The SeaDataNet pan-European marine data management infrastructure

Michèle Fichaut and the SeaDataNet Consortium

SeaTechWeek virtual meeting, Open data session, 14 October 2020

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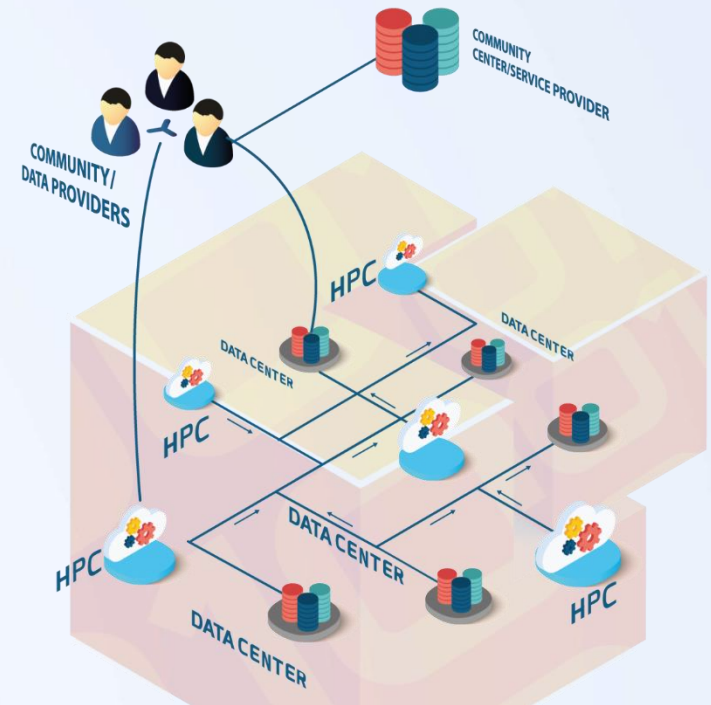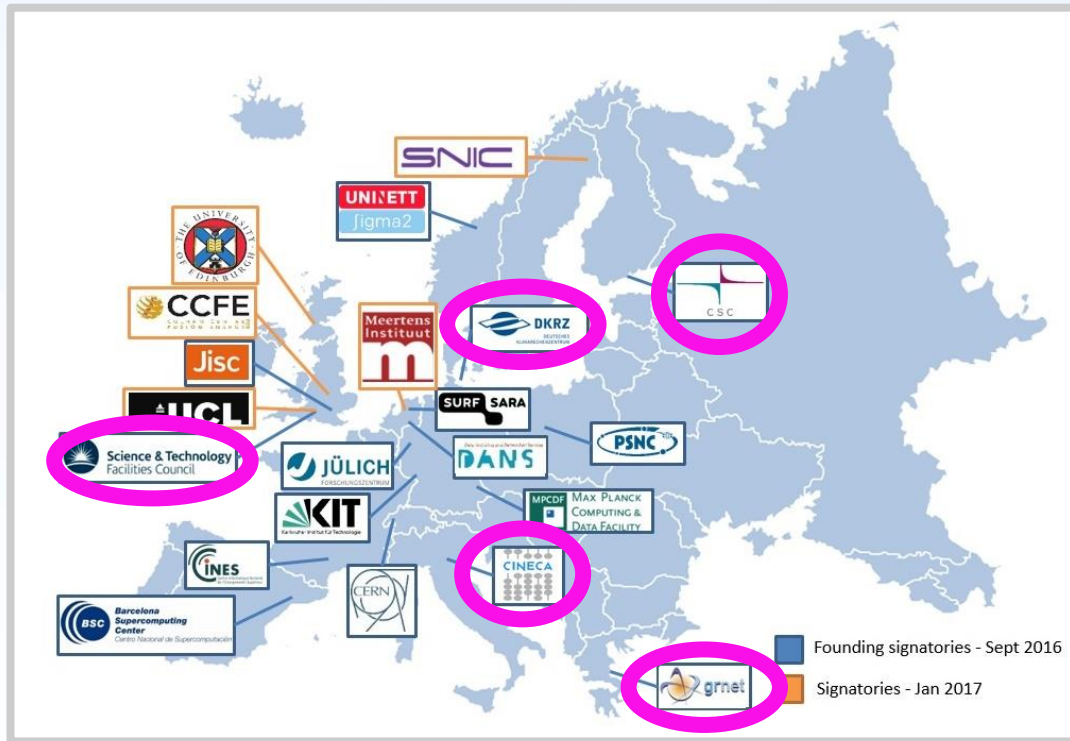
SeaDataNet

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 AISBL

A legal entity : SeaDataNet AISBL for sustainability of the Consortium (same structure than EuroGoos)

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

SeaDataCloud – Link to EOSC via with EUDAT

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

5 EUDAT members are partners of SeaDataCloud

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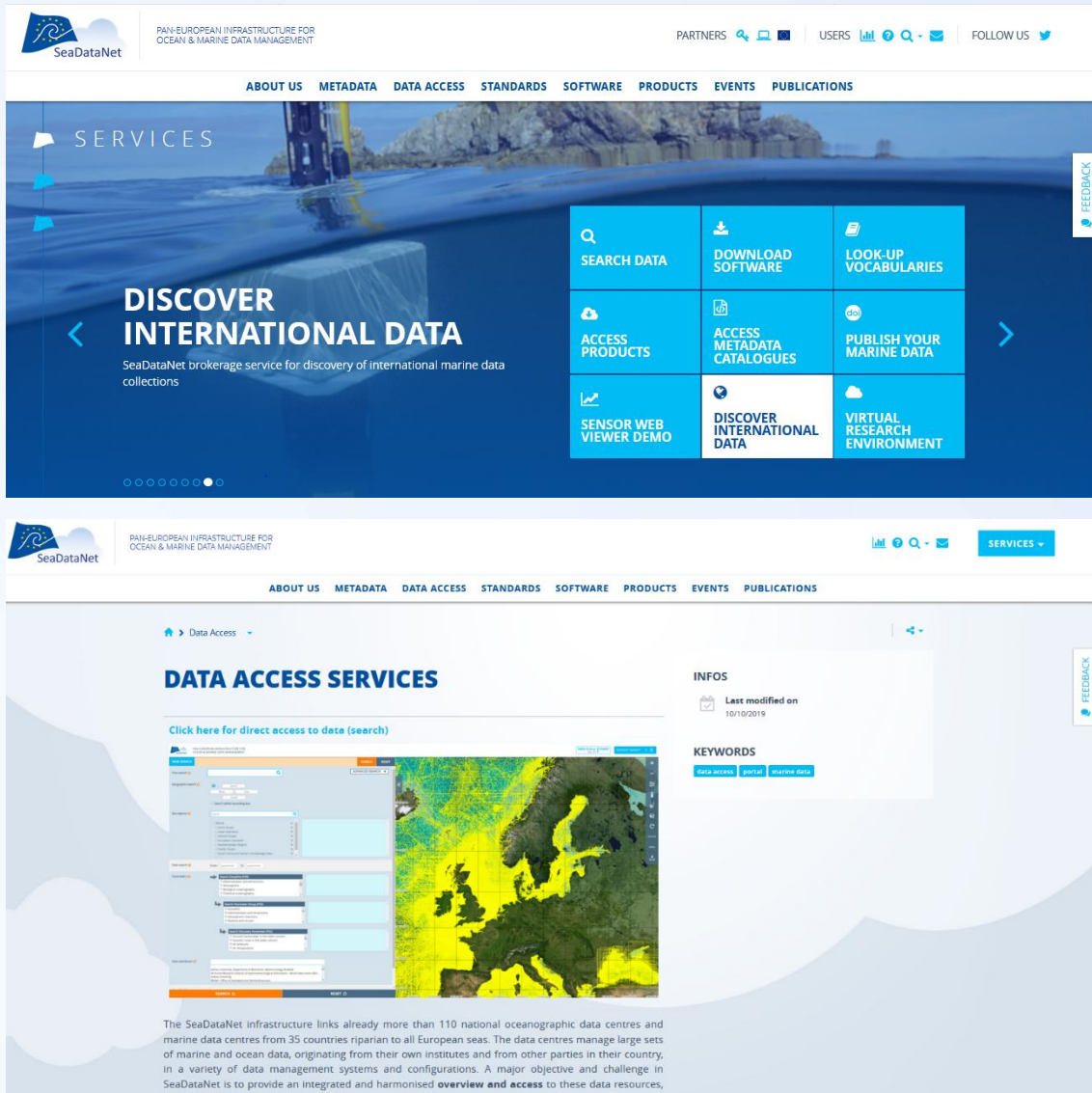
SeaDataNet portal

With access to services

- Standards
  & common vocabularies

- Software tools both for data centres and users

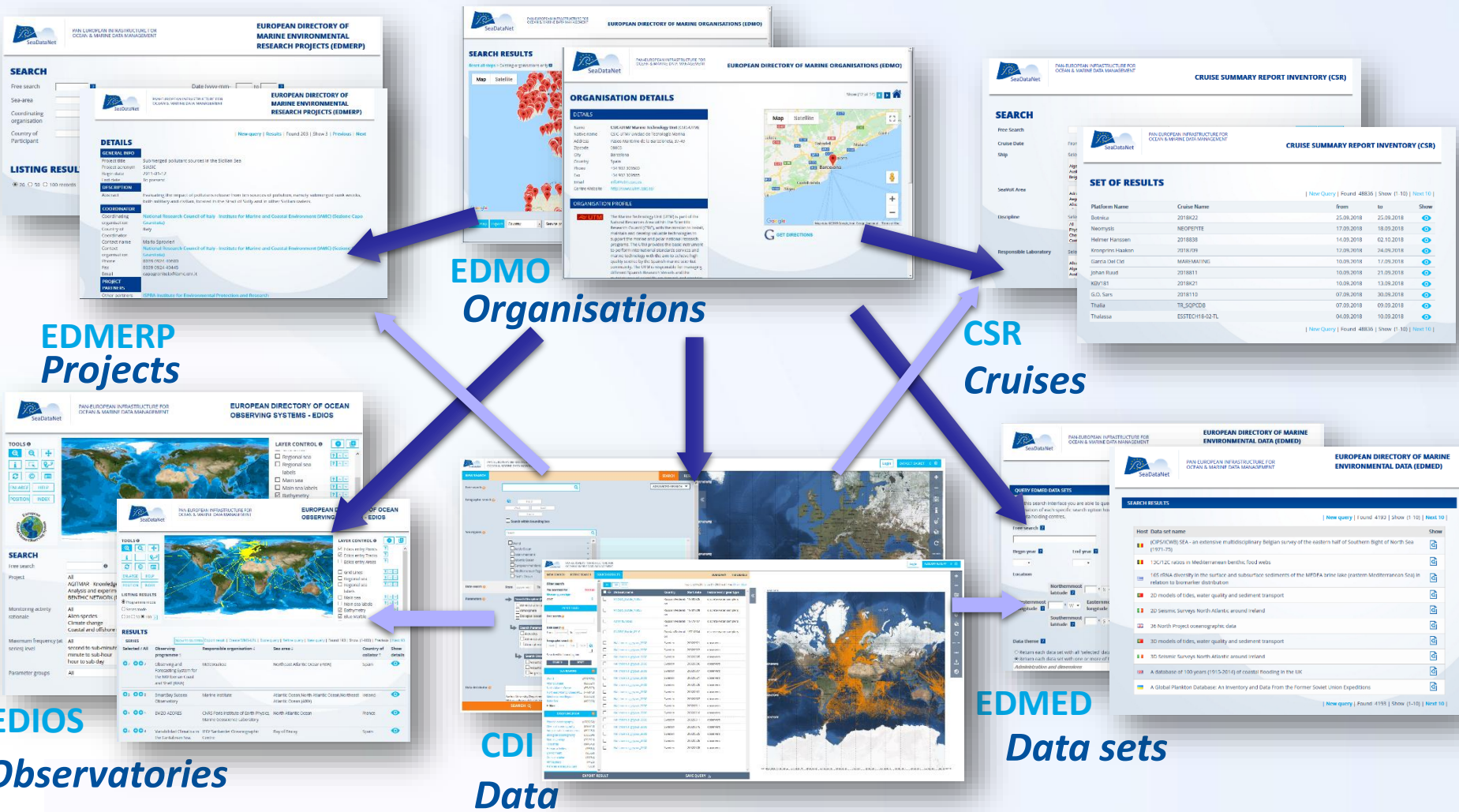- Data and metadata catalogues

- Data products

http://www.seadatanet.org

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

SeaTechWeek virtual meeting, Open data session, 14 October 2020

SeaDataNet metadata directories
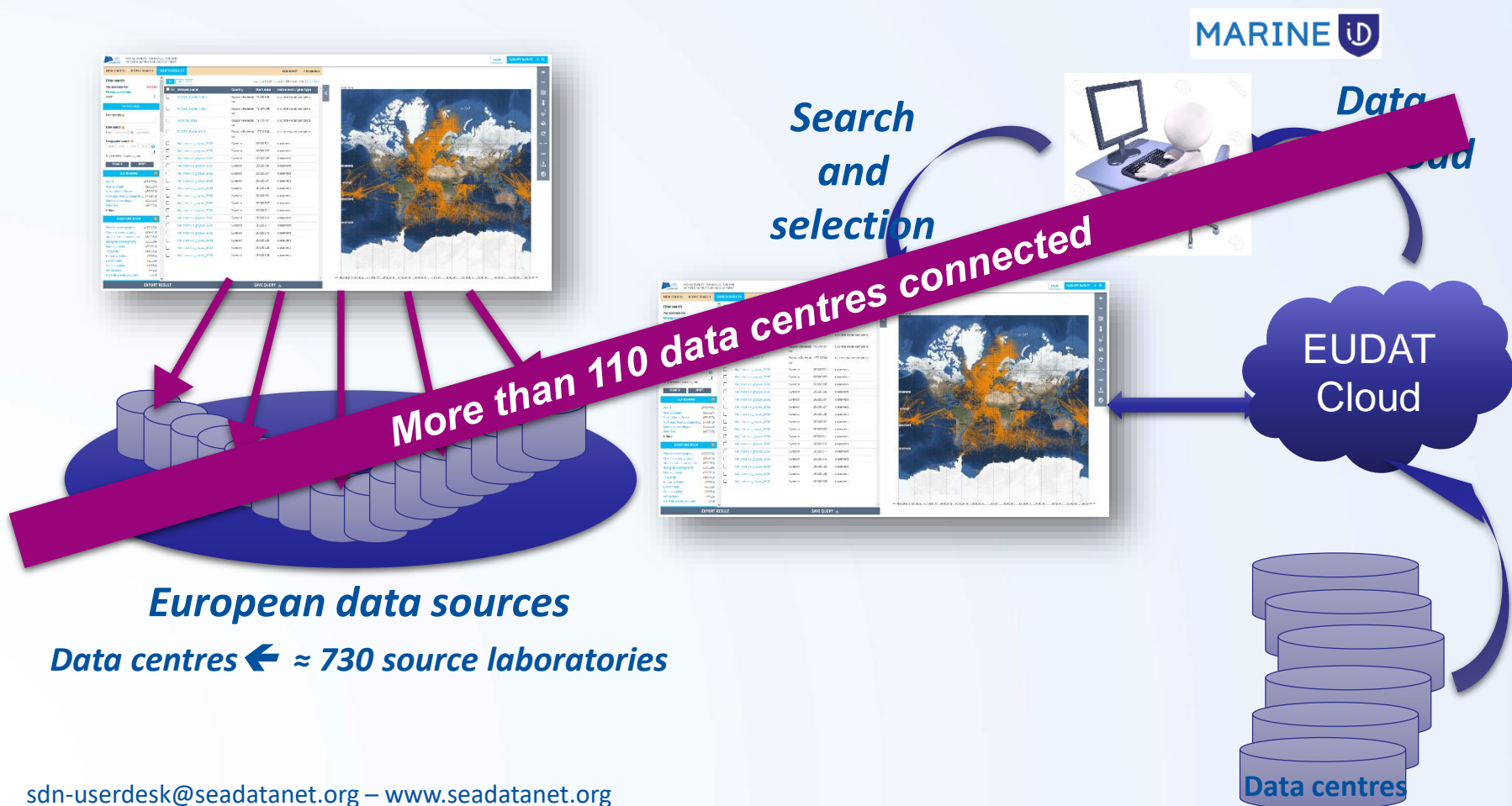
EDMERP *Projects*

EDMO *Organisations*

CSR *Cruises*

EDIOS *Observatories*

CDI *Data*

EDMED *Data sets*

CDI catalogue: discovery and access to data

Search and selection

More than 110 data centres connected

EUDAT Cloud

European data sources

Data centres ⬅ ≈ 730 source laboratories

Data centres

Service for discovery and unified data access

cdi.seadatanet.org/search

Trajectories

Vertical profiles or time series

- since 1800 ➔ 2020

- 2.5 M of CDIs for physical, chemical, biological, bathymetry, geosciences data

- 89 % of unrestricted or SDN license data

SeaDataNet products

CENTRAL CDI

Analysis of data anomalies

Data harvesting

SeaDataNet Quality Checks Strategy (QCS)

QC analysis

File and parameter aggregation

QC-Loop
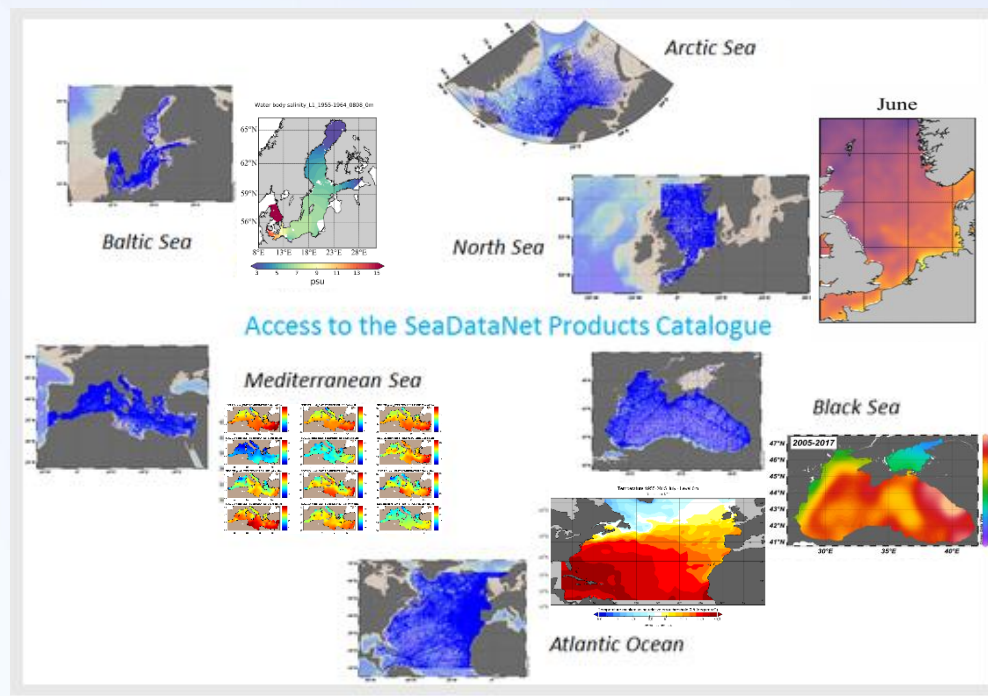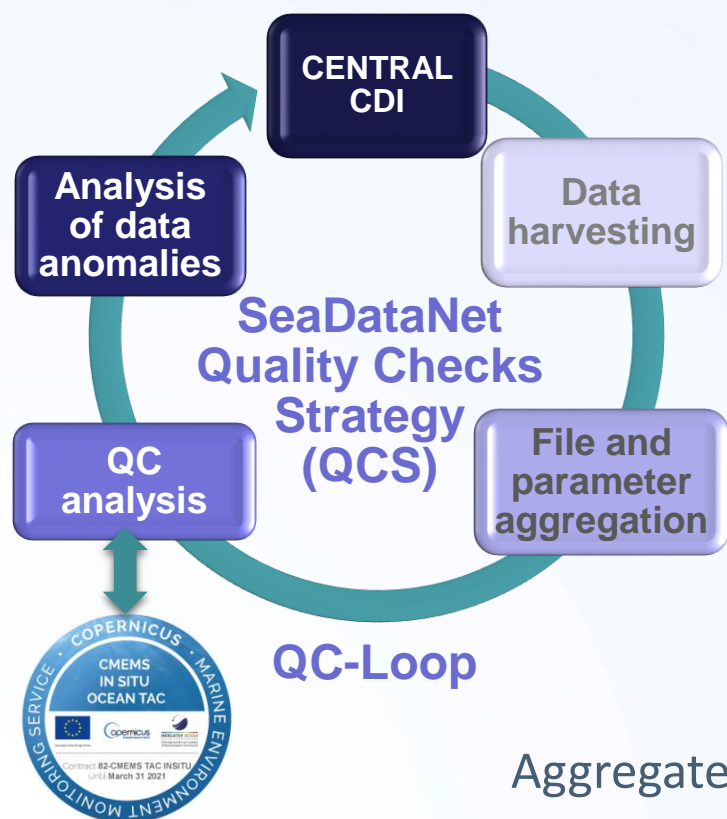
Aggregated datasets and climatologies

Improvement of the data quality

SeaDataNet standards

- Metadata formats for all catalogues
  - ISO19115 and ISO19139
- SeaDataNet data transport formats
  - ASCII (Ocean Data View, and MedAtlas)
  - NetCDF (CF compliant)

➔ Relying on controlled vocabularies governed by NOC-BODC (UK) and used in many other international or national initiatives
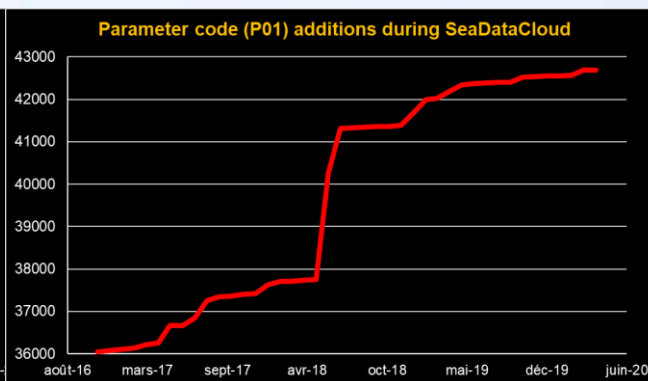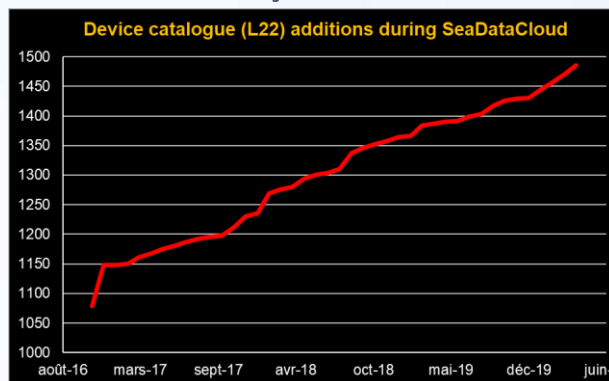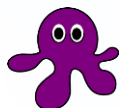
SeaDataNet software Tools (1)

- ## Tools for the data centres – data managers

  - To be connected to the infrastructure and to be able to duplicate data in the cloud (Replication manager)

  - To follow the data downloading by users : MySeaDataCloud

  To generate the metadata at the SDN standards : MIKADO

  To generate the data files at the SeaDataNet standards : NEMO

  To check the compliance of the data files : OCTOPUS

  To quality check the data: ODV

SeaDataNet software Tools (2)

- Tools for the users and data scientist
  - All catalogue search interfaces
  - To visualise data, plot, analyse : ODV
  - To interpolate data : DIVA
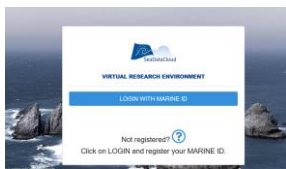 - to publish your data using Sensor Web standards
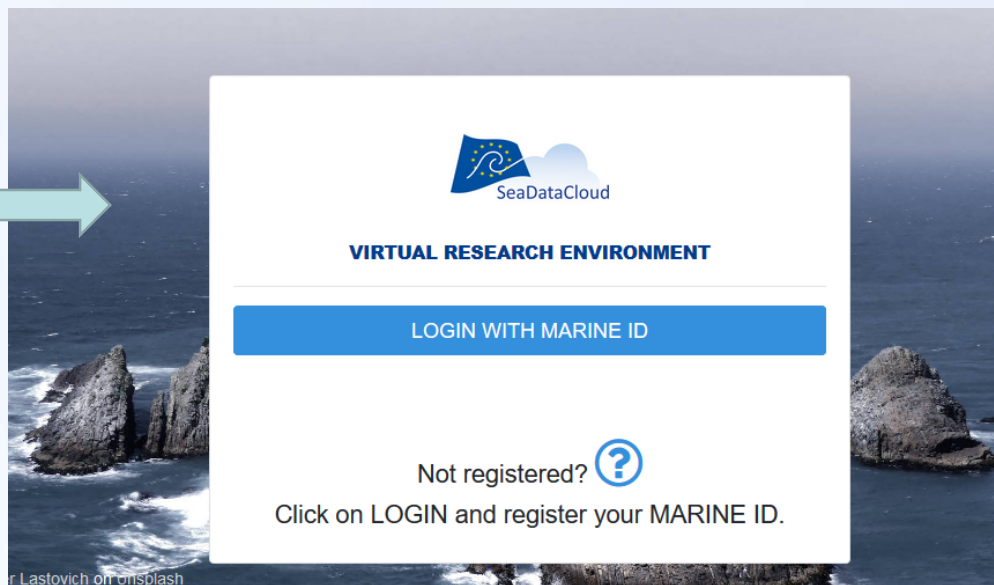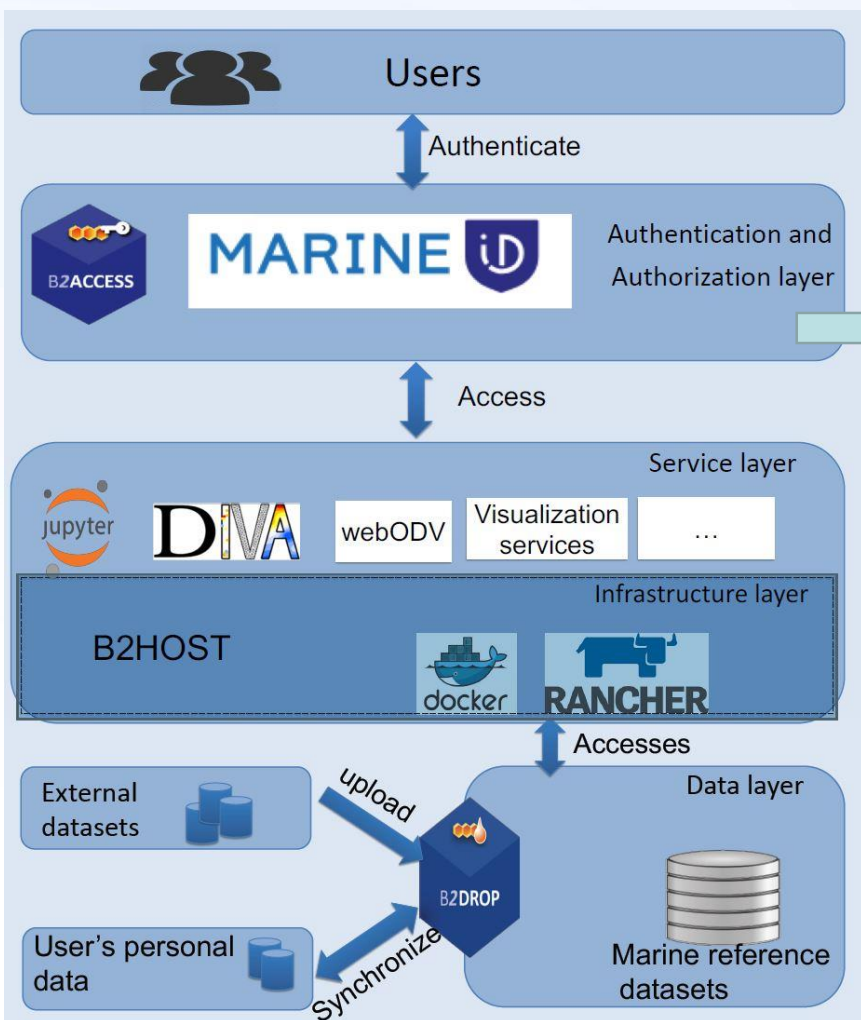 - To publish marine data (and get a DOI) : SEANOE
  - To work on datasets in the cloud environment : **SDN Virtual Research Environment (VRE)**
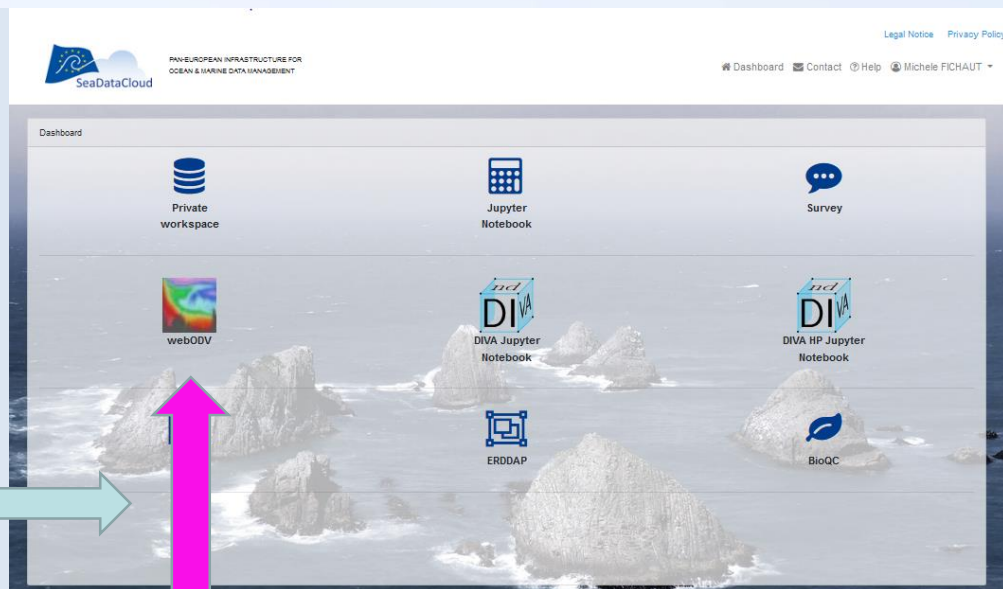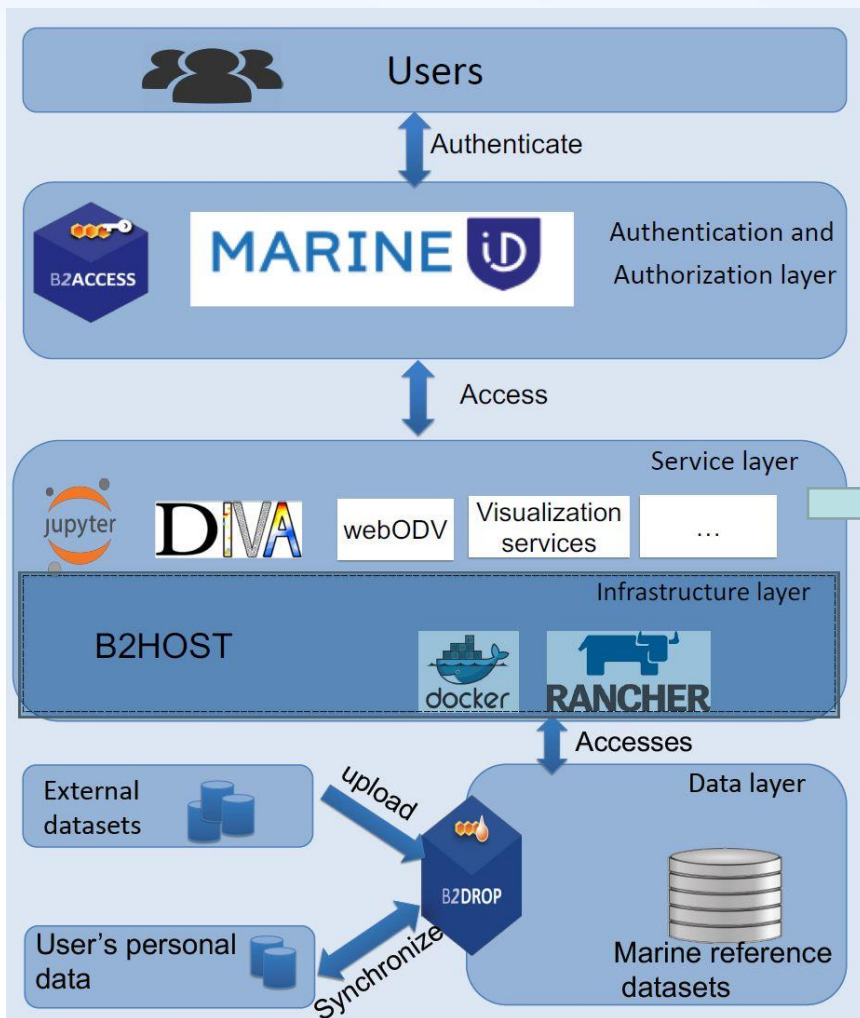   - prototype available and used by the regional product leaders of the SDC project

# Overall architecture for the SDC VRE (1)

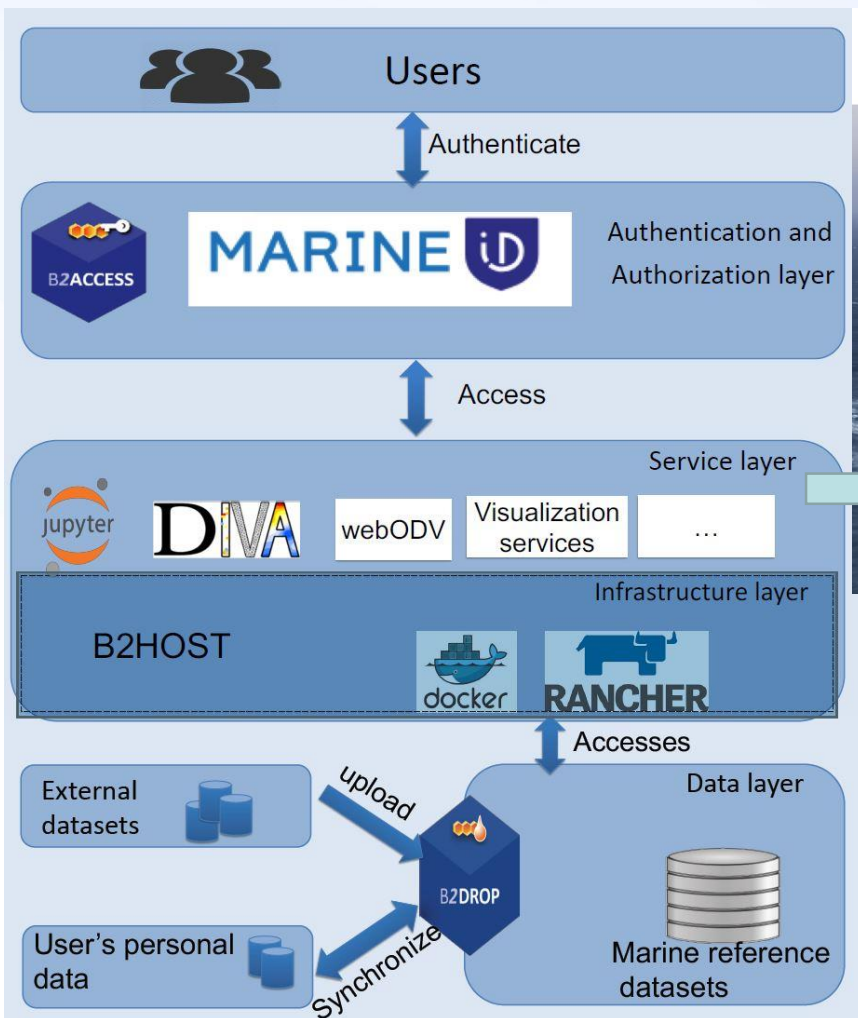# Services in the SDC VRE pilot

WebODV consists of a suite of online services based on Ocean Data View (ODV), designed to interactively perform analysis, exploration and visualization of ocean data.

webODV allows users to aggregate large numbers of SeaDataNet data files and perform quality control.
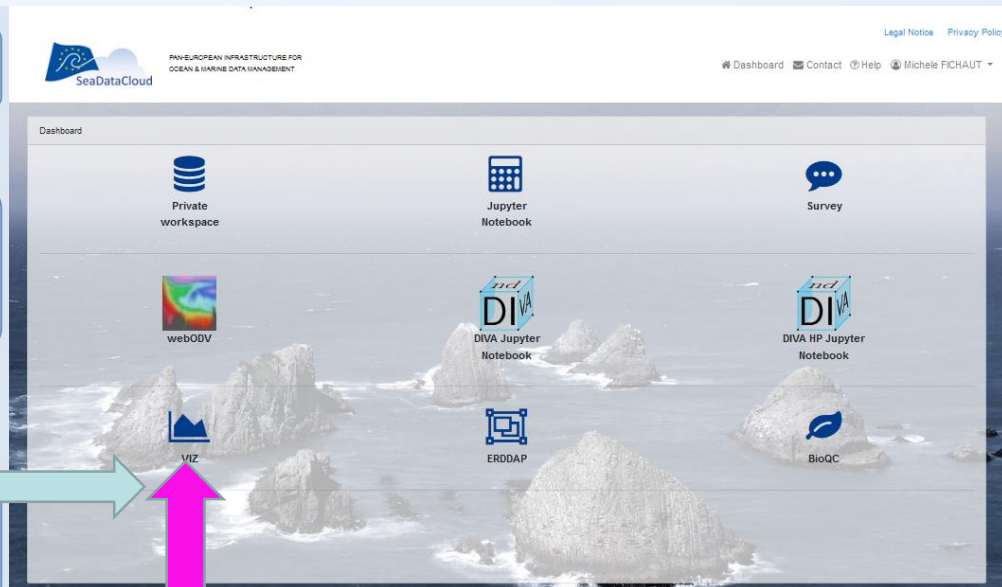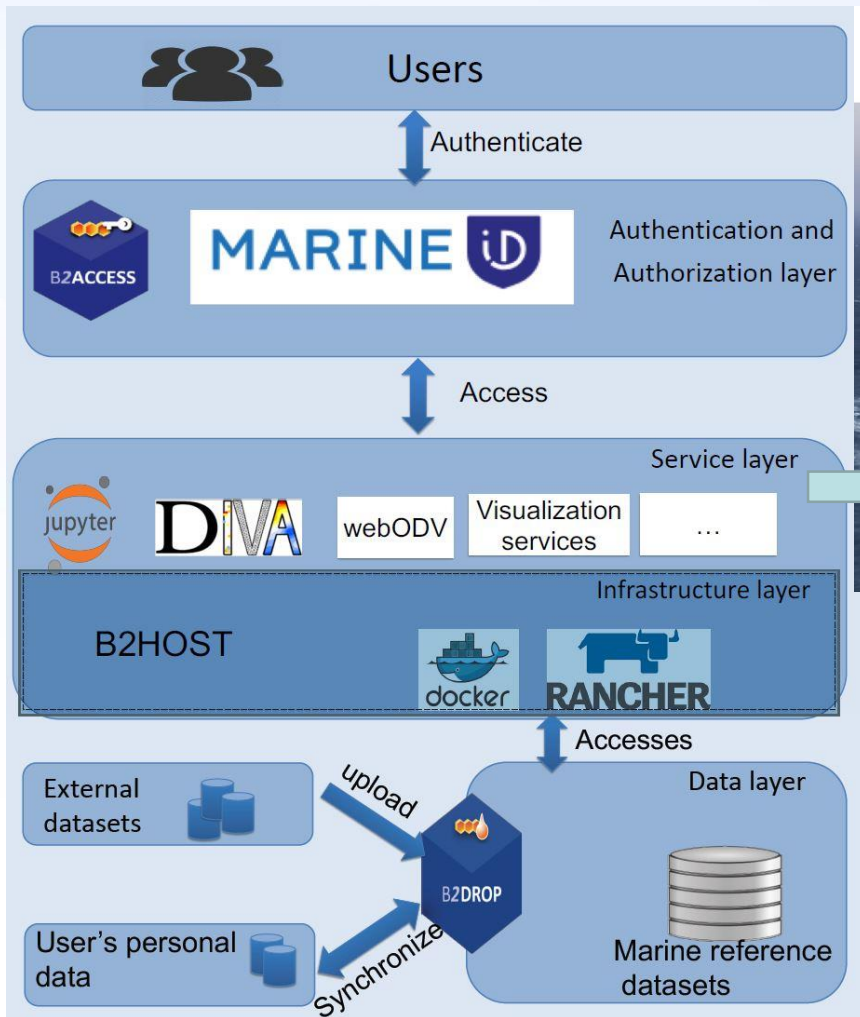
Services in the SDC VRE pilot
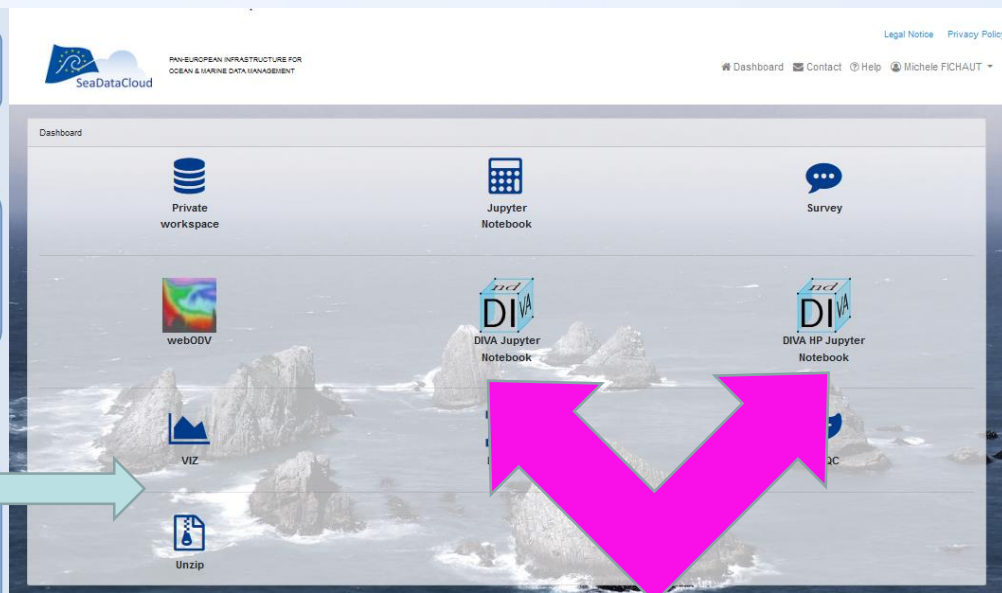
DIVAnd (Data Interpolating Variational Analysis in n dimensions) software tool designed to interpolate in-situ observations onto a regular grid.

A set of Jupyter notebooks provides a guideline to the user on how to prepare the data, optimise the analysis parameters and perform the interpolation

Services in the SDC VRE pilot

VIZ is a modern and dynamic visualisation service to explore datasets on a map.

By clicking on data points, the users see a plot of the full profile prepared with WebODV, and metadata of the input ODV files. A time selector permits to limit the data for the period of interest. Additionally, the visualisation service provides the possibility to explore 4D gridded products prepared with DIVAnd

Services in the SDC VRE pilot
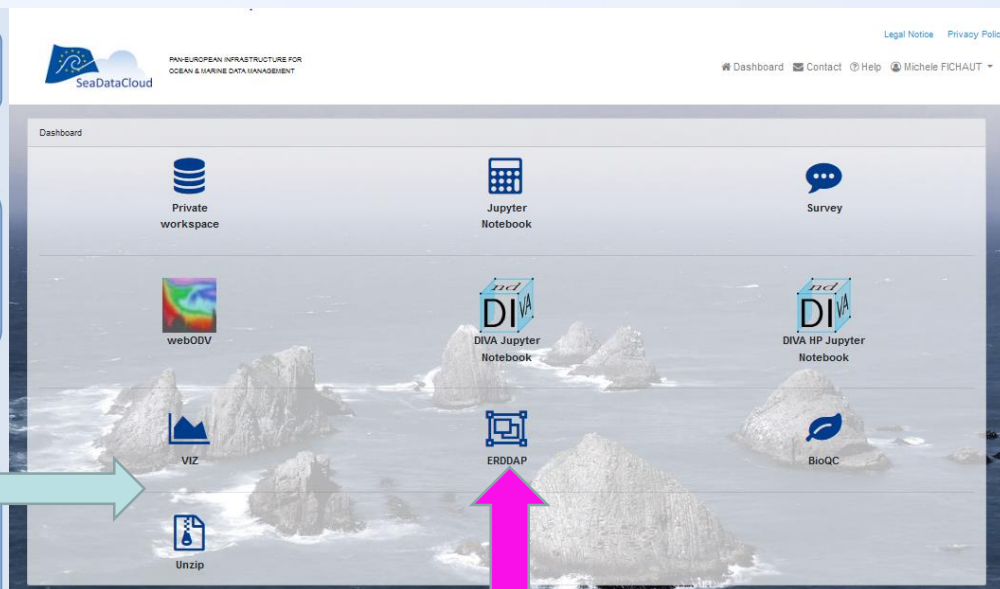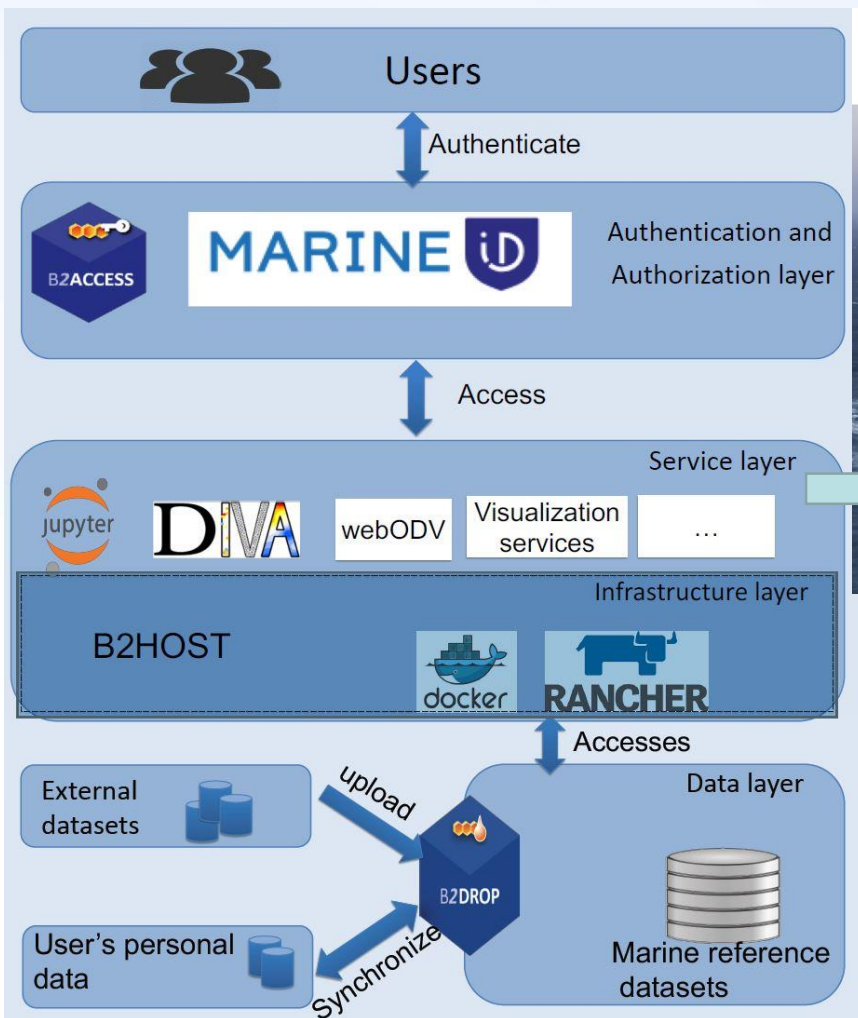
The **Subsetting service** based on ERDDAP strives to make data access easier, by providing services to subset, download and plot data.

It returns datasets in various data formats such as CSV, MATLAB, netCDF, ODV and more. An interactive visualization tool allows users to discover and browse through the subset results with modern web technologies.

Services in the SDC VRE pilot
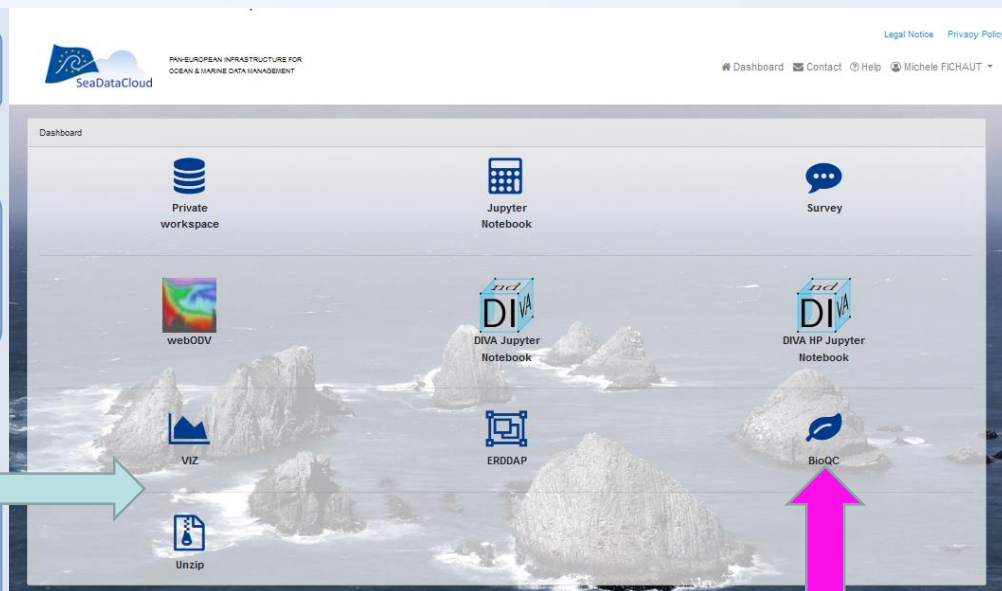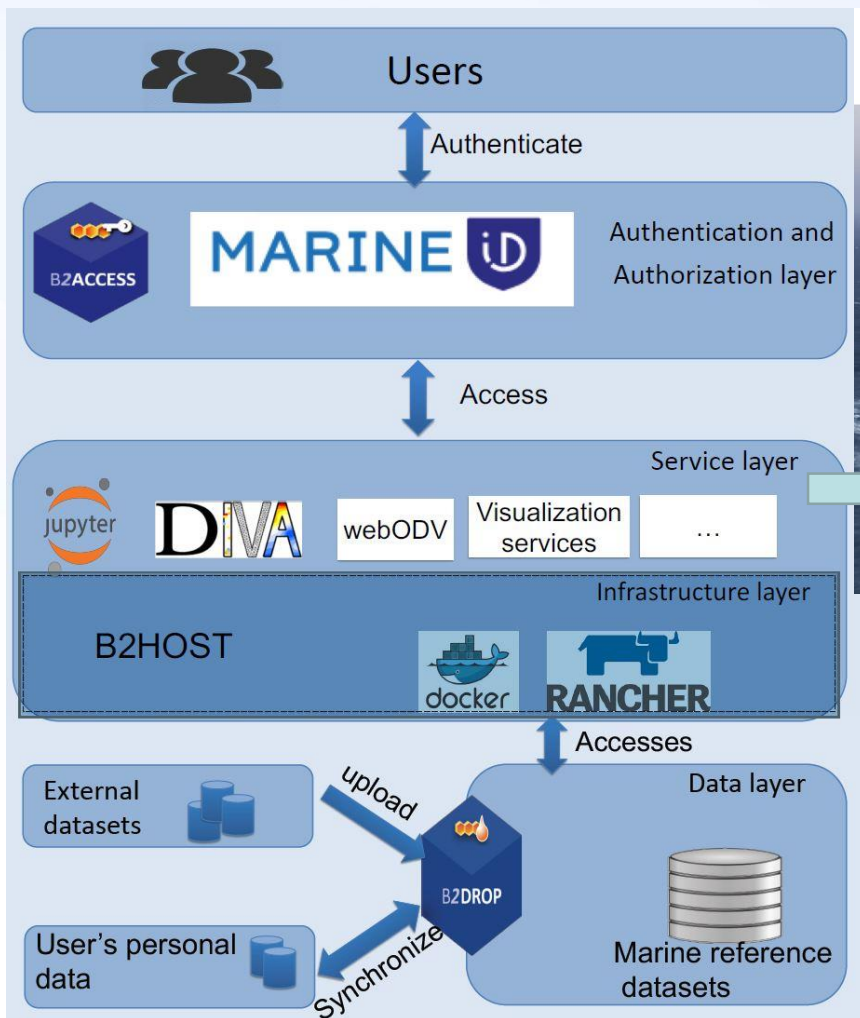
BioQC is a tool to process and to run quality control on biological datasets.

The tool returns the input file with quality information attached for each occurrence record and a detailed report. This result file will enable the users to filter for suitable records.

FAIRNESS of SeaDataNet infrastructure

- SeaDataNet is actively improving the FAIRness of SDN data and metadata by
  - Enriching metadata with information on data collection, on QA-QC activities….
  - Applying Linked Data principles to all services (SPARQL, RDF for example)
  - Harmonising the URLs of the SDN services (GUI and SPARQL end-points)
  - Ensuring SDN data file format conformance
  - Pushing data centres to open their data

THANK YOU