Royal Netherlands Institute for Sea Research

Data Management within a specific domain:

an oceanographic use case in the frame of the SeaDataNet infrastructure

Taco de Bruin

NIOZ

NIOZ is an institute of the Netherlands Organisation for Scientific Research (NWO) in cooperation with Utrecht University

“A scholar’s positive contribution is measured by the sum of the original data that he contributes.

Hypotheses come and go, but data remain.”

Santiago Ramón y Cajal
(Spanish Nobel Prize winner, 1906)

What is data management?

All activities concerning data to maximize the value of data by preserving data for future (re-)use.

The primary objective of data management is to ensure timely, efficient and open access to the best possible data, metadata and associated products, for use and re-use throughout their life-cycle and to prevent loss of data and associated information

POGO – Partnership for Observing the Global Ocean

'*End-to-end data management*'

Royal Netherlands Institute for Sea Research

The National Oceanographic Institute of The Netherlands




NIOZ is an institute of the Netherlands Organisation for Scientific Research

in cooperation with Utrecht University

NIOZ mission



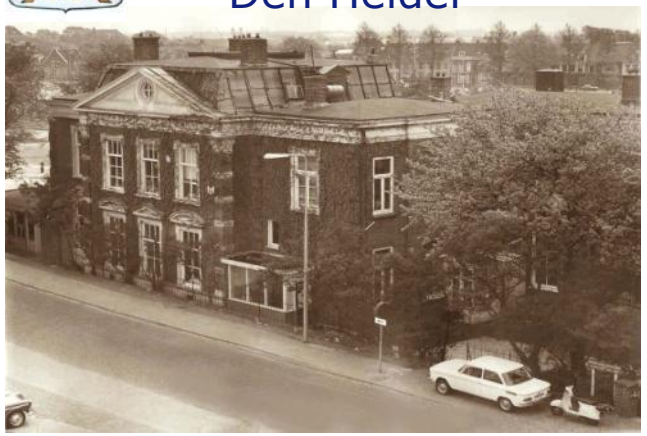
1. To gain and to spread fundamental and frontier-applied scientific knowledge on estuaries, deltas, continental seas and open oceans for the understanding of our blue planet and its sustainable management
2. To maintain marine research facilities for the national scientific community
3. To support marine education programmes in the NL and in Europe




NIOZ history in a nutshell

1876 Zoological Station - NDV

1890 – 1970
Den Helder



1970 Island of Texel

2001

2012 Yerseke - Zeeland

Current Research Vessels

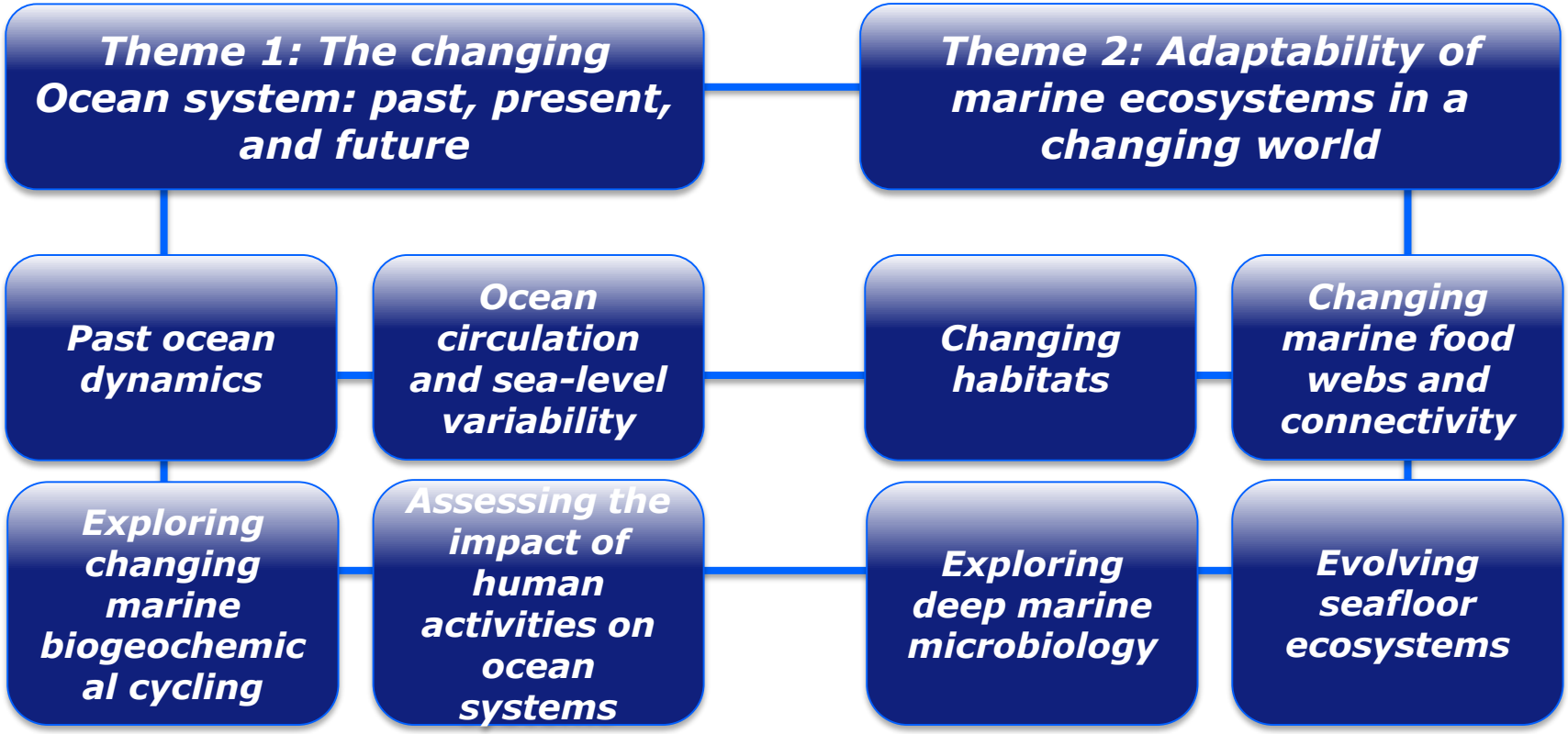
Pelagia (1991)– oceans & open North Sea

Navicula (1981)– Coastal areas

Stern

Vlet ’t Horntje (1961)

Science plan 2016-2020:
2 themes and 8 key topics

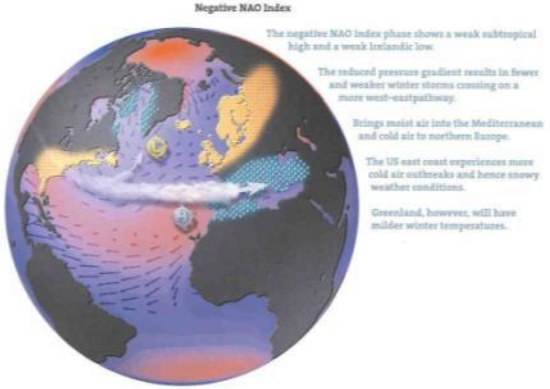


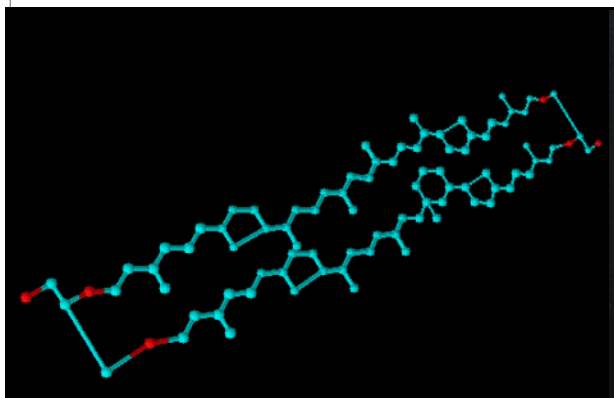
NIOZ science plan: interlinked themes and topics 2014-2020

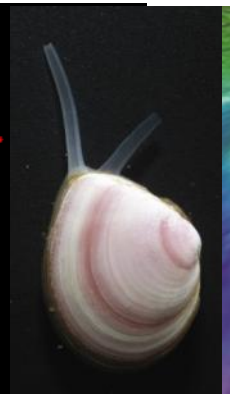
Multidisciplinary oceanographic research

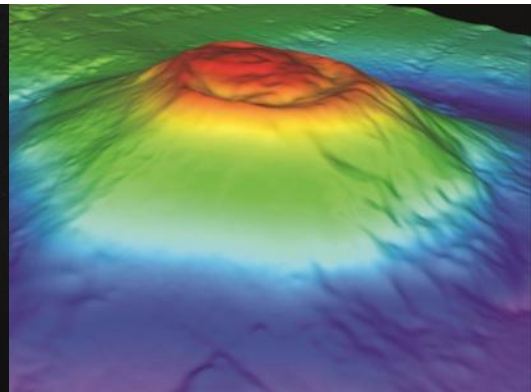
- Physical Oceanography
- Chemical Oceanography
- Marine (Micro)Biology /Ecology
- Marine Geology
- Biogeochemistry

and their interactions in an integrated, multidisciplinary way



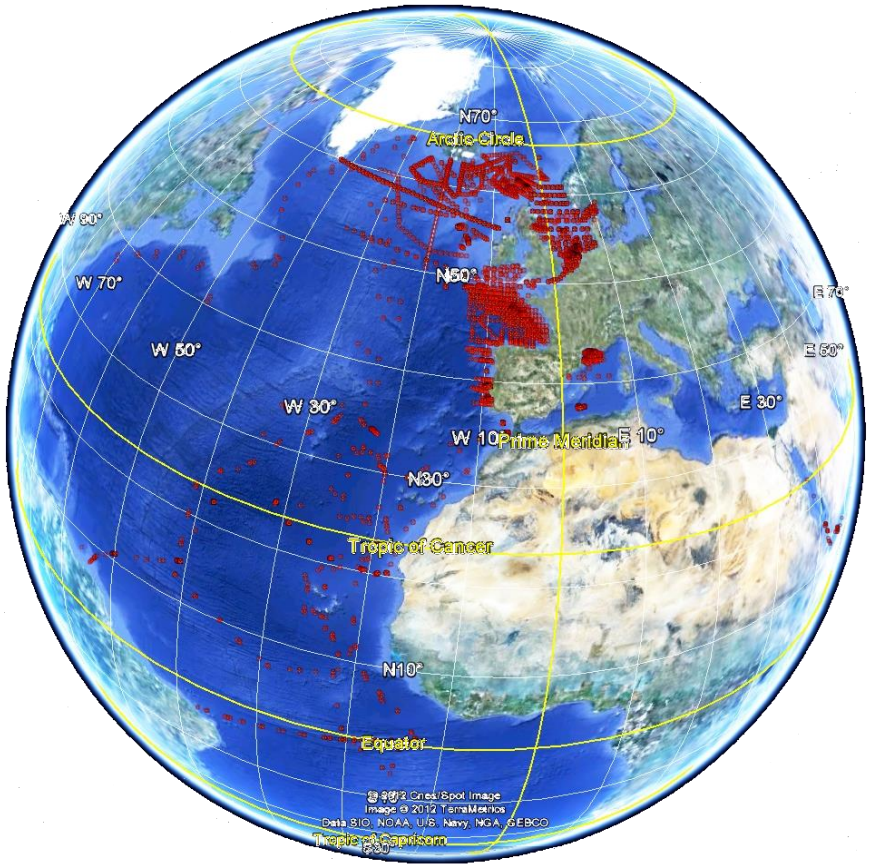


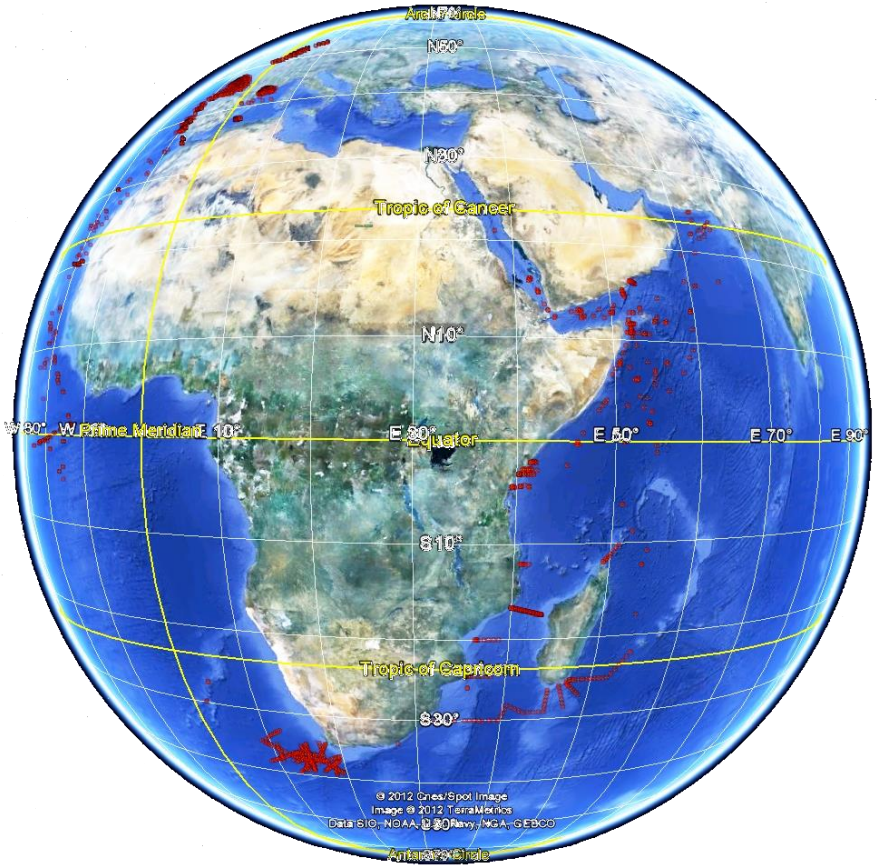




Data from multidisciplinary research

- Tens of different measuring methods
- Hundreds of parameters
- Around the globe





Data from multidisciplinary research

Two ‘different’ data streams:

- Directly from the ship (in digital format)
- After analysis in the lab

Many data types:

- Digital formats
- Photos/pictures
- Videos
- Water-, sediment-, tissue-samples

Date types are not:

- privacy sensitive
- interviews, oral history
- literary texts

Value of data

- Intrinsic value for science: Without data no science
- Economic value – costs of data collection
    - Project planning
    - Data collection phase
    - Data processing
- Uniqueness of observations
    - Time-varying processes
        - Can’t measure a 2004-temperature in 2017
    - Logistical reasons – cruise or field campaign planning

Value of data

- Detecting trends: Data as reference or base-line values
  - Global change
  - Ozone hole
- Interdisciplinary use, both within and outside science

  Met Office (KNMI):
  - Weather forecasts
  - Research
  - Insurance companies
- Policy makers: Effect of eutrophication measures

Because of this great value of data

- Data should be safely archived (in a data centre)
- Data should be available and accessible for (future) users, also across disciplines
  - Legal obligations, e.g. the Antarctic Treaty, Article III-1c:

    Scientific observations and results from Antarctica shall be exchanged and made freely available
    (1959 !!)
  - Requirements by funding agencies (NWO, EU)
  - Programme and project agreements

Data Management Guiding Principles

- Starting point always is the user: scientist

  Make life as easy as possible for the scientists

- Data should be (safely) archived in a data centre

  Backup procedures, durable formats, media transfer, complete metadata, standards

- Data centre close to where expertise on data types is

  Highest possible quality of data, better service

Data Management Guiding Principles

- Store data in one data centre,
  to prevent the occurrence of near-identical datasets
- Make the existence of the data known as widely as possible, by publishing metadata in multiple catalogs
- Build distributed archive systems, linked via central catalogs
- “Do not re-invent the wheel”

NIOZ Data archiving (raw data)

- Data should always be safe against all possible hazards
- Full control over the data
- Physically separated Netapp archive system(s)
- Not online
- Original media (USB-HD) are stored
- Storage in fire proof rooms and -safes
- For NIOZ: Storage above sea level

NIOZ Data archiving (raw data)

- Documentation (metadata)
- For cruises:
  - Semi-automated logging system to keep track of all measuring activities on board
- For lab work:
  - Two (legacy) Laboratory Information Management Systems (LIMS)
  - May be replaced by in-house developed Data Archiving System (DAS)
  - Lab notebooks are archived
- Future
  - Sensor Web Enablement (SWE)
  - Internet of things

Data processing

- Researchers
- Data (or instrument) specialists
  - instruments used during most cruises for ‘background’ data

Final or definitive data

- Archived
- Used in publications
- Made available, following NIOZ Data Policy
- In on-line RDBMS
- Since early 1990s

- >= 20 years ago, a need existed for FAIR data
- Has been addressed by the oceanographic community

Connecting Hundreds of Oceanographic Data Sources from Tens of Countries in and around Europe into One Big Data Network

Implementation of FAIR ‘avant la lettre’

Taco de Bruin – NIOZ Royal Netherlands Institute for Sea Research

Dick Schaap – MARIS Marine Information Services

Problem:

Some 20 years ago, potentially enormous amount of valuable, oceanographic data in Europe

- Very scattered
- Many sources
    - · Government agencies, research institutes, universities, NGOs
    - · Online and offline
- Many different formats (and systems)
- Variable data quality (depending on purposes for which the data were collected)

- Not Findable, no overview
- No (easy) Access
- Not Interoperable, conversion of data formats
- Estimated 50-80% of time ‘wasted’ on finding and converting data

Solution: **SeaDataNet**

an operational pan-European infrastructure for managing marine and ocean data

SeaDataNet connects National Oceanographic Data Centres (NODCs) and oceanographic data focal points from 35 countries bordering European seas

SeaDataCloud (2016-2020) is the culmination of a concerted approach, with a series of European oceanographic data management projects, starting in the 1990s

SeaDataNet in 1 slide (or perhaps 2)

F

Overview

A

One point of access, CDI as Persistent Identifier

I

Interoperability: use of community governed standards and vocabularies

R

Distributed, making use of existing systems and connecting local and national networks

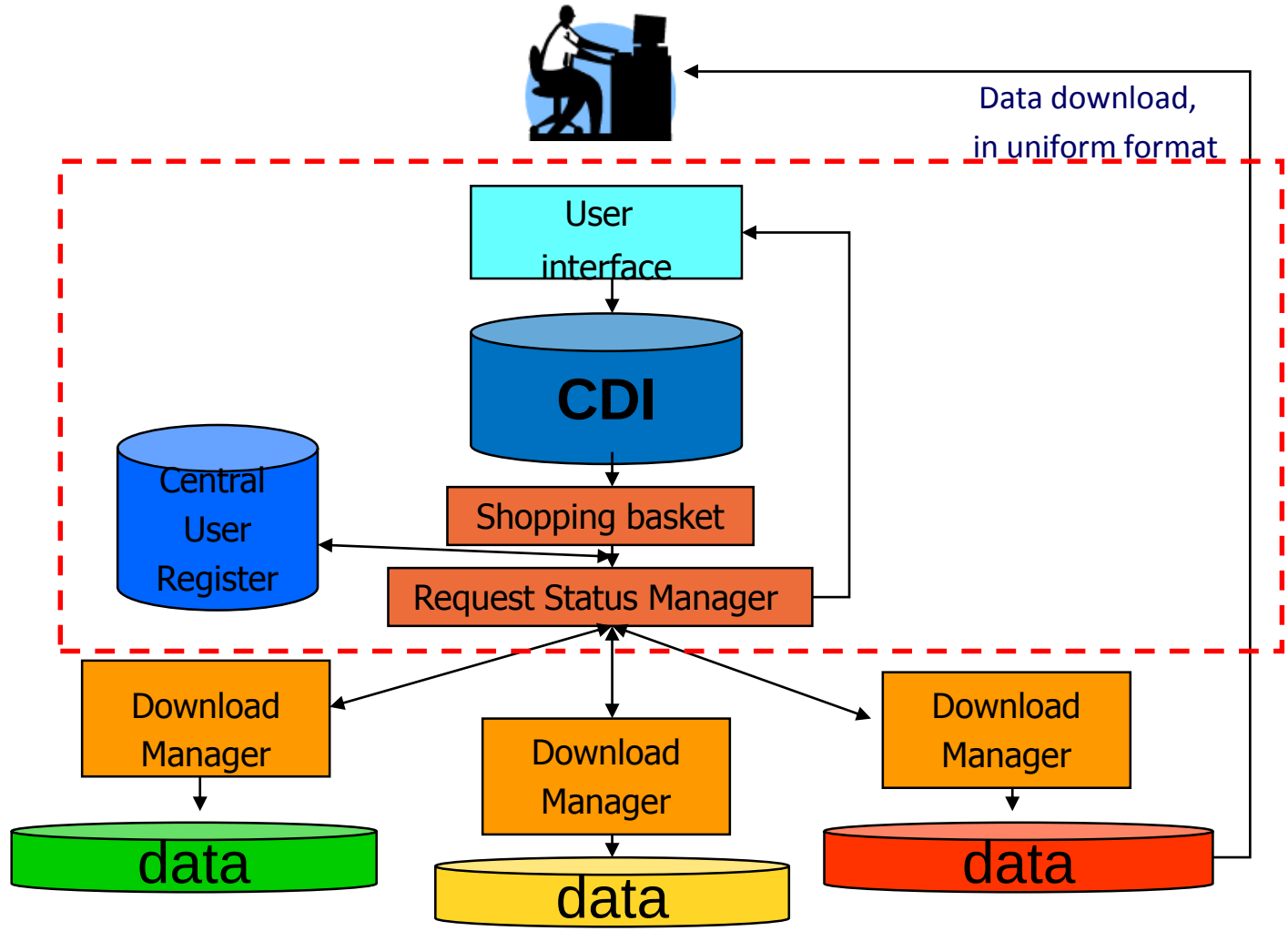
SeaDataNet in 2 slides

F: Overview

A: One point of access

I: Interoperability

R: Distributed networks

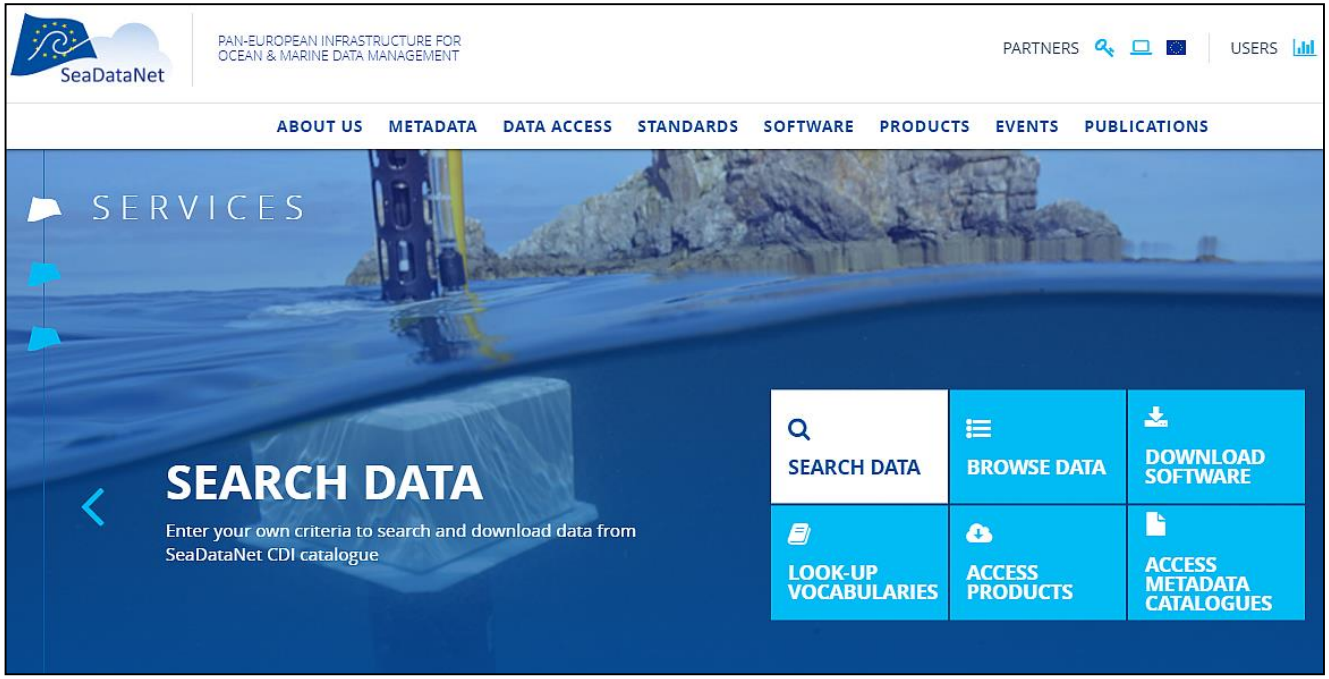


Achievements:

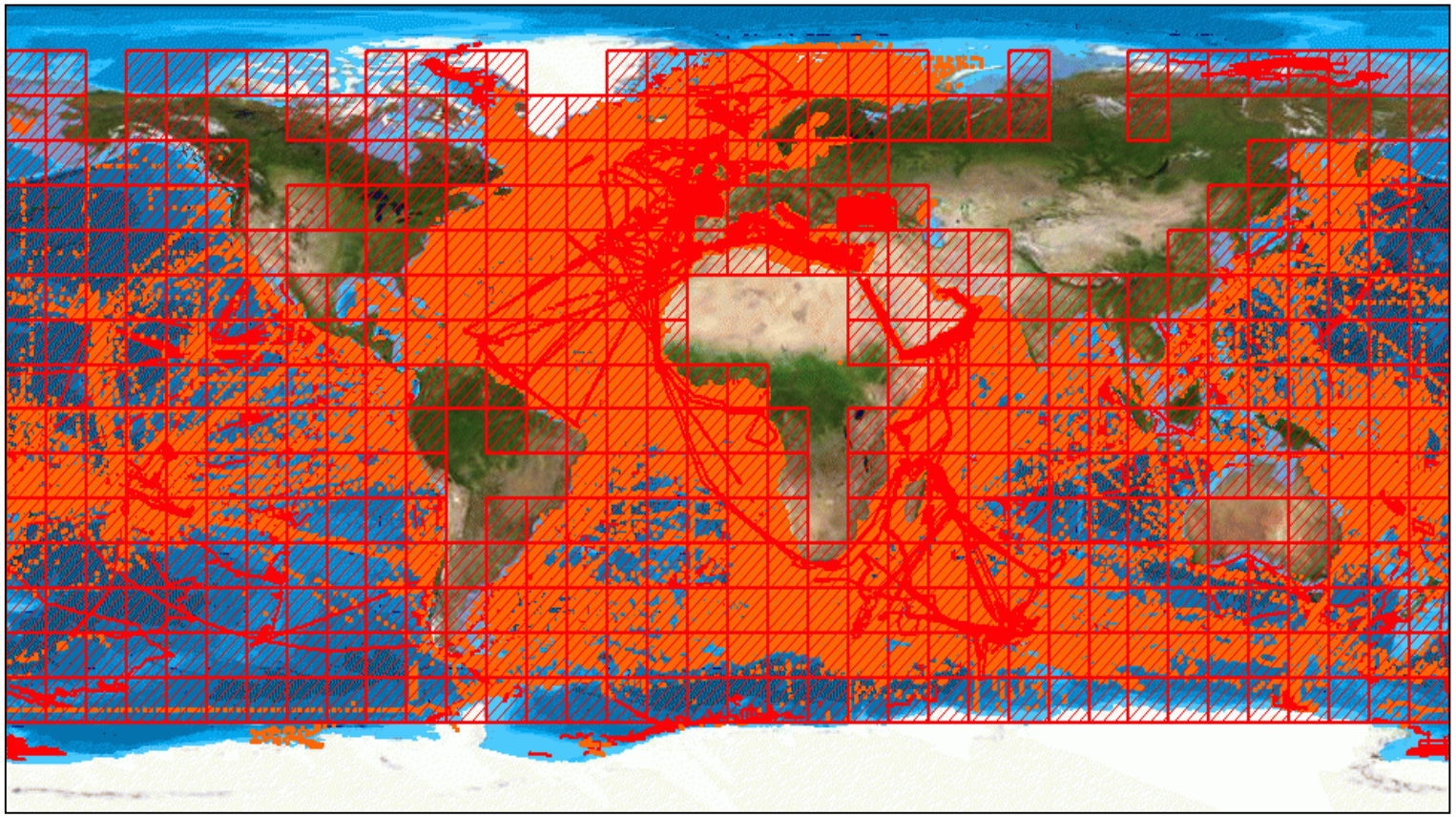
Operational since 2009

Portal providing interoperable data access, central overview, access to data products and tools

CDI (Common Data Index) service for discovery and unified access to data



http://www.seadatanet.org



Current situation: **> 1,97 million** CDI entries from **103** data centres in **34** countries and **509** originators. Data from physics, chemistry, geology, geophysics, bathymetry and biology. **84%** of the data is unrestricted or under SeaDataNet licence

Pan-European directory services

Overview of organizations in Europe with their involvement in marine projects, data sets, research cruises and monitoring activities

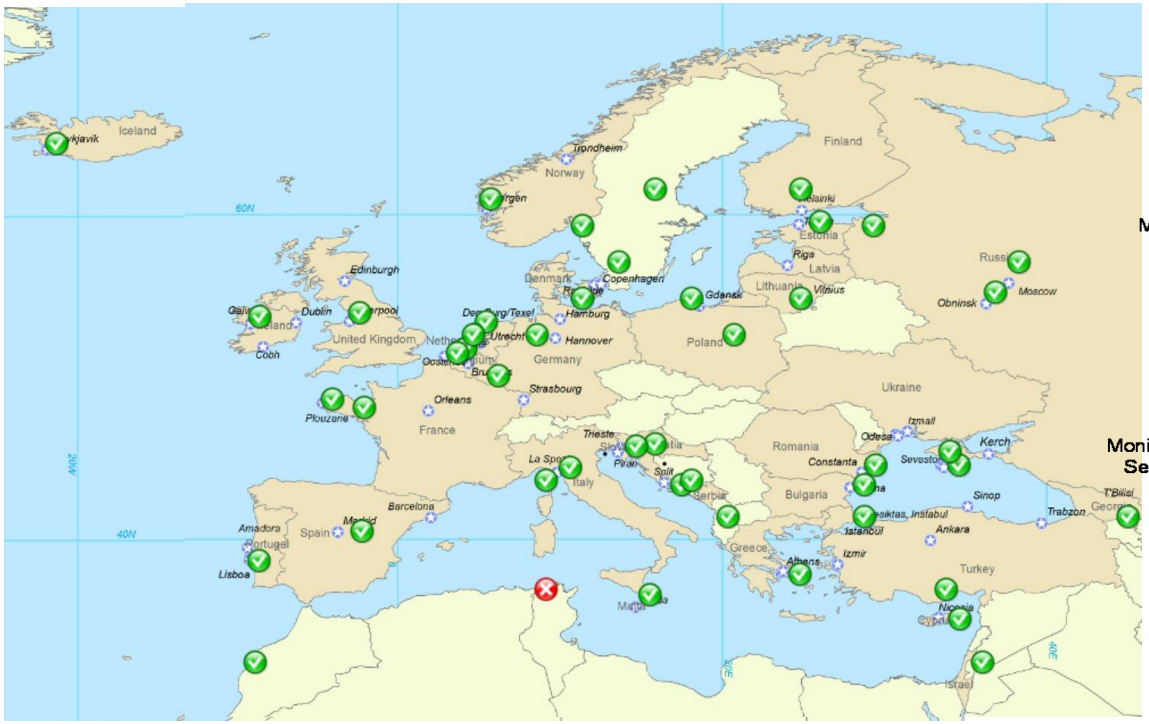
- **EDMO :** European Directory of Marine Organisations (>2900 entries)
- **EDMED** : European Directory of Marine Environmental Datasets (>3900 entries from >700 data holding centres)
- **EDMERP** : European Directory of Marine Environmental Research projects (>2800 entries)
- **CSR** : Cruise Summary Reports (>44000 entries)
- **EDIOS** : European Directory of Ocean Observing Systems (programmes and stations) (360 programmes; >16000 stations)
- **CDI** : Common Data Index (>1.97 million data sets; > 100 data centres; > 500 data originators)

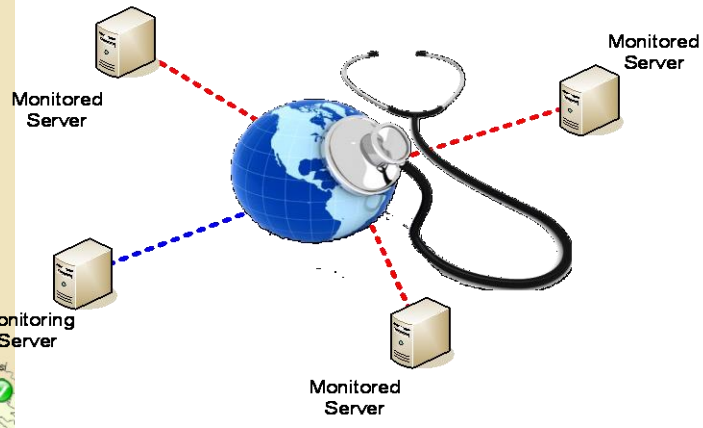
Additional achievements

- Set of common standards for metadata and data formats for the marine domain, adopting ISO and OGC standards and achieving INSPIRE compliance
- Controlled vocabularies for the marine domain (> 160.000 terms over > 60 lists)
- Quality controlled data after rigorous QC in cooperation with the MyOcean project
- Common SeaDataNet Data Policy and SeaDataNet Licence
- Powerful set of software tools (a.o.):
  - ODV: Ocean Data View adapted to SeaDataNet needs
  - DIVA: interpolation software for product generation adapted to SeaDataNet needs

Monitoring of infrastructure

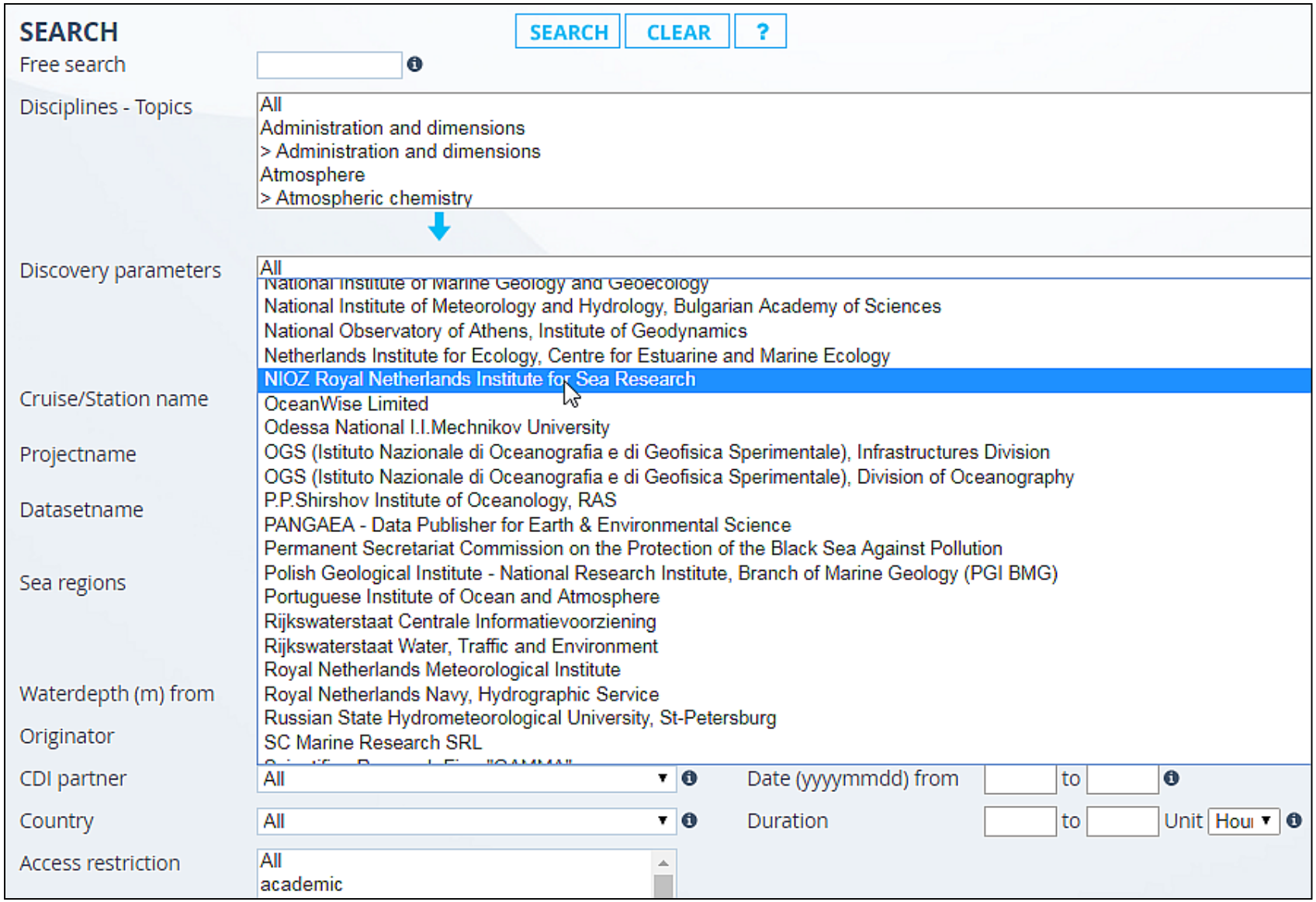
The operational availability of both the central as well as the local core services of the infrastructure is continuously being monitored.





Search for NIOZ data via NIOZ system

Search for NIOZ data via SeaDataNet system



Search for NIOZ data via SeaDataNet system

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

SEADATANET COMMON DATA INDEX (CDI) V3

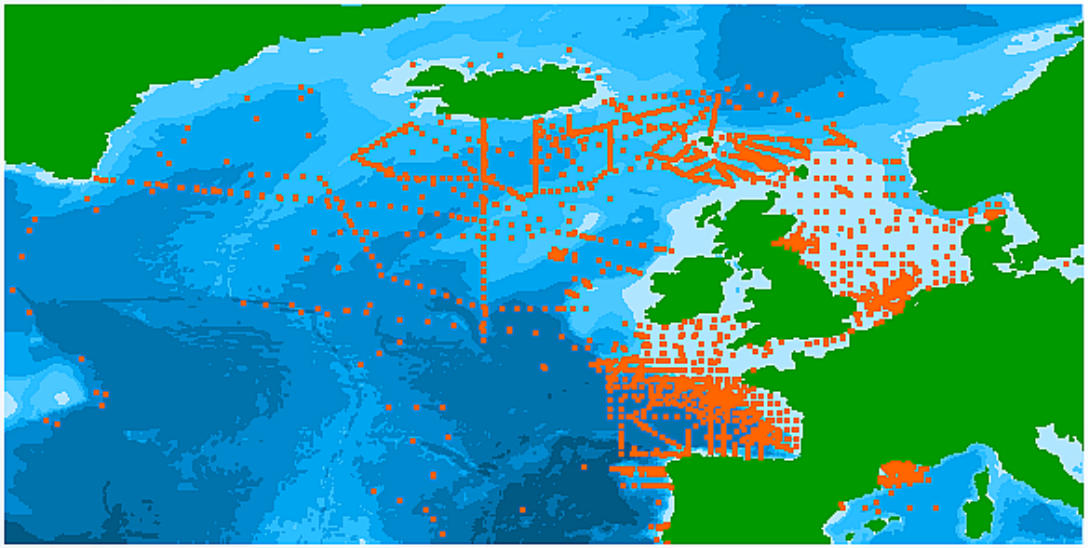
TOOLS

ENLARGE HELP

POSITION INDEX

Datasets 0

BASKET RESET



LAYER CONTROL

- CDI entry Points
- CDI entry Tracks
- CDI entry Areas
- Grid Lines
- Regional sea
- Regional sea labels

Display all selected records

Only selected records in results list

LISTING RESULTS

20 100 1000 records

ADD TO BASKET | TIMESERIES ON | SUMMARY | ZOOM TO SELECTED | EXPORT RESULT | STORE QUERY

| Refine query | New query | Found 5594 | Show (1-20) | Previous | Next 20

#	Data set name	DC country	Start date	Disciplines - Topics	Instrument / gear type	Show
	64PE390_1_29	Netherlands	20140614	Biological oceanography > Pigments Physical oceanography > Optical properties	CTD	

Search for NIOZ data via SeaDataNet system

SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE FOR OCEAN & MARINE DATA MANAGEMENT

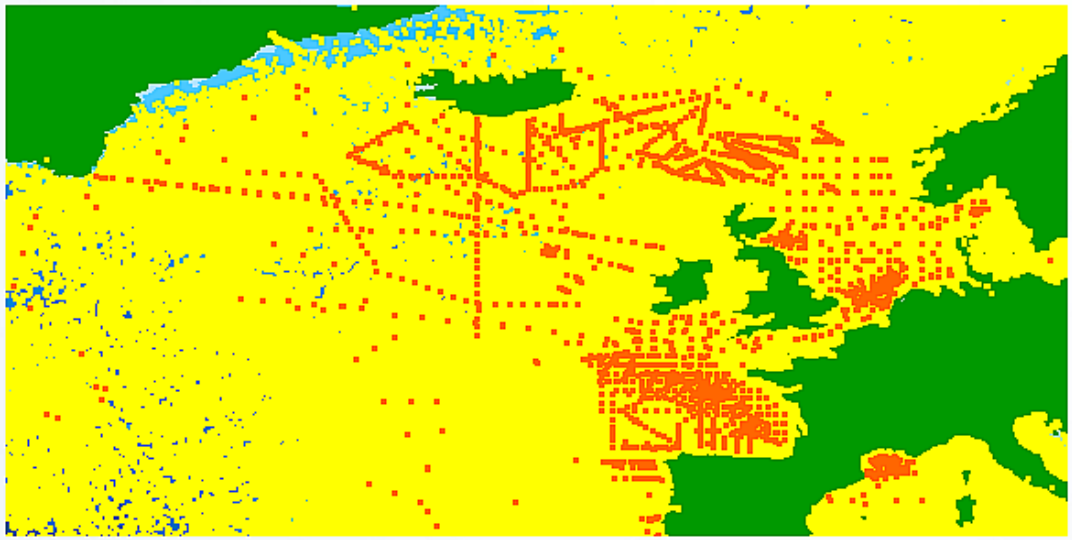
SEADATANET COMMON DATA INDEX (CDI) V3

TOOLS

ENLARGE | HELP | POSITION | INDEX

Datasets 0

BASKET | RESET



LAYER CONTROL

- [x] CDI entry Points
- [ ] CDI entry Tracks
- [ ] CDI entry Areas
- [ ] Grid Lines
- [ ] Regional sea
- [ ] Regional sea labels

(•) Display all selected records
( ) Only selected records in results list

LISTING RESULTS

(•) 20 ( ) 100 ( ) 1000 records

ADD TO BASKET | TIMESERIES ON | SUMMARY | ZOOM TO SELECTED | EXPORT RESULT | STORE QUERY

| Refine query | New query | Found 5594 | Show (1-20) | Previous | Next 20

#	Data set name	DC country	Start date	Disciplines - Topics	Instrument / gear type	Show
	64PE390_1_29	Netherlands	20140614	Biological oceanography > Pigments Physical oceanography > Optical properties > Other physical oceanographic	CTD	

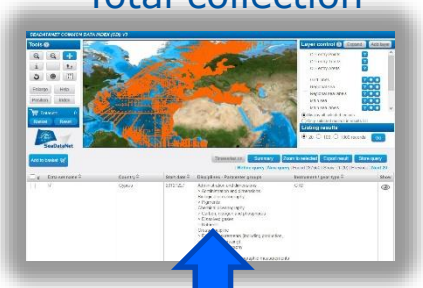
A pyramid of global oceanographic data infrastructures

- SeaDataNet is the data infrastructure for European Marine Observation and Data Network (EMODnet), which produces thematic data products

- SeaDataNet CDIs can be retrieved from the UNESCO/IOC/IODE Ocean Data Portal (ODP)

- SeaDataNet CDIs can be retrieved from the Group on Earth Observations System of Systems portal

Total collection



Data discovery and access

SeaDataNet

> 100 data centres

NODCs; HOs; GEOs; BIOs; ICES; PANGAEA

> 500 European data originators

GEOSS portal

IODE ODP portal

Aggregated collection

Black Sea portal

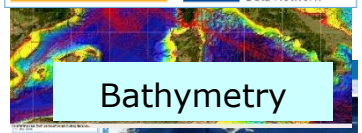
Caspian portal

Geo-Seas portal

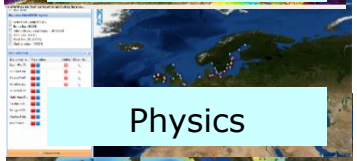
Regional subsets

Thematic portals

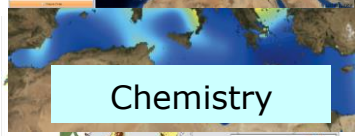
EMODnet European Marine Observation and Data Network



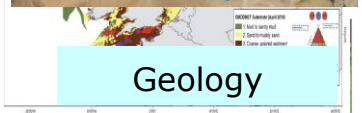
Bathymetry



Physics



Chemistry



Geology

Biology

CDI Data Discovery and Access service