SeaDataCloud

Project overview

Michèle Fichaut, IFREMER
SeaDataCloud project coordinator

Overview

- Context and history

- SeaDataNet components

- Objectives of SeaDataCloud

- Activities and partner's roles
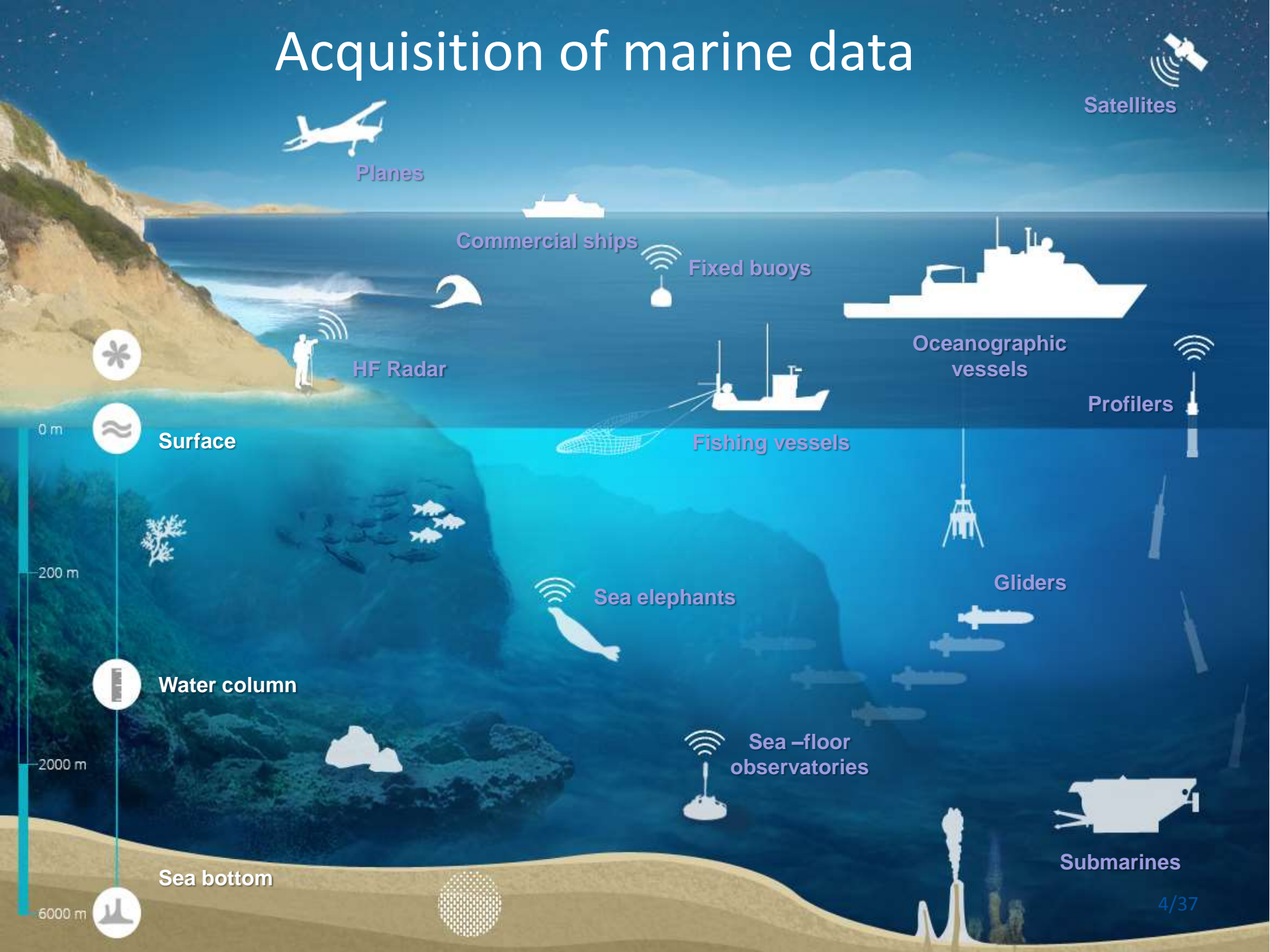
- Zoom on coordination activities (WP2)

Importance of marine data

- Marine data are relevant for many uses
  - Scientific research to gain knowledge and insight
  - Monitoring and assessment (water quality, climate status, stock)
  - Coastal Zone management
  - Modelling (including hindcast, now-cast, forecast)
  - Dimensioning and supporting operations and activities at sea (shipping, offshore industry, dredging industry, ..)
  - Implementation and execution of marine conventions for protection of the seas
  - Implementation of international Directives, such as in Europe directives for water (WFD), marine strategy (MSFD)

Users originate from government, science sector, and industry

Acquisition of marine data

Satellites

Planes

Commercial ships

Fixed buoys

HF Radar

Oceanographic vessels

Profilers

Surface

Fishing vessels

Sea elephants

Gliders

Water column

Sea –floor observatories

Submarines

Sea bottom

0 m

200 m

2000 m

6000 m

SeaDataCloud

Economy of data acquisition

- Data are collected by governments, research institutes and private industry (more than 1,000 organisations in Europe)

- Data from various disciplines such as physics, geophysics, meteorology, chemistry, biology, geology, bathymetry…

- Acquisition of marine data is expensive : annual cost in Europe estimated at 1.4 Billion € (1 for in-situ data, 0.4 for satellite data)

Professional data management is required with agreements on standardisation, quality control procedures, long term archiving, catalogue and access

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, and beginning of 2000s	Data and metadata EDMED - EDIOS EURONODIM Medar/MedAtlas
2002-2005	Metadata directories Sea-Search (FP5)
2006-2011	SeaDataNet (FP6)
2011-2015	SeaDataNet II (FP7)
2016-2020	SeaDataCloud (H2020)

SeaDataNet principles

- SeaDataNet is based on a **semi-distributed** system (centralised metadata catalogues and distributed data). It relies on the existing Data centre network and participate to the enhancement of these data centres

- The SeaDataNet RI enables a data user to see the data from all the data centres connected as managed by a **virtual unique data centre**, delivering integrated data, metadata and products

The 3 phases of SeaDataNet RI

- SeaDataNet (2006-2011, FP6, 10 M€)
  - Creation of the SDN infrastructure: standards, vocabs, metadata catalogues, search and downloads service….
  - At the end : around 50 data centres connected, 1 M data

- SeaDataNet 2 (2011-2015, FP7, 6 M€)
  - Increase the Robustness and the efficiency (machine to machine)
  - More data centres (102), more data (1.8 M)
  - Add new data types (biological data)

- SeaDataCloud (2016-2020, H2020, 10 M€)
  - Move to the cloud (faster, more capacities for data processing)
  - New services and Improved existing ones
  - Online tools in a VRE, MySeaDataCloud
  - New data types: gliders, HF-radar, Flow cytometry
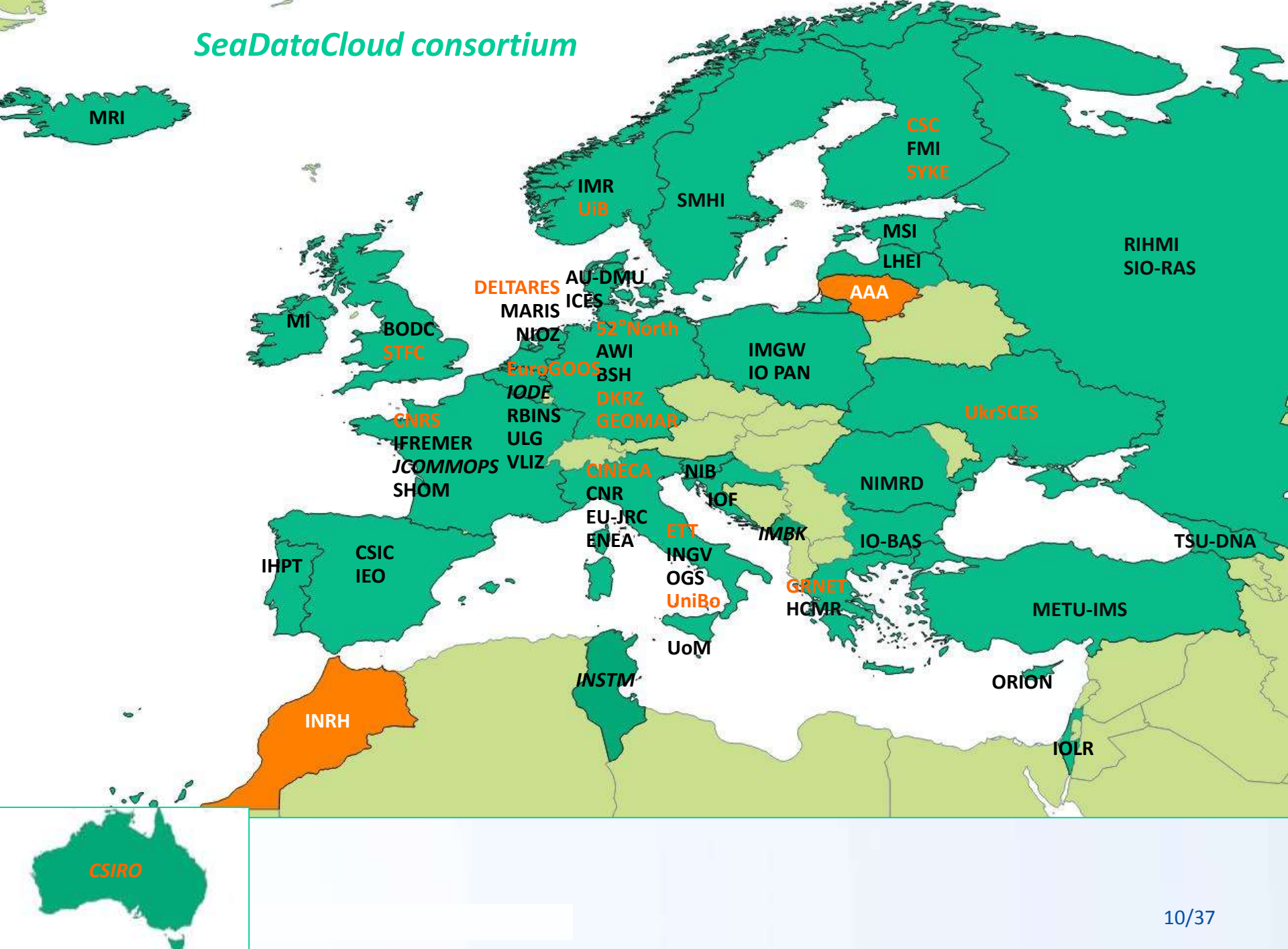 - More data centres (115), more data (2,3 M)

SeaDataCloud – Key figures

- Start 1st of November 2016

- Duration 4 years

- Budget 10 M€

- 56 partners

- 5 subcontractors

- Total of 32 countries represented

- 1110.5 person/month

SeaDataCloud consortium

MRI

CSC
FMI
SYKE

IMR
UiB

SMHI

MSI
LHEI

RIHMI
SIO-RAS

AU-DMU
DELTARES
ICES
MARIS
NIOZ
52°North

AAA

MI

BODC
STFC

AWI
EuroGOOS BSH

IMGW
IO PAN

IODE
RBINS
ULG
VLIZ

DKRZ
GEOMAR

UkrSCES

CNRS
IFREMER
JCOMMOPS
SHOM

CINECA
CNR
EU-JRC
ENEA

NIB
IOF

NIMRD

IMBK

IHPT

CSIC
IEO

ETT
INGV
OGS
UniBo

IO-BAS

TSU-DNA

GRNET
HCMR

METU-IMS

UoM

ORION

INSTM

INRH

IOLR

CSIRO

SeaDataNet standards

- Set of common standards for the marine domain, adapting ISO and OGC standards and achieving INSPIRE compliance

  – **Adoption of ISO 19115 – 19139 standard for describing metadata** on data, data sets, research cruises, monitoring networks, and research projects => marine metadata profiles, schemas, schematron rules

  – **Controlled vocabularies** for the marine domain (>84,000 terms in 111 lists), with international governance and web services

  – **Standard data exchange formats** : ODV ASCII and NetCDF (CF) and MedAtlas ASCII fully supported by controlled vocabularies

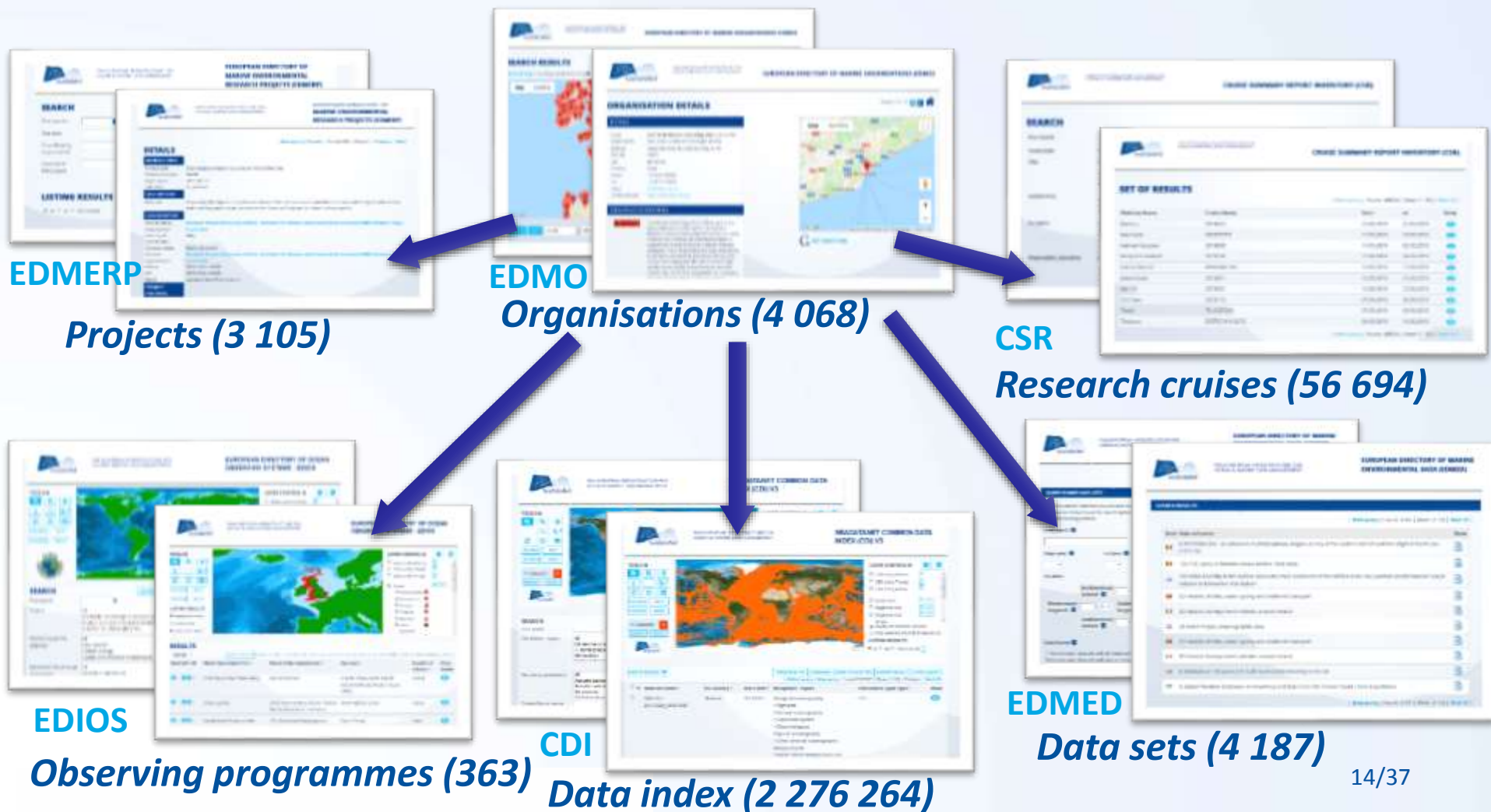- Maintenance and dissemination of standard QA-QC procedures, together with IOC/IODE and ICES (QC-loop)

SeaDataNet standards SDN2

- Set of common standards for the marine domain, adapting ISO and OGC standards and achieving INSPIRE compliance

  – **Adoption of ISO 19115 – 19139 standard for describing metadata** on data, data sets, research cruises, monitoring networks, and research projects => marine metadata profiles, schemas, schematron rules

  – **Controlled vocabularies** for the marine domain **(>84,000 terms in 111 lists)**, with international governance and web services

  – **Standard data exchange formats** : ODV ASCII and **NetCDF (CF)** and MedAtlas ASCII fully supported by controlled vocabularies

- Maintenance and dissemination of standard QA-QC procedures, together with IOC/IODE and ICES **(QC-loop)**

SeaDataNet standards SDC

- Set of common standards for the marine domain, adapting ISO and OGC standards and achieving INSPIRE compliance

    – **Adoption of ISO 19115 – 19139 standard for describing metadata** on data, data sets, research cruises, monitoring networks, and research projects => marine metadata profiles, schemas, schematron rules

    – **Controlled vocabularies** for the marine domain **(>84,000 terms in 111 lists)**, with international **governance and web services**

    – **Standard data exchange formats** : ODV ASCII and NetCDF (CF) and MedAtlas ASCII fully supported by controlled vocabularies ➜ **INSPIRE**

- Maintenance and dissemination of standard QA-QC procedures, together with IOC/IODE and ICES **(QC-loop)**

Centralised SeaDataNet metadata directories

EDMERP
Projects (3 105)

EDMO
Organisations (4 068)

CSR
Research cruises (56 694)

EDIOS
Observing programmes (363)

CDI
Data index (2 276 264)

EDMED
Data sets (4 187)

CDI discovery and access service

SeaDataNet portal

Search and Shop

Data

Presently 115 data centres are connected, some are still underway

European data sources

data centres ← ≈ 650 originators

Metadata

+ transaction data

Data centres

SeaDataNet tools

- **Set of tools** to be used each data centre and freely available from the SeaDataNet portal: metadata editor (MIKADO), data conversion software (NEMO and **OCTOPUS**), data analysis software (ODV), data interpolation software (DIVA), navigation sampling (EndsAndBends), DownLoad Manager (DM) ➔ **Replication Manager**

SOFTWARE

A major objective and challenge in SeaDataNet is to provide an integrated and harmonised overview and access to data resources, managed by distributed data centres. Moreover it is an objective to provide users common means for analysing and presenting data and data products. Therefore the Technical Task Team of SeaDataNet has designed an overall system architecture, and is developing common software tools for data centres and users.

Common software tools are being developed and freely made available to **Data Centres and/or End Users** for:

- Editing and generating XML metadata entries: **MIKADO javatool**
- Tool for the generation of spatial objects from vessel navigation during observations: **EndsAndBends**
- SeaDataNet file format convertor : **OCTOPUS**
- Conversion of any ASCII format to the SeaDataNet ODV4 ASCII format: **NEMO javatool**
- Connecting systems of Data Centres to the SeaDataNet portal for data access: **Download Manager javatool**

- Analysing and visualising of data sets: **Ocean Data View (ODV) software package**
- Interpolation and variational analysis of data sets: **DIVA software package**

MIKADO | NEMO | OCTOPUS
ODV | DIVA | DOWNLOAD MANAGER
ENDSANDBENDS

SeaDataNet T and S products

SeaDataNet Quality Checks Strategy (QCS)

- CENTRAL CDI
- Data harvesting
- Files and parameters aggregation
- QC analysis
- Analysis of data anomalies
- Copernicus

Access to the SeaDataNet Products Catalogue
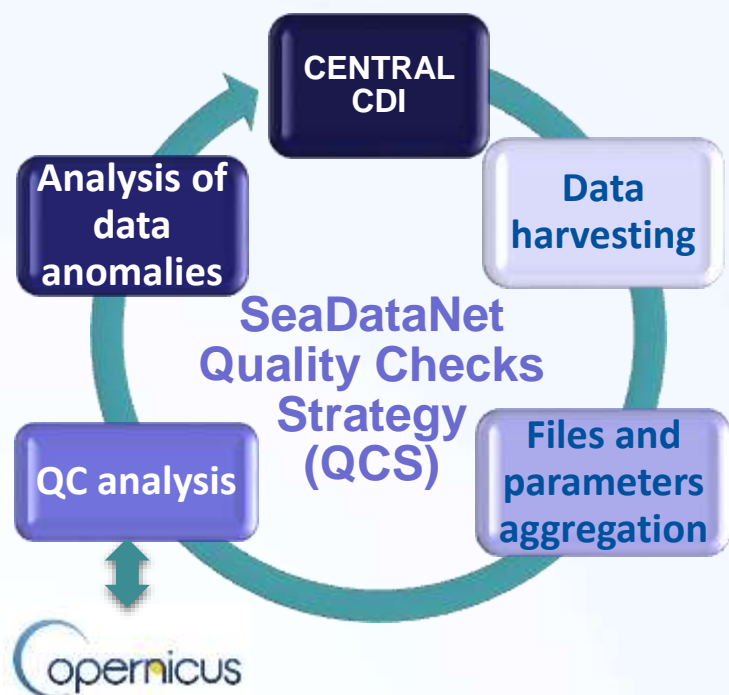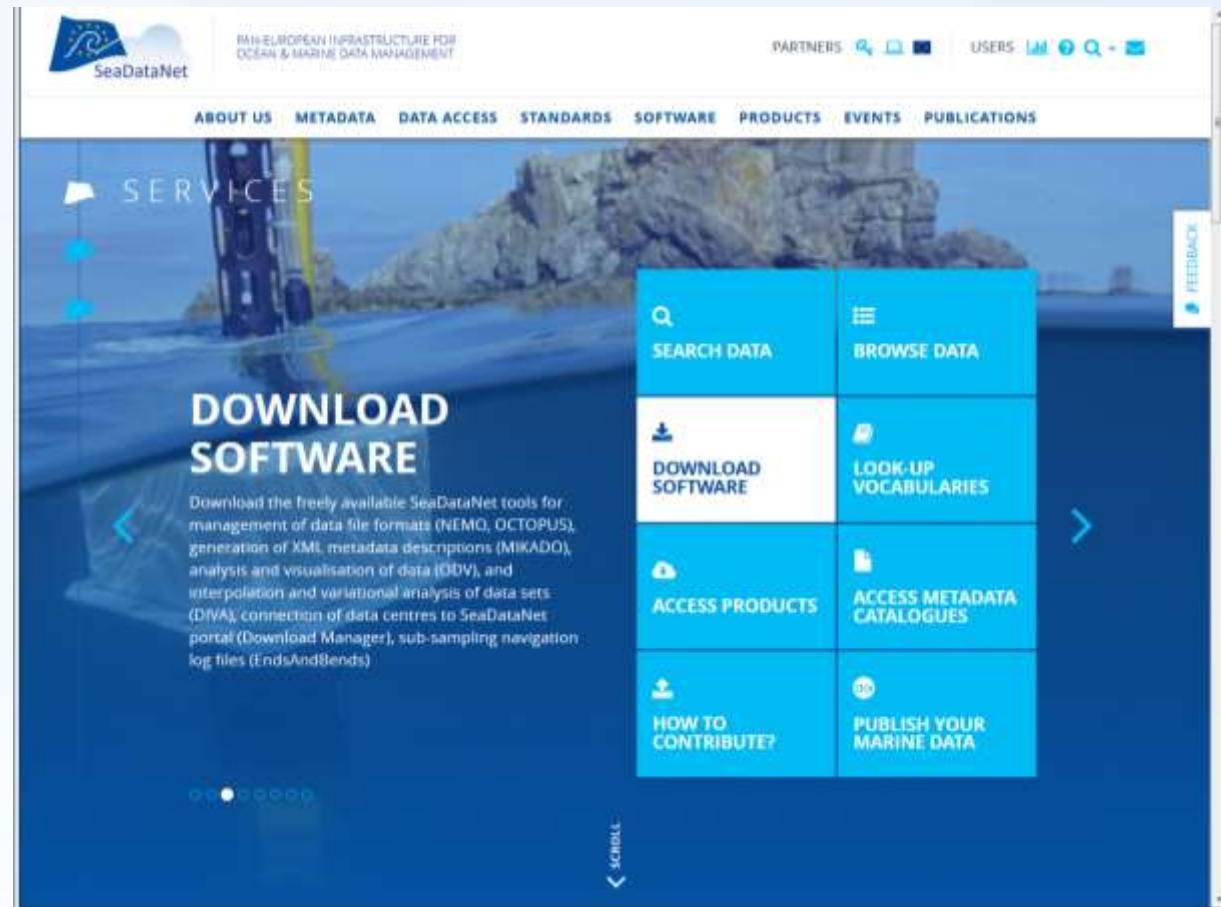
Improvement of the data quality
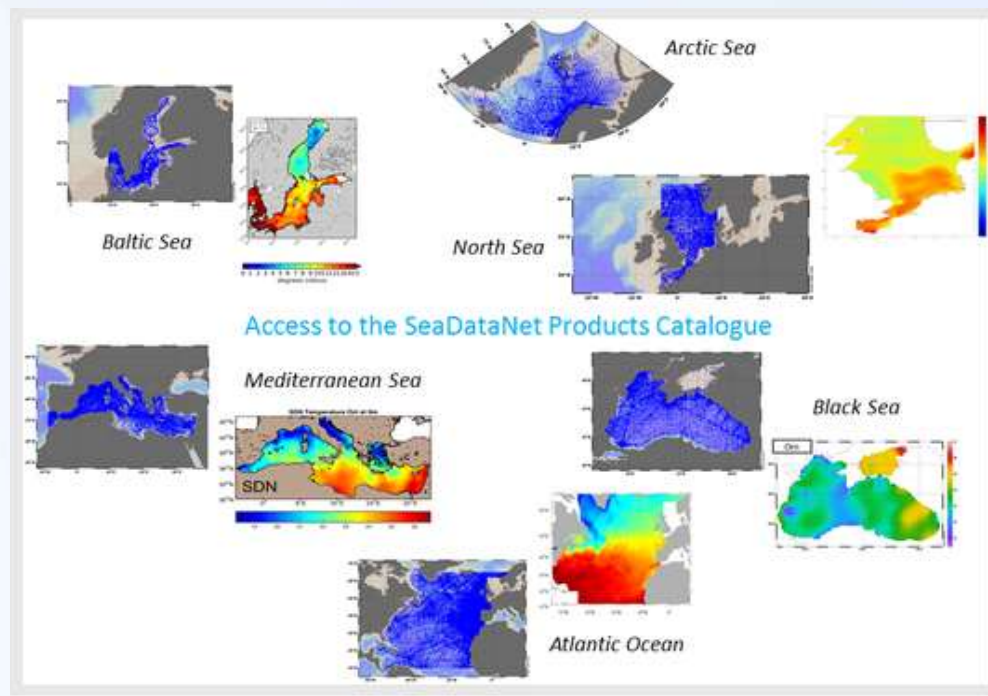
Aggregated datasets and climatologies – 6 regional products

SeaDataNet portal (totally redesigned for SDC)

Giving access to

- Standards, tools both for data centres and other users

- Data and metadata

- Products

Issues with the present CDI service

- **Usage of the discovery and access services lags behind expectation**; major obstacle is that users have to undertake multiple download transactions in case of shopping baskets with data from multiple data centres.

- **Performance issues** mainly linked to the different machine capacities of the data centres connected to the infrastructure which can give 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 etc => there are different versions installed, because upgrading can give issues

Solutions: new developments done or on-going in SeaDataCloud

- Make use of the cloud environment to improve the performance of the CDI data access and to improve capacities for data processing including additional checks (WP9, WP10)

- Develop online services like WebODV and DIVA-online (WP9)

- Upgrade of the SeaDataNet vocabularies services and governance (WP8)

- Apply linked data principles for SDN directories (WP8)

- Develop a Virtual Research Environment to facilitate individual and collaborative research by users (WP10)

- Improve the data quality by implementing more checks on the data file and by improving the QC loop when generating the SeaDataNet products (WP11)

- SeaDataNet architecture without the Cloud
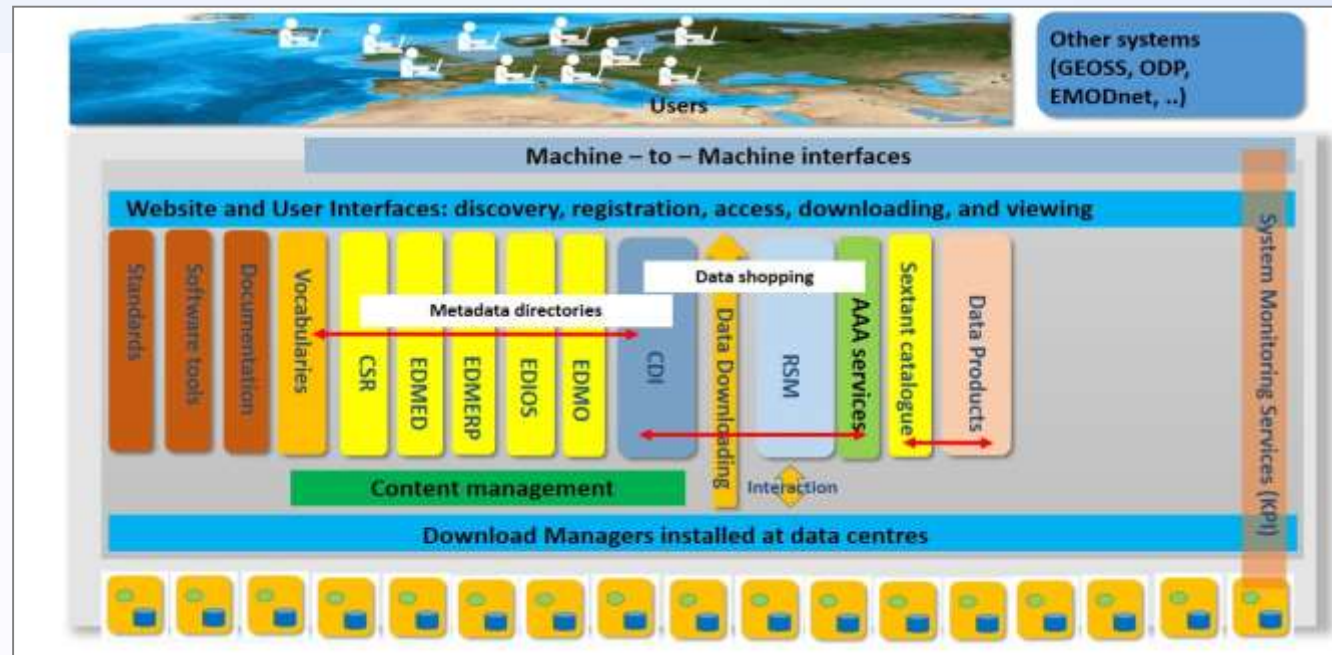
- SeaDataNet architecture with the cloud
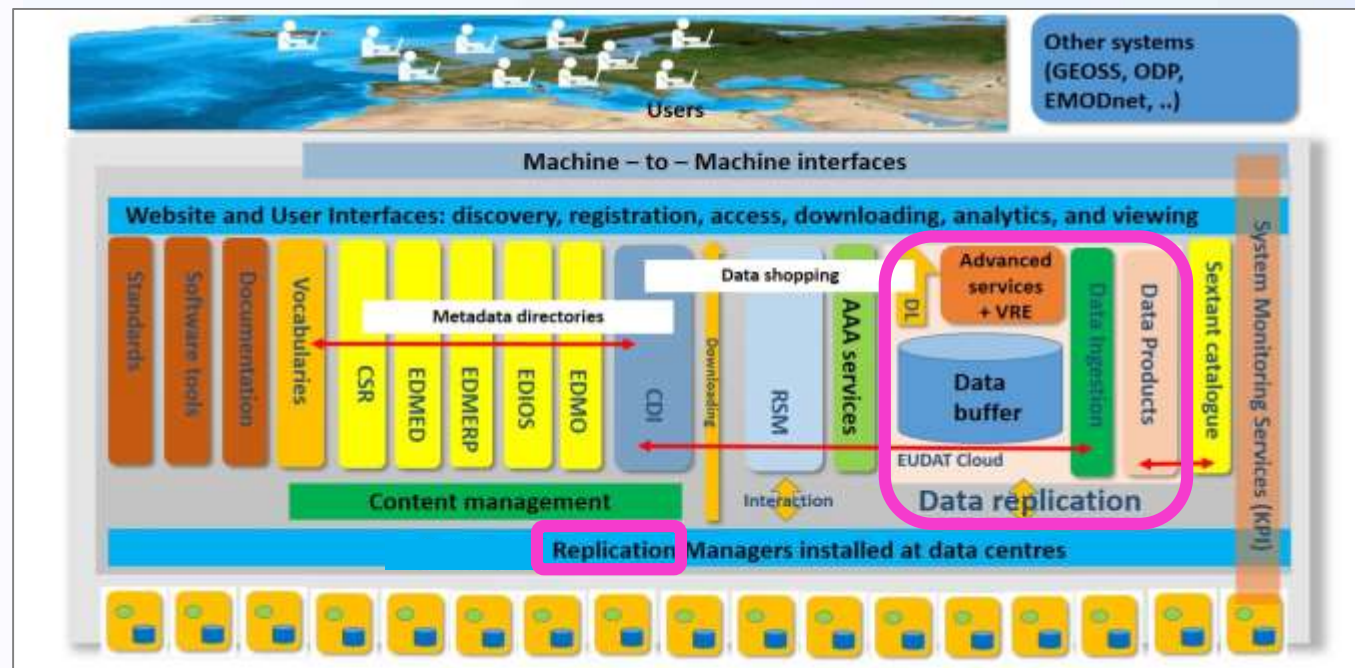
SeaDataCloud : 3 types of activity – 11 WP (1)

- – WP1 : Project management (IFREMER) (22 p/m)
- Networking activities, 4 work packages (352.5 p/m)
  - – WP 2: NA1- Project network coordination (IFREMER)
  - – WP3: NA2 - Training of data providers (RBINS)
  - – WP4: NA3 - Communication, dissemination and exploitation development (ENEA)
  - – WP5: NA4 - Expansion and governance of metadata and data content (HNODC)
- Virtual access activities, 1 work package (71 p/m)
  - – WP6: VA1 - Core and advanced services (OGS)

# SeaDataCloud : 3 types of activity – 11 WP (2)

- Joint Research Activities, 5 work-packages (665 p/m)
  - WP7: JRA1 - Tuning of requirements and overall integration (EU-JRC)
  - WP8 : JRA2 - Governance of standards and development of common services (NERC-BODC)
  - WP9 : JRA3 - Development of upstream services (MARIS)
  - WP10: JRA4 - Development of downstream services (IFREMER)
  - WP11: JRA5 – Development, update and publication of data products for European sea regions (INGV)

Management – Structure (1)

- Key-functions :
    - EU project officer : Agnès Robin
    - Project coordinator (PC - IFREMER)
    - Technical coordinator (TC - MARIS)
    - Scientific coordinator (SC – EC-JRC)
    - Work-package leader
    - Task Leader
    - Regional coordinator for products (RC)

Management – Structure (2)

- Key-committees
  - Steering Group (SG)
    - PC, TC, SC, WP leaders, RC
    - ICES (MSFD : Marine Strategy Framework Directive), VLIZ (Marine biological data management), BSH (management of CSR catalogue), CSC (representing the EUDAT consortium), ULG and AWI (development of major tools)
  - Project office (PO for supporting PC and SG)
  - Technical Task Group (TTG)
  - Scientific Committee (SC)
  - Advisory Group (group of external experts)
  - Project coordination group of all partners and subcontractors
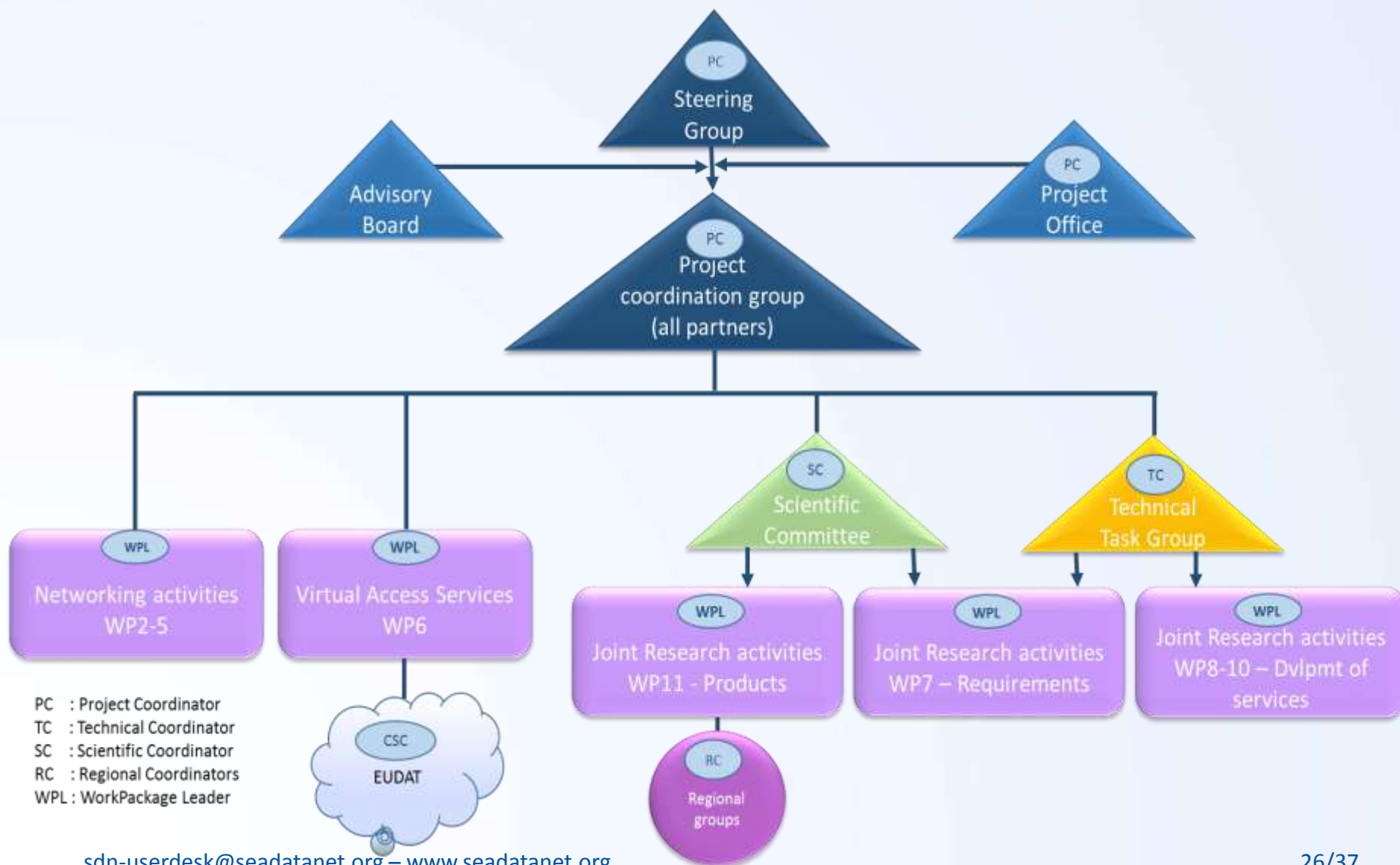
Partners' roles

- Project management and coordination
  - Project coordinator (1)
  - Technical coordinator (1)
  - Work package Leaders (9)

- Specific roles
  - Central catalogues and services managers including the cloud (10)
  - Regional coordinators for products (6)
  - Software and standard developers (30)
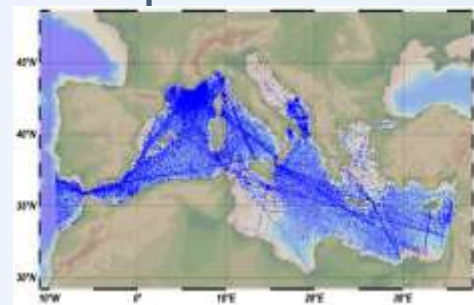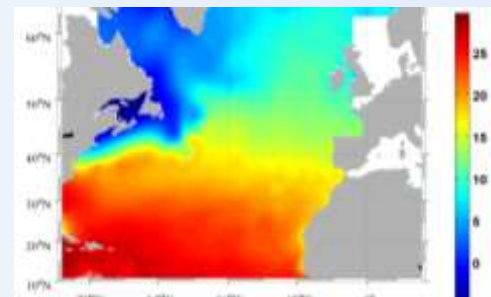 - Data centres (44)
  - Scientific advisers (6)

Central catalogue managers' role

- Maintain and Upgrade the tools for the catalogues (Content Management System for manual input, loading software, extracting software, discovery interface)

- Helpdesk

- Keep partners informed on the new developments

- Load the new XML metadata files coming from partners

- Feedback to partners

- Training of partners

Regional coordinators' role

- Aggregate data sets and control overall quality

- Generate a set of generic data products such as aggregated datasets and climatologies

- Validate products and report on the quality and the possible limitations of the data sets used

- Prepare documentation about the products

- Harmonise products between sea basins

- Make products and documentation available online

- Report errors and gaps to data centres – follow-up of the data corrections

- Participate to the DIVA workshops

Data centres' role

- Keep the SDN system available 'continuously'

- Regularly update the catalogues content with new data and metadata and/or updates of existing records

- Manage the user data requests for restricted data

- Attribute roles to the SDN users at the national level

- Upgrade the SDN tools each time new deliveries are available

- Attend the SDC training sessions

Coordination activities – WP2

- Organisation of 3 plenary meetings
  - Kick-off, Riga, Latvia, 30 November – 1 December 2016 *(2 days)*
  - 1st annual, Athens, Greece, 18-20 October 2017 *(3 days)*
  - 2nd annual, Barcelona, Spain, 8-9 November 2018 *(1.5 day)*

- Organisation of 5 Steering Group meetings
  - SG1, Riga, Latvia, 28 November 2016 *(0.5 day)*
  - SG2, Bologna, Italy, 2-3 May 2017 *(2*0.5 days)*
  - SG3, Athens, Greece, 16 October 2017 *(0.5 day)*
  - SG4, Sopot, Poland, 24 April 2018 *(1 day)*
  - SG5, Barcelona, Spain, 9 November 2018 *(0.5 day)*

- Organisation of 5 Technical Task Group meetings
  - TTG1, Riga, Latvia, 29 November 2016 *(1 day)*
  - TTG2, Bologna, Italy, 3-4 May 2017 *(2*0.5 days)*
  - TTG3, Athens, Greece, 16-17 October 2017 *(1.5 day)*
  - TTG4, Sopot, Poland, 24 April 2018 *(1.5 day)*
  - TTG5, Helsinki, Finland, 10-11 September 2018 *(1.5 day)*
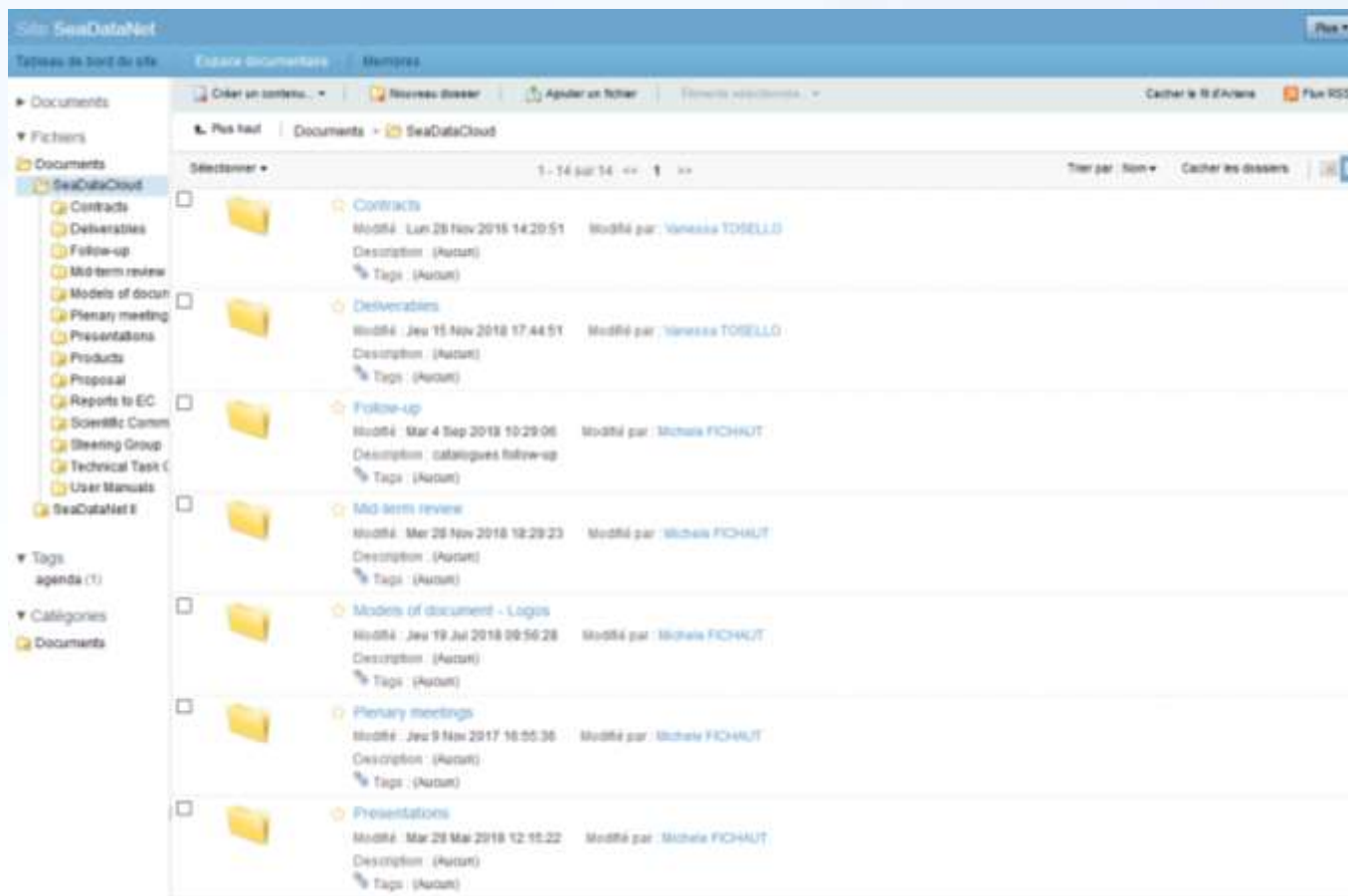
31/37

Coordination tools (1) – Mailing lists

- 12 mailing lists specific to the project
  - For project groups: sdc-admin, sdc-advisoryboard, sdc-all, sdc-legal, sdc-partners, sdc-po, sdc-product-lead, sdc-sciencecom, sdc-stcom, sdc-subcontractors, sdc-tech, sdc-vre

- 6 mailing lists for users
  - Helpdesk: sdn-userdesk
  - Software users: one list per software (DM, MIKADO, NEMO, OCTOPUS, EndsAndBends)

Coordination tools (2) – Extranet

• For the shared documents

Coordination tools (3) – Web site
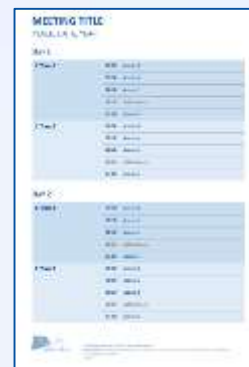
Coordination tools (4) – Models of documents

- For the powerpoint presentations

- For the deliverables

- For the meeting agendas

- Available on the extranet

Presentation

Report

Agenda