SeaDataCloud

Delivering marine data from the cloud using the SeaDataCloud Discovery and Access service

Peter Thijsse (MARIS, NL), on behalf of the CDI team

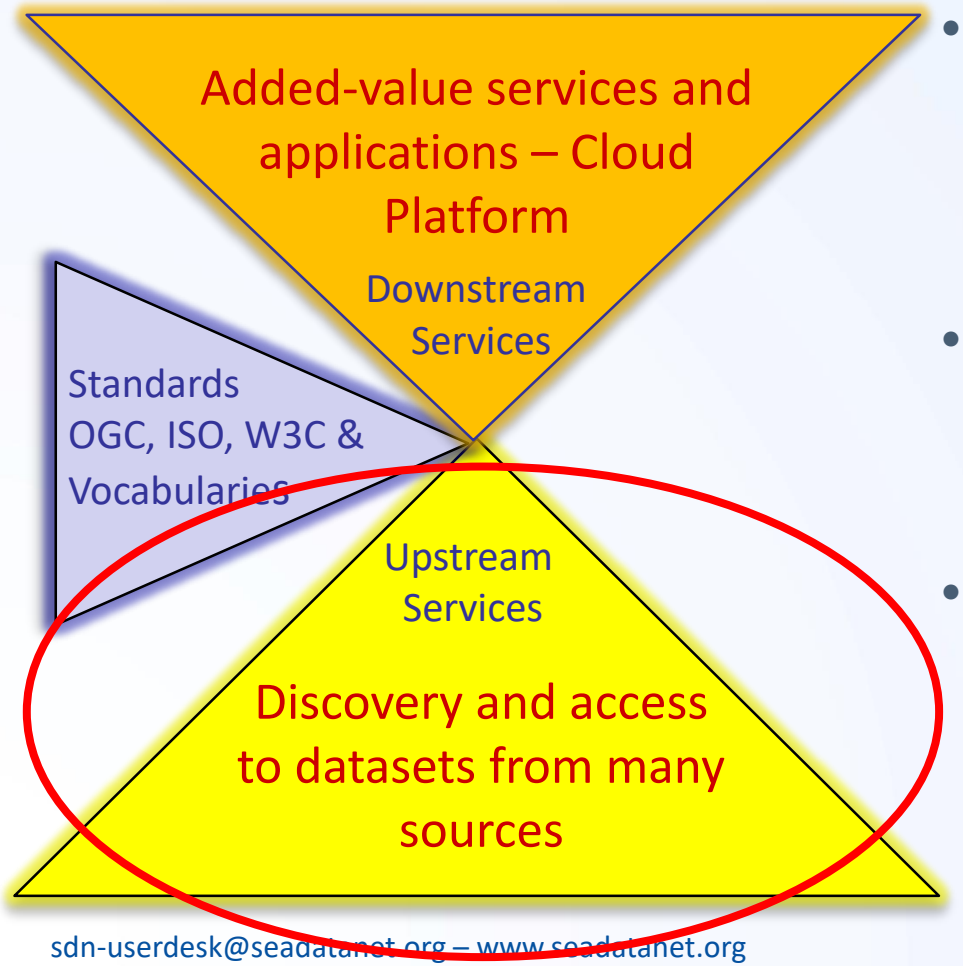
SDC Webinar - 17 September 2020

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

CC BY

1. DATA ACCESS CONCEPT IN SEADATACLOUD

Leading concept for SeaDataCloud

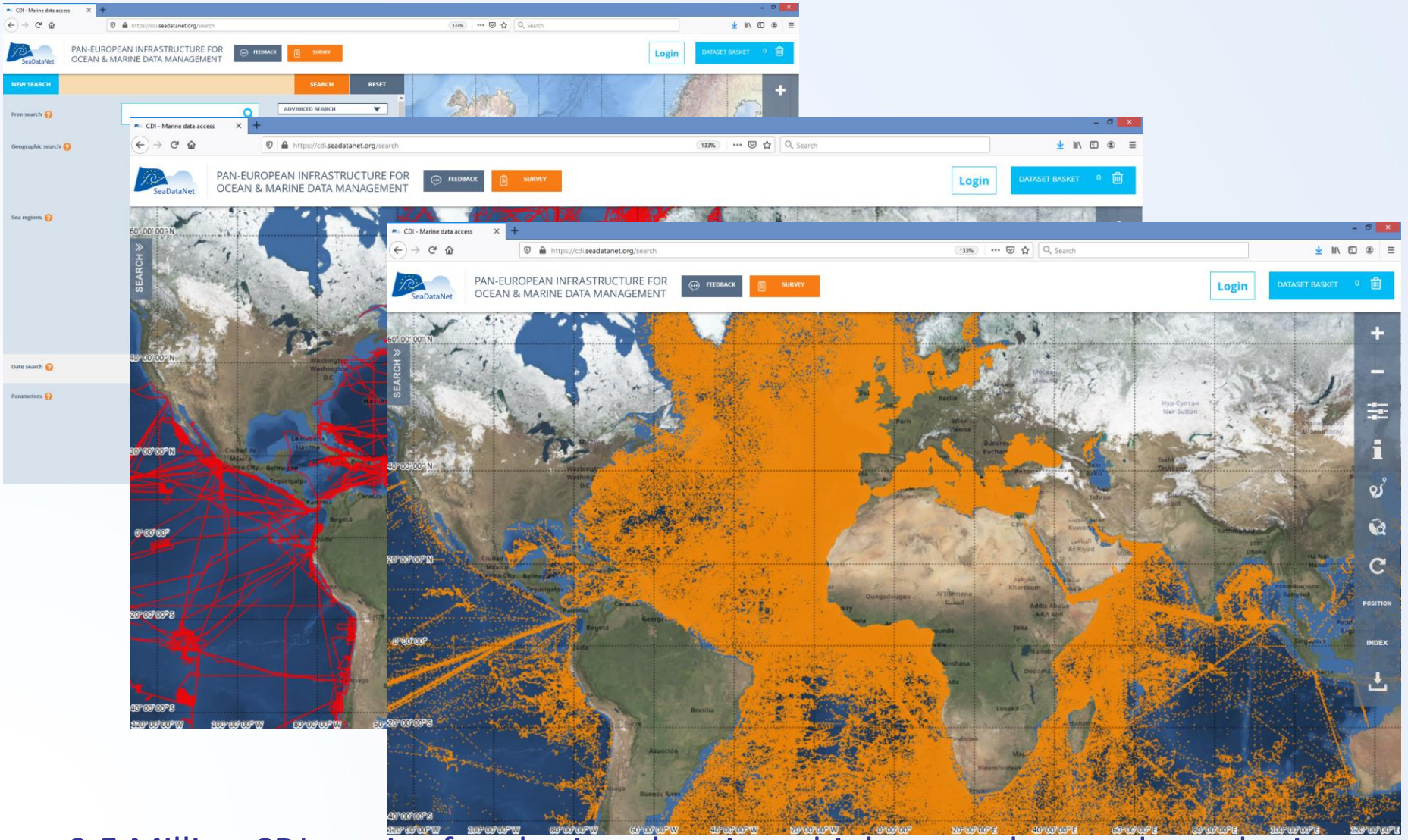


- Providing a cloud platform with common services for data pre-processing, subsetting, analyses, visualizations, publishing, DOIs…
- Applying common standards and interoperability solutions for providing harmonised data and metadata
- Providing harmonised discovery and access to data output from multiple sources, such as European research and monitoring data gathering, but also from other European and international data infrastructures

CDI Data Discovery and Access service

- It is one of the core services of the SeaDataNet infrastructure
- It provides a highly detailed insight and unified access to the large volumes of marine and oceanographic data sets managed by the distributed data centres
- It is a fine-grained index (ISO 19115 – ISO 19139) to individual data measurements (such as a CTD cast or moored instrument record)
- Current content:
  - Over **2.5M** metadata records linked to datasets
  - More than 110 data centers connected
  - Providing data from over **730 data originators**
  - Around **89% available directly from the EUDAT cloud**.

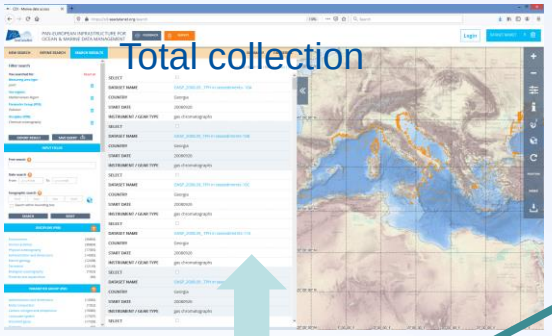
CDI service with global coverage



> 2.5 Million CDI entries for physics, chemistry, biology, geology and geophysics

CDI service contributes to many portals

SeaDataCloud



Total collection

Data discovery and access

SeaDataNet

> 110 data centres

NODCs; HOs; GEOs; BIOs; ICES; PANGAEA

≈ 650 European data originators

GEOSS portal

IODE ODP portal

Aggregated collection

Black Sea portal

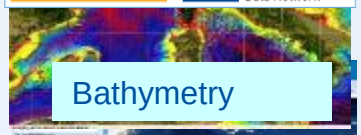
Caspian portal

Geo-Seas portal

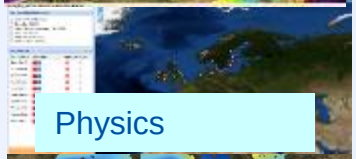
Regional subsets

Thematic portals

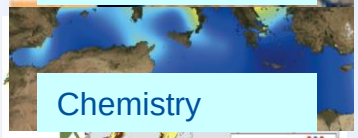
EMODnet European Marine Observation and Data Network



Bathymetry

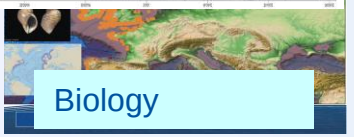


Physics



Chemistry

Geology



Biology

CDI Data Discovery and Access service

sdn-userdesk@seadatanet.org – w

Installed base of CDI nodes (> 110)

2. SDC THEME: LINK TO EOSC VIA EUDAT

European Open Science Cloud (EOSC)

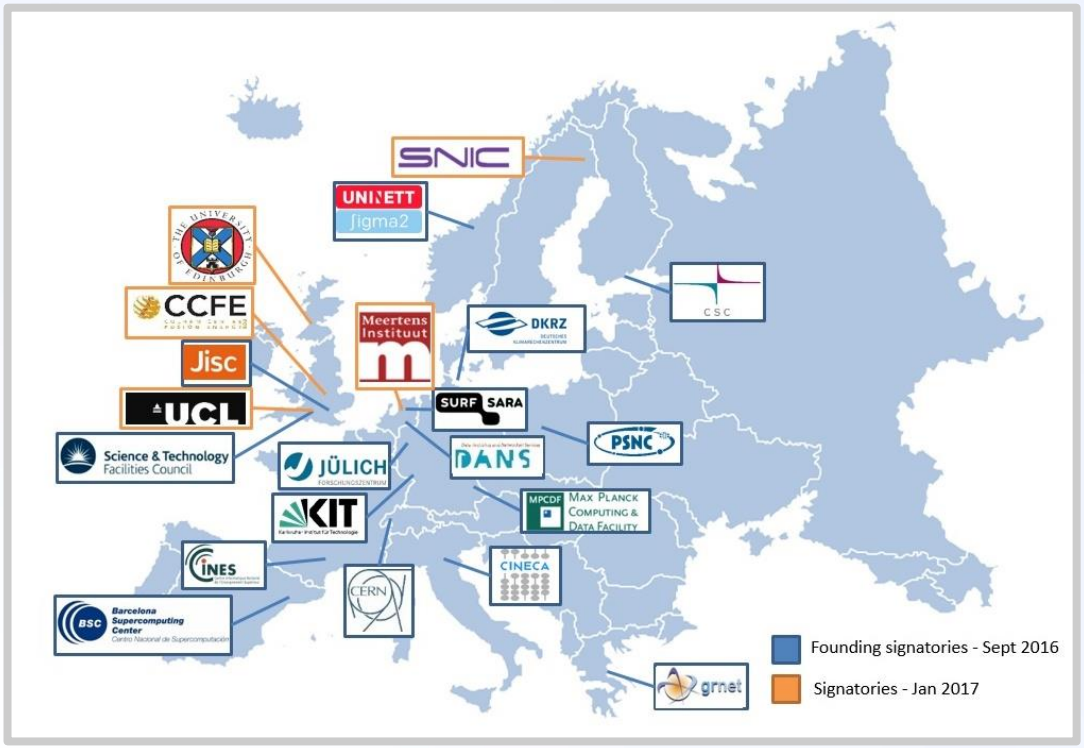
- **6 May 2015** the EU adopted the Digital Single Markets strategy and announced the launch of a **cloud for research data**
- A **High Level Expert Group** was established to advise on scientific services to be provided on the cloud and its governance structure
- **16 April 2016** the EU published a package of measures for digitising European industry, including a communication on the **European Cloud Initiative**
- **12 June 2017** the **European Open Science Cloud Summit** took place bringing together key players from across Europe, to make the '***EOSC a reality by 2020***'
- **14 March 2018** the EU adopted the **Implementation Roadmap for the European Science Cloud**

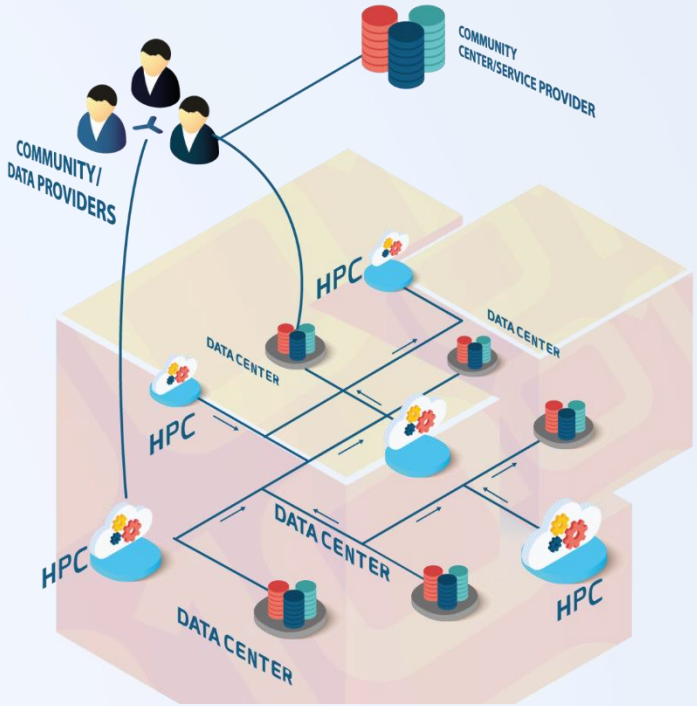
EOSC should to give the EU a global lead in research data management and ensure that European scientists reap the full benefits of data-driven science. It also foresees setting up a European Data Infrastructure, with high-capacity cloud solutions with super-computing capacity.

SeaDataCloud was anticipating EOSC

- In anticipation of these EOSC planning activities, SeaDataNet Started **in December 2016** the **SeaDataCloud project** with the following aims:
  - Further **developing skills, standards, tools and services** for dealing with marine data, such as handling data from new instruments, INSPIRE compliance, interoperability with other data infrastructures, vocabularies governance, and adopting new technical approaches such as 'Linked Data', and 'Sensor Web Enablement',
  - Exploring the trend towards **cloud storage and cloud computing**, also taking into account **'big data' challenges**, for instance by analysing an architecture for a Virtual Research Environment (VRE) and developing a VRE pilot
  - Improving and expanding services and tools for **data providers** for connecting and ingesting data AND for **users** to make it easier and more performing to find, access, and use datasets
  - ***Better joining in an early stage and riding the EOSC wave than losing ourselves in the undertow***

Cooperation with the EUDAT network of e-infrastructure providers





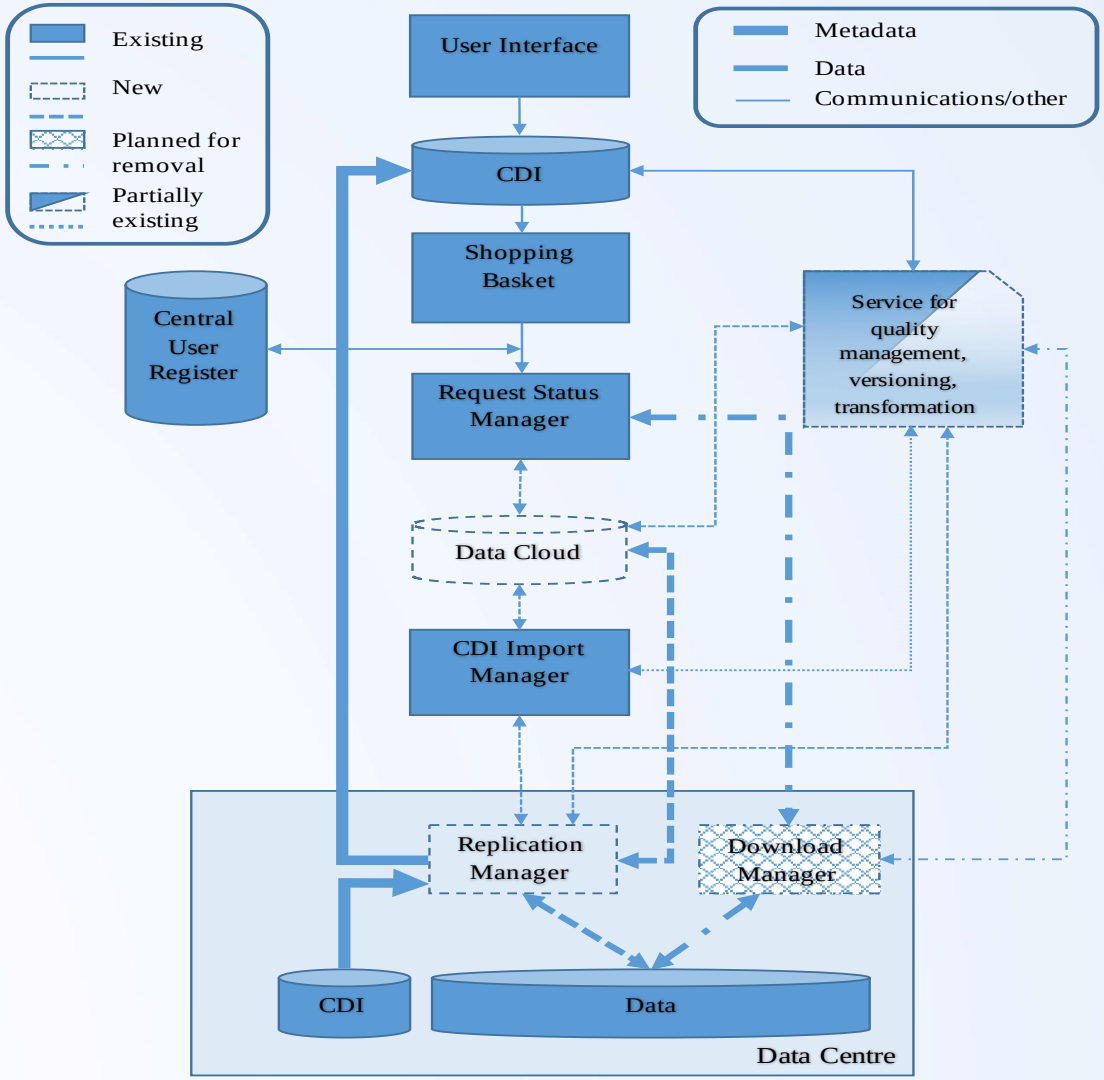
5 EUDAT members are partners of SeaDataCloud :
CINECA, CSC, DKRZ, GRNET and STFC

3. THE NEW CDI SERVICE

CDI service is now using the cloud

- To configure and maintain a **CLOUD buffer** to host **copies of unrestricted data resources**
- Exchange by dynamic **replication** from the individual data centres, following their updating of the CDI catalogue service
- In the cloud buffer:
  - **checking overall quality of metadata and data,** as extra check on top of local QA-QC by data centres
  - checking integrity of data files and metadata relations.
  - results of checks to be reported back to data centres for corrections
- Include **transformation services** for converting data sets to SeaDataNet ODV and NetCDF formats and relevant INSPIRE data models.
- **Introduce PID and versioning** of metadata and data as part of provenance

New CDI service architecture



Where are we now in back-end?

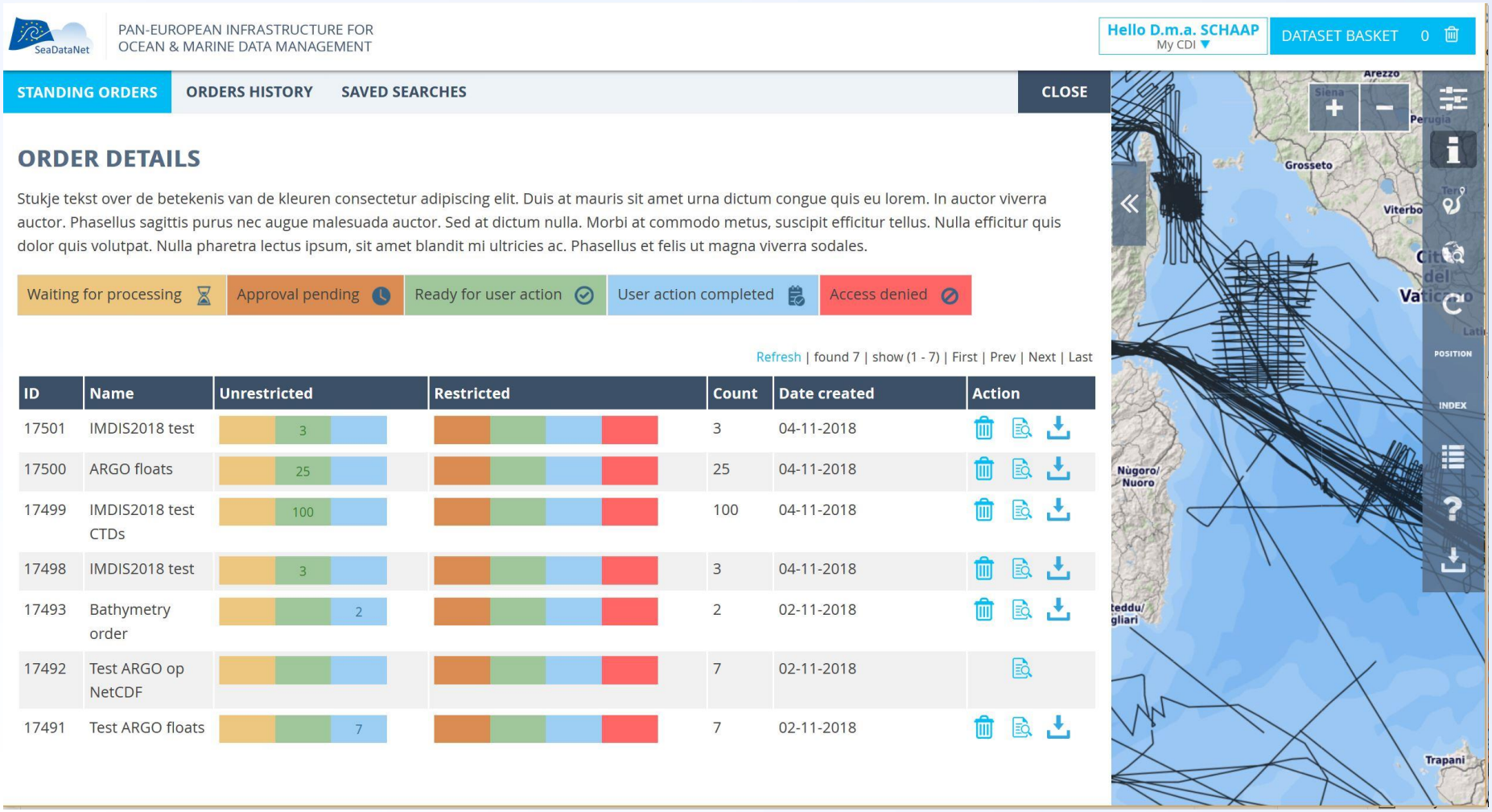
- SeaDataCloud project in final year: CDI Data access service fully live!
- **Cloud**:
  - Cloud system set up using EUDAT components B2Stage (API), B2Handle (PIDs), B2Access (secure), B2Safe (storage)
  - Development environment (Cineca), test environment, production environments (both at CSC)
  - All unrestricted data (+/- 90%) cached
- **Import manager:**
  - Fully working, connected to cloud and Replication Managers
  - loaded with all metadata
- **Replication Manager:**
  - Replicates data to cloud for connected partners
  - Communicates to Import Manager and cloud
- **Machine2Machine services and increasing FAIRness:**
  - SparQL endpoint available to support linked data
  - API in development
  - Metadata expansion for more QC and provenance information (e.g. for the Sensor data SWE)

Front-end – All new GUI

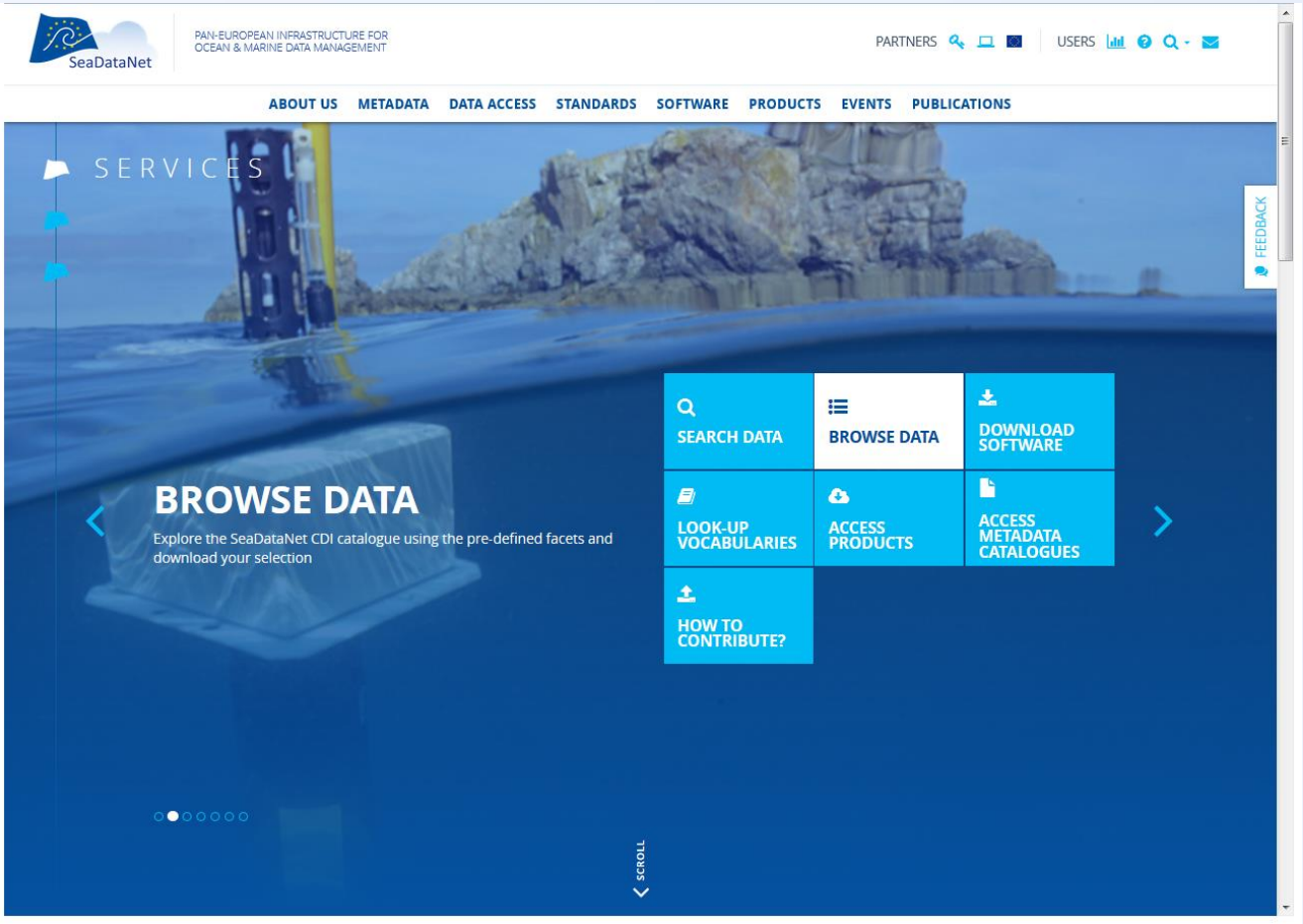
- Quicker access to data
- Very fast, using **Elastic Search** for search and indexing, and **GeoServer** for latest mapping technology
- Easier, including full text search next to controlled terms
- Modern design with large map and sliding windows
- **MySeaDataNet** integrated in GUI for customized services, such as SSO, data access, search profiles, prepared for VRE data pooling

- Access via https://cdi.seadatanet.org

New GUI impressions



SeaDataCloud



Keep following us via: www.seadatanet.org

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