



SeaDataCloud

Upgrading the CDI Data Discovery and Access service

Intro, background and components by Dick MA Schaap –
SeaDataCloud Technical Coordinator

Training workshop, June 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org



SeaDataCloud rationale

- Standards and information technology are always evolving, there is a move towards cloud storage and cloud computing, and the SeaDataNet infrastructure must stay up-to-date to maintain and further expand its **standards and services** to its lead customers and major stakeholders
- A strategic and operational cooperation between the SeaDataNet consortium of marine and ocean data centres and the EUDAT consortium of e-infrastructure service providers, also with a perspective to EOSC
- SeaDataCloud project, started Nov 2016 with 4 year run

Cooperation with EUDAT



European Computing Infrastructure



A consortium of high performance computing (HPC) / data centres, libraries, scientific communities, data scientists

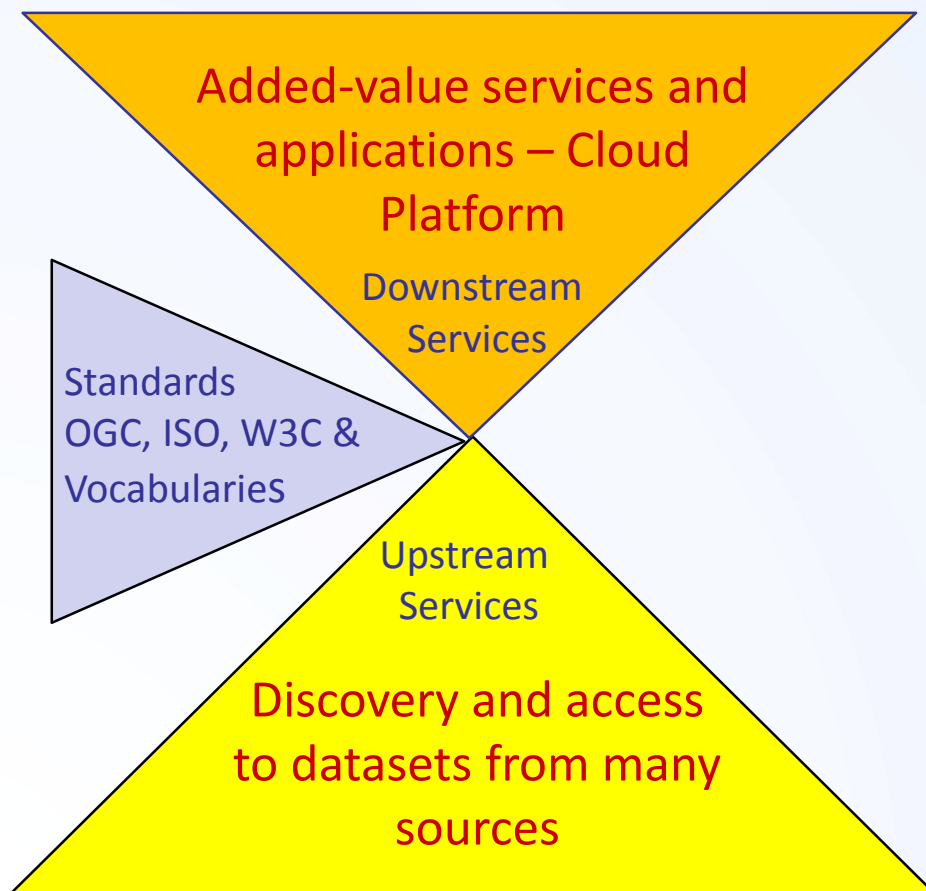




General challenges

- SeaDataCloud is the successor to the SeaDataNet II project
- It is about updating and further developing standards
- It is about improving and innovating services & products
- It is about adopting and elaborating new technologies
- It is about giving more attention to users and putting the user experience in a central position

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

SeaDataNet cooperation

- **Copernicus Marine Environmental Monitoring Services (CMEMS):** providing long-term archives and standards
- **Marine Strategy Framework Directive (MSFD):** providing infrastructure, standards and data collections for several indicators
- **Large ocean monitoring systems and their projects (EuroGOOS, AtlantOS, Euro-ARGO, JERICO-Next, ..):** providing standards and validation + long-term archiving services
- **EU projects, such as Upgrade BlackSeaScene, CaspInfo, Geo-Seas, Eurofleets ...:** adopting and adapting SeaDataNet standards and services for developing marine data management capabilities
- **Ocean Data Interoperability Platform (ODIP):** exploring and demonstrating common standards and interoperability with leading data management infrastructures in USA and Australia
- **GEOSS - EuroGEOSS:** Maintaining the GEOSS portal with SeaDataNet in-situ data collections from large community of European data holders (> 100 data centres; >600 data originators)
- **European Open Science Cloud (EOSC):** shaping the Blue Cloud

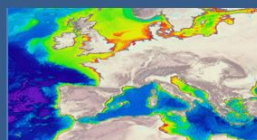
SeaDataNet and EMODnet

- EU initiative for an overarching **European Marine Observation and Data Network (EMODNet)** driven by Marine Knowledge 2020 and Blue Growth
- SeaDataNet qualified as a leading infrastructure for the EMODnet data management component and is driving several thematic portals from the start in 2008
- *'Bottom-up meets top-down'*
- This synergy has resulted in many more data centres adopting SeaDataNet standards and connecting to the SeaDataNet services while it gave a flying start to EMODnet



EMODnet thematic portals

Bathymetry



Minimum cell water depth

Maximum cell water depth

Average cell water depth

Standard deviation of cell water depth

Number of values used for interpolation of cell water depth

Horizontal coordinate reference system

Depth reference system

Lowest Astronomical Tide

Geology



Seabed substrate

Sediment accumulation rate

Sea-floor geology

Seabed lithology

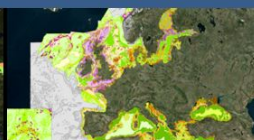
Stratigraphy

Coastline migration

Aggregate resources

Geological events

Seabed Habitats



Data on modelled seabed habitats (depth, seabed substrate, broad scale biological zone, T, S, light, oxygen, energy due to waves & current)

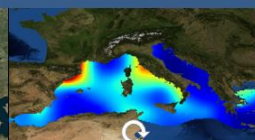
Broad-scale physical habitat map

Detailed habitat maps from surveys

Individual habitat modelling outputs

Habitat point data

Chemistry



Pesticides & Biocides

Antifoulants

Pharmaceuticals

Heavy Metals

Hydrocarbons

Radionuclides

Fertilizers

Acidity

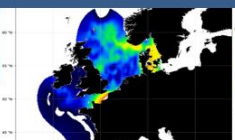
Dissolved Gasses

Plastics

Marine Litter Beach litter Seafloor litter Micro litter

Chlorophyll
Silicates
Organic Matter

Biology



Biomass

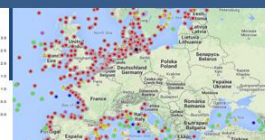
Abundance

Gridded Abundance maps

species groups:

- phytoplankton
- zooplankton
- angiosperms
- macro-algae
- invertebrate bottom fauna
- birds
- mammals
- reptiles
- Fish

Physics



Waves

Water temperature

Water salinity/conductivity/density

Currents

Light attenuation/fluorescence

Sea level

Atmospheric parameters

Wind

Underwater noise

Rivers

Ice

Human activities



Aggregate Extraction Dredging

Fisheries

Hydrocarbon Extraction Main Ports

Mariculture

Ocean Energy Facilities

Pipelines and Cables

Protected Areas

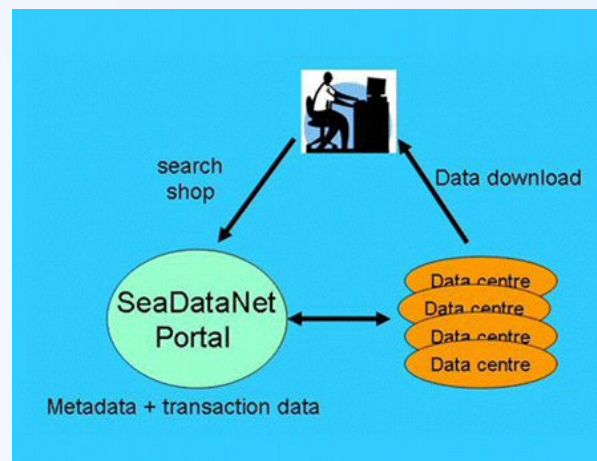
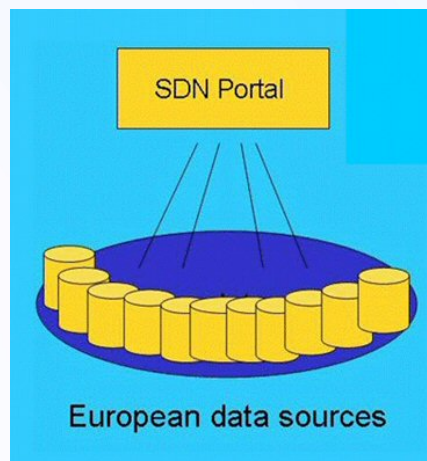
Waste Disposal

Wind Farms

Other Forms of Area Management / Designation

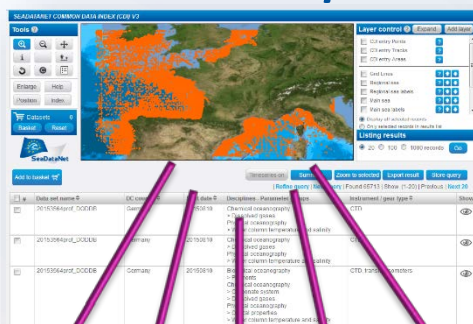
CDI Data Discovery and Access service

- One of the core services of the SeaDataNet infrastructure
- Providing a highly detailed insight and unified access to the large volumes of marine and oceanographic data sets managed by the distributed data centres
- Fine-grained index (ISO 19115 – ISO 19139) to individual data measurements (such as a CTD cast or moored instrument record)
- Supported by Controlled Vocabularies, and Directories (EDMO, EDMERP, CSR, EDMED)

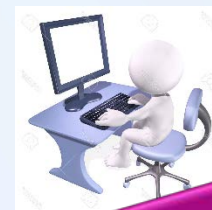


CDI service for discovery and unified data access

SeaDataNet portal

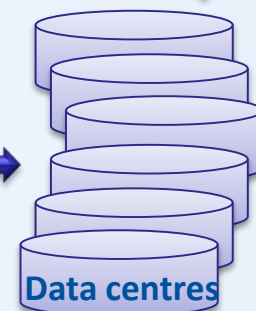
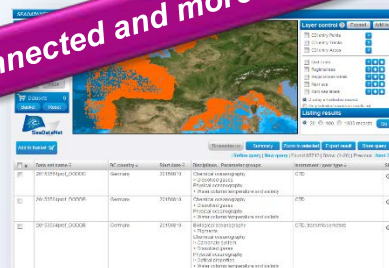


Search
and
Shop



Data
add

Already 110 data centres connected and more underway



Metadata

+ transaction data

European data sources
data centres > 650 originators

Current CDI user interfaces



SEADATANET COMMON DATA INDEX (CDI) V3

TOOLS

- ENLARGE
- HELP
- POSITION
- INDEX
- Datasets
- BASKET
- RESULT

LAYER CONTROL

- ☐ DI entry Points
- ☐ DI entry Tracks
- ☐ DI entry Areas
- ☐ Grid Lines
- ☐ Regional sea
- ☐ Regional sea
- ☐ LAT/LONG
- Upper-left
- Lower-right

SEARCH

Free search

Disciplines - Topics

Discovery parameters

Cruise/Station name

Projectname

Datasetname

Sea regions

Waterdepth (m) from

Originator

CDI partner

Country

Access restriction

Instrument type

Instrument depth (m) from

Platform type

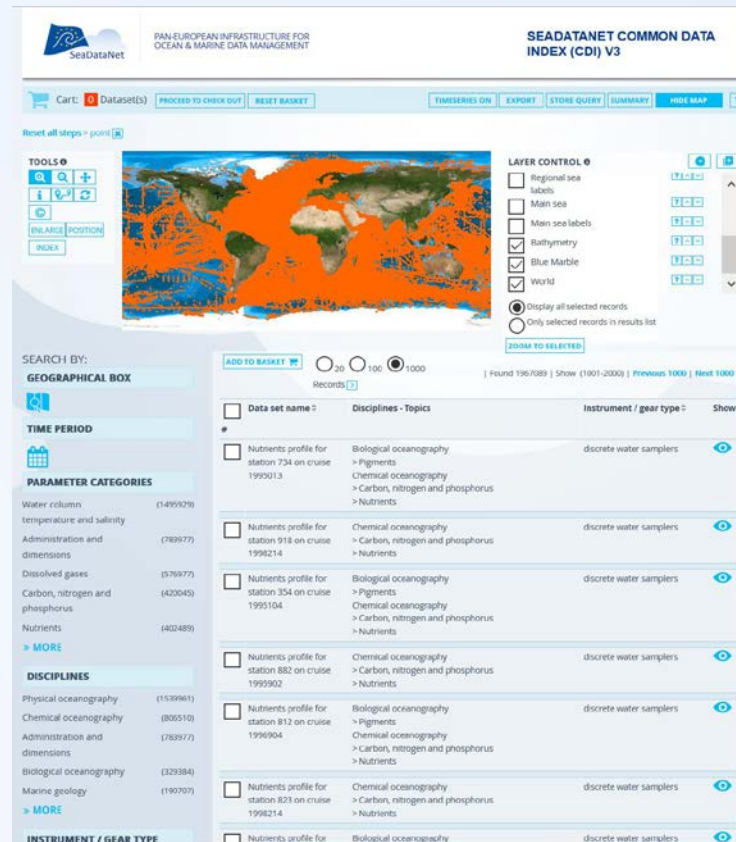
Measuring area type

Temporal resolution

Date (yyyymmdd) from

Duration

Extended Search



SEADATANET COMMON DATA INDEX (CDI) V3

Cart: Datasets

TOOLS

- ENLARGE
- POSITION
- INDEX

LAYER CONTROL

- ☐ Regional sea labels
- ☐ Main sea
- ☐ Main sea labels
- ☐ Bathymetry
- ☒ Blue Marble
- ☐ World
- ☒ Display all selected records
- ☐ Only selected records in results list

SEARCH BY:

GEOGRAPHICAL BOX

TIME PERIOD


PARAMETER CATEGORIES

DISCIPLINES


INSTRUMENT / GEAR TYPE

Data set name	Disciplines - Topics	Instrument / gear type	Show
Nutrients profile for station 734 on cruise 1999013	Biological oceanography > Pigments Chemical oceanography > Carbon, nitrogen and phosphorus > Nutrients	discrete water samplers	Show
Nutrients profile for station 918 on cruise 1999214	Chemical oceanography > Carbon, nitrogen and phosphorus > Nutrients	discrete water samplers	Show
Nutrients profile for station 354 on cruise 1999104	Biological oceanography > Pigments Chemical oceanography > Carbon, nitrogen and phosphorus > Nutrients	discrete water samplers	Show
Nutrients profile for station 882 on cruise 1999902	Chemical oceanography > Carbon, nitrogen and phosphorus > Nutrients	discrete water samplers	Show
Nutrients profile for station 812 on cruise 1999904	Biological oceanography > Pigments Chemical oceanography > Carbon, nitrogen and phosphorus > Nutrients	discrete water samplers	Show
Nutrients profile for station 823 on cruise 1999214	Chemical oceanography > Carbon, nitrogen and phosphorus > Nutrients	discrete water samplers	Show
Nutrients profile for station 823 on cruise 1999214	Biological oceanography	discrete water samplers	Show

Quick (facet) Search



EMODnet
European Marine
Observation and
Data Network

Free search 

[Home](#) [Approach](#) [Metadata & Data](#) [Data products](#) [News](#) [Promotion](#) [Partners](#) [Helpdesk](#)

Bathymetry

© HCMR

providing services for discovery and requesting access to bathymetric data

SURVEY DATA SETS

CDI Data Discovery and Access service


[READ MORE](#)



EMODnet DTM

Bathymetry Viewing and Download service


[READ MORE](#)



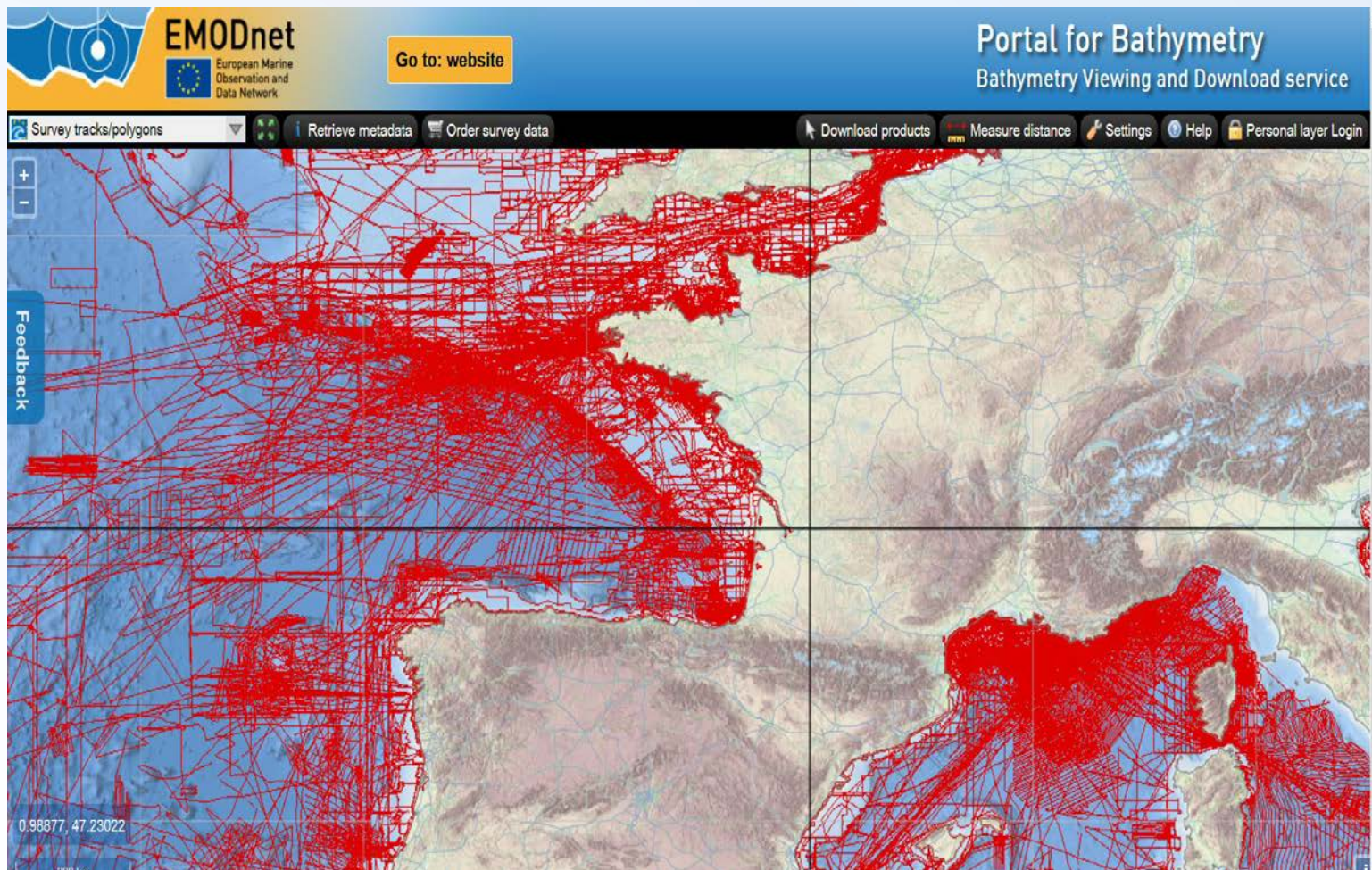
COMPOSITE DTM

Sextant Catalogue service

[READ MORE](#)

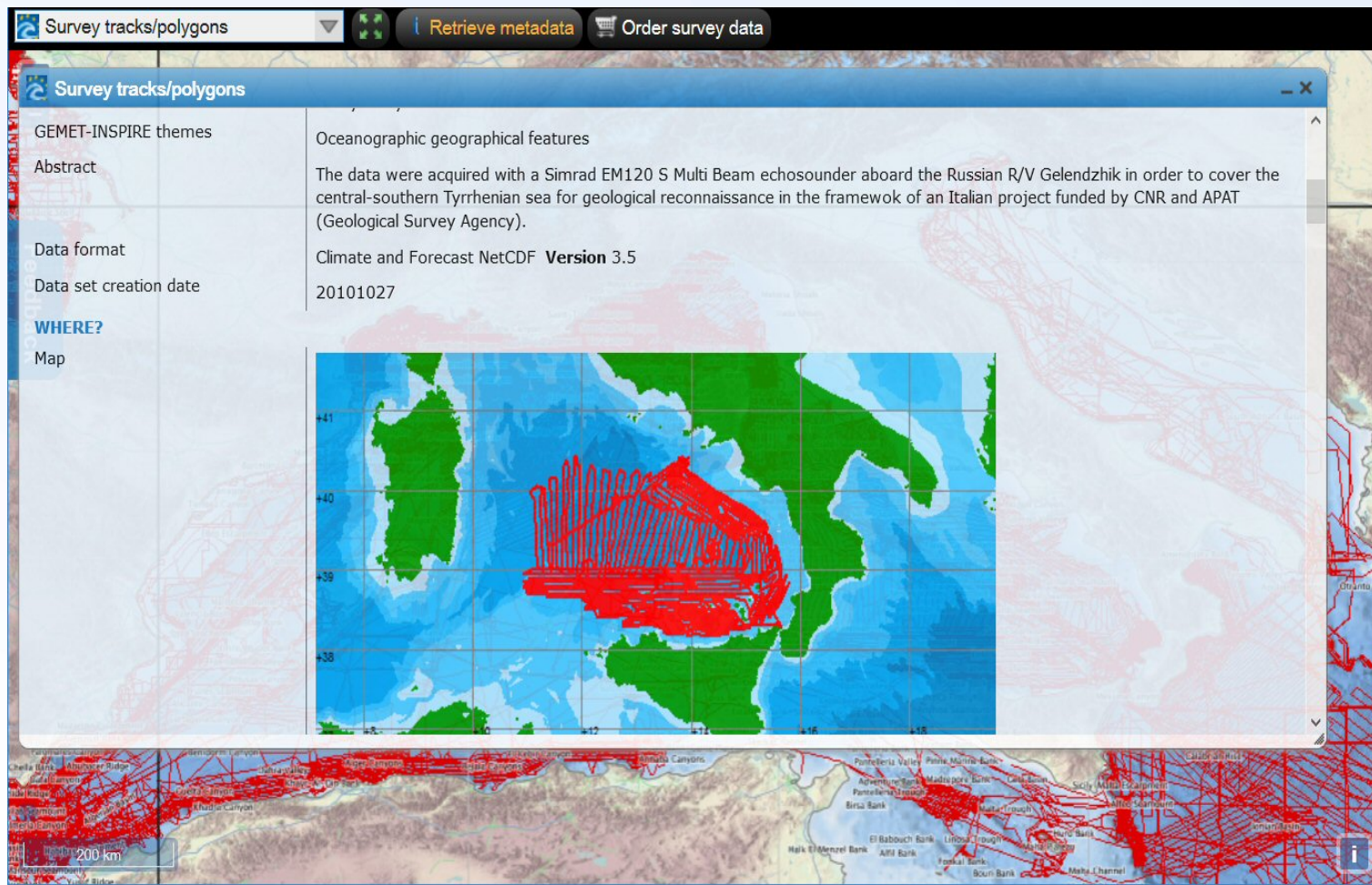


CDI service in EMODnet Bathymetry



Layer with CDI data references

CDI service in EMODnet Bathymetry



Layer with CDI data references

CDI access at EMODnet Physics



PHYSICS

Oceans Physics at your fingertips

[CONTACT US](#)
[SUBMIT DATA](#)
[HOME](#)
[MAP VIEWER](#)
[CATALOGUE](#)
[ABOUT](#)
[HELPDESK](#)
[SUBMIT DATA](#)
[CENTRAL PORTAL](#)
[WAVES](#)
[WATER TEMPERATURE](#)
[WATER SALINITY](#)
[CURRENTS](#)
[OPTICAL PROPERTIES](#)
[SEA LEVEL](#)
[ATMOSPHERIC](#)
[WATER CONDUCTIVITY](#)
[WINDS](#)
[RIVER](#)
[UNDER WATER NOISE](#)

in collaboration with



DATA
INGESTION

PRODUCTS

THREDDS

ERDDAP

GEOSERVER

API REST
SOAP

WMS
WFS

DASHBOARD

GITHUB

VIDEOS



Pillars under EMODnet Physics



The European Global Ocean Observing System, association and its regional components (ROOSs)

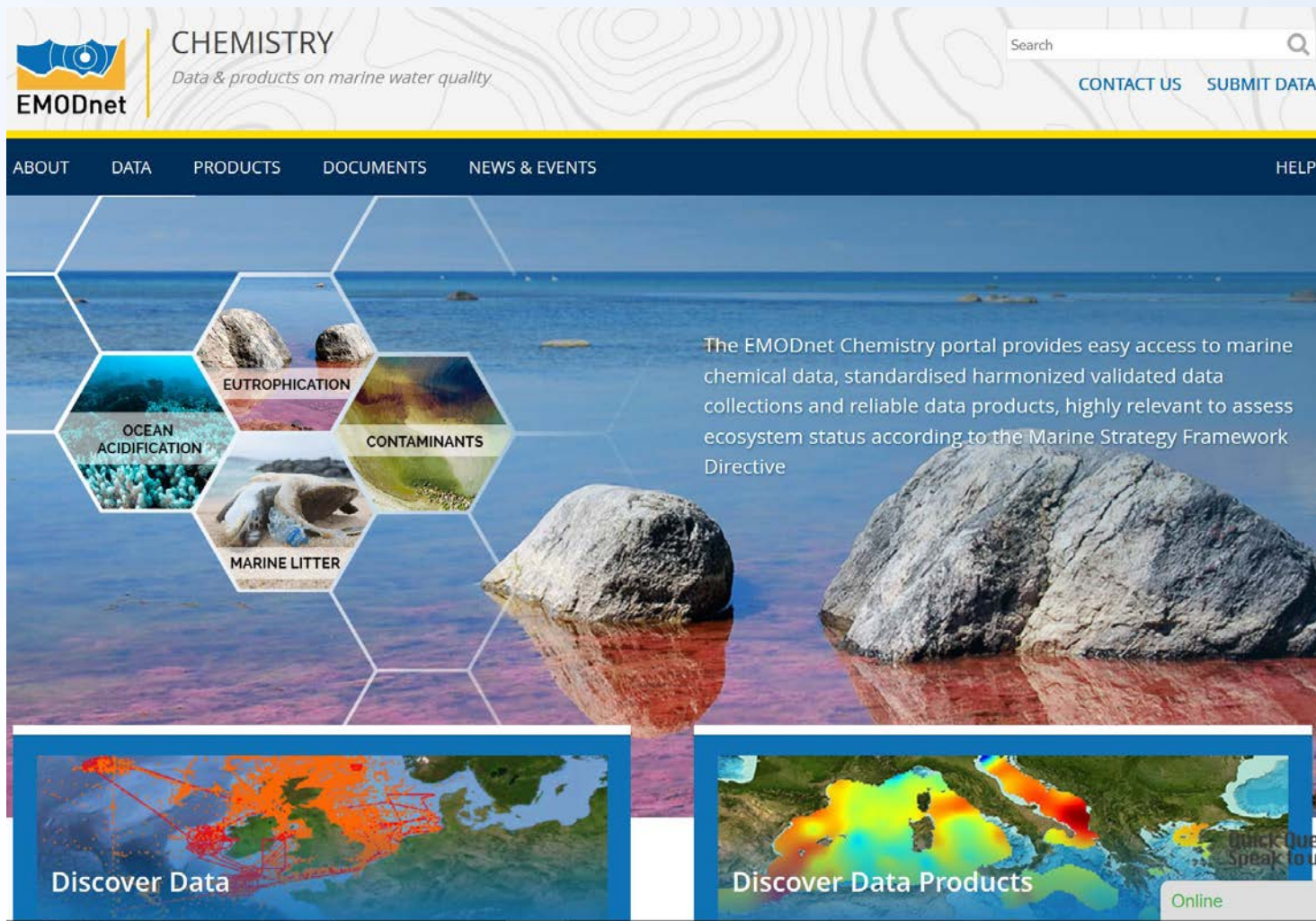


Copernicus Marine Environment Monitoring System (CMEMS)



SeaDataNet, pan-European marine data management infrastructure and network of NODCs

CDI service in EMODnet Chemistry



The screenshot displays the EMODnet Chemistry portal. At the top left is the EMODnet logo, and next to it is the text "CHEMISTRY" with the tagline "Data & products on marine water quality". To the right is a search bar and links for "CONTACT US" and "SUBMIT DATA". Below this is a navigation bar with links for "ABOUT", "DATA", "PRODUCTS", "DOCUMENTS", "NEWS & EVENTS", and "HELP". The main content area features a large background image of a rocky coastline. Overlaid on this is a hexagonal grid containing five smaller images: "OCEAN ACIDIFICATION", "EUTROPHICATION", "CONTAMINANTS", "MARINE LITTER", and "OCEAN ACIDIFICATION". To the right of this grid is a text block stating: "The EMODnet Chemistry portal provides easy access to marine chemical data, standardised harmonized validated data collections and reliable data products, highly relevant to assess ecosystem status according to the Marine Strategy Framework Directive". At the bottom, there are two panels: "Discover Data" on the left, showing a map of Europe with red and orange data points, and "Discover Data Products" on the right, showing a map of Europe with a color-coded data distribution. A "Quick Query" button is visible in the bottom right corner of the "Discover Data Products" panel.

EMODnet

CHEMISTRY
Data & products on marine water quality

Search

CONTACT US SUBMIT DATA

ABOUT DATA PRODUCTS DOCUMENTS NEWS & EVENTS HELP

OCEAN ACIDIFICATION

EUTROPHICATION

CONTAMINANTS

MARINE LITTER

The EMODnet Chemistry portal provides easy access to marine chemical data, standardised harmonized validated data collections and reliable data products, highly relevant to assess ecosystem status according to the Marine Strategy Framework Directive

Discover Data

Discover Data Products

Quick Query
Speak to us

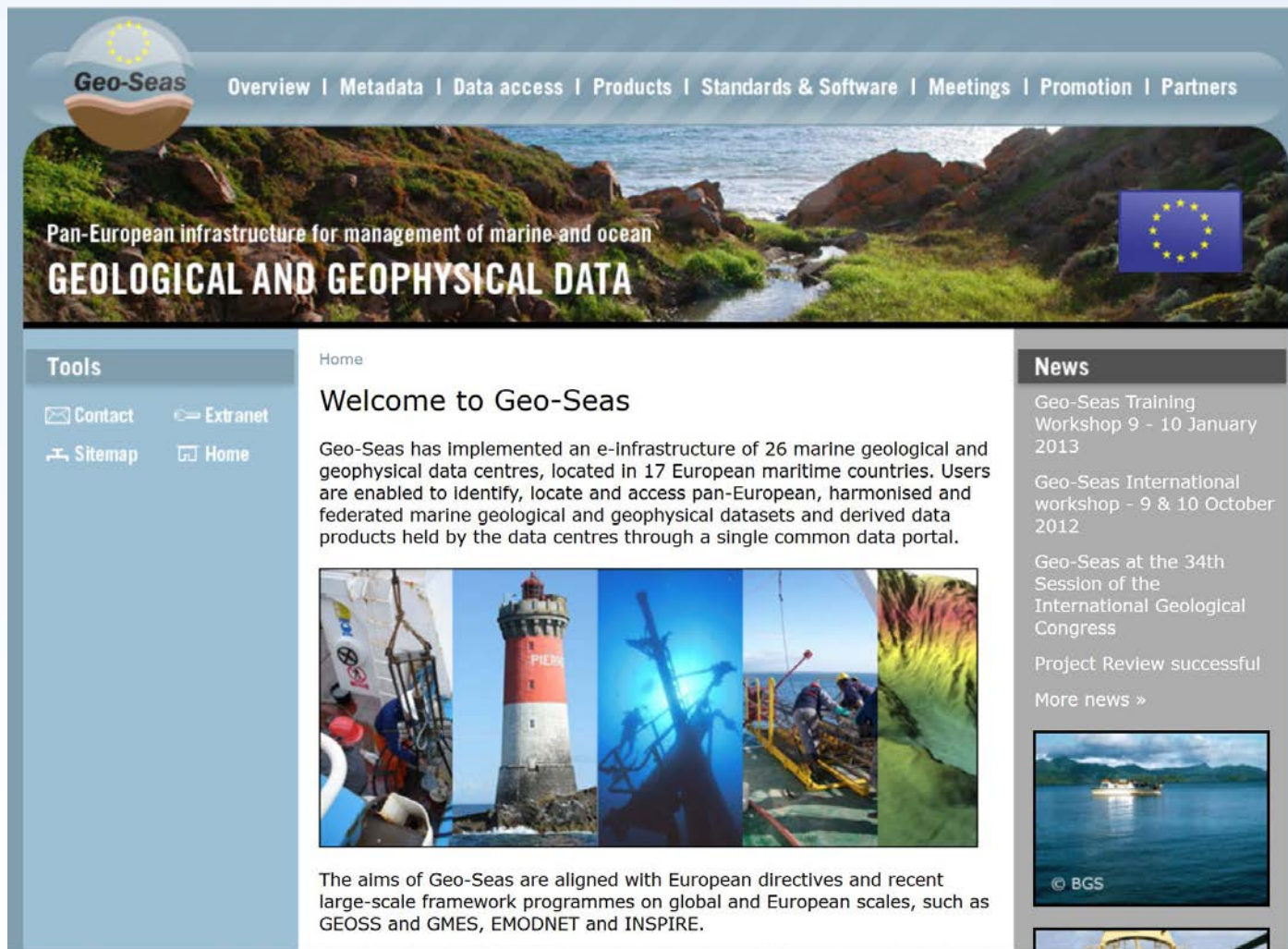
Online

CDI service in EMODnet Chemistry

MSFD regions

Group of Variables	Baltic Sea	Iberian peninsula - Macaronesia - Bay of Biscay	Greater North Sea - Celtic Sea - Faroes	Arctic Ocean - Norwegian Sea - Greenland Sea - Barents Sea - Icelandic Waters	Mediterranean Sea	Black Sea - Sea of Azov - Marmara Sea
Acidity ⁱ	■	■	■	■	■	■
Antifoulants ⁱ	■	■	■		■	■
Chlorophyll ⁱ	■	■	■	■	■	■
Dissolved gasses ⁱ	■	■	■	■	■	■
Fertilisers ⁱ	■	■	■	■	■	■
Hydrocarbons ⁱ	■	■	■	■	■	■
Heavy metals ⁱ	■	■	■	■	■	■
Organic matter ⁱ	■	■	■	■	■	■
Polychlorinated biphenyls ⁱ	■	■	■		■	■
Pesticides and biocides ⁱ	■	■	■	■	■	■
Radionuclides ⁱ	■		■	■	■	■
Silicates ⁱ	■	■	■	■	■	■
<div> <div>■ 1-50</div> <div>■ 251-1000</div> <div>■ 2501-5000</div> <div>■ 10001-25000</div> <div>■ 51-250</div> <div>■ 1001-2500</div> <div>■ 5001-10000</div> <div>■ >25000</div> </div>						

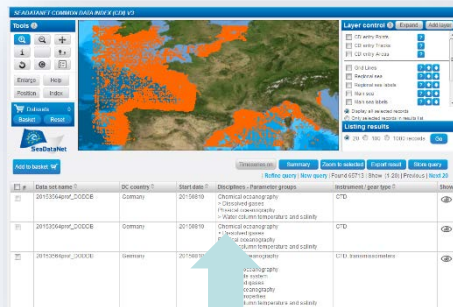
Legend - number of measurement data sets for each variable per marine region



The screenshot shows the Geo-Seas website interface. At the top, there is a navigation bar with the following links: Overview | Metadata | Data access | Products | Standards & Software | Meetings | Promotion | Partners. Below this is a large banner image of a rocky coastline with the text "Pan-European infrastructure for management of marine and ocean GEOLOGICAL AND GEOPHYSICAL DATA" and the European Union flag. On the left side, there is a "Tools" section with links to Contact, Extranet, Sitemap, and Home. The main content area is titled "Home" and "Welcome to Geo-Seas". It contains a paragraph describing the infrastructure: "Geo-Seas has implemented an e-Infrastructure of 26 marine geological and geophysical data centres, located in 17 European maritime countries. Users are enabled to identify, locate and access pan-European, harmonised and federated marine geological and geophysical datasets and derived data products held by the data centres through a single common data portal." Below this text is a row of five small images: a ship's deck, a lighthouse, a blue underwater scene, a person on a boat, and a colorful geological map. At the bottom of the main content area, there is a paragraph: "The aims of Geo-Seas are aligned with European directives and recent large-scale framework programmes on global and European scales, such as GEOSS and GMES, EMODNET and INSPIRE." On the right side, there is a "News" section with three items: "Geo-Seas Training Workshop 9 - 10 January 2013", "Geo-Seas International workshop - 9 & 10 October 2012", and "Geo-Seas at the 34th Session of the International Geological Congress". Below these items is a link "Project Review successful" and a "More news »" link. At the bottom right, there is a small image of a boat on the water with the text "© BGS" below it.

Geological and geophysical data sets

CDI service as driver



Data discovery
and access



> 110 data centres



NODCs; HO; GEOs; BIOs; ICES; PANGAEA

≈ 650 European data
originators



GEOSS portal



IODE ODP portal

Aggregated collection



Black Sea portal



Caspian portal

Regional subsets



Geo-Seas portal

Thematic portals



Bathymetry

Physics

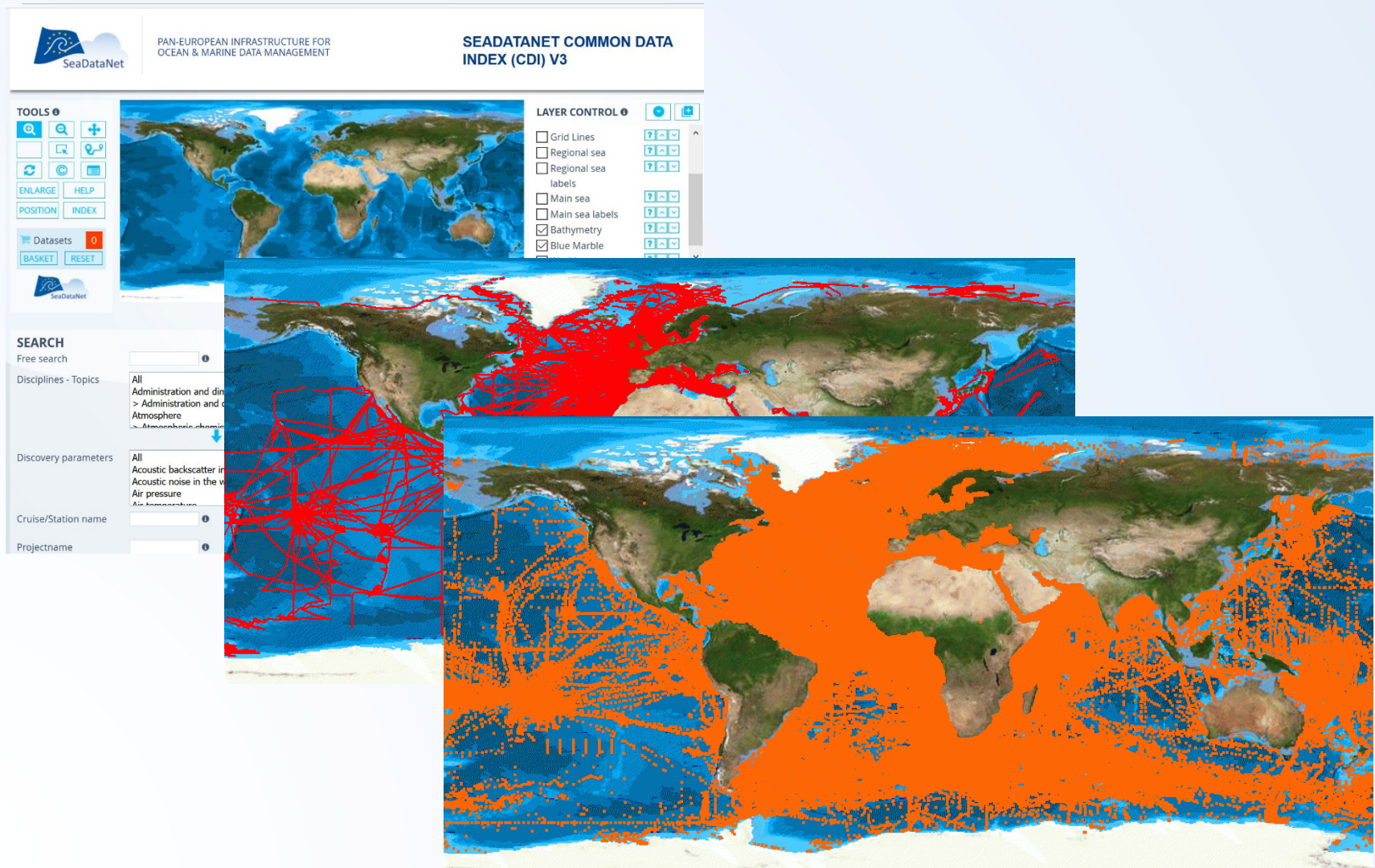
Chemistry

Geology

Biology

CDI Data Discovery and
Access service

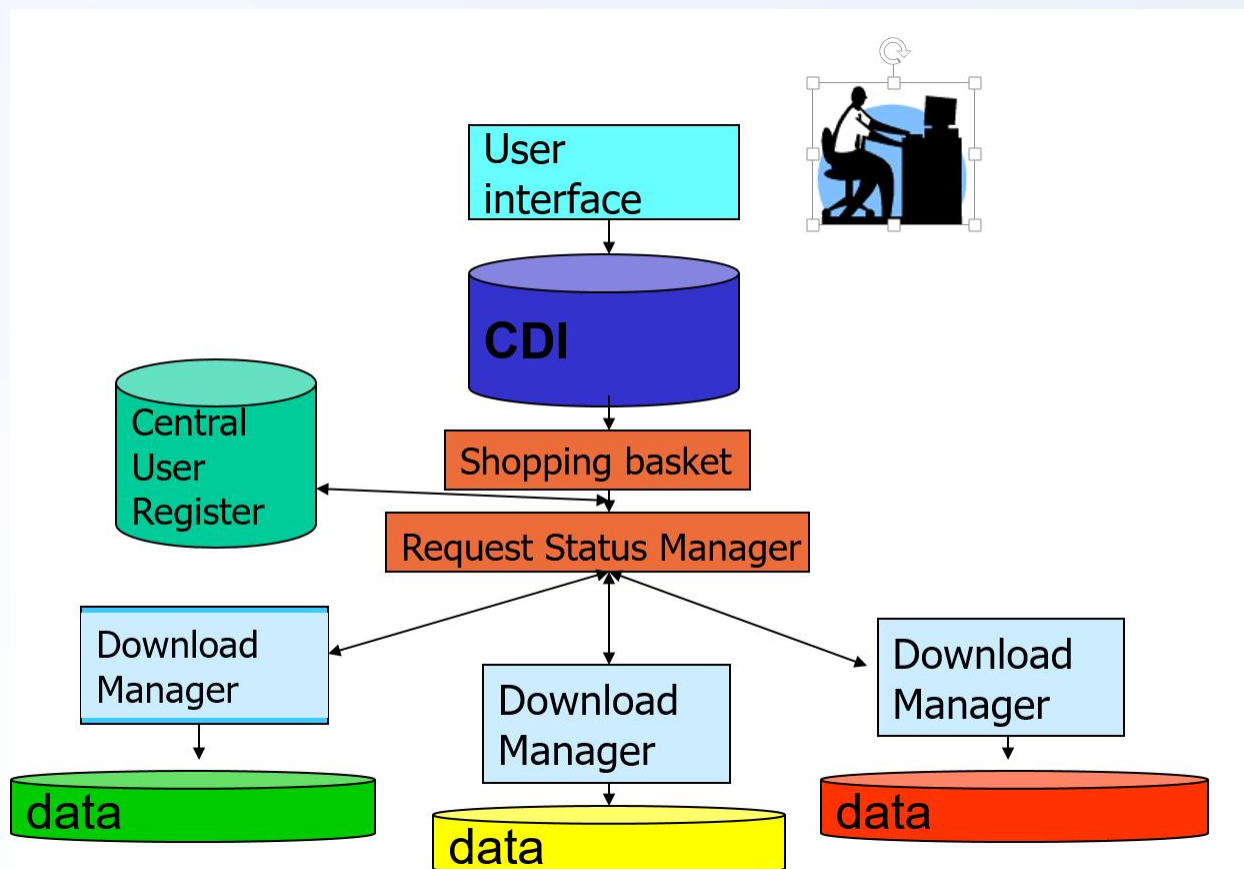
CDI service with global coverage



The screenshot displays the SEADATANET COMMON DATA INDEX (CDI) V3 web interface. The interface includes a header with the SeaDataNet logo and the text "PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT". Below the header, there is a "TOOLS" section with buttons for "ENLARGE", "HELP", "POSITION", "INDEX", "BASKET", and "RESET". A "SEARCH" section is also present, featuring a "Free search" input field, a "Disciplines - Topics" dropdown menu, and "Discovery parameters" including "Acoustic backscatter in", "Acoustic noise in the v", "Air pressure", and "Air temperature". The main content area shows a global map with data coverage. The "LAYER CONTROL" section on the right lists various data layers: "Grid Lines", "Regional sea", "Regional sea labels", "Main sea", "Main sea labels", "Bathymetry", and "Blue Marble". The map displays a dense network of red lines representing data tracks and a large area of orange dots representing data points, indicating global coverage.

> 2.1 Million CDI entries for physics, chemistry, biology, geology and geophysics

Current CDI service architecture



Issues with current CDI service

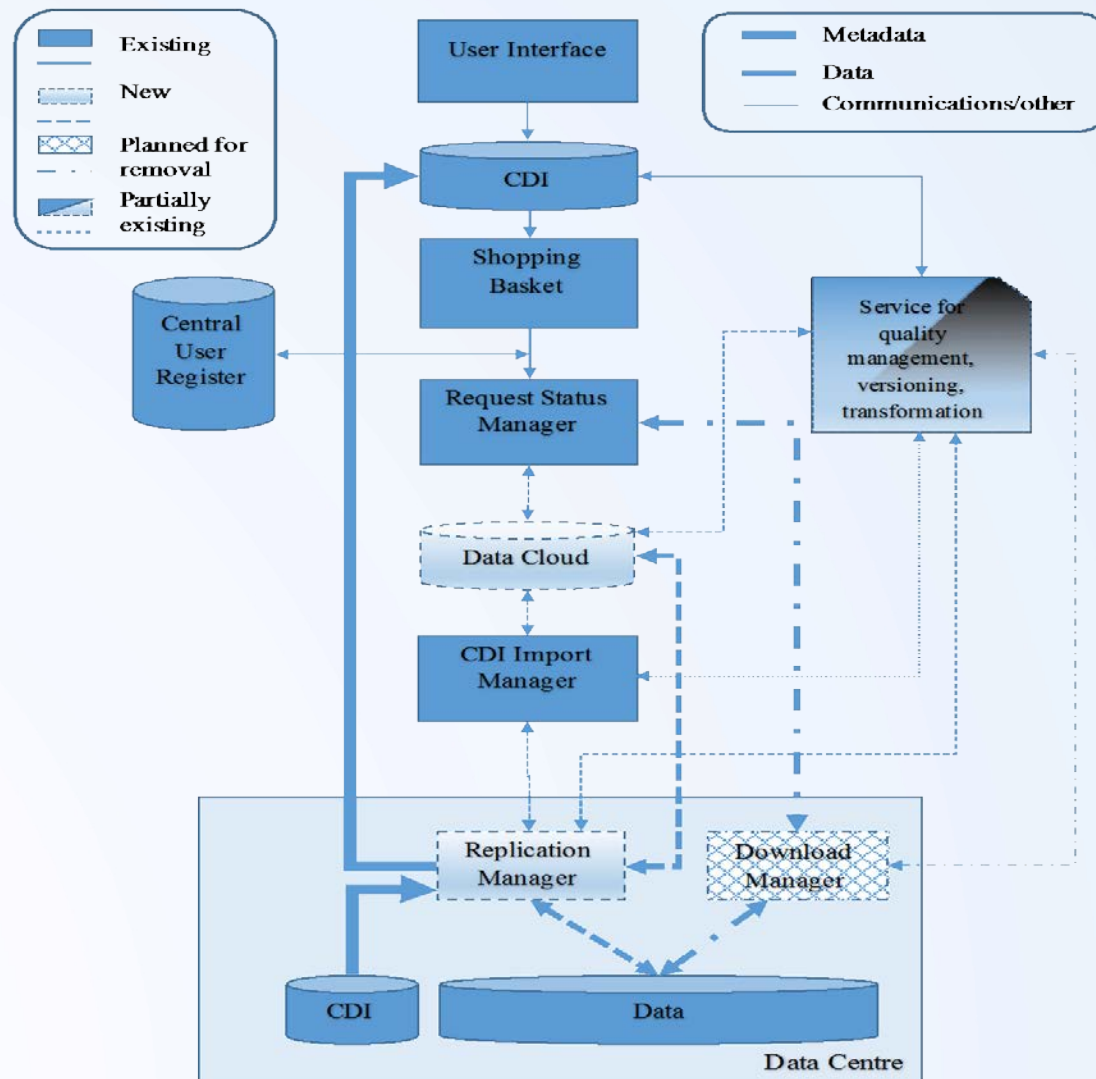
- **performance for users:** CDI data access service interacts with the distributed data collections and databases at the connected data centres.
 - user can submit a shopping basket with requests for data from multiple data centres.
 - user must await the automatic data preparation by each of these data centres
 - user must download resulting data sets through the RSM as packages directly from each data centre, which implicates multiple download transactions
- **performance for users:** data centres are not always online, operational and have different machine capacities which might give extra delays
- **quality issues:** concerning formats of data files (ODV + NetCDF) and their consistency with CDI metadata.
- **installation and configuration** of the Download Manager software can be challenging due to different configurations, firewalls etc., which in practice results in having different versions installed



Principles for upgrading the CDI service using the cloud

- To configure and maintain a **CLOUD** environment with High Performance Computing (HPC) facilities 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 amendments of their submissions and/or local configurations for mapping data and metadata.
- Include transformation services for converting data sets to SeaDataNet ODV and NetCDF formats and relevant INSPIRE data models.
- Introduce versioning of metadata and data as part of provenance

New CDI service architecture



Potential benefits for users

- The performance will be speeded up, discovery and data requests improved, and downloading made more easy as each shopping request will provide one integrated download package instead of multiple packages from multiple data centres.
- Overall quality and coherence (data – metadata) will improve
- Tracking and tracing of data transactions will continue to be administered by an upgraded and much faster RSM service to oversee shopping requests and deliveries. The user RSM will be integrated as MySeaDataCloud service in the CDI user interface.
- Versioning of metadata and data will facilitate repeated analysis of e.g. environmental assessments in MSFD context after many years, and for scientific papers.

Potential benefits for data centres

- Data centres will have a **Replication Manager** module and an **Import Dashboard** to trigger and control themselves the import of new and updated metadata and data sets (unrestricted) into the CDI service
- Data providers can oversee all relevant transactions for their data centre in the upgraded and much faster RSM system and generate relevant reports
- The system will also support handling restricted data sets
- Data centres will be outfitted with a Replication Manager (RM) replacing the Download Manager. The RM has less complexity and is easier to configure.
- Alternatively, Data centres can make use of the ‘interim solution’ which will be provided with improved functionality, handling both unrestricted and restricted data sets

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

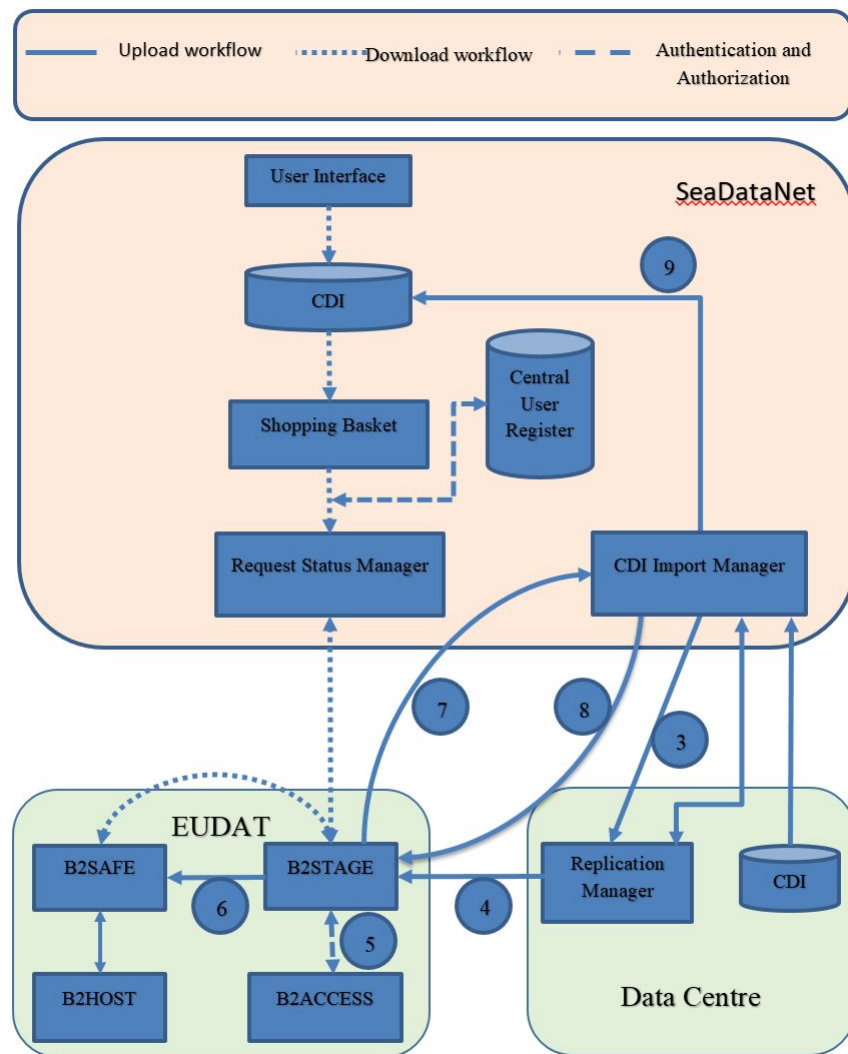
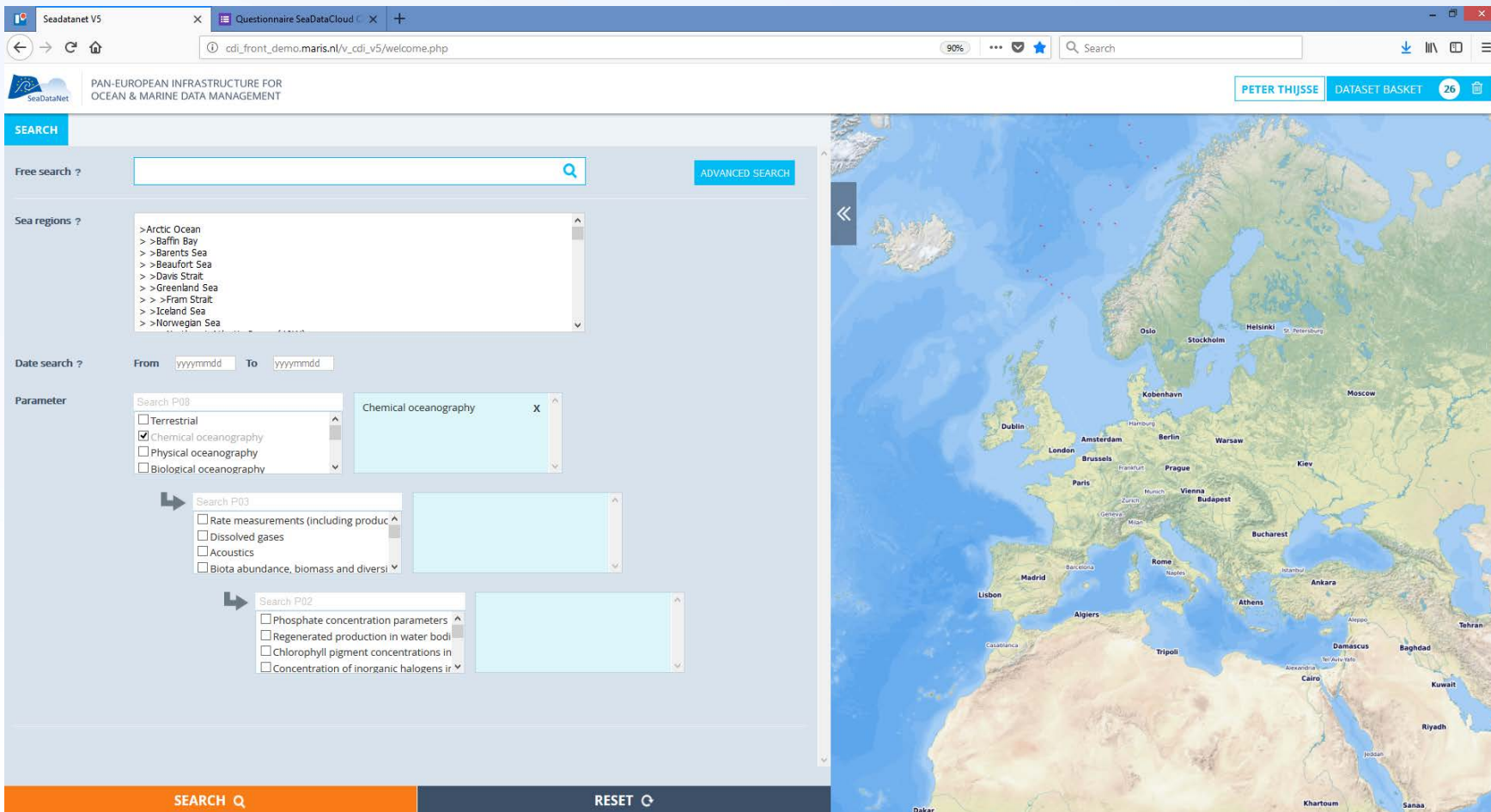


Figure: High level architecture

New CDI interface



The screenshot displays the SeaDataCloud CDI interface. The top navigation bar includes the SeaDataCloud logo, the text "PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT", and a user profile section for "PETER THIJSSE" with a "DATASET BASKET" containing 26 items. The main interface is divided into a search panel on the left and a map on the right.

Search Panel:

- Free search ?**: A text input field with a search icon.
- Sea regions ?**: A dropdown menu showing a list of regions: > Arctic Ocean, > Baffin Bay, > Barents Sea, > Beaufort Sea, > Davis Strait, > Greenland Sea, > Fram Strait, > Iceland Sea, > Norwegian Sea.
- Date search ?**: Fields for "From" and "To" with the format "yyyymmdd".
- Parameter**:
 - Search P00**: A dropdown menu with options: ☐ Terrestrial, ☒ Chemical oceanography, ☐ Physical oceanography, ☐ Biological oceanography. A tag "Chemical oceanography" is visible.
 - Search P03**: A dropdown menu with options: ☐ Rate measurements (including produc), ☐ Dissolved gases, ☐ Acoustics, ☐ Biota abundance, biomass and divers. A tag is visible.
 - Search P02**: A dropdown menu with options: ☐ Phosphate concentration parameters, ☐ Regenerated production in water bodi, ☐ Chlorophyll pigment concentrations in, ☐ Concentration of inorganic halogens ir. A tag is visible.

At the bottom of the search panel are buttons for "SEARCH Q" and "RESET".

Map: A map of Europe and surrounding regions, showing various cities and geographical features. The map is interactive, with a zoom level of 90% indicated in the top right corner.