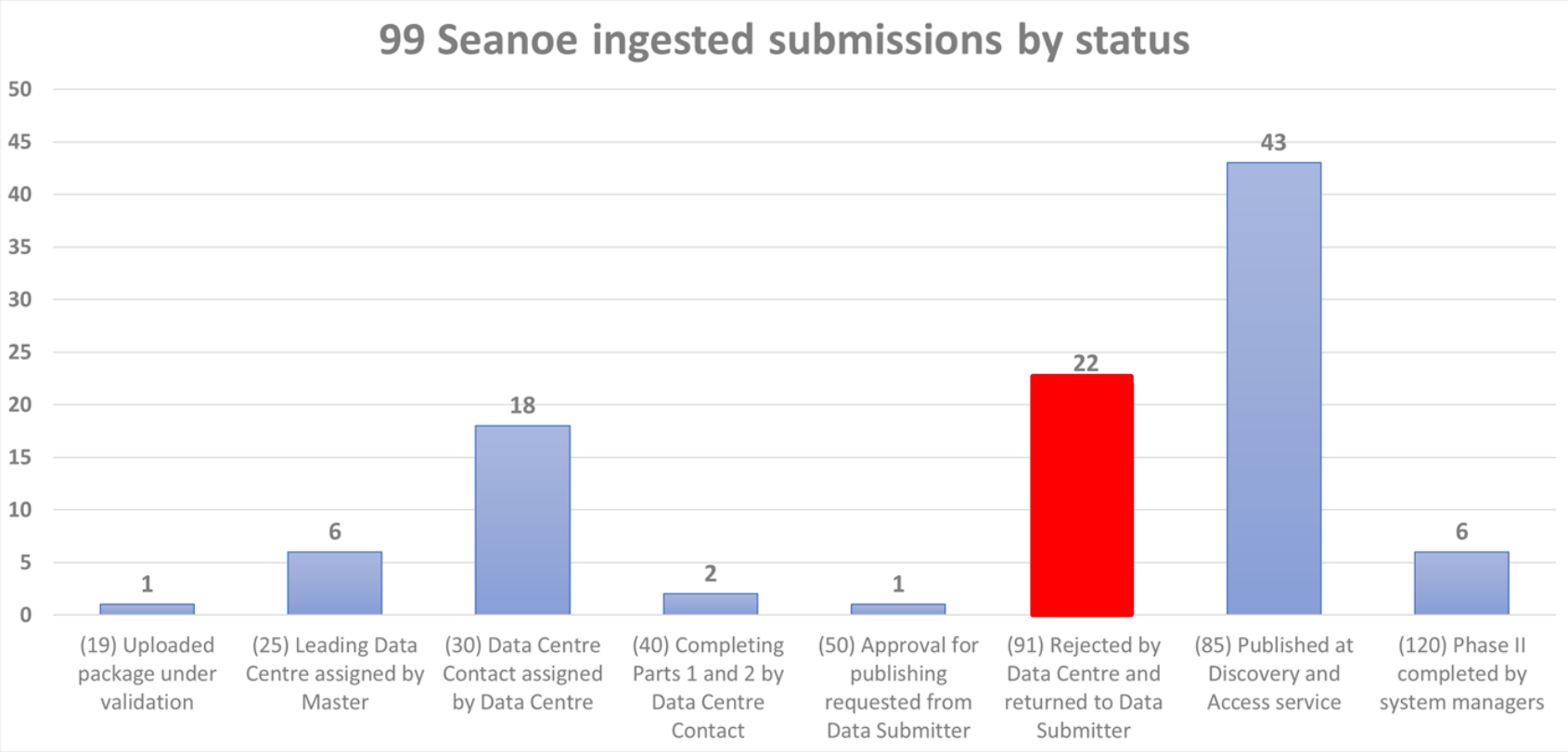
- **SEANOE service** for the publication of marine science research data as citable resources (DOI); at SeaDataNet homepage
- Started in 2015; Today 686 datasets of which 314 under CC-By
- Link set-up between **SEANOE and EMODnet Ingestion**
- So far 99 Data sets forwarded, which are manually selected to be fit for EMODnet Ingestion and in particular, further elaboration
- Learning curve to get the selection right





SEANOE DOI publishing service

- Status of processing so far in EMODnet Ingestion by assigned data centres



- This has resulted in further improvements in the overall procedure
- Will be reported in **D9.16** by IFREMER

Other earlier activities in WP9

- Developing integrated online services for ingesting autonomous observatory data => resulting in the **SeaDataNet SWE Toolkit and demonstrator**

- Expanding SeaDataNet capability for handling different data types
  - Ingesting, validating, long-term storage and access of **HF Radar data**
  - Ingesting, validating, long-term storage and access of **Flow Cytometer data**
  - Ingesting, validating, long-term storage and access of **Glider data**

- Integration of external datasets from international programmes and organisations => **SeaDataNet data brokerage service**

- Develop a preconfigured and pre-built **virtual appliance system** as a complete solution to new data centres to connect to the CDI service => alternative for the Download Manager

Summary

- All targets for WP9 have been achieved and have delivered great results, which are useful, in use, and (being) adopted as basis for further developments and projects

- Remaining:
  - Finalising last bits

  - Contributing to Final Report, for all WP9 technical activities and results

  - Preparing final Deliverables:
    - **D9.8 -** SdnToInspire Transformation service by IFREMER
    - **D9.16** - SEANOE – EMODnet Ingestion, by IFREMER

A great thanks to all the colleagues, both in the TTG and the SeaDataCloud overall project, for the inspiring cooperation and team spirit!