



SeaDataCloud, further developing the SeaDataNet pan-European infrastructure for marine and ocean data management

By

Dick M.A. Schaap – Technical Coordinator SeaDataCloud

Hobart – Australia, 7th March 2017, 7th ODIP II Workshop

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

2006-2011	Metadata directories MEDAR/MedAtlas
2002-2005	Sea-Search (FP5)
2006-2011	SeaDataNet (FP6)
2011-2015	SeaDataNet II (FP7)
2016-2020	SeaDataCloud (H2020)

SeaDataCloud – a new opportunity

- Standards are always evolving and the SeaDataNet network must stay up-to-date to maintain and further expand its services to its lead customers and major stakeholders:
 - the European and international ocean and marine research community
 - Copernicus Marine Environmental Monitoring Service (CMEMS)
 - the European Marine Observation and Data network (EMODnet) portals by which SeaDataNet services and output are used for serving stakeholders in the Marine Strategy Framework Directive (MSFD) and actors in the marine and maritime economy
 - large marine observing networks (EuroGOOS, AtlantOS, Euro-ARGO, EMSO, JERICO-Next, and others).
- The **SeaDataCloud project** (10 Meuro) takes place in the framework of the EU HORIZON 2020 programme for further developing the SeaDataNet infrastructure and associated standards in the coming 4 years. (61 members from 32 countries; start 1 November 2016)

- SDC is about updating and further developing standards
- SDC is about improving and innovating services & products
- SDC is about adopting and elaborating new technologies
- SDC is about giving more attention to users and putting the user experience in a central position
- Moreover, it is about implementing a strategic and operational cooperation between the SeaDataNet consortium of marine and ocean data centres and the EUDAT consortium of e-infrastructure service providers



SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE
FOR OCEAN & MARINE DATA
MANAGEMENT

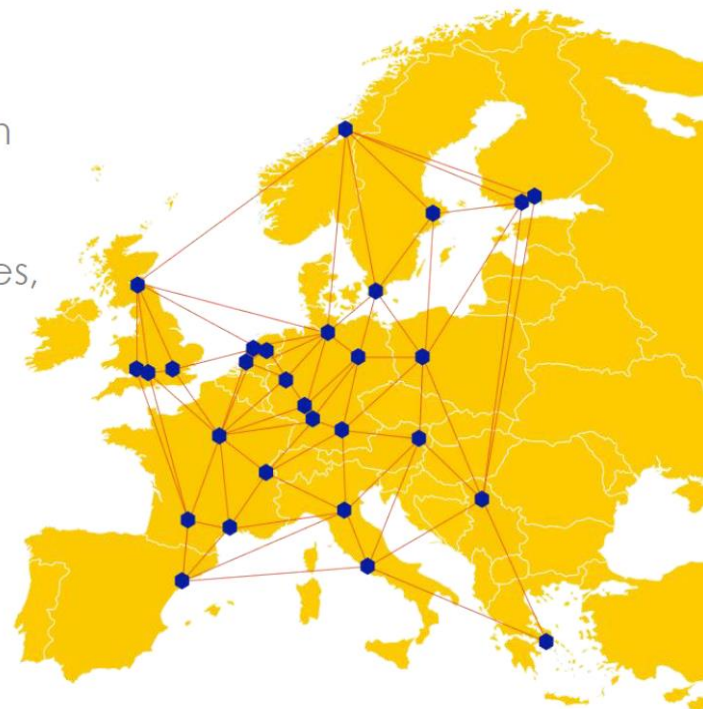
SeaDataCloud – cooperation with EUDAT



European Data Infrastructure



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



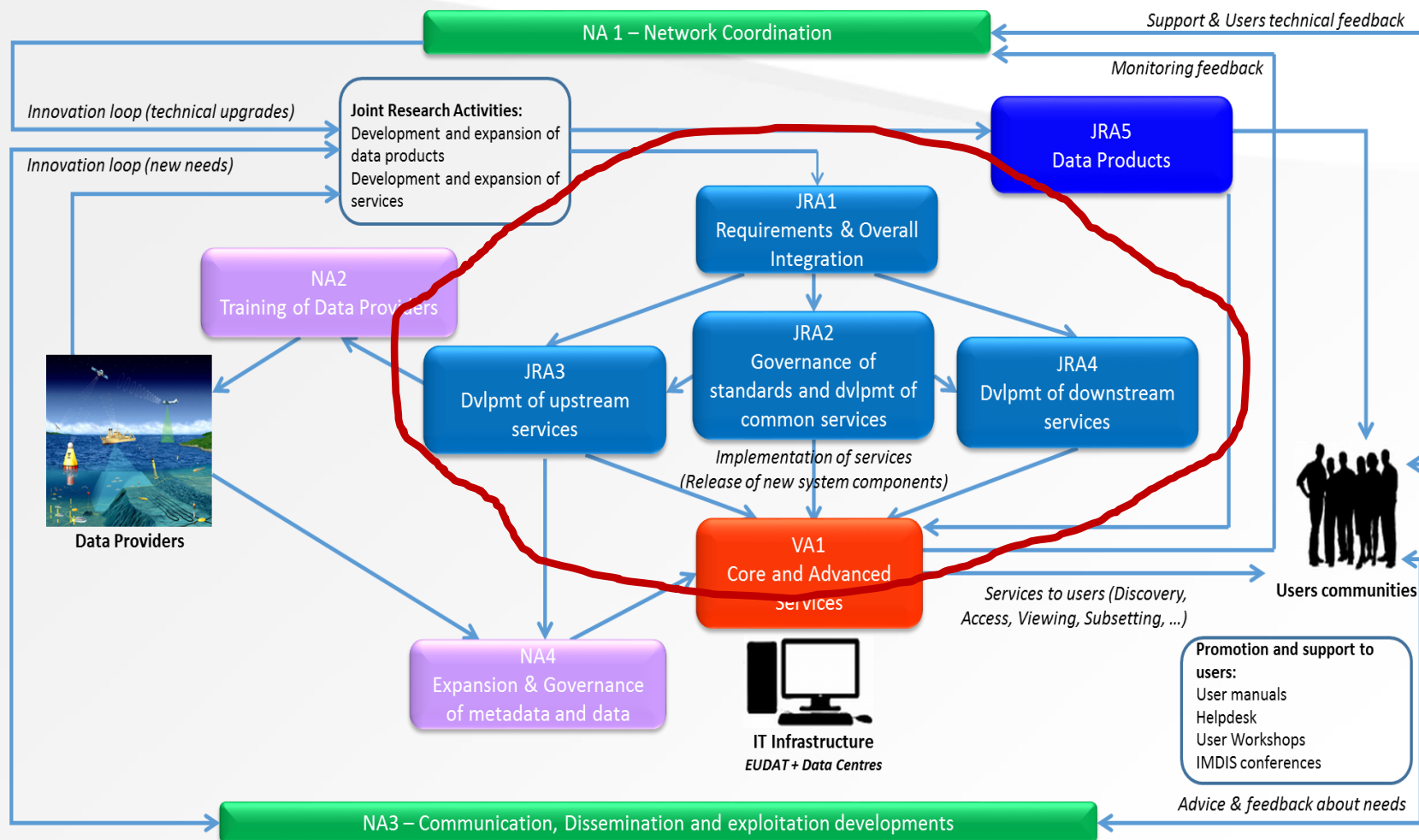


SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE
FOR OCEAN & MARINE DATA
MANAGEMENT

SeaDataCloud Pert diagram

WP1 – Project Management

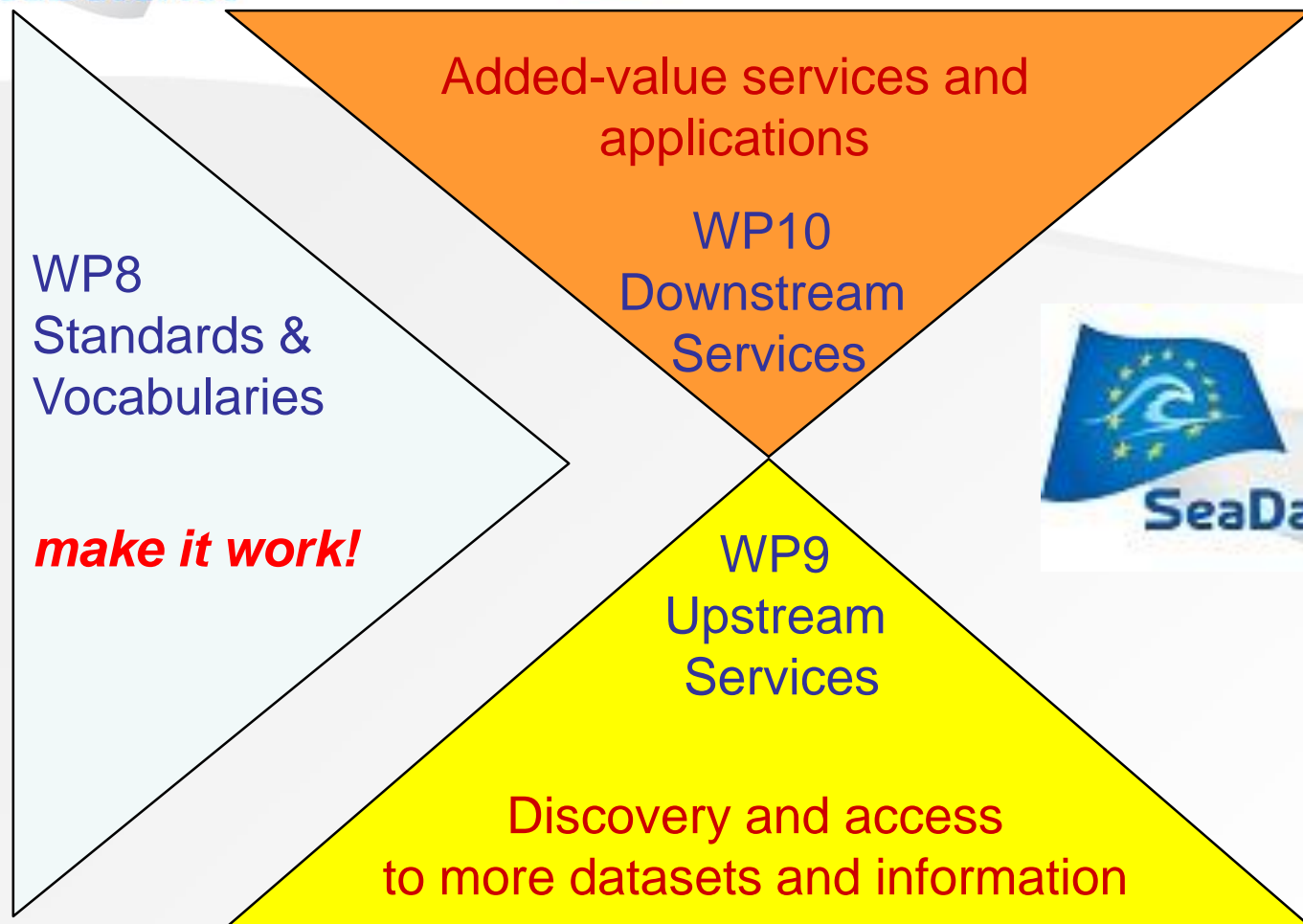


sdn-userdesk@seadatanet.org – www.seadatanet.org



SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE
FOR OCEAN & MARINE DATA
MANAGEMENT



JRA2 - 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 AAI services with GEANT/eduGAIN and social networks,
- To upgrade the SeaDataCloud monitoring service.

JRA3 - Developments of upstream 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.

JRA4 - Developments of downstream service

- 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.

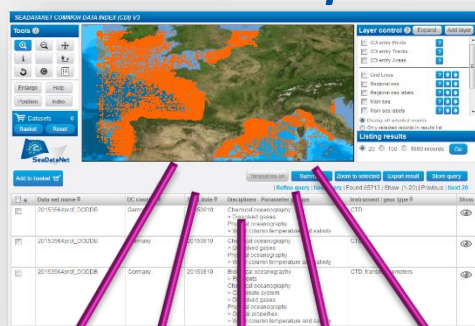


SeaDataNet

PAN-EUROPEAN INFRASTRUCTURE
FOR OCEAN & MARINE DATA
MANAGEMENT

CDI service for discovery and unified data access

SeaDataNet portal



Search
and
Shop



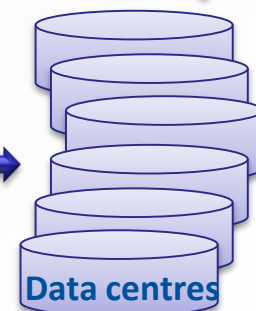
Data
add

Already > 100 data centres connected and more underway

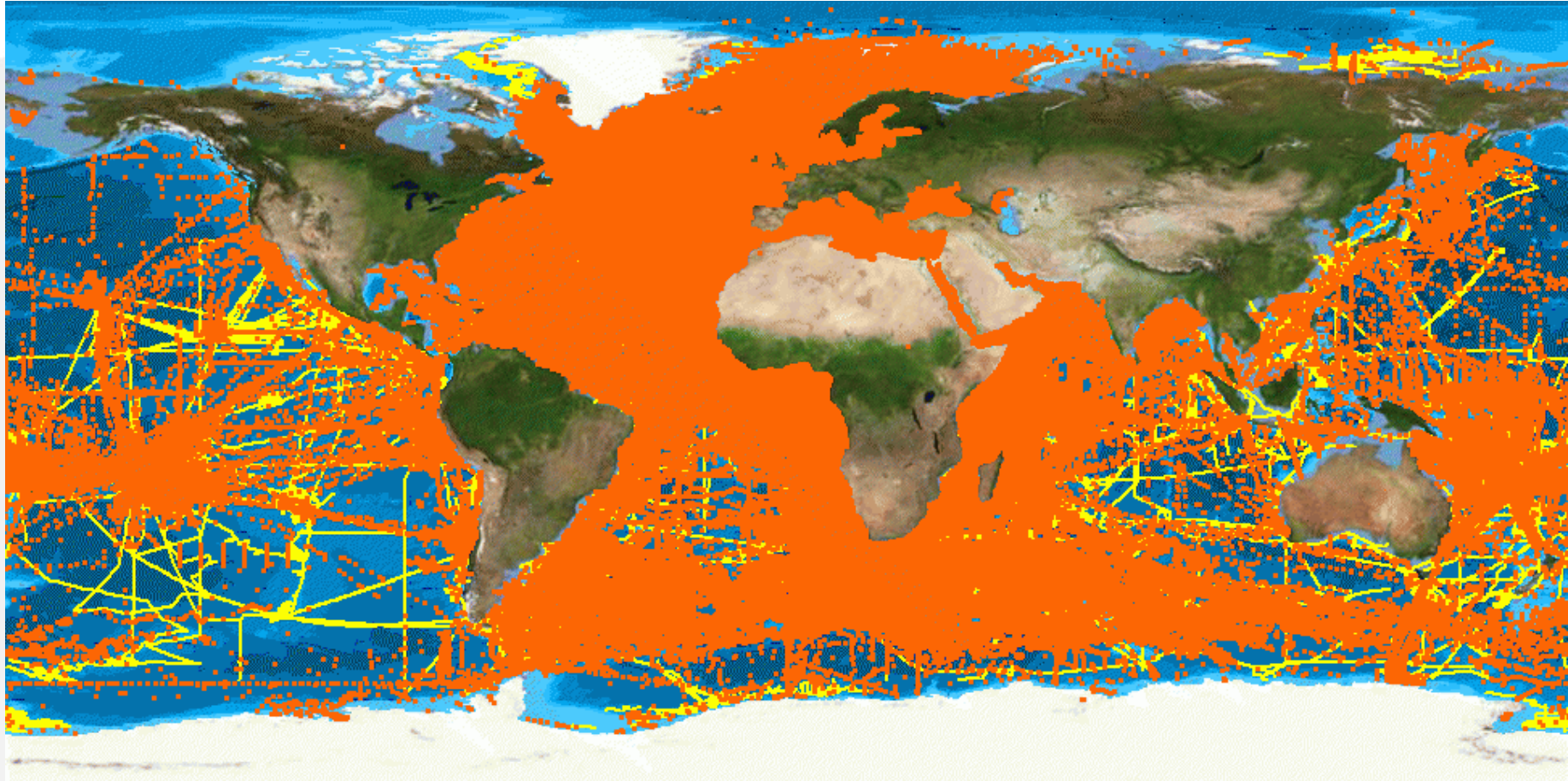


Metadata

+ transaction data



European data sources
data centres ← ≈ 600 originators



1.94 million CDI entries from ***34*** countries, ***102*** data centres and ***607*** originators for physics, chemistry, geology, geophysics, bathymetry and biology; from ***1805 to 2017***; ***86%*** unrestricted or under SDN License

Issues with present CDI service

- Population and uptake by Data Centres of the CDI service is very successful with > 100 Data Centres and > 1.9 million CDIs
- However:
 - 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.
 - there are performance issues that data centres are not always online, operational and different machine capacities; this gives extra delays
 - there are 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

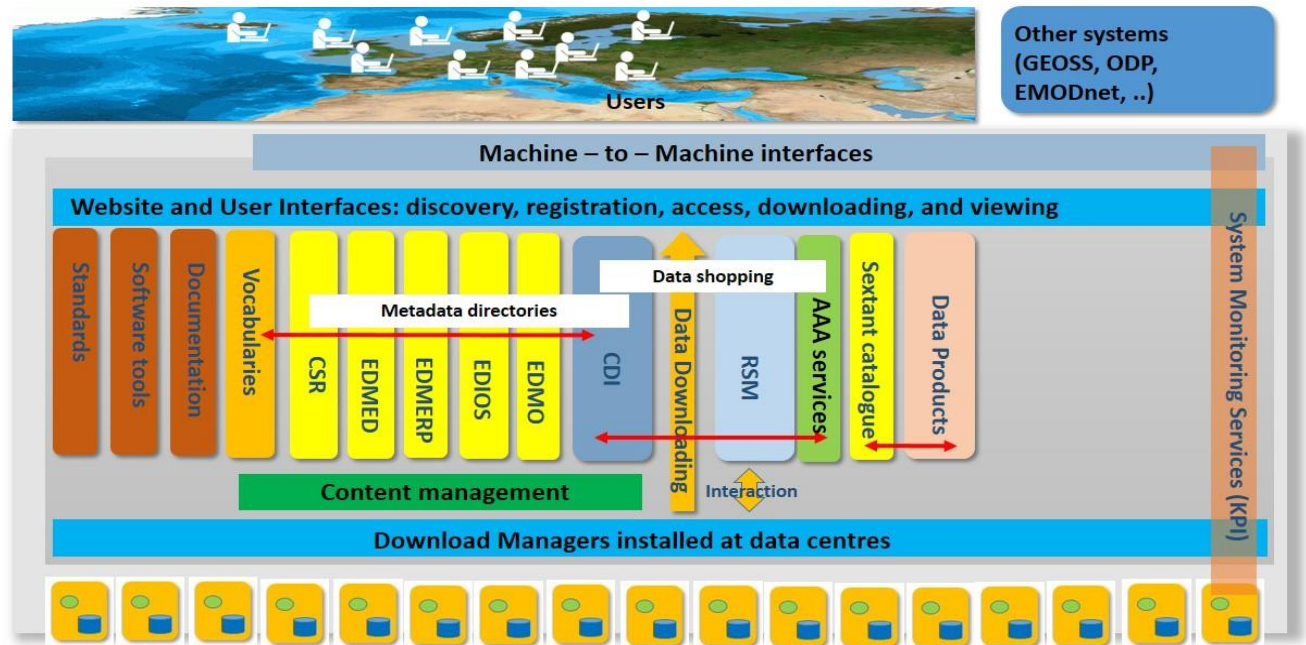
Upgrading CDI service using the cloud

- To configure and maintain a cloud environment as a 'cache' to host **copies of all data resources**
- Exchange by dynamic **replication** from the individual data centres, following their updating of the CDI catalogue service
- In the cloud buffer new functions:
 - 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.
- Include **transformation services** for harmonizing data sets to common parameters and units, and converting data sets to other required output formats such as SeaDataNet NetCDF and relevant INSPIRE data models.

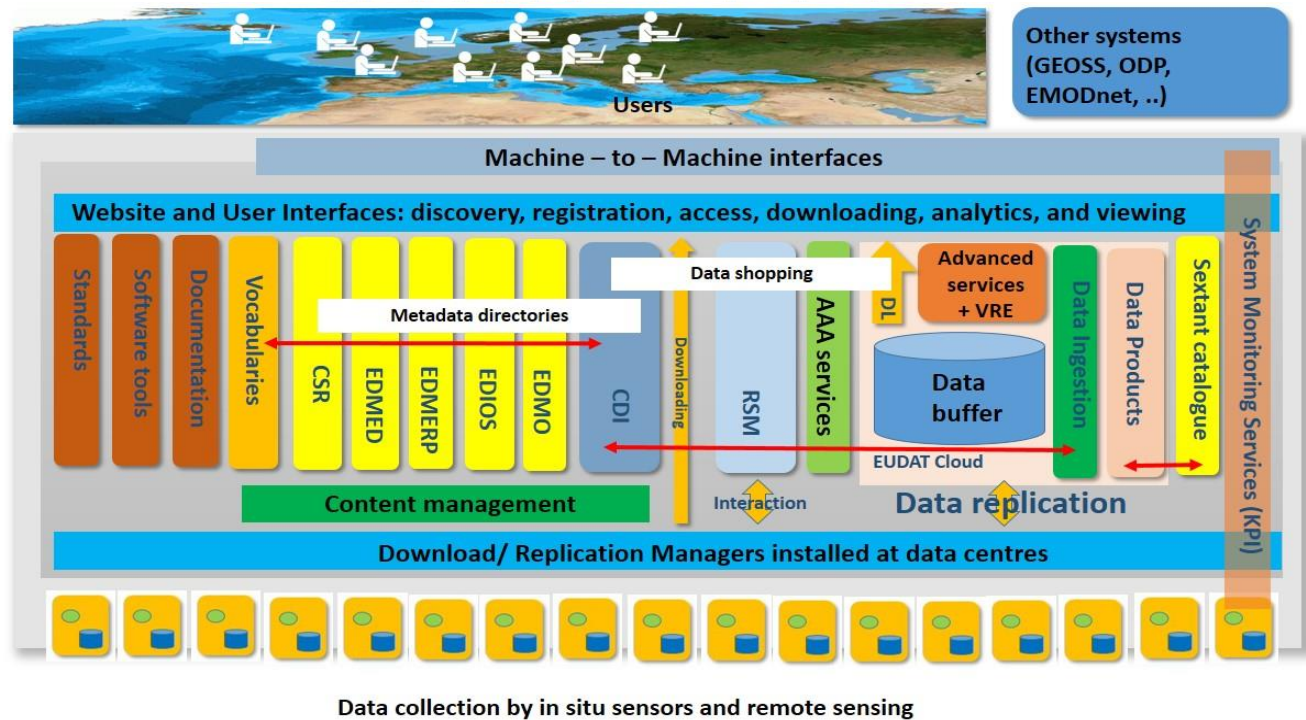
Benefits for CDI service and its users

- The cloud buffer in combination with the CDI service will speed up the performance, expand discovery and ease of use of the data access and downloading and will provide users with one integrated download package instead of multiple packages from multiple data centres.
- Overall quality and coherence (data – metadata) will improve
- Tracking and tracing of transactions will continue to be administered by the upgraded RSM service, so that data providers can oversee all relevant transactions for their data centre and users can oversee their requests and deliveries.
- Data replication will be triggered per data centre by CDI updates. The **replication module** might have less complexity than the present Download Manager module
- A system of **versioning** will be introduced which is required in the context of the MSFD for facilitating repeated analysis of environmental assessments after many years, and for scientific papers.

Present SeaDataNet architecture



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





PAN-EUROPEAN INFRASTRUCTURE
FOR OCEAN & MARINE DATA
MANAGEMENT



www.emodnet.eu



www.seadatanet.org



www.odip.org