



SeaDataCloud

SeaDataCloud technical challenges

Dick M.A. Schaap, MARIS, SDC Technical Coordinator

SDC Review, Brussels – Belgium, 6 December 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org



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**

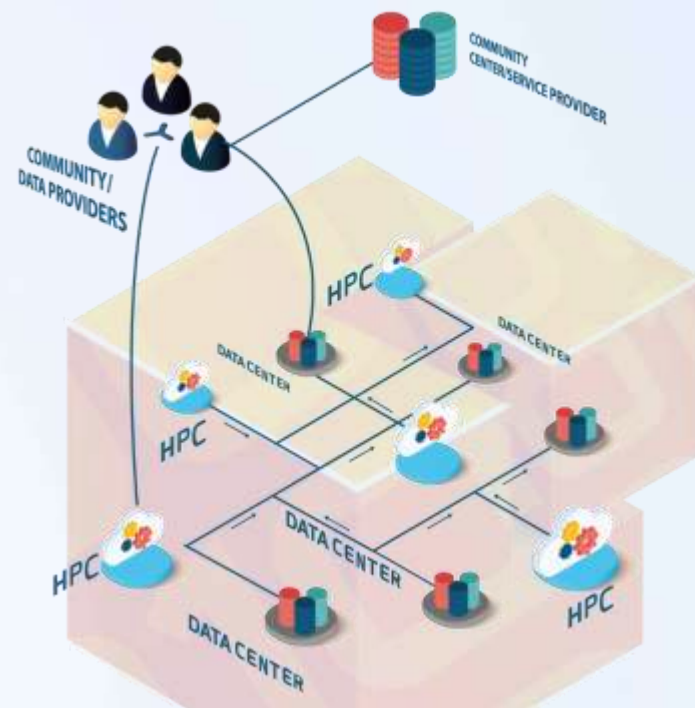
EOSC should to give the EU a global lead in research data management and ensure that European scientists reap the full benefits of data-driven science. It also foresees setting up a European Data Infrastructure, with high-capacity cloud solutions with super-computing capacity.

SeaDataNet anticipating EOSC

- In anticipation of these EOSC planning activities, SeaDataNet submitted in **March 2016** the **SeaDataCloud proposal** 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***

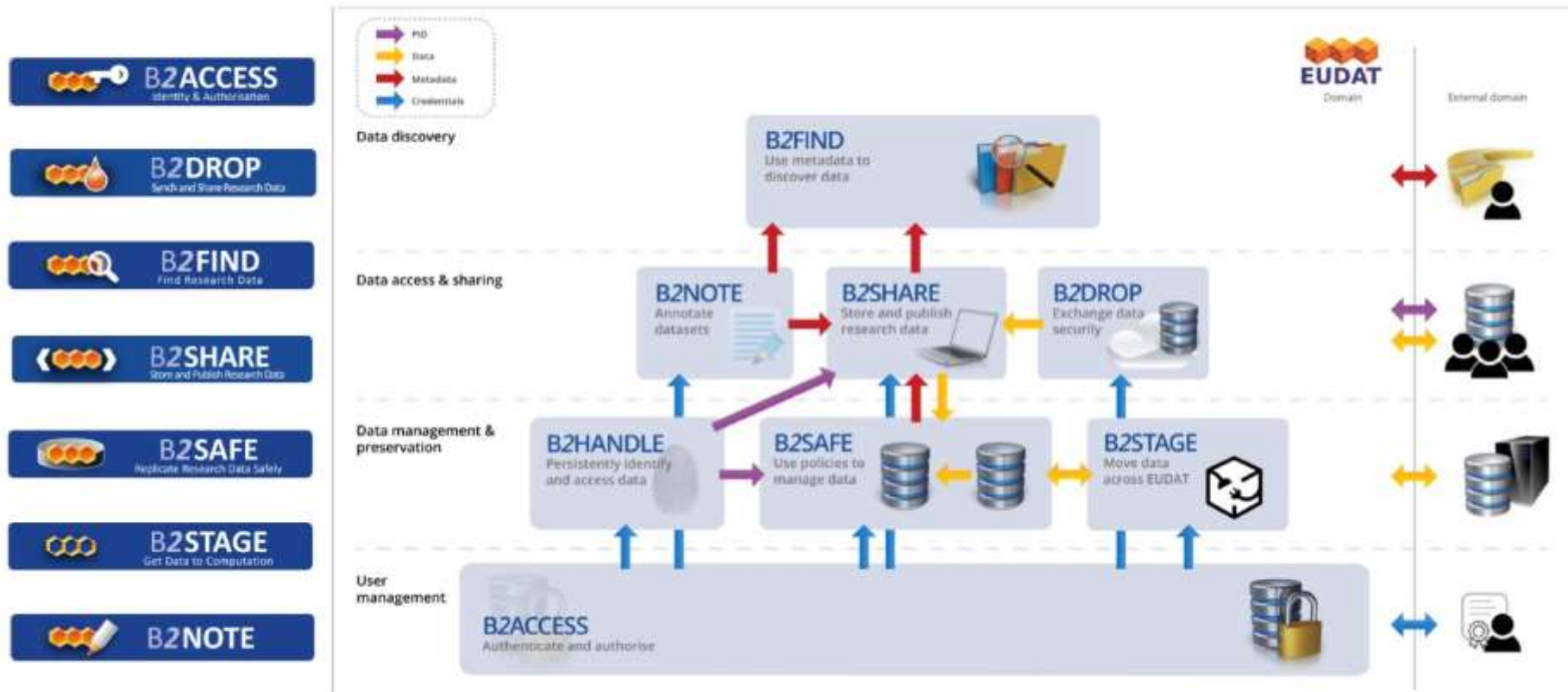


Cooperation with EUDAT



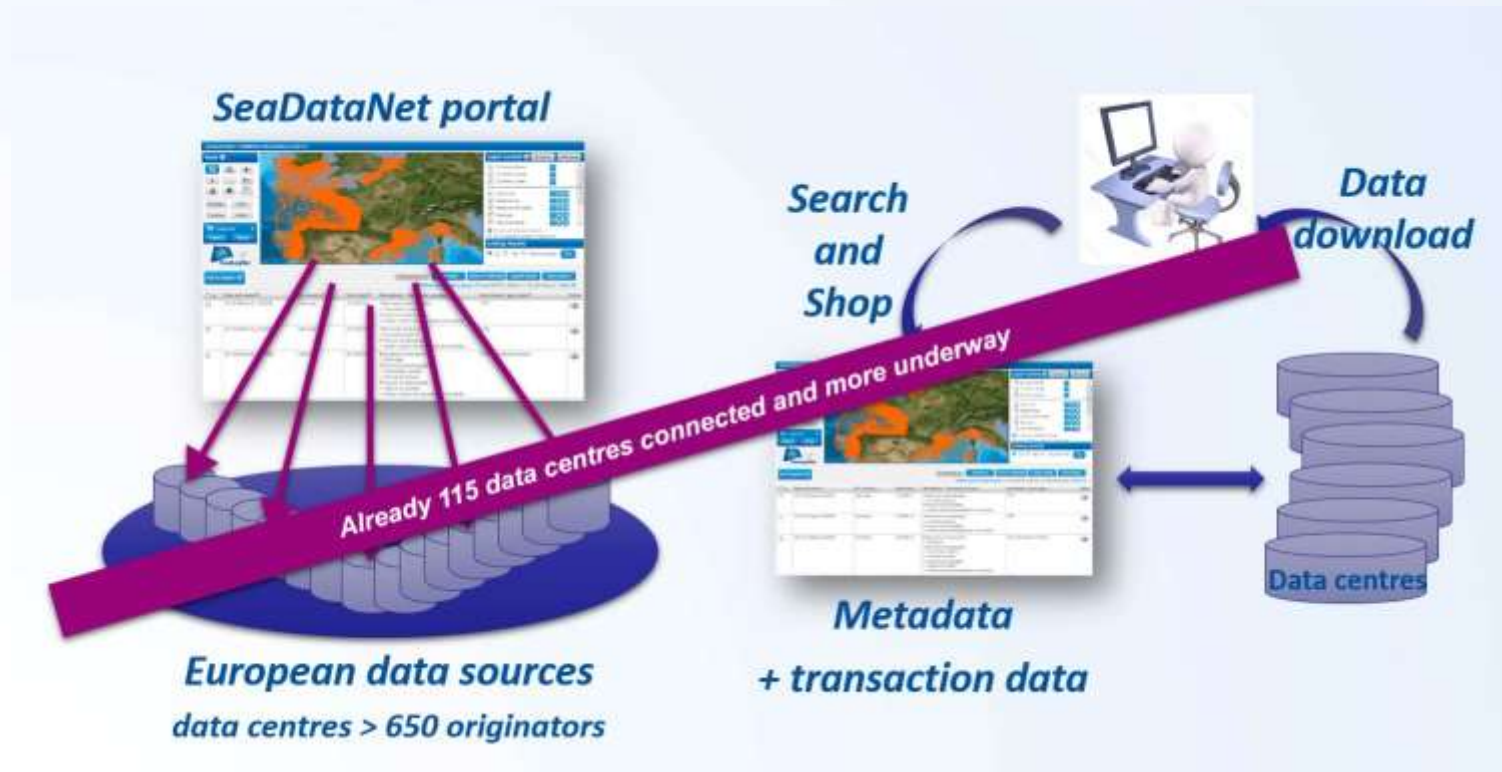
5 EUDAT members are partners of SeaDataCloud :
CINECA, CSC, DKRZ, GRNET and STFC

Cooperation with EUDAT

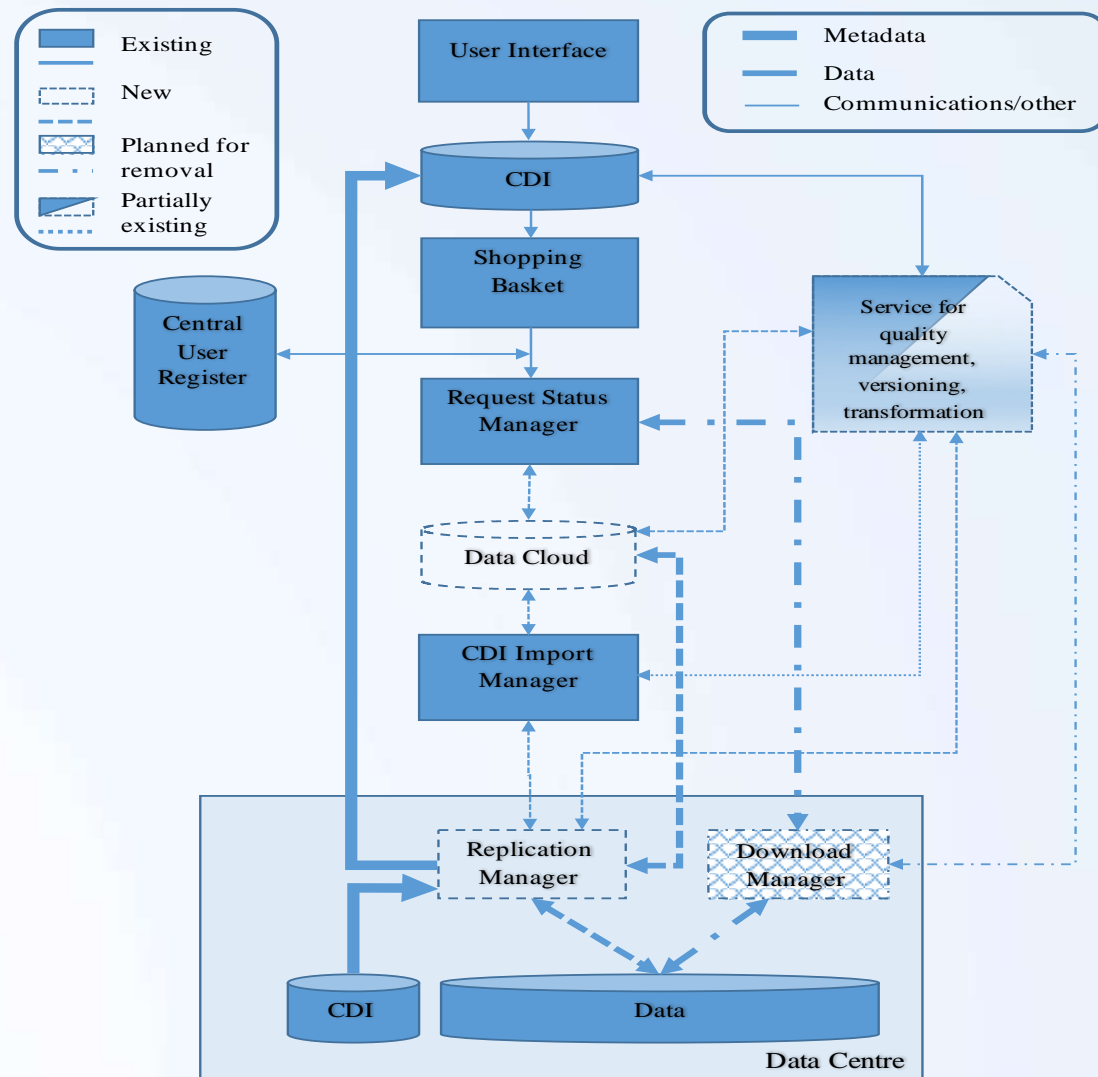


SeaDataCloud technical challenges

- SDC is about **improving and innovating services** such as:
 - upgrading the CDI Data Discovery & Access service by adopting the cloud
 - developing MySeaDataCloud custom services

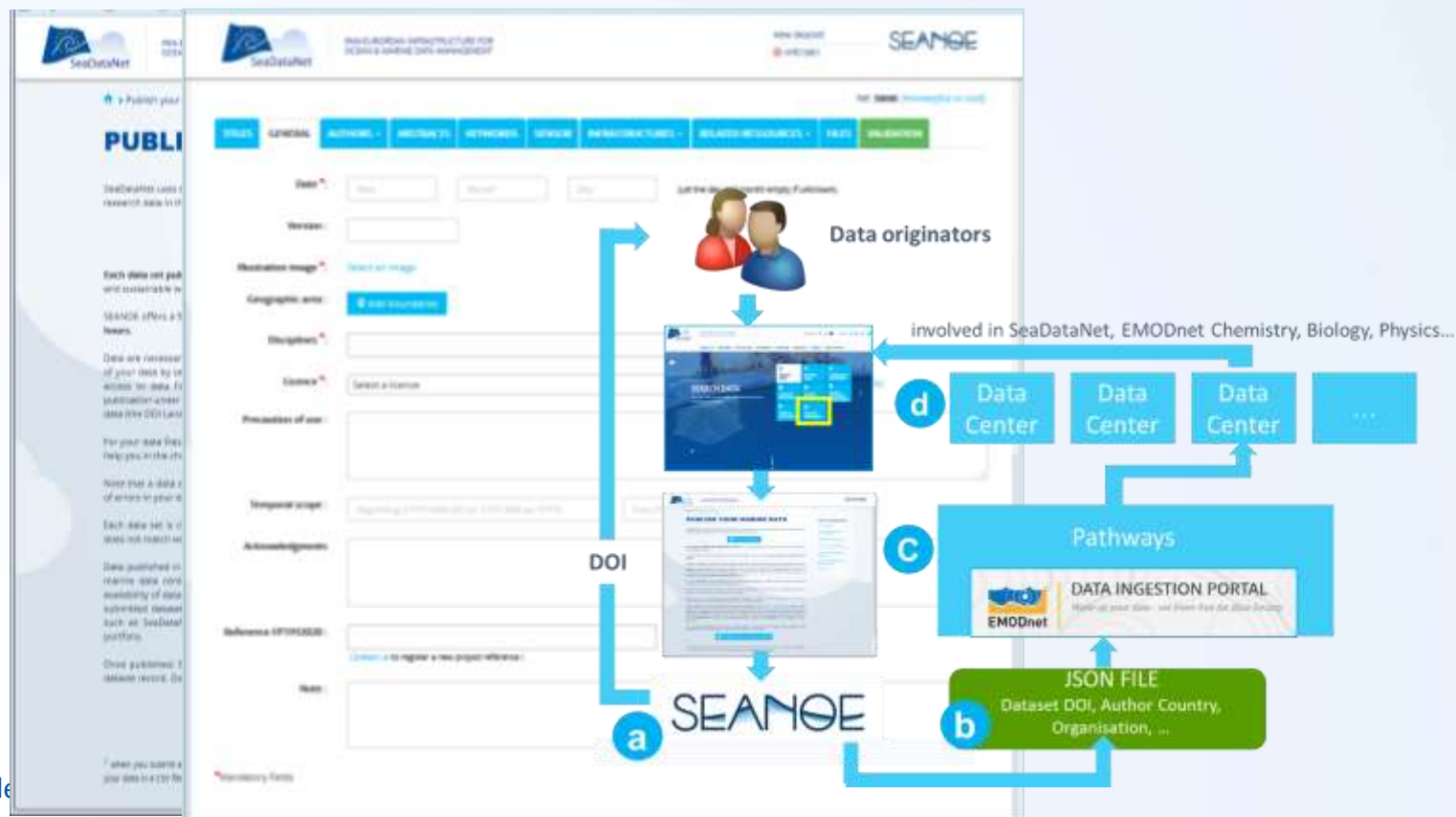


New CDI service architecture



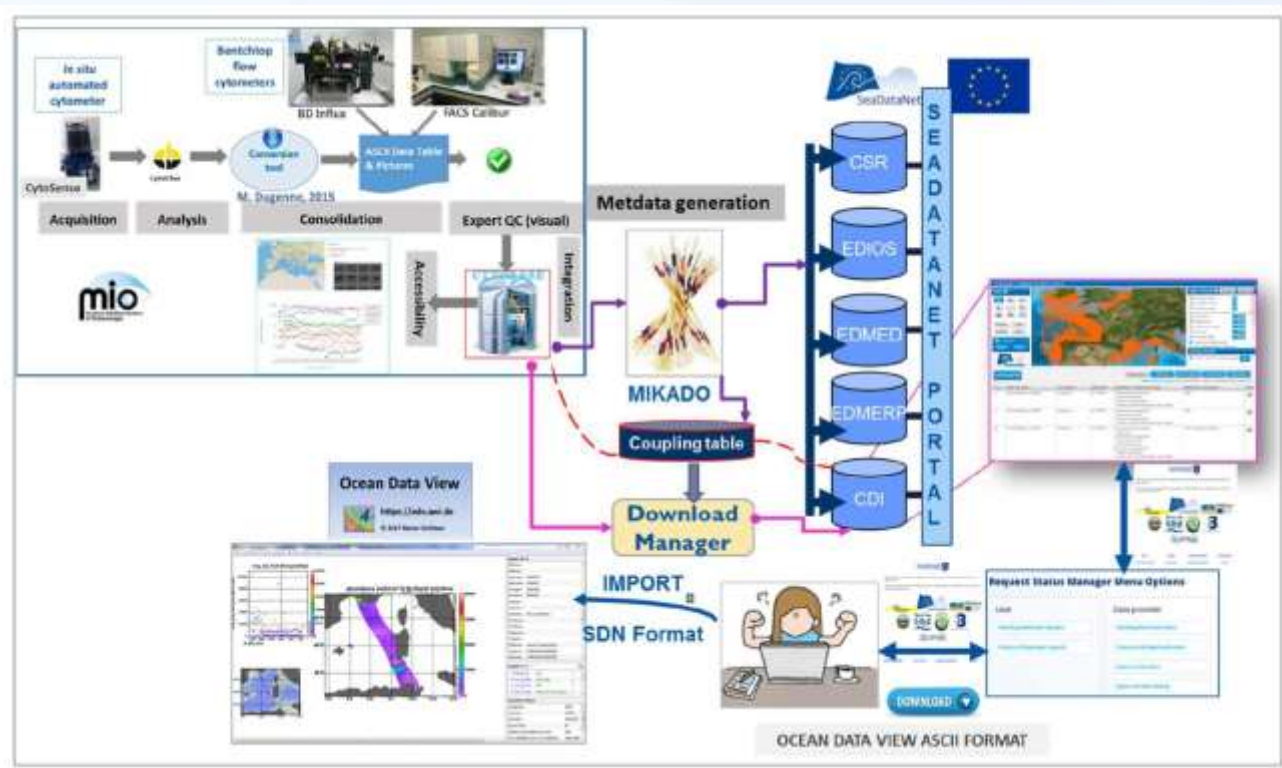
SeaDataCloud technical challenges

- SDC is about **improving and innovating services** such as:
 - upgrading the monitoring of the SeaDataNet system
 - Deploying a DOI minting service in combination with EMODnet Ingestion



SeaDataCloud technical challenges

- SDC is about **updating** and **further developing** standards, such as:
 - expanding vocabularies
 - linked data principle for richer metadata and searches
 - handling different data types such as HF Radar, Gliders, Flow cytometer data



SeaDataCloud – INSPIRE compliance

Environmental Monitoring
Facilities **EF** theme (Annex III)

Oceanographic geographical
Features **OF** theme (Annex III)

*Location
environ
facilities*



MarineVessel :EnvironmentalMonitoringFacility

inspireId = <http://vocab.ndg.nrc.ac.uk/term/P173/0/45CE>
name = Celtic Explorer
additionalDescription = The RV Celtic Explorer is 55.5m length marine research vessel and accommodates 35 personnel, including 19-21 scientists
legalBackground = Marine Institute Act, 1992
responsibleParty = Marine Institute
onlineResource = <http://www.marine.ie/home/services/researchvessels/explore/home.htm>
purpose = <http://vocab.ndg.nrc.ac.uk/term/C77/3/M06>
measurementRegime = <http://inspire.ec.europa.eu/codeList/MeasurementRegimeValue/periodicDataCollection>
mediaMonitored = <http://inspire.ec.europa.eu/codeList/MediaValue/water>
mobile = True
specialisedEMFType = <http://vocab.ndg.nrc.ac.uk/term/L06/7/31>

+broader
+narrower



Automated Sampling Device :EnvironmentalMonitoringFacility

inspireId = <http://vocab.ndg.nrc.ac.uk/.../xxx>
additionalDescription = Rosette® Multi bottle Water Sampling Systems can accommodate 12 or 24 bottles attached to an ...
measurementRegime = <http://inspire.ec.europa.eu/codeList/MeasurementRegimeValue/periodicDataCollection>
mediaMonitored = <http://inspire.ec.europa.eu/codeList/MediaValue/water>
mobile = True
specialisedEMFType = <http://vocab.ndg.nrc.ac.uk/term/L06/7/45>
resultAcquisitionSource = <http://inspire.ec.europa.eu/codeList/ResultAcquisitionSourceValue/exSitu>

+broader
+narrower

WaterBottle :EnvironmentalMonitoringFacility

inspireId = <http://vocab.ndg.nrc.ac.uk/.../yyy>
additionalDescription = One water bottle wfrom the Rosette® Multi bottle Water Sampling Systems
measurementRegime = <http://inspire.ec.europa.eu/codeList/MeasurementRegimeValue/periodicDataCollection>
mediaMonitored = <http://inspire.ec.europa.eu/codeList/MediaValue/water>
mobile = True
specialisedEMFType = <http://vocab.ndg.nrc.ac.uk/term/L051/26/30>
resultAcquisitionSource = <http://inspire.ec.europa.eu/codeList/ResultAcquisitionSourceValue/exSitu>

Meta

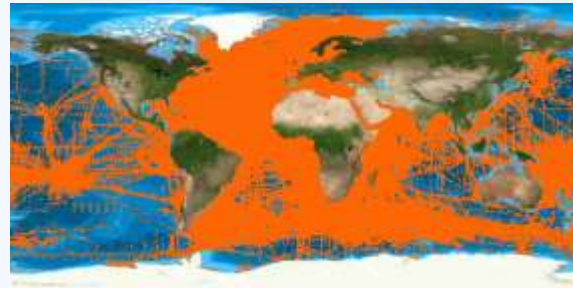
- [
- [
- [

Relevant

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

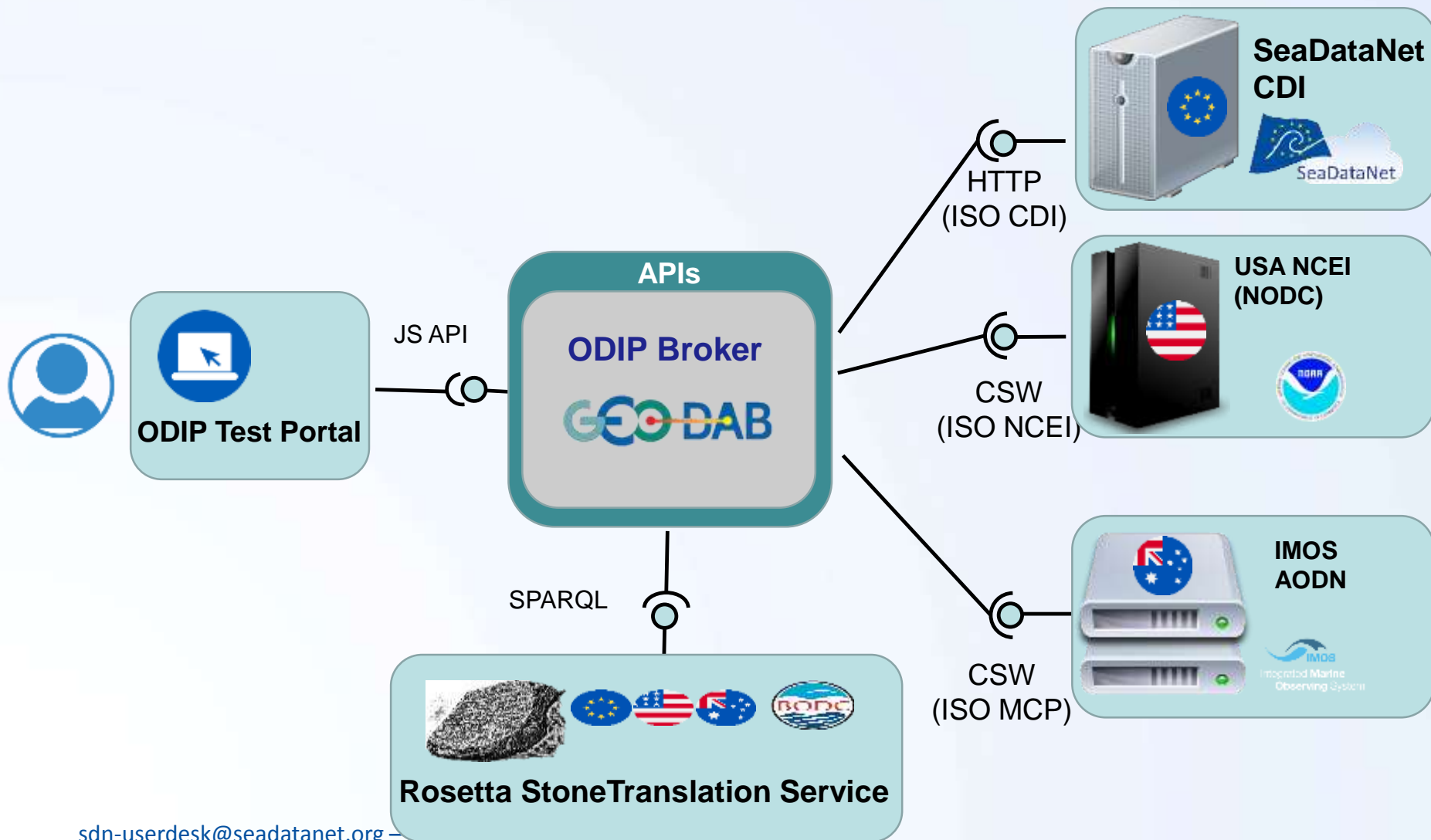
Integration of multiple data streams

- Further expansion of the number of data sets populated in the **CDI service** by an increasing number of connected data centres, in particular through SeaDataCloud, EMODnet, AtlantOS, and other projects.

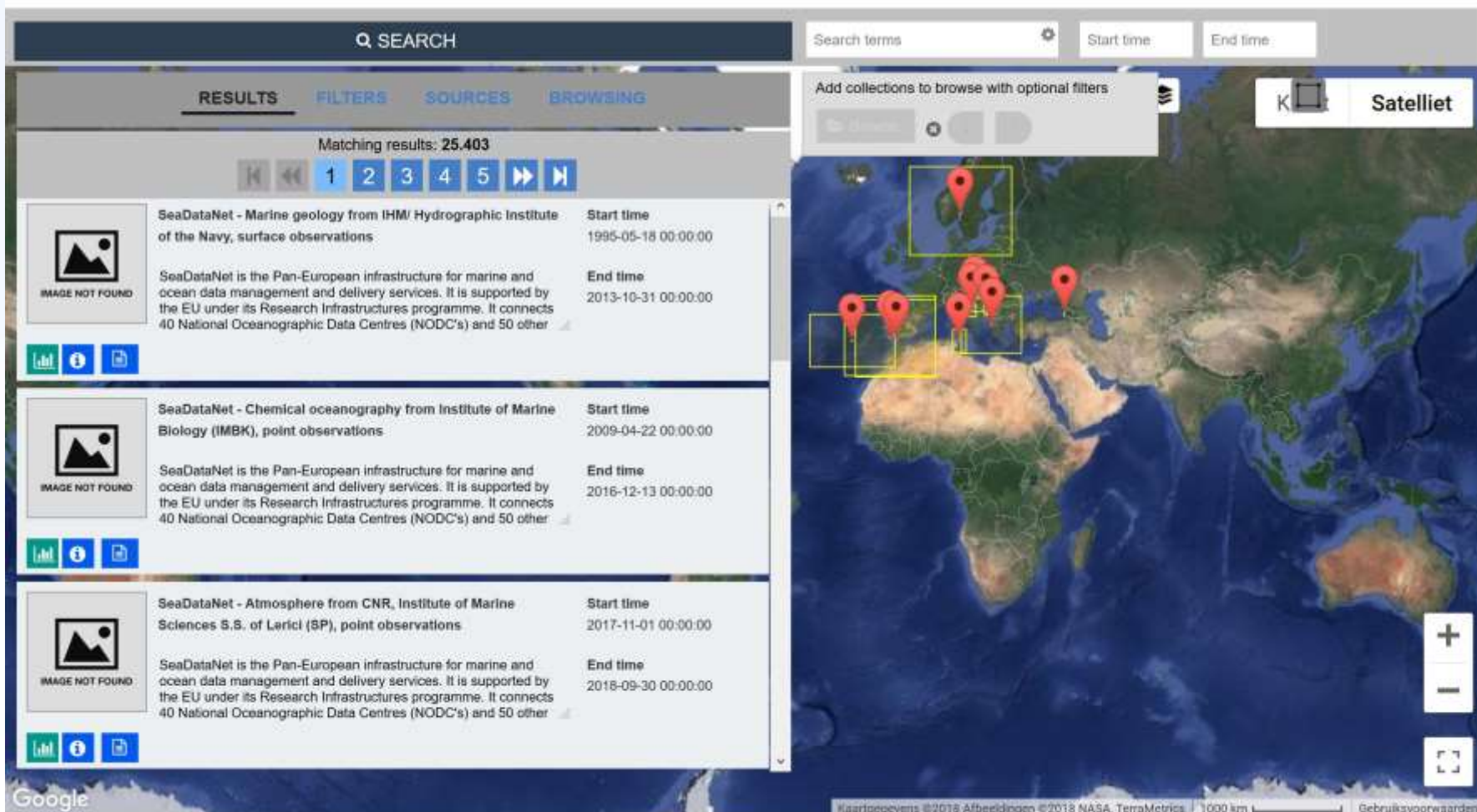


- Integrating and operating a **Brokerage Service** following ODIPII (Ocean Data Interoperability Platform) for connecting other major international marine data discovery and access services such as US NCEI, Australian AODN, and others. This is done through seeking interoperability. SDN users will be offered an extra facility for discovering available data sets in those other systems and given their links and services to access associated data sets;

Brokering Prototype – Semantic Discovery



ODIP Brokerage Service



The screenshot displays the ODIP Brokerage Service search interface. At the top, there is a search bar with the text "SEARCH" and a magnifying glass icon. Below the search bar, there are tabs for "RESULTS", "FILTERS", "SOURCES", and "BROWSING". The "RESULTS" tab is active, showing a list of search results. The results are displayed in a grid format, with each result showing a thumbnail image (labeled "IMAGE NOT FOUND"), a description, and a date range (Start time and End time). The results are filtered by "Satelliet" and "K". The search results show 25,403 matching results. The results are displayed in a grid format, with each result showing a thumbnail image (labeled "IMAGE NOT FOUND"), a description, and a date range (Start time and End time). The results are filtered by "Satelliet" and "K".

Search terms: [Search terms] Start time: [Start time] End time: [End time]

Add collections to browse with optional filters: [Add collections to browse with optional filters]

Matching results: 25,403

RESULTS FILTERS SOURCES BROWSING

SeaDataNet - Marine geology from IHM/ Hydrographic Institute of the Navy, surface observations

Start time: 1995-05-18 00:00:00

End time: 2013-10-31 00:00:00

SeaDataNet - Chemical oceanography from Institute of Marine Biology (IMBK), point observations

Start time: 2009-04-22 00:00:00

End time: 2016-12-13 00:00:00

SeaDataNet - Atmosphere from CNR, Institute of Marine Sciences S.S. of Lerici (SP), point observations

Start time: 2017-11-01 00:00:00

End time: 2018-09-30 00:00:00

Google

Kaartgegevens ©2018 Afbeeldingen ©2018 NASA, TerraMetrics 1000 km Gebruiksvoorwaarden

<http://odip-prototype.essi-lab.eu/broker/search>

Integration of multiple data streams

- Using **EMODnet Ingestion portal** where third parties can ingest data collections. Submissions are assigned to expert data centres.

A screenshot of the EMODnet Data Ingestion Portal website. The header features the EMODnet logo, the title "DATA INGESTION PORTAL", a search bar, and a "CONTACT" link. A navigation menu includes links for ABOUT, DATA SUBMISSION, OPERATIONAL DATA, SUBMISSIONS, GUIDELINES, DATA WANTED, HELP, PROMOTION, and CENTRAL PORTAL. The main content area starts with a "Home" link and a welcome message: "Welcome to the EMODnet Data Ingestion portal". It describes the European Marine Observation and Data Network (EMODnet) and its mission. A "READ MORE" button is present. Below this is a large orange banner with the text "EMODNET INGESTION AND HOW DOES IT WORK". The page is divided into three columns. The first column, titled "Submit your data files", shows a hand clicking a blue "submit" button on a keyboard. The second column, titled "Ingest operational data", shows a blue background with a line graph and data points. The third column, titled "View submissions", shows a blue background with the text "DATA ACCESS". Each column has a brief description of the service.

EMODnet

DATA INGESTION PORTAL
Wake up your data - set them free for Blue Society

Search

CONTACT

ABOUT DATA SUBMISSION OPERATIONAL DATA SUBMISSIONS GUIDELINES DATA WANTED HELP PROMOTION CENTRAL PORTAL

Home

Welcome to the EMODnet Data Ingestion portal

The European Marine Observation and Data Network (EMODnet) consists of more than 160 organisations that together work on assembling, harmonising and making marine data, products and metadata more available to public and private users. This Data Ingestion portal facilitates additional data managers to ingest their marine datasets for further processing, publishing as open data and contributing to applications for society.

[READ MORE](#)

EMODNET INGESTION AND HOW DOES IT WORK

Submit your data files

The online Data Submission service facilitates you to submit marine datasets by completing a form and uploading your data as a file package. The service

Ingest operational data

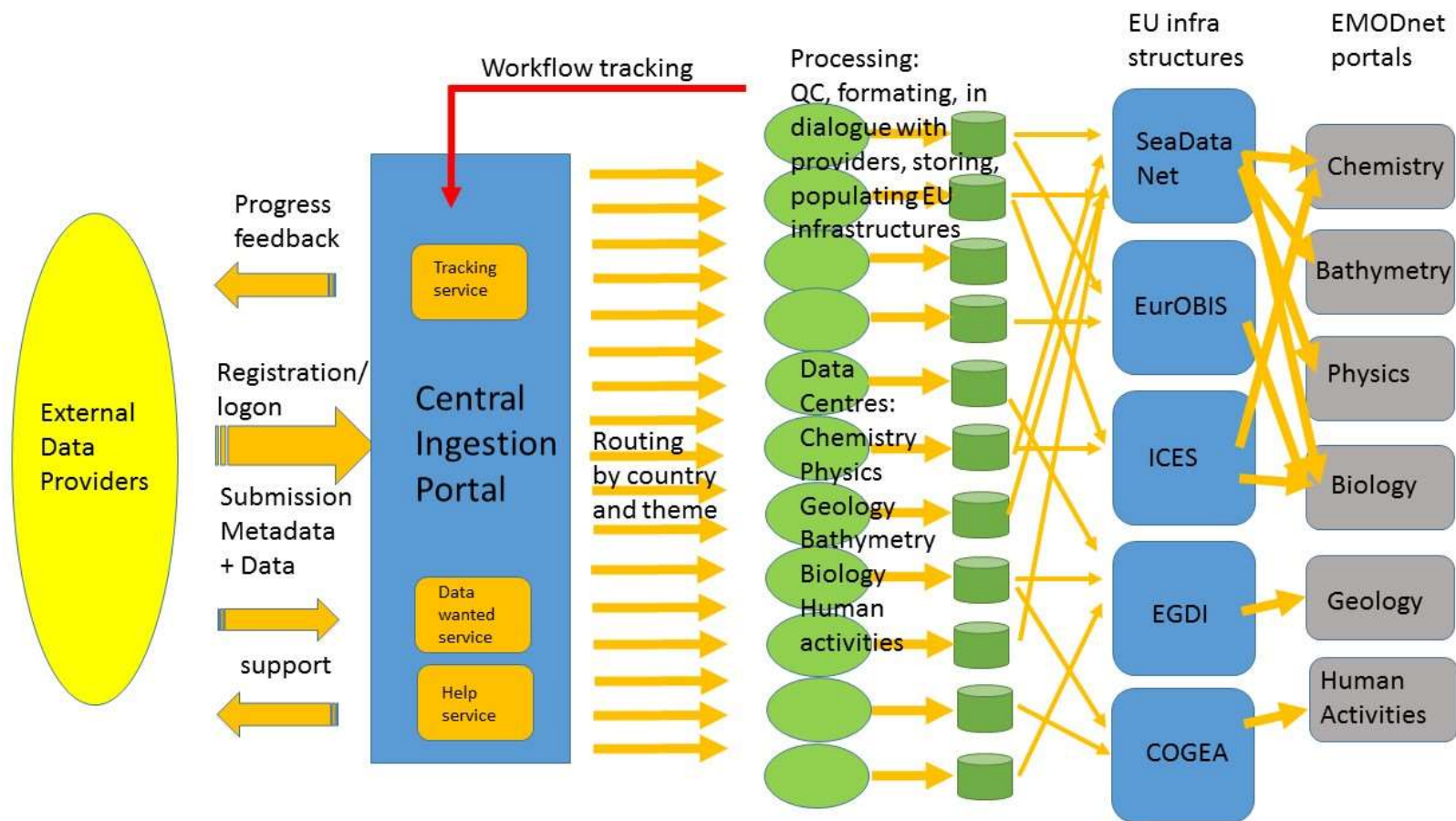
We are also interested in (Near) Real-Time ((N)RT) data streams from fixed and autonomous ocean observing platforms. This section explains how you

View submissions

View, search and download datasets that have been submitted by data providers using the Data Submission service.

Integration of multiple data streams

- EMODnet Ingestion Pathways for working up the ingested data sets.



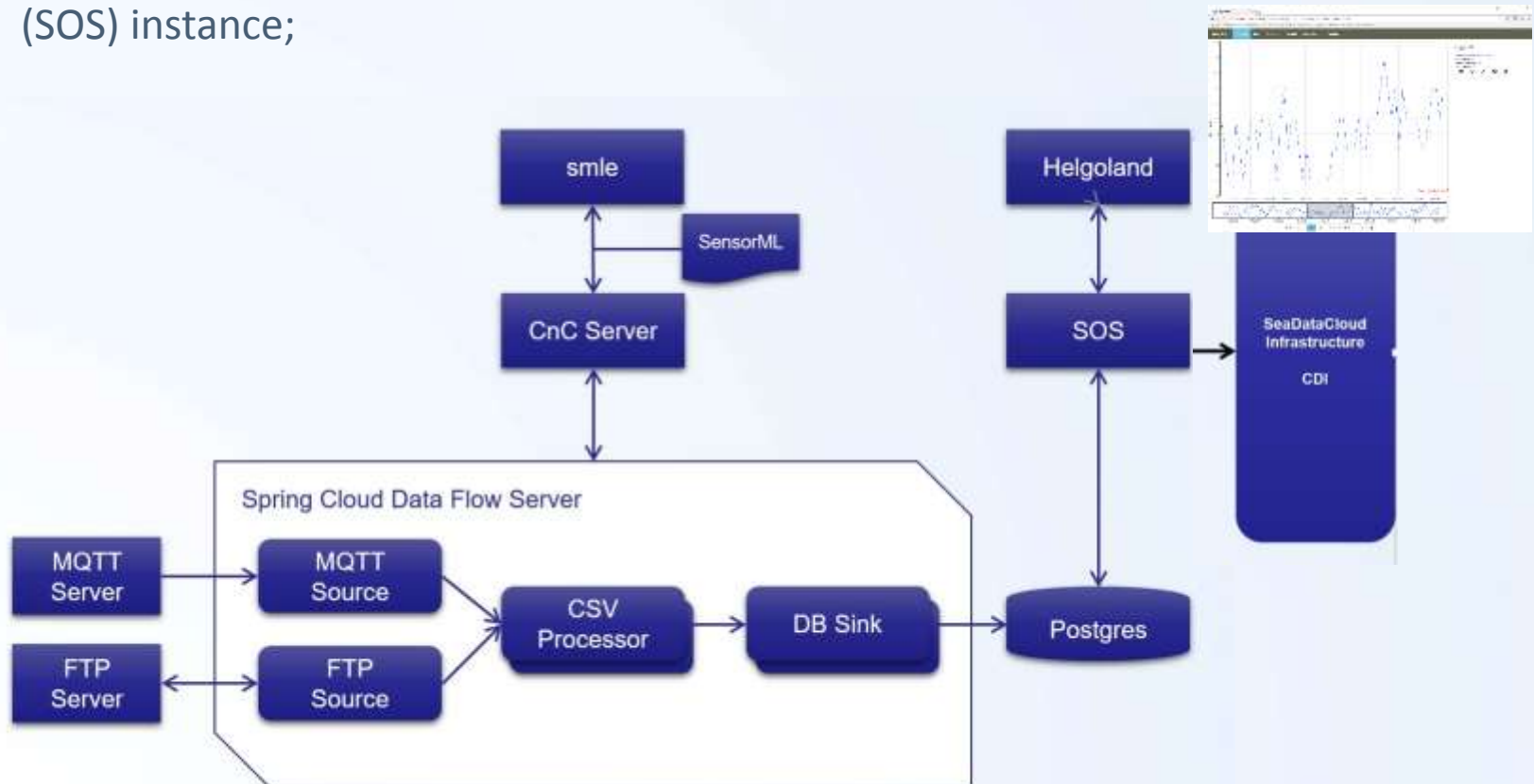
Integration of multiple data streams

Integrating the **online Sensor Web Enablement (SWE) Ingestion service (52N)** for ingesting near real time data sets from operational oceanography sensor networks:

- Closing the gap between operational oceanography and delayed mode validation and archiving. The SWE ingestion service will facilitate streamlining the data and metadata transfer from operational networks to a database buffer from which assigned SDN data centres can pick up the data timeseries for further elaboration and later population into the CDI service
- Enriching and expanding the CDI service for its users with access to historical long timeseries AND the latest NRT data timeseries from the operational sensors linking to the 52°North “Helgoland” viewer. As part of SeaDataCloud, Helgoland will be further extended in close cooperation with users. This comprises especially the support of further data types as well as usability improvements.

Integration of multiple data streams

- a backend implementation to perform the collection of data from different sources, the interpretation and conversion into an internal data model (based on the ISO/OGC Observation and Measurements (O&M) standard, as well as the publication of the collected data into the database of an interoperable OGC Sensor Observation Service (SOS) instance;

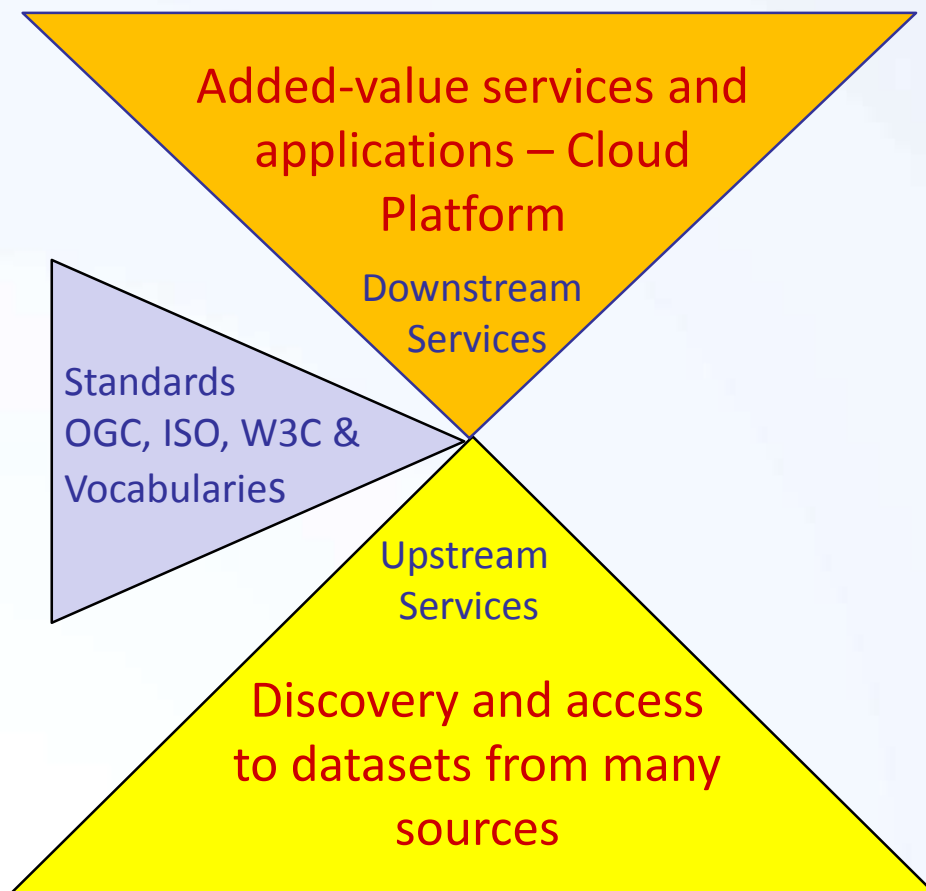


SeaDataCloud technical challenges

- SDC is about **adopting and elaborating new technologies**, such as:
 - developing extra tools for data managers, such as OCTOPUS for SeaDataNet file format compliance checking and format conversions
 - developing a **Virtual Research Environment (VRE)** for researchers with dedicated cloud-based tools for processing and analyzing data



Leading concept => towards a Blue Cloud

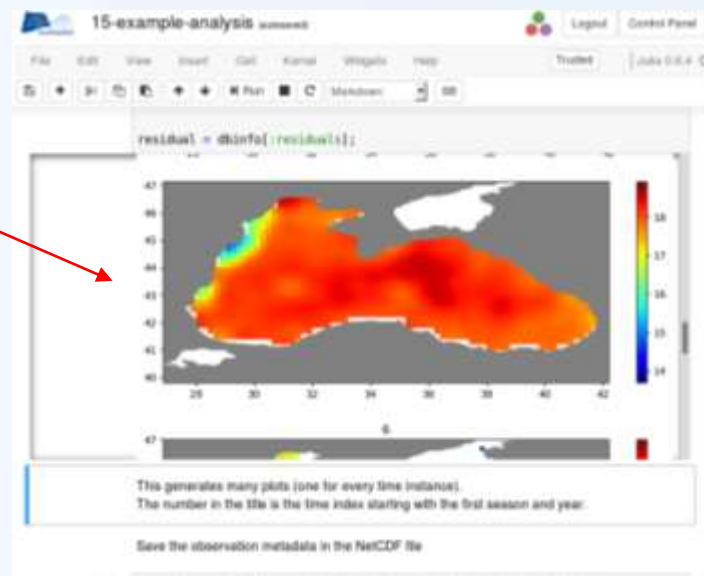
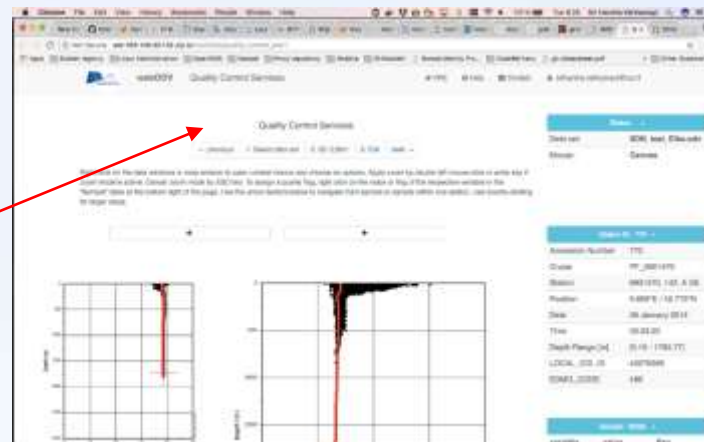
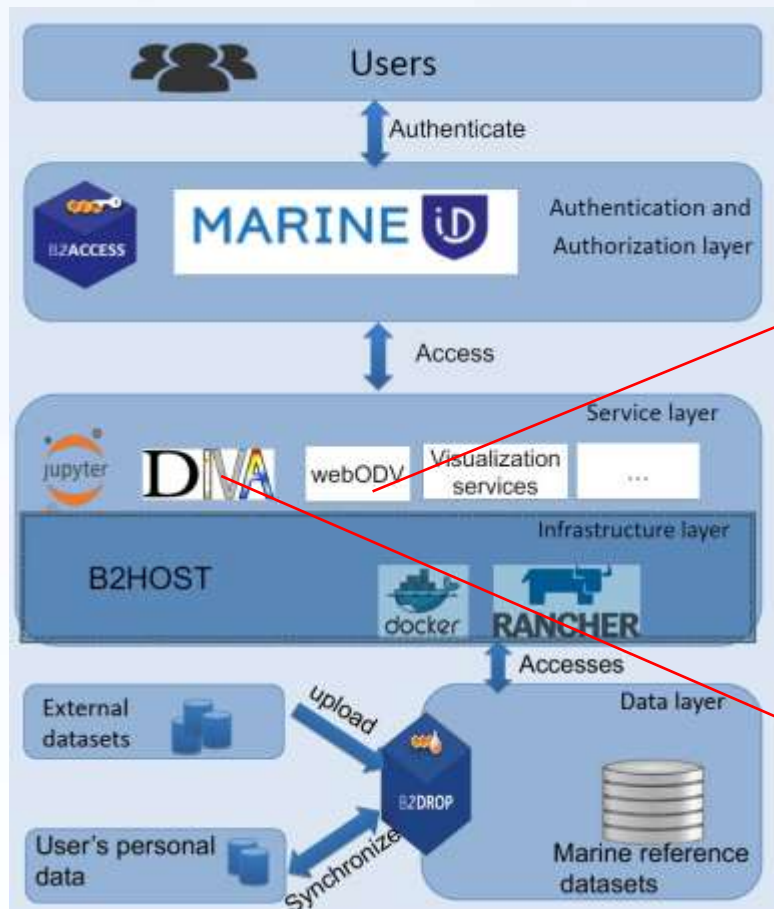


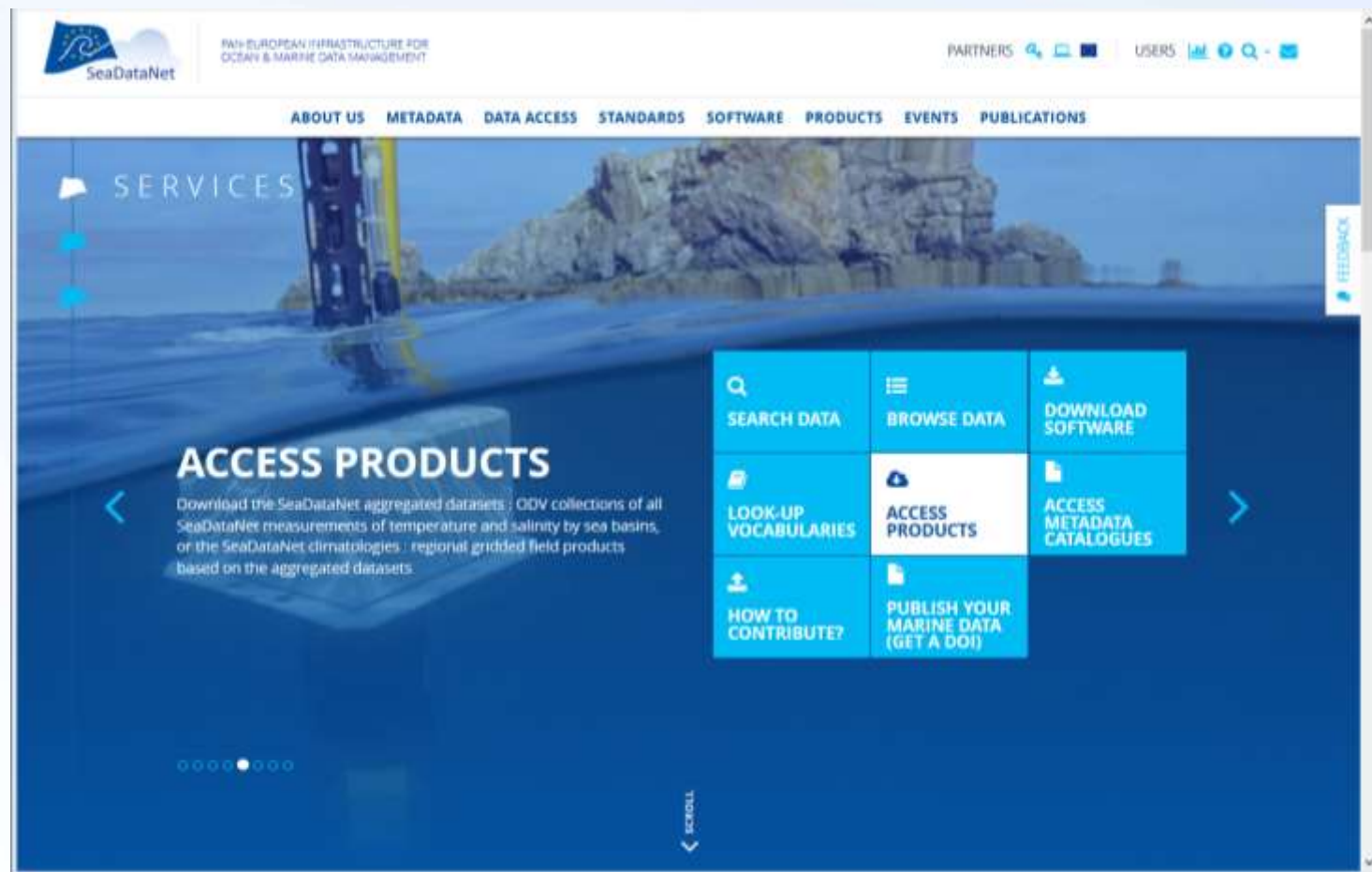
- Cloud platform with common services for data pre-processing, analyses, visualizations, publishing, DOIs...
- Applying common standards and interoperability solutions for providing harmonised data and metadata
- Providing harmonised discovery and access to data output from multiple sources, European and international

Virtual Research Environment (VRE)

- Reviewing of existing VRE developments in Australia, USA and Europe, also as part of the Ocean Data Interoperability Platform (ODIP) activities
- Formulating a ToR for the VRE overall specifications
- Analysing and defining an overall architecture for the SDC VRE, considering EUDAT's generic services and embedding dedicated SDN services, custom and generic
- Analysing the workflow and specific functions required for supporting the SDC pilot: **generating a SeaDataNet Temperature & Salinity climatology**
- Developing VRE components and integrating these into the SDC VRE pilot

Overall architecture for the SDC VRE





www.seadatanet.org