SeaDataCloud

Further developing the SeaDataNet pan-European infrastructure for marine and ocean data management

Michèle Fichaut, project coordinator
IFREMER, France

IODE XXIV, Kuala Lumpur, Malaysia, 27 March 2017
sdn-userdesk@seadatanet.org – www.seadatanet.org

What is SeaDataNet?

A pan-European infrastructure set up and operated for managing marine and ocean data in cooperation with the NODCs and data focal points of 34 countries bordering the European seas

90s	Metadata directories Medar/MedAtlas
2002-2005	Sea-Search (FP5)
2006-2011	SeaDataNet (FP6)
2011-2015	SeaDataNet II (FP7)
2016-2020	SeaDataCloud (H2020)

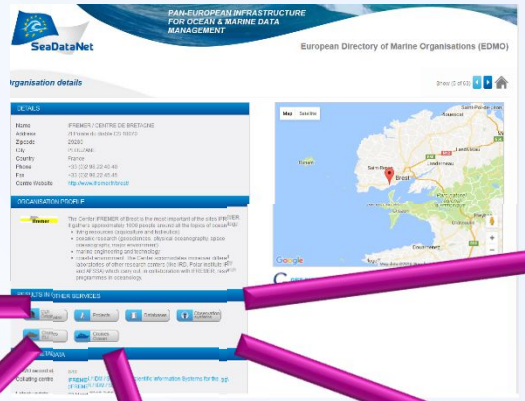
SeaDataNet portal

With access to

- Standards, tools both for data centres and other users
- Data
- Products

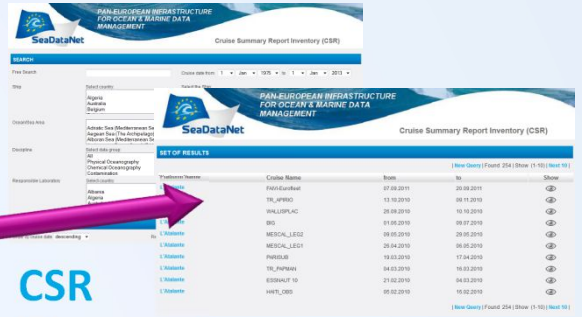
SeaDataNet metadata directories

EDMO ***Organisations***



EDMERP

Projects

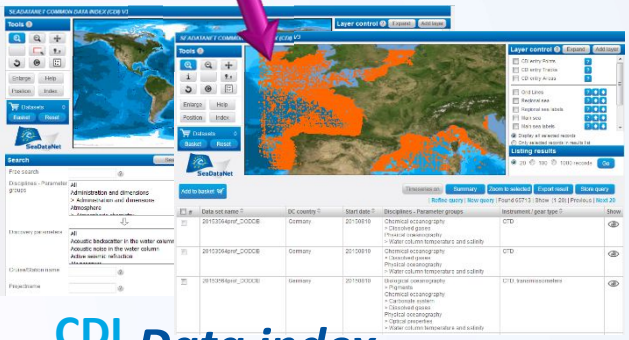


CSR

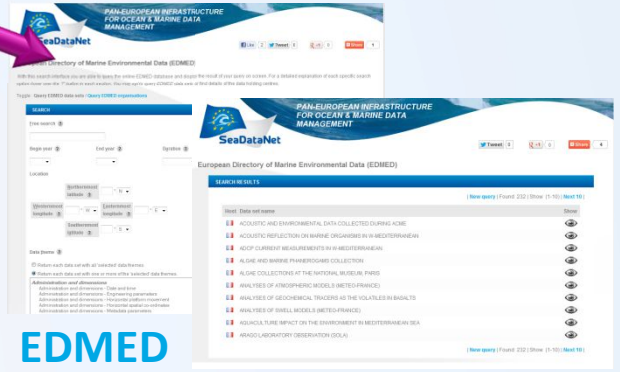
Research cruises

EDIOS

Observing programmes



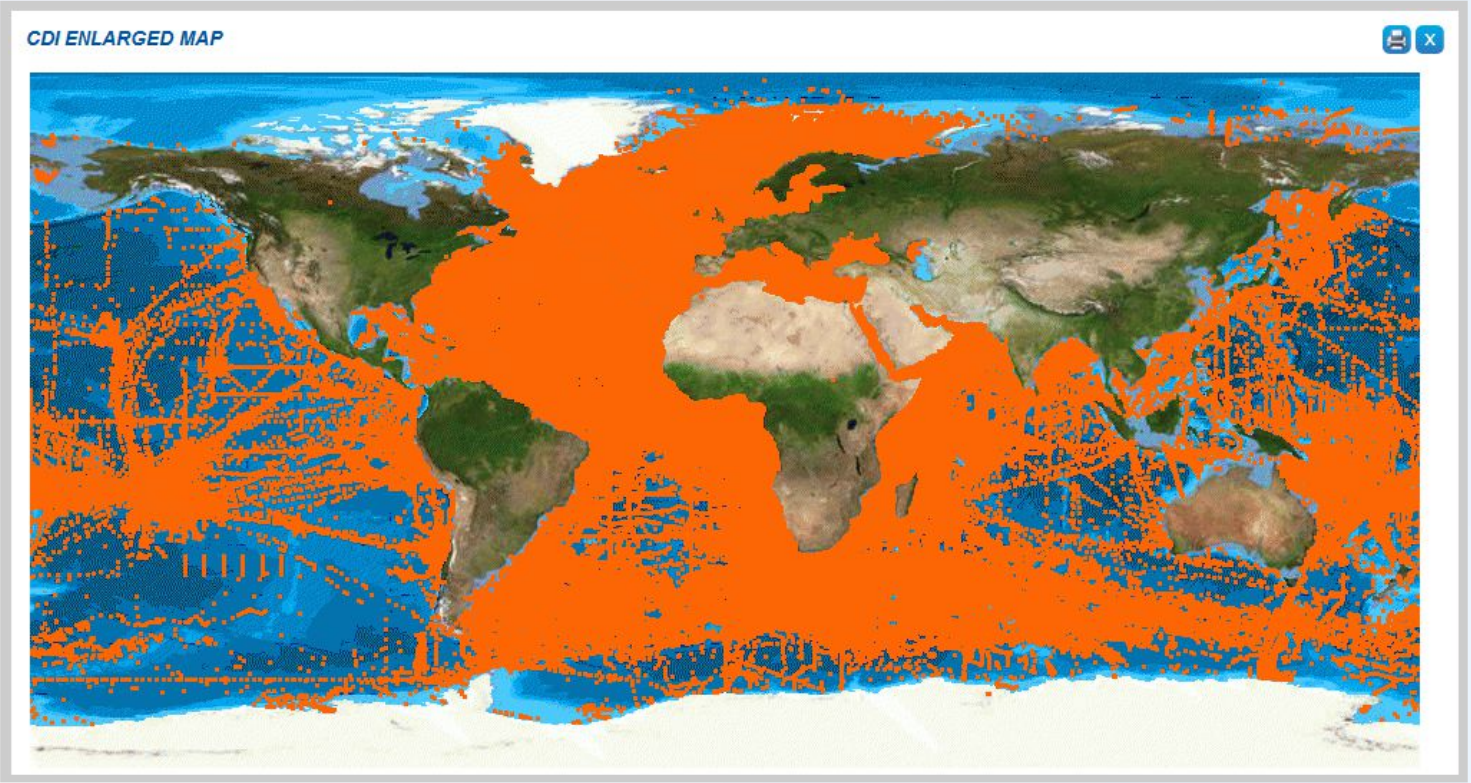
CDI ***Data index***



EDMED

Data sets

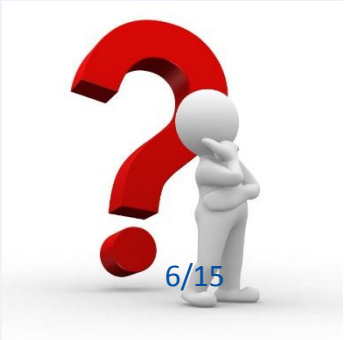
CDI service for discovery and unified data access



+1.95 million CDI entries from 34 countries, 102 data centres and 597 originators for physics, chemistry, geology, geophysics, bathymetry and biology; from 1805 to 2017; 86% unrestricted or under SDN License

Issues with CDI present service

- Usage of the discovery and access services lags behind expectation; major obstacle ➔ **multiple download transactions** in case of shopping baskets with data from multiple data centres.
- **performance issues** because data centres are not always online, operational and have different machine capacities ➔ extra delays
- **Quality issues concerning formats of data files** (ODV + NetCDF) and their consistency with CDI metadata
- **Connecting new data centres can be challenging** due to different configurations, firewalls.. ➔ different versions installed, because upgrading can give issues

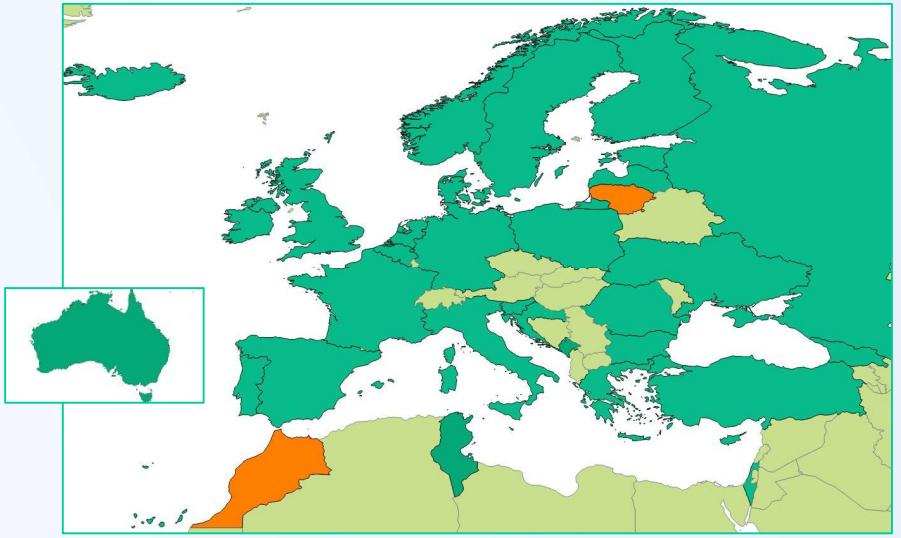


SeaDataCloud a new opportunity

Standards and information technology are always evolving, and the SeaDataNet infrastructure must stay up-to-date to maintain and further expand its services to its leads customers and major stakeholders

SeaDataCloud - Key numbers

- 4 year duration, started 1st November, 2016
- 10 M euros
- 56 partners
- 5 subcontractors
- 32 countries
- 1110.5 man/months

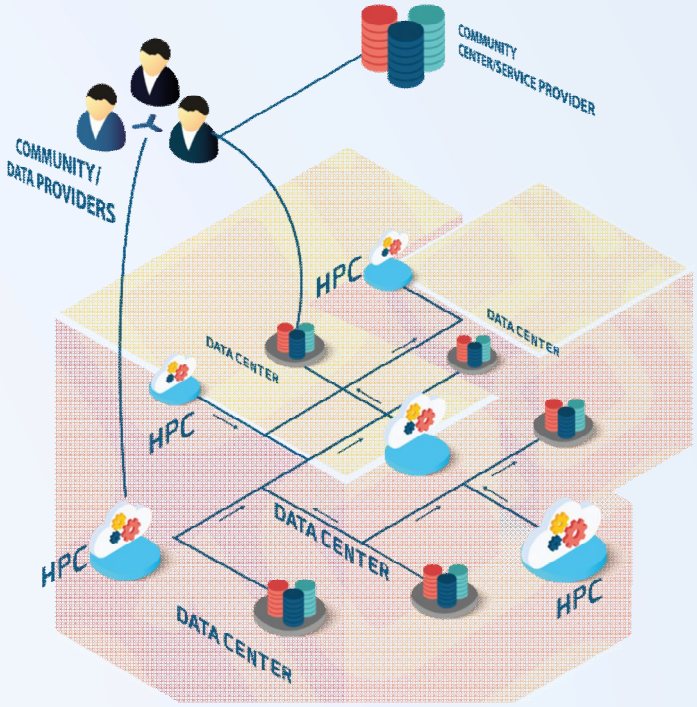


SeaDataCloud

Whatis it About?

- SDC is about **updating and further developing standards**
- SDC is about improving and innovating services & products
- SDC is about adopting and elaborating new technologies
- SDC is about giving more attention to users and putting the user experience in a central position
- Moreover, it is about implementing a strategic and operational cooperation between the SeaDataNet consortium of marine and ocean data centres and the EUDAT consortium of e-infrastructure service providers

SeaDataCloud - Cooperation with EUDAT



A consortium of 20 High Performance Computing (HPC) centres offering also storage resources

5 EUDAT members are partners of SeaDataCloud

Upgrading CDI service using the cloud

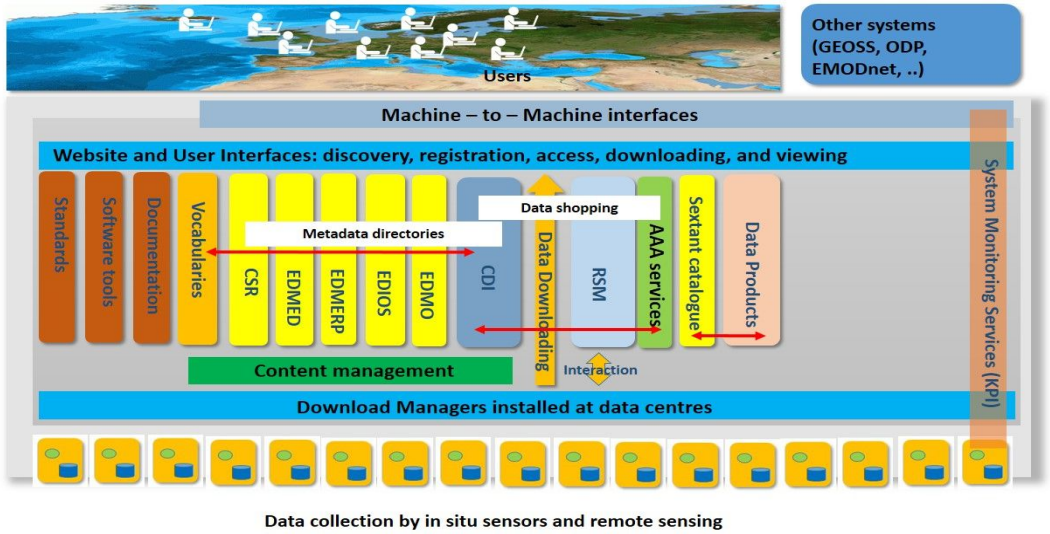
- Configure and maintain a cloud environment as a 'cache' to host copies of all data resources (from the distributed data centres)
- Exchange by dynamic replication from the individual data centres, following their updating of the CDI catalogue service
- In the cloud buffer new functions:
  - checking possible duplicates
  - Checking overall quality of formats
  - Checking integrity of data files and metadata relations.
  - Results of checks reported back to data centres for amendments of their submissions
- Develop a Virtual Research Environment (VRE) to facilitate collaborative and individual research by users
- Provide customised services (MySeaDataCloud) to let users have search profile, receive alerts on new available data, ingest and manage its own dataset

Benefits for CDI service and its users

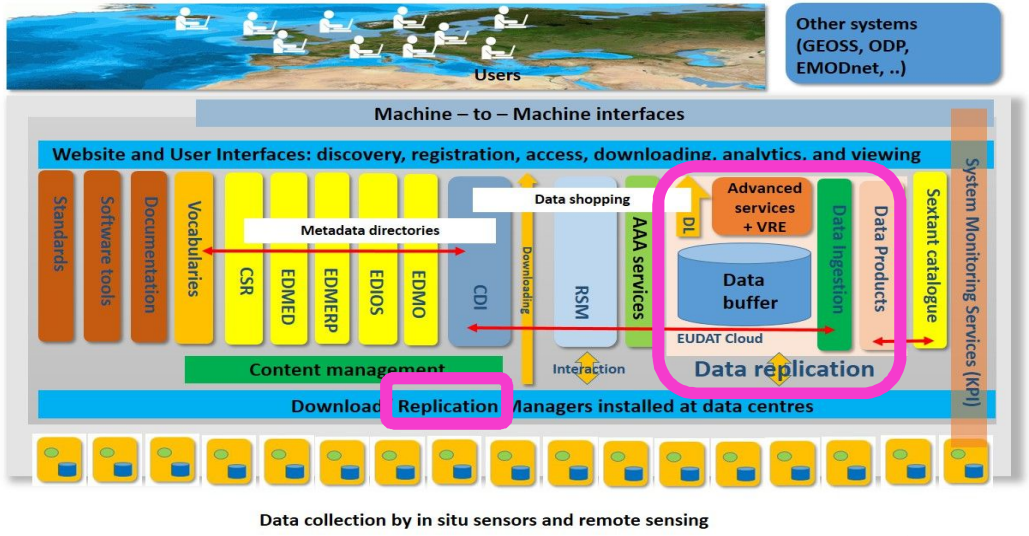
- Cloud buffer in combination with the CDI service will
  - speed up the performance,
  - expand discovery and ease of use of the data access and downloading
  - provide users with one integrated download package instead of multiple packages from multiple data centres.
- Overall quality and coherence (data – metadata) will improve
- Data replication will be triggered per data centre by CDI updates. The replication module might have less complexity than the present Download Manager module
- A system of versioning will be introduced which is required in the context of the MSFD for facilitating repeated analysis of environmental assessments after many years, and for scientific papers.

Architecture

Present SeaDataNet architecture



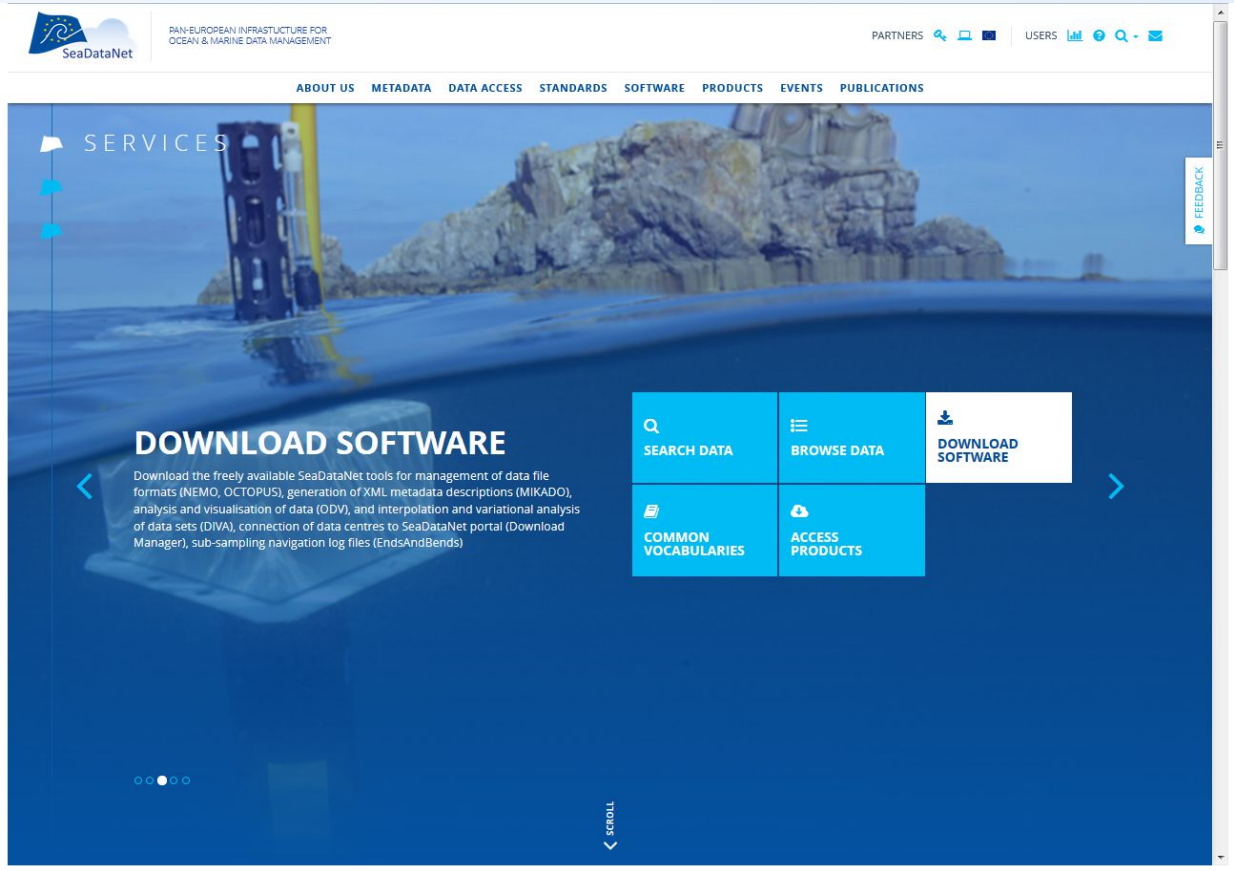
Planned upgraded architecture with data replication, advance services and VRE in the cloud



SeaDataNet new portal

Under development

- Services directly accessible from the home page



IOC/IODE role in SeaDataCloud

- IODE is a member of the SeaDataCloud consortium as subcontractor of IFREMER for the:
  - Organisation and the hosting of the training sessions for data managers and technician as well as for users of the SeaDataNet infrastructure
  - Integration the training material in the Ocean Teacher System
  - Cooperation for the organisation of 2 IMDIS conferences that will be organised during the project in 2018 and 2020