SeaDataNet, a network of distributed oceanographic data centres now going to the cloud

Serge SCORY (RBINS, Belgium), Dick M.A. SCHAAP (MARIS, The Netherlands) &
Michèle FICHAUT (IFREMER, France)
on behalf of the SeaDataNet communities

**International Workshop on Sharing, Citation and Publication of Scientific Data across Disciplines
Tachikawa, Tokyo, Japan, 5–7 December 2017**

- What is SeaDataNet, how does it work?
- On-going developments
- The reasons of success

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 35 countries bordering the European seas

90's	Metadata catalogs: MEDAR/MedAtlas, EDMED (FP4)
1998-2001	Euronodim
2002-2005	Sea-Search (FP5)
2006-2011	SeaDaatNet (FP6)
2011-2015	SeaDataNet II (FP7)
2016-2020	SeaDataCloud (H2020 = FP8)

Already 6 development phases

At the forefront: Portal with standards, tools, and services, both for users and data centres

SeaDataNet standards

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

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

  – Controlled vocabularies for the marine domain (> 65,000 terms and > 80 lists), with international governance and web services

  – Standard data exchange formats: ODV and NetCDF (CF)

SeaDataNet metadata directories
the conceptual backbone

EDMO *Organisations*

EDMERP *Projects*

CSR *Research cruises*

EDIOS *Observing programmes*

CDI *Data index*

EDMED *Data sets*

sdn-userdesk@seadatanet.org –

Vocabularies

- SeaDatanet is using code lists and controlled vocabularies to regulate the population of metadata. This opens up data sets to computer aided manipulation, distribution and long term reuse.

- Example: Parameter Usage Vocabulary (37364 terms!)

Parameter Usage Vocabulary

- Five elements in the semantic model:
  - Measurement property
  - Measurement statistical qualifier
  - Chemical substance
  - Measurement-matrix relationship
  - Matrix

Parameter Usage Vocabulary (P01)

3-layer hierarchy of discovery keywords:

– SeaDataNet Parameter Discovery Vocabulary (P02, 432): fine-grained related groups of measurement phenomena designed to be used in dataset discovery interfaces.

– SeaDataNet agreed Parameter Groups (P03, 70): coarse-grained groupings

– SeaDataNet Parameter Disciplines (P08, 11): topic/theme level

Simple Knowledge Organisation Systems (SKOS) mappings between these vocabularies

Aggregation

Aggregation of data sometimes require semantic interoperability infrastructure E.g. EMODNet chemistry product vocabulary (P35)

'Cadmium concentrations in shellfish'

- The P35 entry is mapped to 'micrograms per kilogram' in P06

- The P35 entry is mapped to the list of P01 entries that represent 'cadmium concentrations in shellfish'

CDI service for discovery and unified data access

SeaDataNet portal
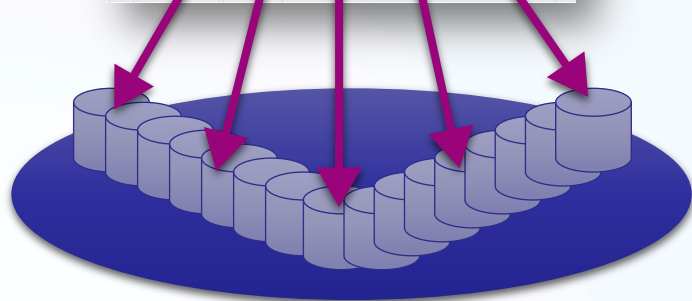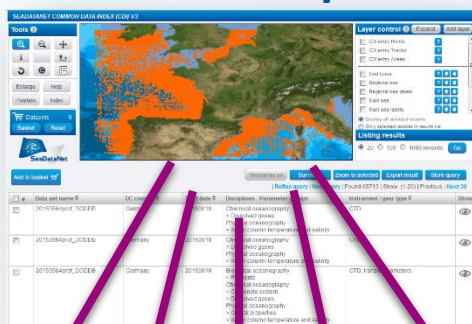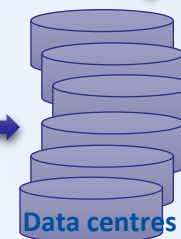

European data sources

109 data centres ← 600+ originators

Search and Shop

Data download

Metadata

+ transaction data

Data centres

SeaDataNet is a semi-distributed infrastructure:

- Central metadata database

- Datasets in distributed data centres

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

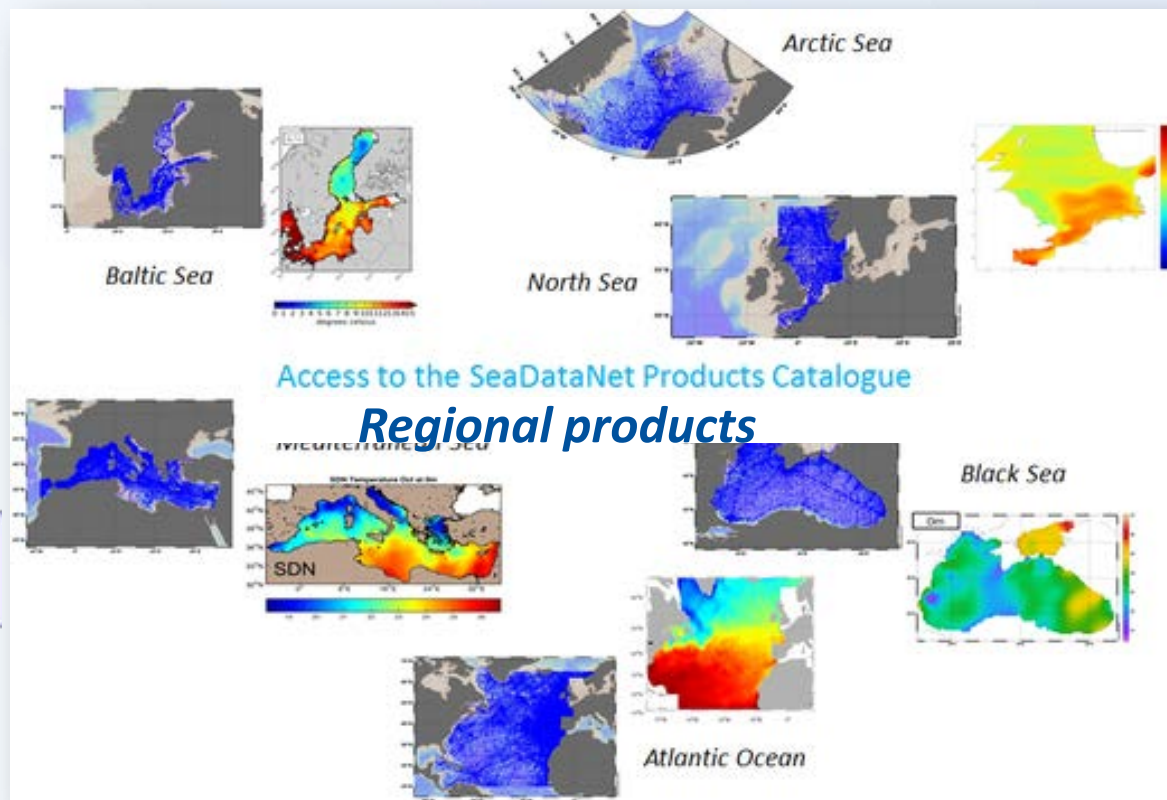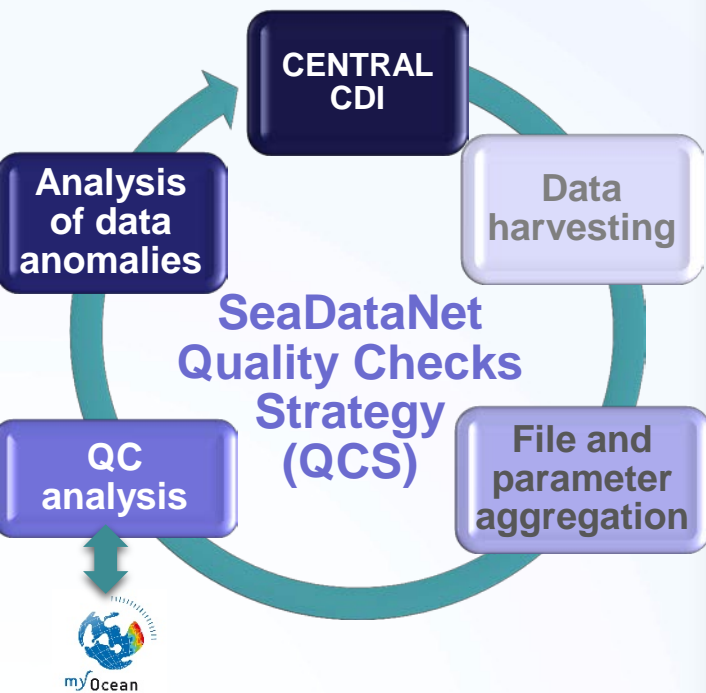# Interoperability with global portals

- CDI is available as OGC CSW, WMS and WFS service for exchange of CDI metadata

- CDI is connected with GEOSS by CSW and IODE

  - Aggregation of SeaDataNet metadata CDI granules to CDI collections (ISO 19115–19139) (1.9 million => 500 collections), conversion to Common Brokerage Model, and harvesting via CS-W and OAI-PMH service
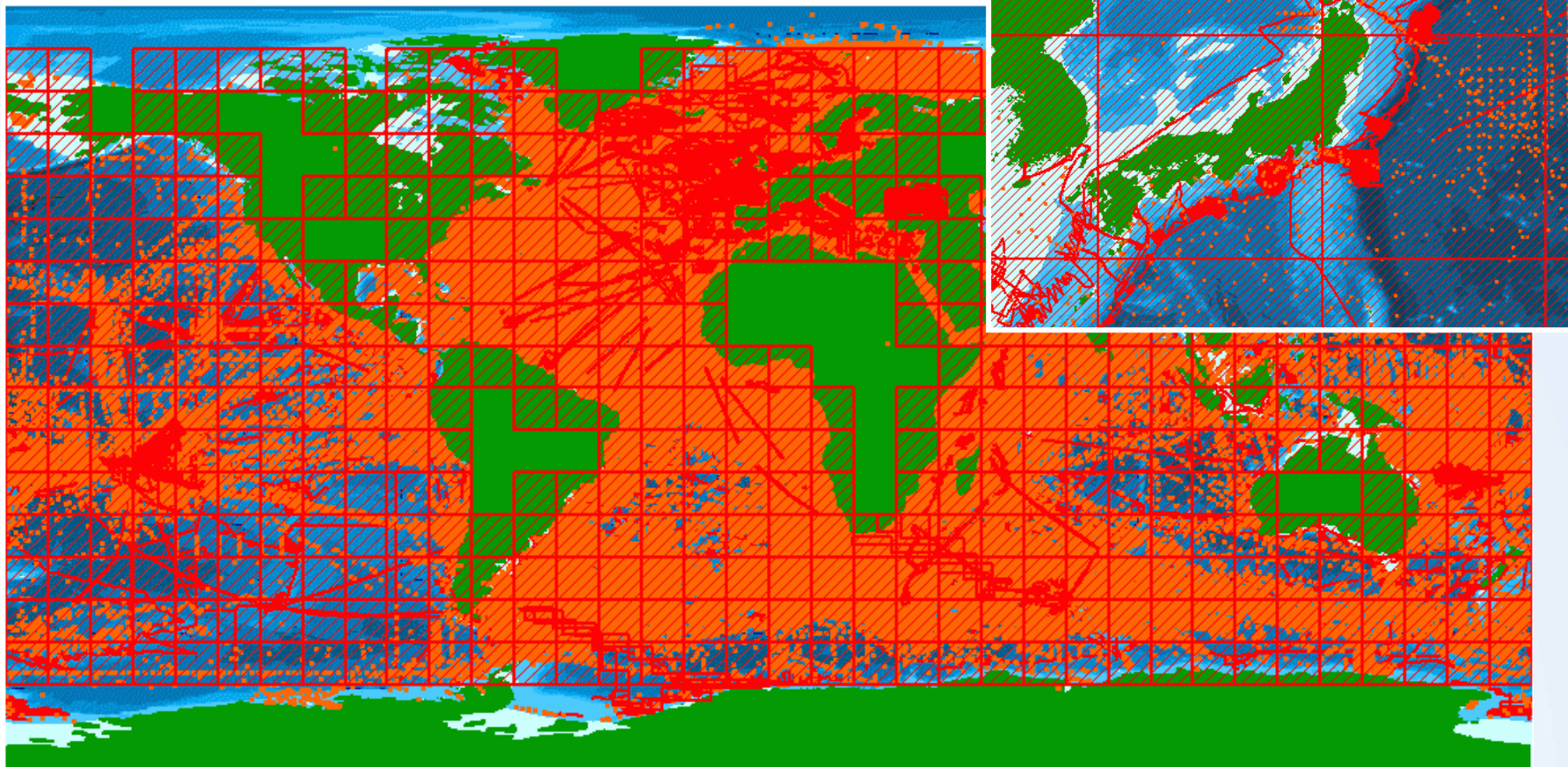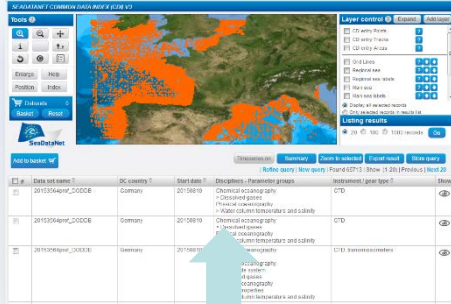
2.1 millions CDI entries from *34* countries, *102* data centres and *612* originators for physics, chemistry, geology, geophysics, bathymetry and biology; from *1805 to 2017*; *87.6%* unrestricted or under SDN License

SeaDataNet products

CENTRAL CDI

Data harvesting

Analysis of data anomalies

SeaDataNet Quality Checks Strategy (QCS)

QC analysis

File and parameter aggregation

myOcean

Access to the SeaDataNet Products Catalogue

Regional products

Baltic Sea

Arctic Sea

North Sea

Mediterranean Sea

Black Sea

Atlantic Ocean

Aggregated datasets and climatologies

Improvement of the data quality

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

SeaDataNet

GEOSS portal

IODE ODP portal

Aggregated collection

Data discovery
and access

Black Sea portal   Caspian portal

Regional subsets

Geo-Seas portal

EMODnet
European Marine
Observation and
Data Network

> 100 data centres

Thematic portals

NODCs; HOs; GEOs; BIOs; ICES; PANGAEA

*≈ 600 European
data originators*

*CDI Data Discovery and
Access service*

sdn-userdesk@seadatanet.org – w

https://youtu.be/p3vwngxyXuo
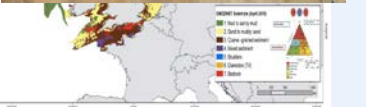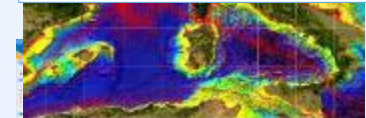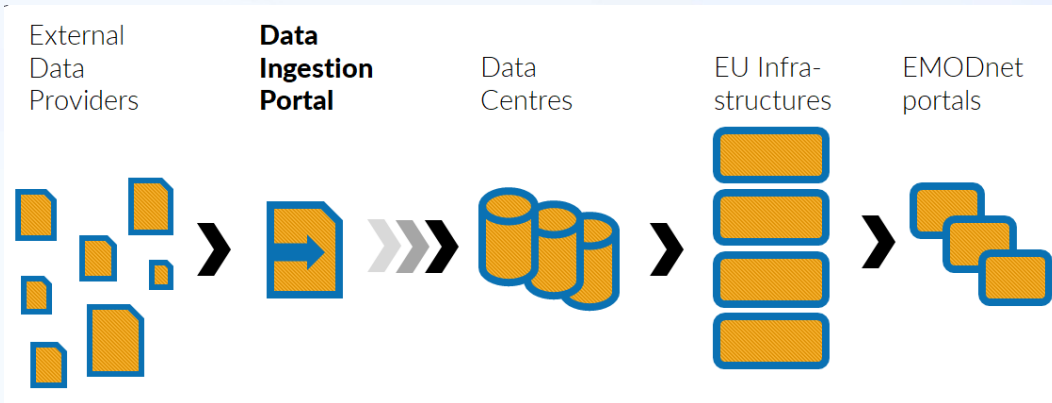
European Union initiative on Marine knowledge:

"Collect once, use many times!"

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

SeaDataCloud – a new opportunity

- Standards and information technology are always evolving, and the SeaDataNet infrastructure must stay up-to-date to maintain and further expand its services

- November 2016 start of H2020 **SeaDataCloud** project for further developing SeaDataNet infrastructure and associated standards: 10 Meuro, 61 members, 32 countries, 4 years

SeaDataCloud – general challenges

- Updating and further developing standards

- Improving and innovating services & products

- Adopting and elaborating new technologies

- Giving more attention to users and putting the user experience in a central position

- Implementing a strategic and operational cooperation between SeaDataNet and EUDAT (consortium of e-infrastructure service providers)

SeaDataCloud – cooperation with EUDAT

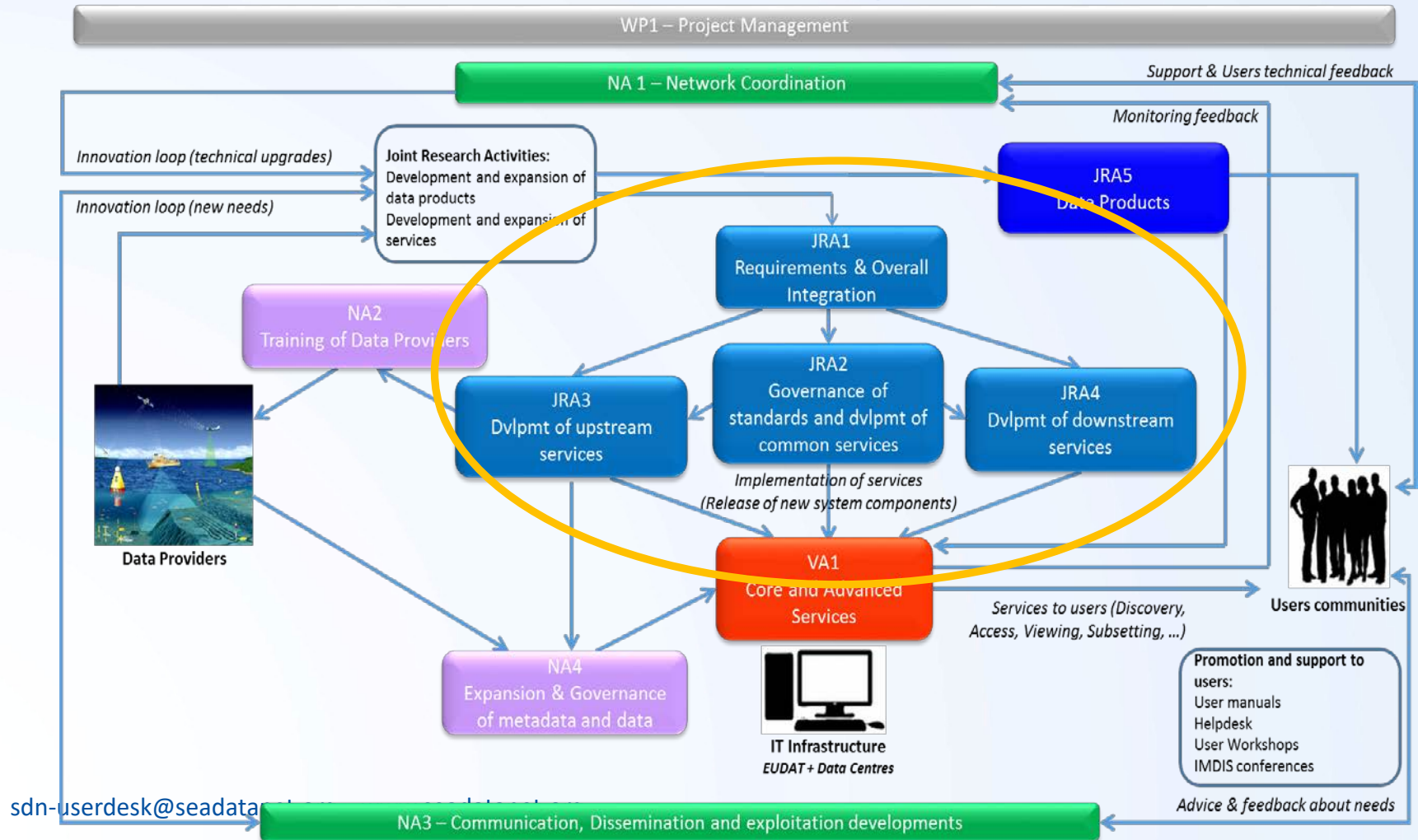**European Computing Infrastructure**

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

SeaDataCloud:
- **Maintaining the infrastructure**
- **Running the infrastructure**
- **Improving the infrastructure**

sdn-userdesk@seadatanet.org

WP8 - Governance of standards and development of common services

- To develop further the SeaDataNet controlled vocabularies and related services,

- To analyse and deploy a pilot for adopting the Linked Data principle for SeaDataNet directories,

- To review and expand the SeaDataNet data formats for achieving INSPIRE compliance,

- To integrate the SeaDataNet authentication services with GEANT/eduGAIN and social networks,

- To upgrade the SeaDataCloud monitoring service.

WP9 - Developments of <u>upstream</u> services

- To upgrade the CDI Data Discovery and Access service making use of the cloud,

- To develop an online SWE ingestion service for operational observing systems,

- To expand SeaDataNet capability for handling different data types,

- To integrate external datasets from international programmes and organisations,

- To develop a solution for a coordinated distributed DataCite DOI minting service.

WP10 - Developments of <u>downstream</u> services

To expand the range of services of the SeaDataNet infrastructure by specifying, developing and deploying a Virtual Research Environment (VRE)

- with advanced e-services to facilitate individual and collaborative research by using, handling, curating, quality controlling, transforming and processing marine and ocean data into value-added analyses, harmonised data collections, and data products
- which can be integrated, visualised and published using OGC and high level visualisation services.

Added-value services
and applications

WP10
Downstream
Services

WP8
Standards &
Vocabularies

make it work!

SeaDataCloud

WP9
Upstream
Services

Discovery and access
to more datasets and
information

sdn-userdesk@seadatanet.org – ww

Main change for improvement: Upgrading the CDI service using the cloud

- To configure and maintain a cloud environment to host **copies of data resources**

- Exchange by dynamic **replication** from the individual data centres, following their updating of the CDI catalogue service

Main change for improvement: Upgrading the CDI service using the cloud

- In the cloud buffer:

  - checking possible duplicates

  - Checking overall quality of formats

  - 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.
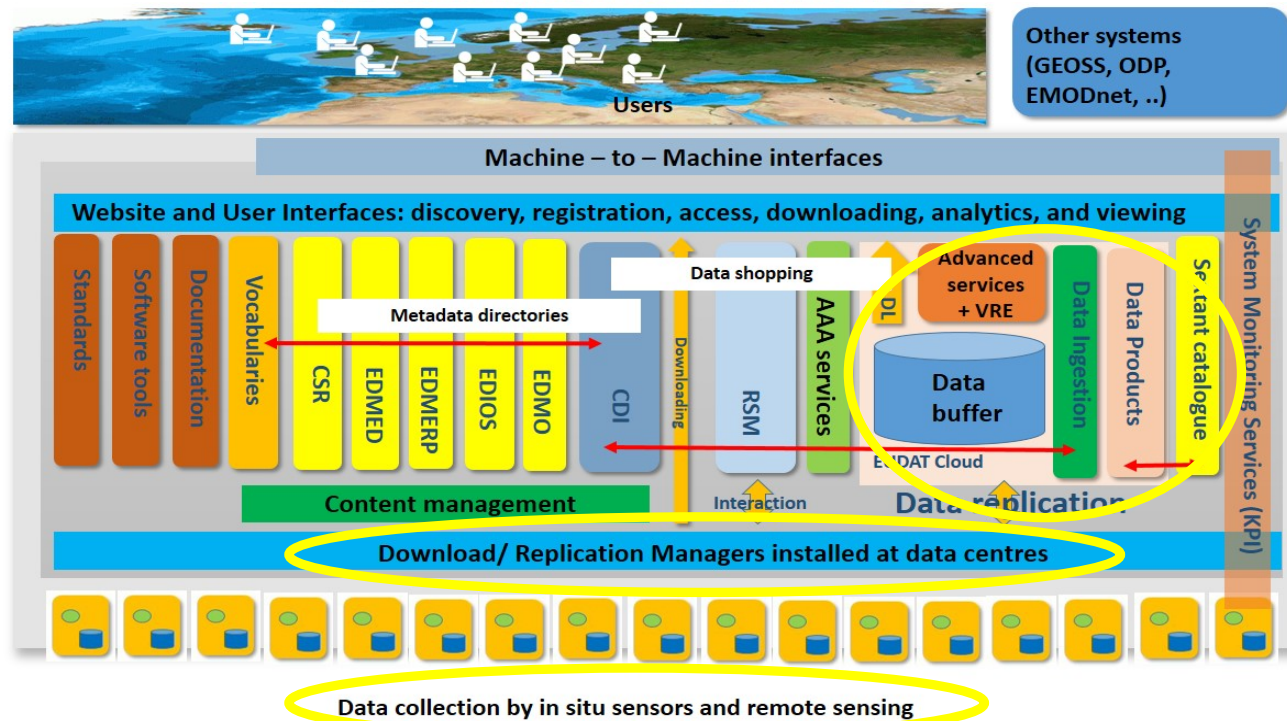
Main change for improvement: Upgrading the CDI service using the cloud

- Include **transformation services** for converting data sets to other required output formats such as SeaDataNet NetCDF and relevant INSPIRE data models.

Present SeaDataNet architecture
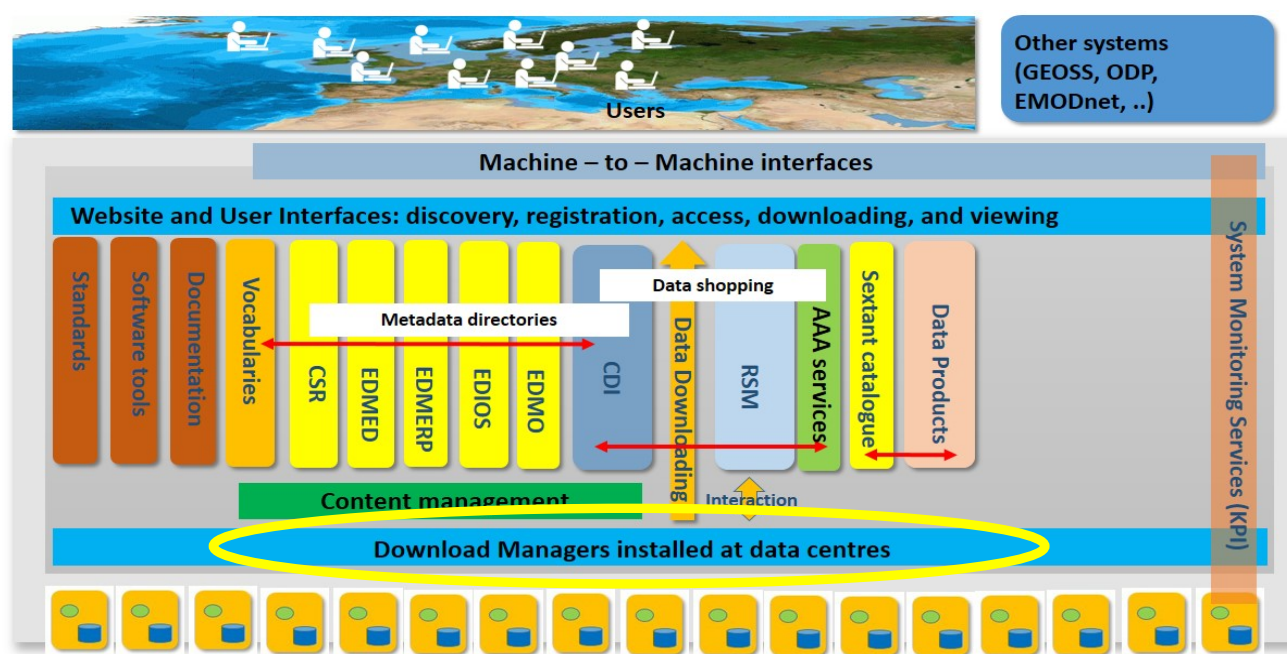
Proposed upgraded architecture with data replication, advance services and VRE in the cloud

Reasons for success?

- Strong motivation of partners, based on people more than on organizations (low concurrence, high collaboration)

- Wise development planning and pace

- Interoperability at various levels

Useful links

- SeaDatanet: www.seadatanet.org

- EMODnet: www.emodnet.eu

- ODIP: www.odip.org

Thank you for your attention!

Questions?