



# SeaDataCloud

Deployment of the upgraded CDI service

Dick MA Schaap – SeaDataCloud Technical Coordinator

SeaDataCloud SC + Plenary Group meeting – Brest –  
France, 16 - 18 October 2019

TTG meeting, May 2019  
[sdn-userdesk@seadatanet.org](mailto:sdn-userdesk@seadatanet.org) – [www.seadatanet.org](http://www.seadatanet.org)

# New CDI service has been launched!

SeaDataCloud PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

Login DATASET BASKET 0

NEW SEARCH REFINER SEARCH SEARCH RESULTS SUMMARY TIMESERIES

Filter search  
You searched for: **argo**  
Free search:

INPUT FIELDS

Free search

Date search  From  To

Geographic search  North West East South Search within bounding box

SEARCH RESET

SEA REGIONS

- World (229694)
- Atlantic Ocean (142535)
- North Atlantic Ocean (97347)
- Northeast Atlantic Ocean (40W) (74947)
- South Atlantic Ocean (45193)
- Indian Ocean (33098)
- + More

DISCIPLINE (P08X)

- Physical oceanography (228593)
- Chemical oceanography (30111)
- Biological oceanography (21740)
- Administration and dimensions (70)
- Marine geology (28)
- Environment (13)
- Human activities (13)
- Terrestrial (2)

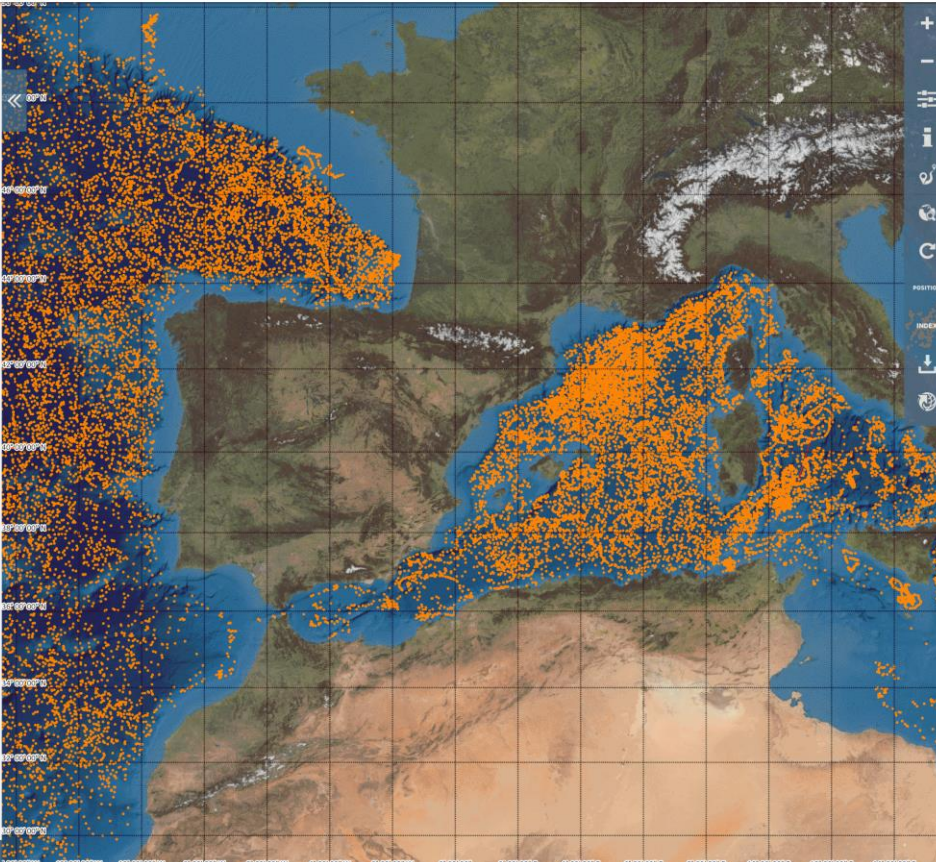
PARAMETER GROUP (P03)

- Water column temperature a... (228593)
- Dissolved gases (29230)
- Pigments (21739)
- Other organic chemical meas... (711)
- Carbon, nitrogen and phosph... (185)

Found 229695 | Show (1 - 20) | First | Prev | Next | Last

Dataset name	Country	Start date	Instrument / gear type
Argos_Cruise_1998_01_H71	Sweden	19980118	thermosalinographs
Argos_Cruise_1998_02_H71	Sweden	19980126	thermosalinographs
Argos_Cruise_1998_03_H71	Sweden	19980216	thermosalinographs
Argos_Cruise_1998_04_H71	Sweden	19980223	thermosalinographs
Argos_Cruise_1998_05_H71	Sweden	19980316	thermosalinographs
Argos_Cruise_1998_09_H71	Sweden	19980602	thermosalinographs
Argos_Cruise_1998_10_H71	Sweden	19980622	thermosalinographs
Argos_Cruise_1998_11_H71	Sweden	19980719	thermosalinographs
Argos_Cruise_1998_12_H71	Sweden	19980821	thermosalinographs
Argos_Cruise_1998_13_H71	Sweden	19980831	thermosalinographs
Argos_Cruise_1998_14_H71	Sweden	19980921	thermosalinographs
Argos_Cruise_1998_16_H71	Sweden	19981019	thermosalinographs
Argos_Cruise_1998_17_H71	Sweden	19981109	thermosalinographs
Argos_Cruise_1998_18_H71	Sweden	19981123	thermosalinographs
Argos_Cruise_1998_19_H71	Sweden	19981207	thermosalinographs
Argos_Cruise_1999_01_H71	Sweden	19990118	thermosalinographs
Argos_Cruise_1999_02_H71	Sweden	19990125	thermosalinographs
Argos_Cruise_1999_03_H71	Sweden	19990222	thermosalinographs
Argos_Cruise_1999_04_H71	Sweden	19990301	thermosalinographs
Argos_Cruise_1999_05_H71	Sweden	19990325	thermosalinographs

EXPORT RESULT SAVE QUERY

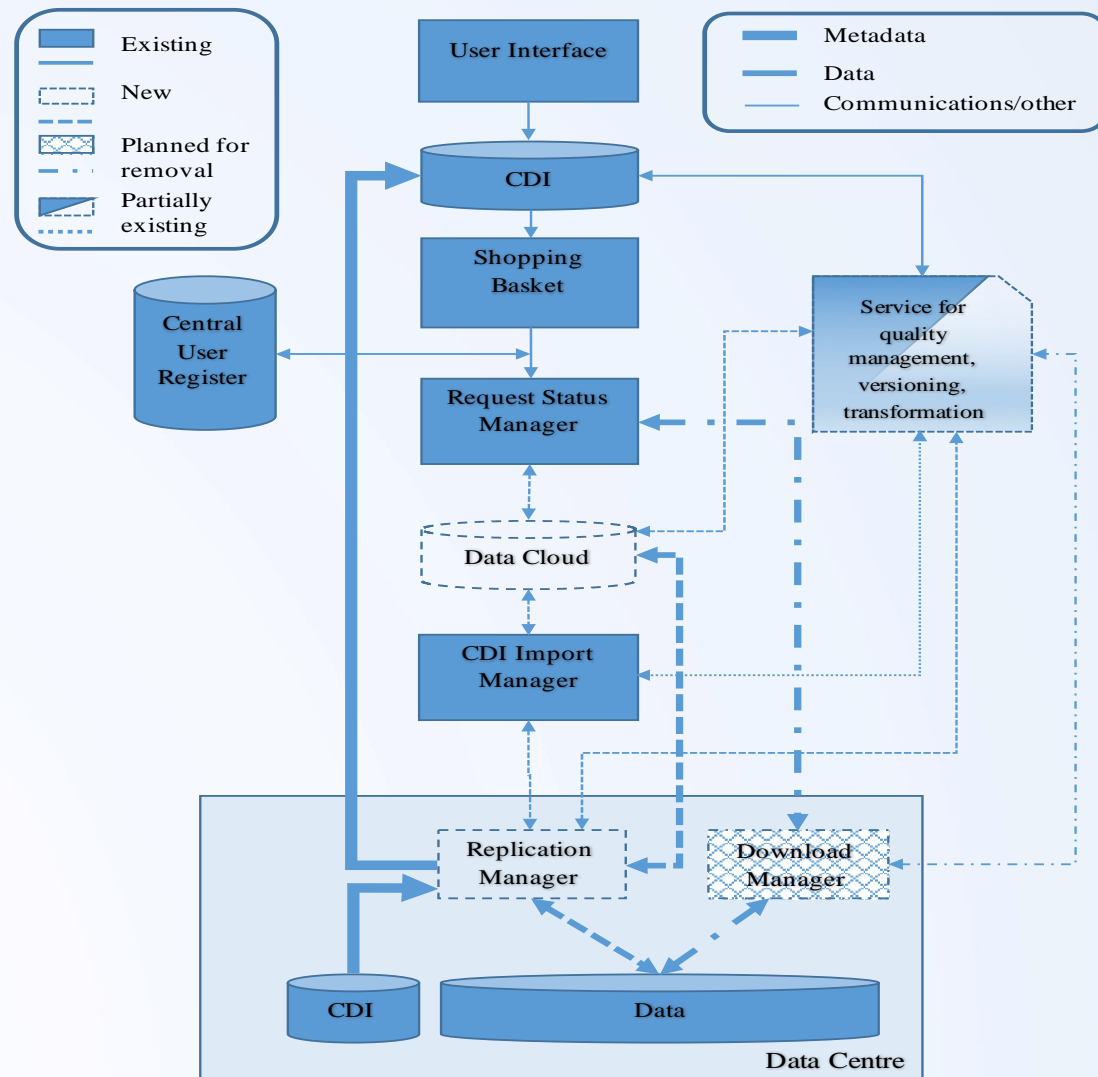


<https://cdi.seadatanet.org/search>

# Features of GUI

- Combination of search criteria from pull down lists + search by facets + powerful 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

# New CDI service architecture



# New CDI service components

- **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

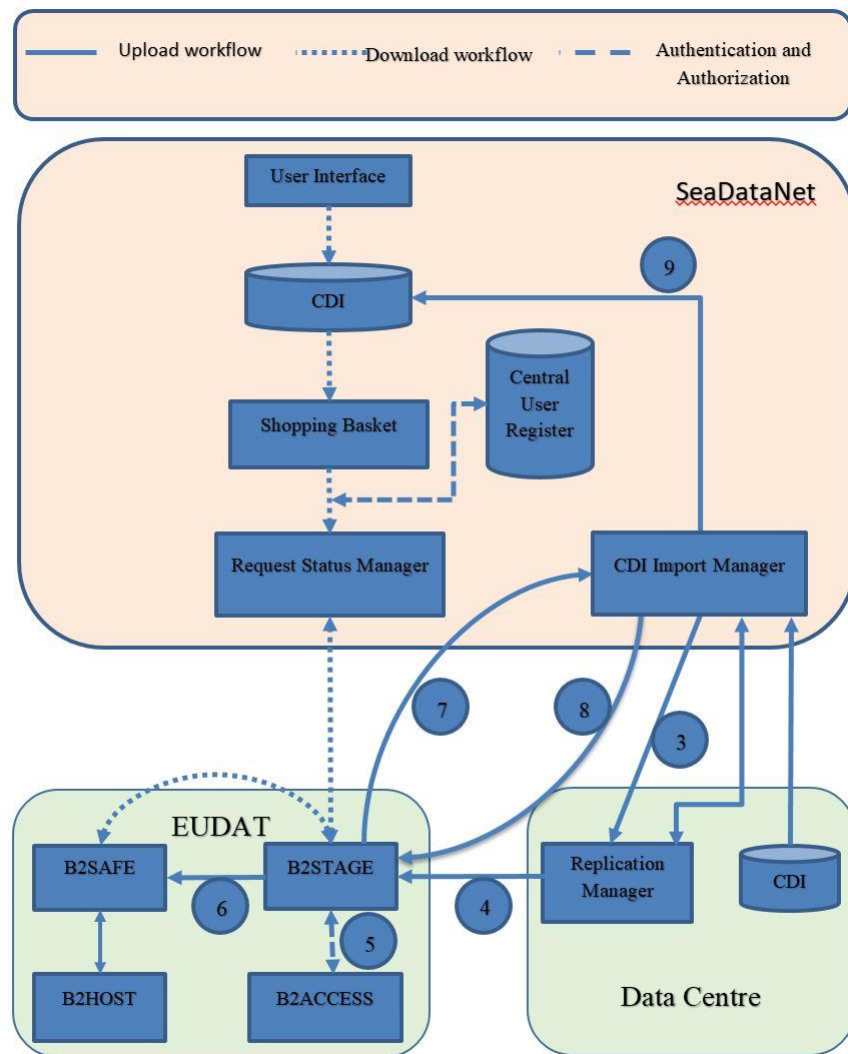


Figure ...: High level architecture

## Activities undertaken

- Finalising technical developments for a robust and operational system with versions for development, testing, import, and production (like current CDI system)
  - Connecting all SeaDataNet nodes to the upgraded CDI service
  - Validating and working up the central CDI database and associated unrestricted data collection, including completing different formats.
  - Integrating all components for public launch
  - Amending selected website pages
- [sdn-userdesk@seadatanet.org](mailto:sdn-userdesk@seadatanet.org) – [www.seadatanet.org](http://www.seadatanet.org)



# Connecting SeaDataNet CDI nodes



# Connecting SeaDataNet CDI nodes

- 2 models:
  - **Replication Manager:** 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
- **73 CDI nodes** have agreed to install the Replication Manager (RM)
  - **57 nodes** have latest RM (1.0.39) installed; some still to test
  - **8 nodes** have to upgrade to latest RM (takes 30 min incl test)
  - **8 nodes** still have to start with RM installation
- **25 CDI nodes** have agreed to function with Interim solution
- **8 CDI nodes** have indicated not to continue; however their unrestricted data sets will remain part of the shop collection



# 8 RM nodes to upgrade to latest RM

EDMO	NAME
732	Karadeniz Technical University, Faculty of Marine Sciences (TR)
961	National Institute of Meteorology and Hydrology, BAS (BG)
1169	Odessa National I.I.Mechnikov University (UA)
1703	National Observatory of Athens, Institute of Geodynamics (GR)
2176	Ankara University (TR)
2276	OGS - Infrastructures Division (IT)
3009	ISPRA-Institute for Environmental Protection and Research (IT)
3078	Mediterranean Institute of Oceanography - LUMINY (MIO) - UMR (FR)

# 8 CDI nodes who have to undertake RM installation and test

EDMO	NAME
565	Centre de Données Géophysiques (CDG) / CNRS - Université de Strasbourg (FR)
723	State Oceanographic Institute (SOI) (RU)
1002	Aquitaine Observatory of Sciences of the Universe – University of Bordeaux) (FR)
1022	BRGM / Office of Geological and Mining Resources (FR)
2172	UB/ Barcelona University (ES)
2231	Danube Hydro-meteorological Observatory (UA)
3092	Institute of Geology & Mineral Exploration (GR)
632	Royal Netherlands Meteorological Institute (NL)

# 8 CDI nodes who will not maintain their collection, but data in shop

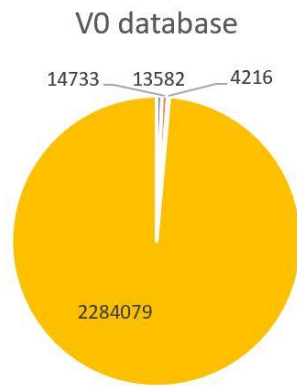
EDMO	NAME
924	Russian State Hydrometeorological University, St-Petersburg (RU)
2228	Institute of Geological Sciences – IGS (UA)
3234	PANGAEA - Data Publisher for Earth & Environmental Science (DE)
1543	Geological Survey of Denmark and Greenland (GEUS) (DK)
731	Dep of Navigation and Hydrography and Oceanography, Turkish Navy (TR)
3393	SOVON Dutch Centre for Field Ornithology (NL)
635	TNO Geological Survey of the Netherlands (NL)
45	Marine Biological Association of the United Kingdom (UK)



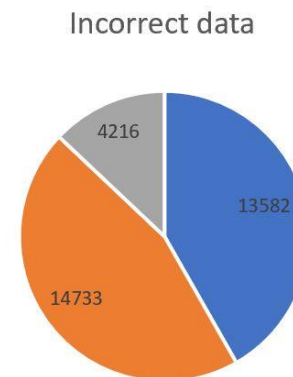
# Preparing the V0 unrestricted data cloud

- MARIS has checked all unrestricted ODV files with OCTOPUS (after downloading to local buffer)
- List of data files with issues; large part corrected by partners or MARIS
- CDIs for remaining error files de-active in V0 CDI catalogue
- Using OCTOPUS, MARIS has generated NetCDF files for all ODVs (where possible – few exceptions)
- All validated ODV + NetCDF data sets and other non-ODV data sets uploaded to Central data cloud at EUDAT with PIDs (Handle system)
- **Remaining error files to be corrected by CDI nodes**

# Preparing the V0 unrestricted data cloud



■ Deactivated ■ Incorrect ODV to be corrected ■ To be corrected by MARIS ■ Correct



■ Deactivated ■ Incorrect ODV to be corrected ■ To be corrected by MARIS

- Most data (98,5%) was corrected before the data collection was frozen
- Circa 2.3 million CDIs and > 4 million data files
- 1,5% needed to be corrected. Undertaken by MARIS or partners. Still part remaining.
- 13582 records de-activated because data no more available



# Remaining corrections

EDMO	Name	Octopus 5.0.8	Octopus 5.1
43	BODC (UK)	1319	319
120	OGS, Division of Oceanography (IT)	20	20
144	CNR - ISMAR) – Ancona (IT)	87	87
193	IMGW MB (PL)	1158	1158
353	IEO/ Spanish Oceanographic Institute	173	173
681	RIHMI-WDC (RU)	34	34
693	TSU-DNA (GE)	24	24
698	Latvian Institute of Aquatic Ecology (LV)	1369	66
700	IOF (HR)	372	372
1104	SYKE (FI)	802	802
1229	NIB (SI)	8	8
1578	RBINS (BE)	836	836
4537	ORION (CY)	3	3
Total		6205	3902

# CDI nodes that indicated to have corrected data files

EDMO	Name	Octopus 5.0.8
96	BSH (DE)	3
269	HCMR/HNODC (GR)	30
486	IFREMER / IDM / SISMER (FR)	1973
729	AU-DEA (DK)	3490
730	ICES (INT)	3002
Total		8498

# Actions SeaDataNet CDI nodes

## **Urgent request to all outstanding RM nodes:**

- to install and configure missing RM (if applicable)
- to upgrade to latest RM version (1.0.39) (if applicable)
- to perform import test run ((if applicable)
- To move to RM production

## **Urgent request to CDI data centres with data format errors:**

- to correct the incorrect data formats
- to populate the corrected CDI – data sets via RM or CDI support

## **Urgent request to all CDI data centres:**

- to pick up the process of further populating the CDI catalogue

- [illegible]

[illegible]

## Remaining actions on short term

- Upgrading other portals with CDI engine: Geo-Seas, Harmonia, BlackSeaScene, CaspInfo, NL-NODC, WALTER, RO-NODC
- Upgrading exchange services to EMODnet Physics and EMODnet Biology
- Finetuning OGC WMS – WFS services for EMODnet Chemistry, EMODnet Bathymetry and IHO + NOAA
- Upgrading CDI aggregate service to GEODAB brokerage because of CDI CSW + OAI-PMH + OpenSearch services (via CNR)
- Upgrading API for IP-CHEM

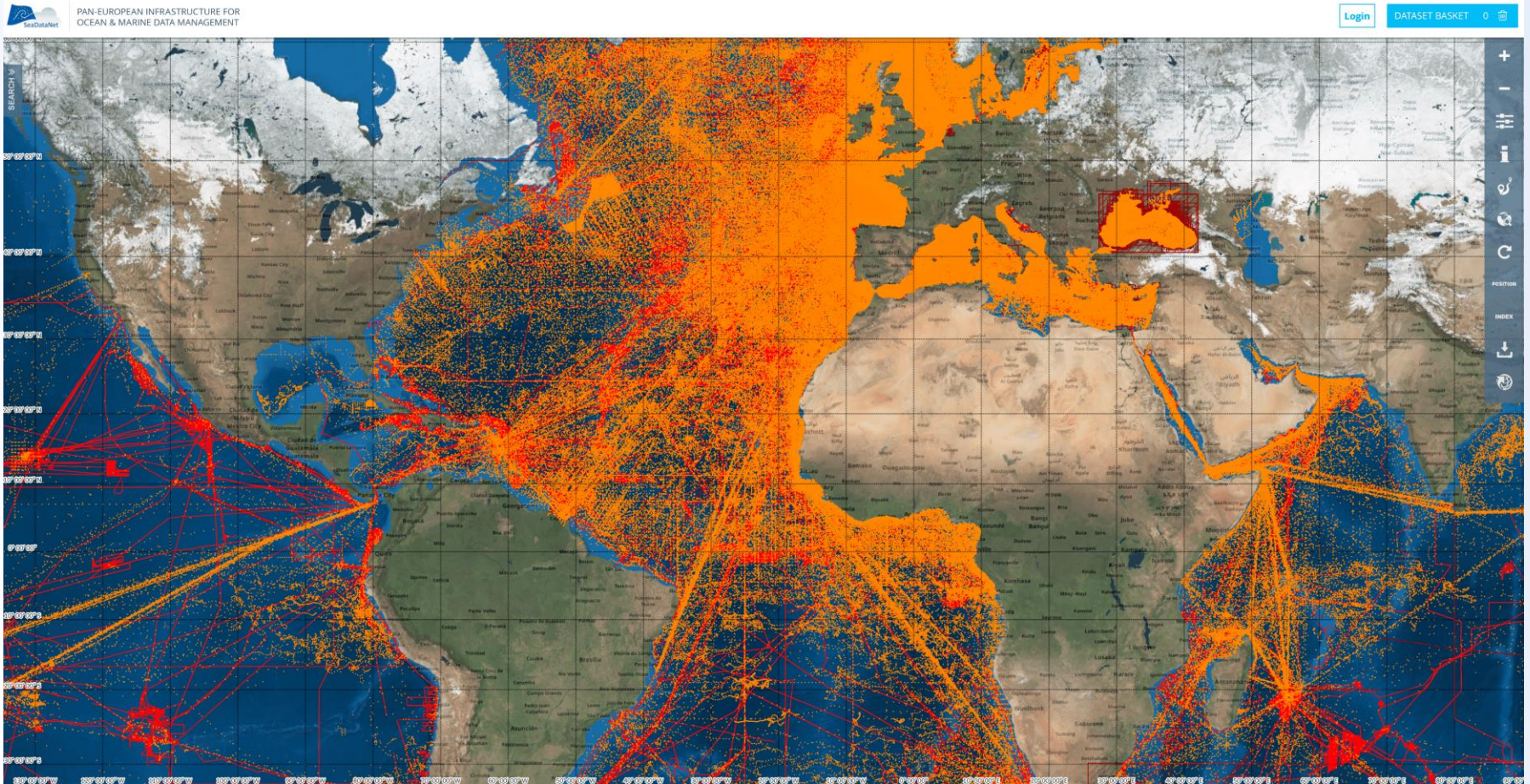


# EUDAT actions

- EUDAT is participating with 5 processing centres: CSC (FI), STFC (UK), DKRZ (DE), CINECA (IT) and GRNET (GR)
- The **Import process** is configured at one EUDAT centre – **CSC** in order to lower risk as the import is complex and EUDAT centres have different infrastructures. This means that all RMs will exchange data with CSC, controlled by the Import Manager service which runs at MARIS => **EUDAT to make the system more robust, so that it will function on an operational basis 24/7.**
- The **CDI shopping system** also communicates only with CSC as manager of **central data cloud** => **to keep this configuration for the coming 6 months to build up experience and to make it operationally robust**
- Later, next year, EUDAT might establish horizontal replication and load balancing between the 5 EUDAT centres. Users should not notice.

# Monitoring actions

- Monitoring should be adapted to the upgraded CDI service and its multiple components:
  - At connected CDI nodes
  - At EUDAT
  - At MARIS
- Not only NAGIOS, but also monitoring of full data shopping cycle



# Questions?

[sdn-userdesk@seadatanet.org](mailto:sdn-userdesk@seadatanet.org) – [www.seadatanet.org](http://www.seadatanet.org)



Del no	Title	Lead	Delivery date	Delivered
D9.1	Specification of the SeaDataNet European cloud environment	CSC	5/7/2017	Y
D9.2	Enhanced EUDAT B2SAFE replication software	CINECA	11/3/2017	Y
D9.3	Adapted EUDAT B2HOST service operational	DKRZ	11/3/2017	Y
D9.4	Upgraded CDI service operational	MARIS	6/27/2018	Y
D9.5	Cloud service for ensuring integrity and conformity of the CDI cloud data resources operational	IFREMER	5/31/2018	Y
D9.6	Upgraded directory maintenance services and protocols operational	MARIS	5/6/2018	y
D9.7	Upgraded versions of MIKADO and Download Manager	IFREMER	4/11/2018	y
<b>D9.8</b>	<b>Integrated data format transformation services operational</b>	<b>IFREMER</b>		
D9.9	Specification of the SWE ingestion service, including SWE profiles and architecture	52°North	10/12/2017	Y
D9.10	SWE ingestion service and user interfaces operational	52°North	5/31/2018	Y
D9.11	Pilot directory of monitoring platforms and sensors operational and feeding into EDIOS	IFREMER	4/23/2019	Y
D9.12	SeaDataNet data management protocols for HF Radar data	ETT	5/28/2019	Y
D9.13	SeaDataNet data management protocols for Flow Cytometer data	VLIZ	7/9/2019	Y
D9.14	SeaDataNet data management protocols for glider data	UKRI-BODC	14/10/2019	Y
D9.15	List of relevant and possible external data sets	EuroGOOS	7/10/2017	Y
<b>D9.16</b>	<b>External data sets ingested in the EMODnet Ingestion service</b>	<b>IFREMER</b>		
D9.17	Preconfigured and pre-built virtual appliance system to connect to the CDI service	ENEA	11/30/2018	Y
D9.18	Coordinated distributed DataCite DOI minting service operational	IFREMER	11/30/2018	Y