SeaDataCloud

Progress of the SeaDataCloud project, its data discovery and access services and linked data

Michele Fichaut (IFREMER, FR - Coordinator), Peter Thijsse, Dick M.A. Schaap (MARIS, NL, Tech coordinator)

EGU 2018, Vienna, Austria, 10 April 2018

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

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

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 sets, research cruises, monitoring networks, and research projects => marine metadata profiles, schemas, schematron rules
  - **Controlled vocabularies** for the marine domain (>65,000 terms in 82 lists), with international governance and web services
  - **Standard data exchange formats** : ODV ASCII and NetCDF (CF) fully supported by controlled vocabularies
- Maintenance and dissemination of standard QA-QC procedures, together with IOC/IODE and ICES

SeaDataNet services and tools

- **Set of tools** to be used each data centre and freely available from the SeaDataNet portal: metadata editor, data conversion software, data analysis software (ODV), data interpolation software (DIVA)

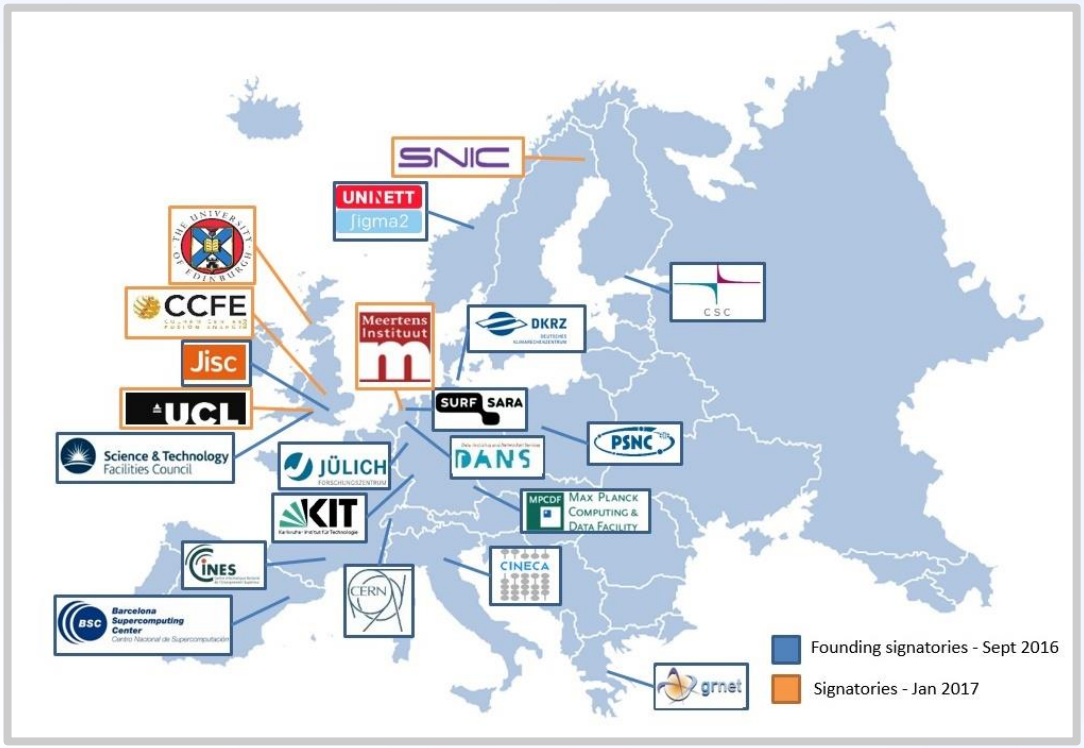
- **Capacity building** by training workshops for uptake of standards and tools by the data centres in order to achieve standardisation

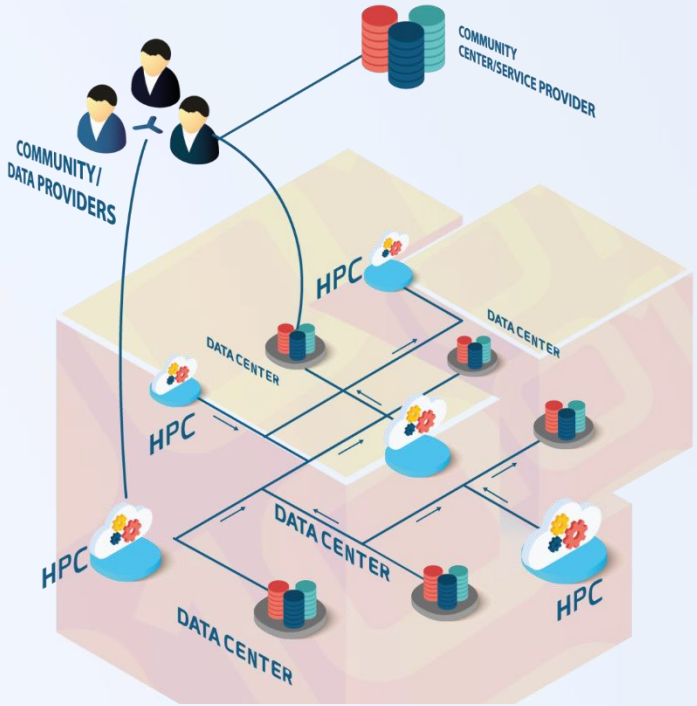
- **Pan-European services** for harmonised discovery, access, visualisation of data and data products
- Common SeaDataNet **Data Policy** and License

SeaDataCloud a new opportunity

- Standards and information technology are always evolving, there is a move **towards cloud storage and cloud computing (VRE)**, and the SeaDataNet infrastructure must stay up-to-date to maintain and further **expand its services for data access and data discovery** to its lead customers and major stakeholder
- SeaDataCloud project, started Nov 2016 with 4 year run and 10 Meuro funding
- A strategic and operational **cooperation between the SeaDataNet consortium** of marine and ocean data centres **and the EUDAT consortium** of e-infrastructure service providers

Cooperation with EUDAT





5 EUDAT members are partners of SeaDataCloud :
CINECA, CSC, DKRZ, GRNET and STFC

IMPROVED DATA DISCOVERY: LINKED DATA

Taken from: rob Thomas, Adam Leadbetter (MI), Alexandra Kokkinaki (BODC), et al.

Pan-European metadata directories

EDMERP
Projects

EDMO
Organisations

CSR
Research cruises

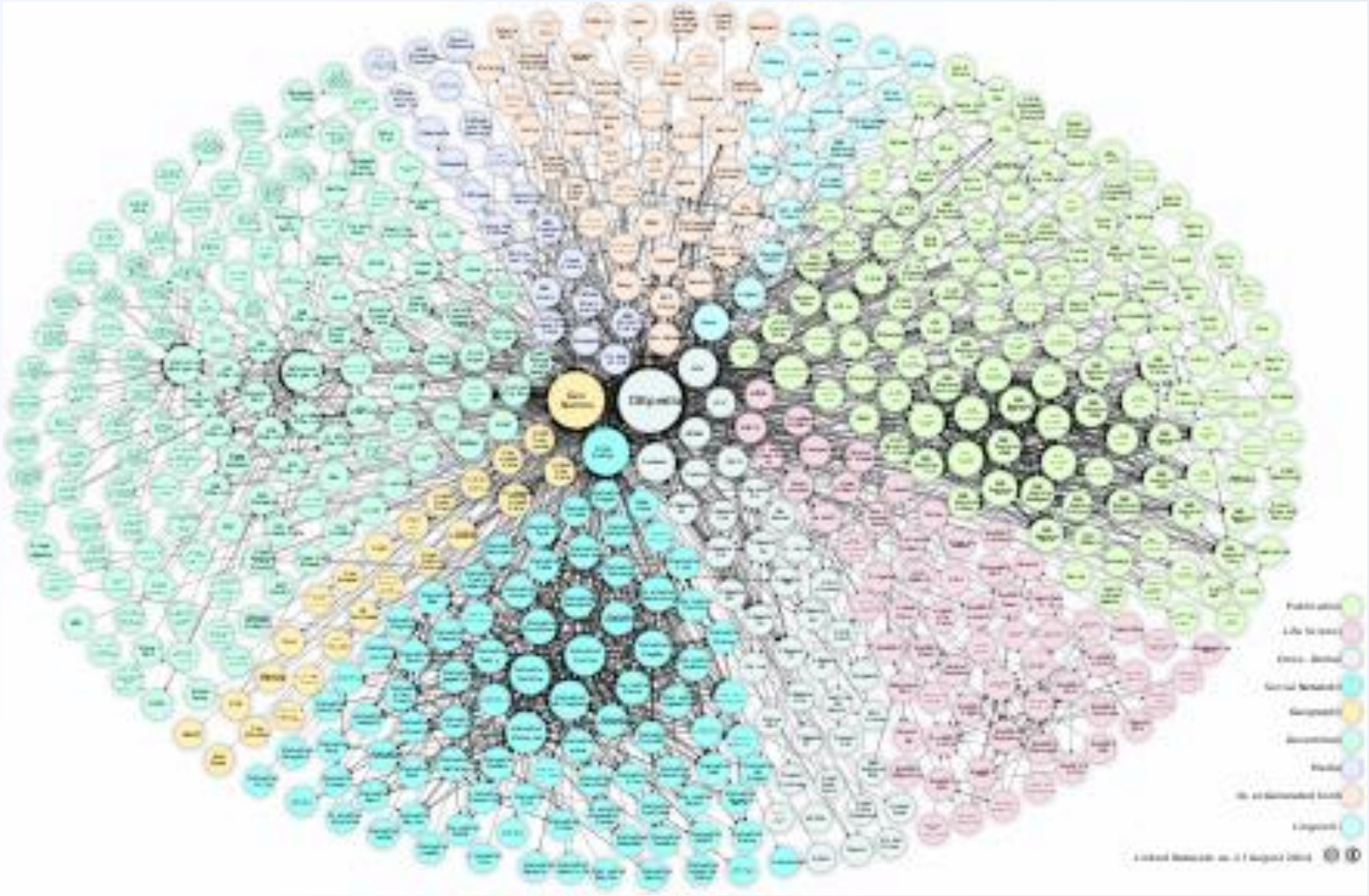
EDIOS
Observing programmes

CDI ***Data index***

EDMED
Data sets

Why Linked Data for SeaDataNet?

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



https://commons.wikimedia.org/wiki/File:LOD_Cloud_2014-08.svg#/media/File:LOD_Cloud_2014-08.svg

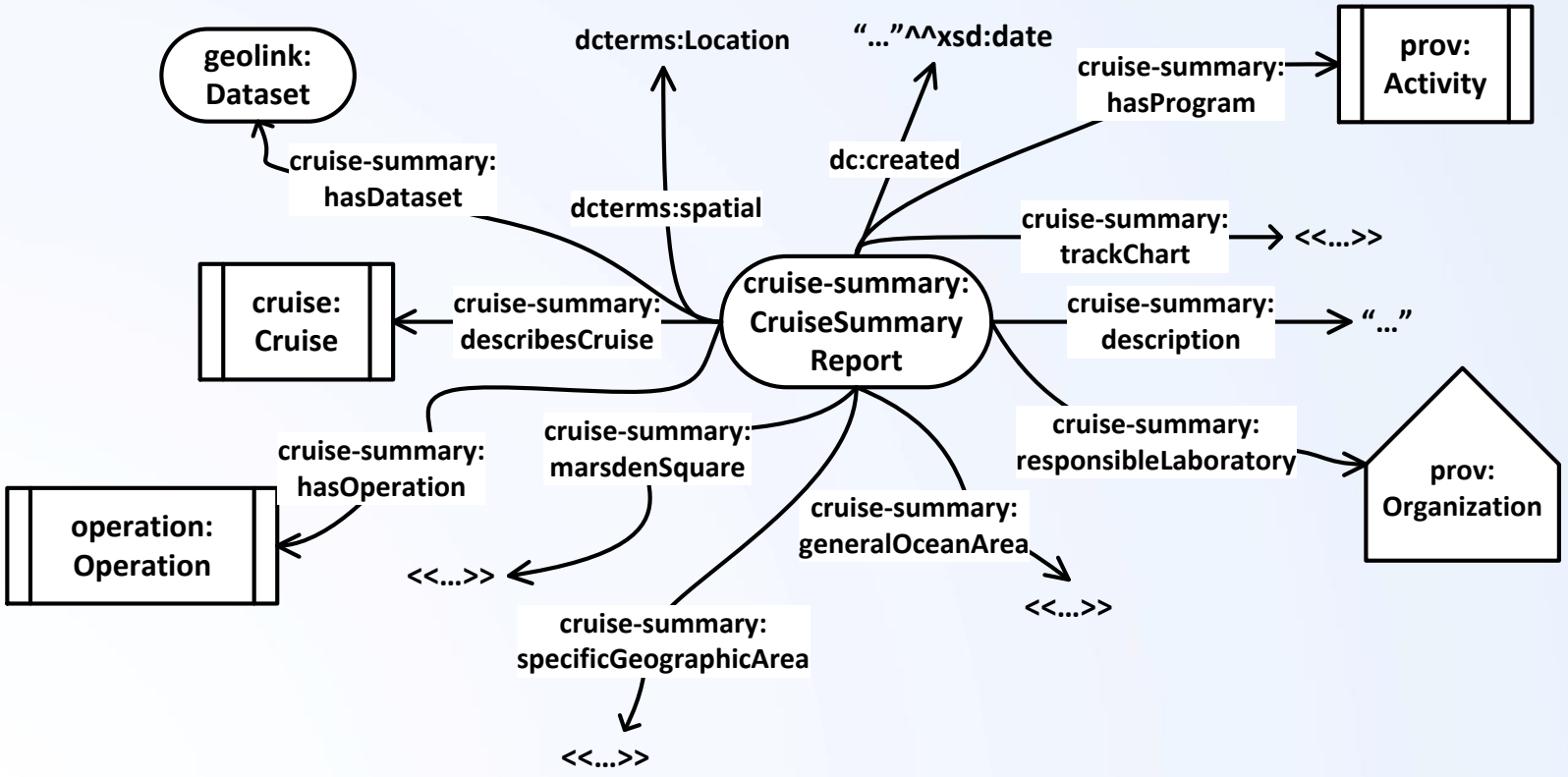
SeaDataCloud

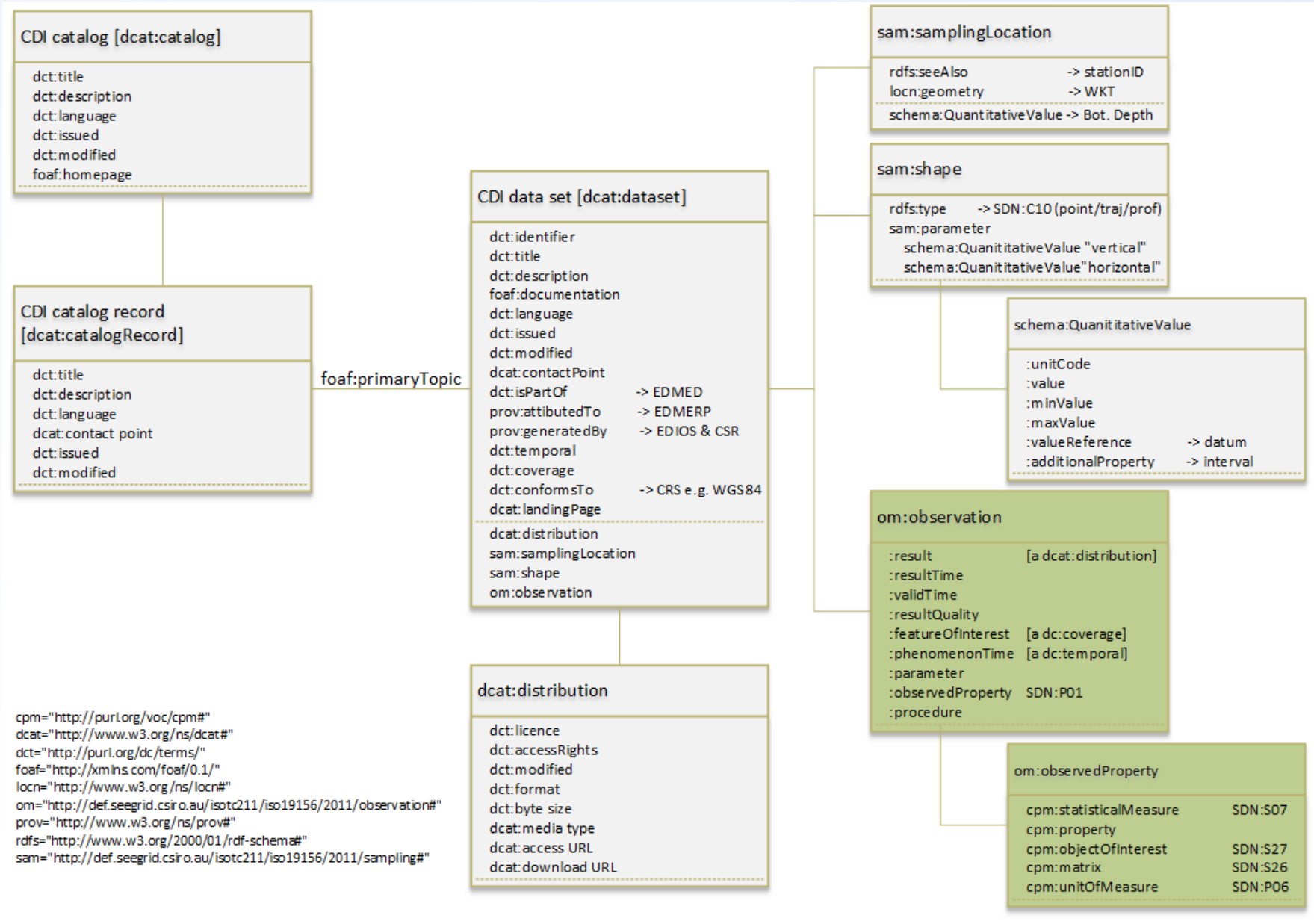
How we approach it

Reusing existing patterns:

- EDMO – W3C Organisation
- EDMED – W3C DCAT
- EDMERP – W3C Prov / DBPedia Research Project
- CDI – W3C DCAT
  - ODV metadata to INSPIRE / ISO O&M
- CSR – new patterns
- EDIOS – INSPIRE Environmental Monitoring Facilities

CSR Patterns



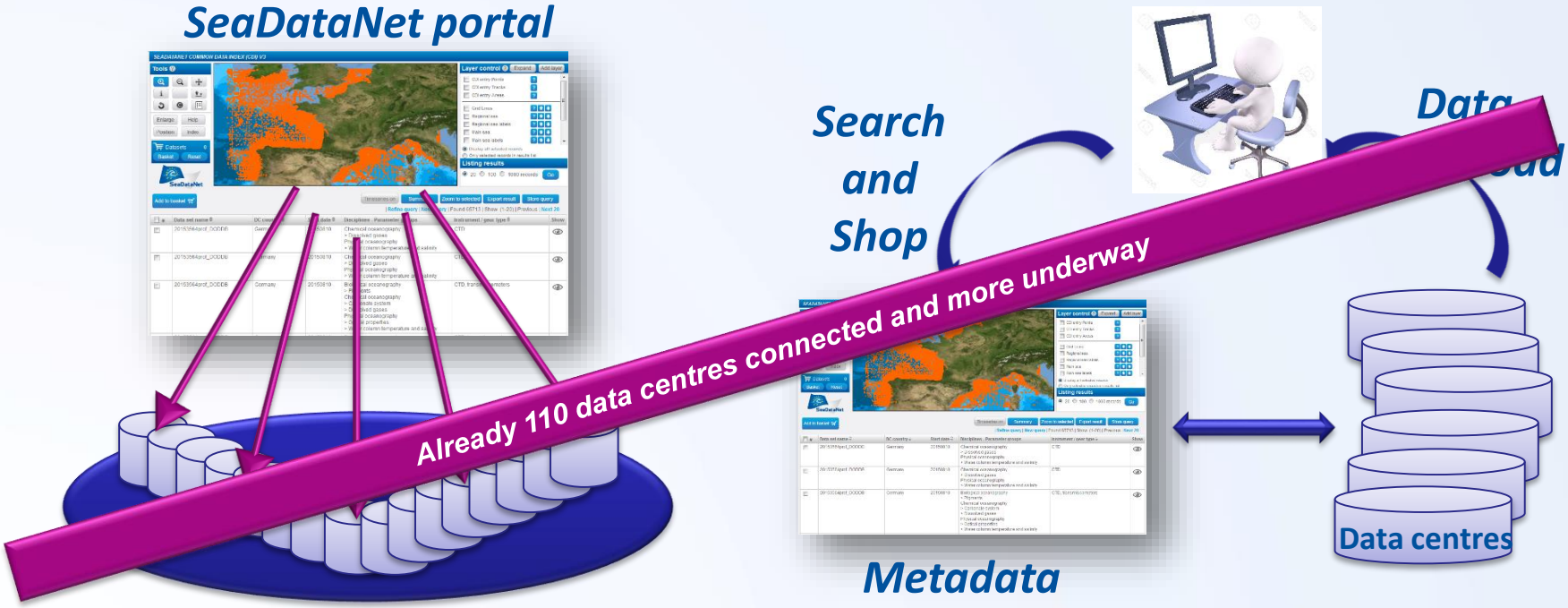

CDI catalog [dcat:catalog]
dct:title
dct:description
dct:language
dct:issued
dct:modified
foaf:homepage
CDI catalog record [dcat:catalogRecord]
dct:title
dct:description
dct:language
dcat:contact point
dct:issued
dct:modified
foaf:primaryTopic
CDI data set [dcat:dataset]
dct:identifier
dct:title
dct:description
foaf:documentation
dct:language
dct:issued
dct:modified
dcat:contactPoint
dct:isPartOf -> EDMED
prov:attibutedTo -> EDMERP
prov:generatedBy -> EDIOS & CSR
dct:temporal
dct:coverage
dct:conformsTo -> CRS e.g. WGS84
dcat:landingPage
dcat:distribution
sam:samplingLocation
sam:shape
om:observation
dcat:distribution
dct:licence
dct:accessRights
dct:modified
dct:format
dct:byte size
dcat:media type
dcat:access URL
dcat:download URL
sam:samplingLocation
rdfs:seeAlso -> stationID
locn:geometry -> WKT
schema:QuantitativeValue -> Bot. Depth
sam:shape
rdfs:type -> SDN:C10 (point/traj/prof)
sam:parameter
schema:QuanititativeValue "vertical"
schema:QuanititativeValue"horizontal"
schema:QuanititativeValue
:unitCode
:value
:minValue
:maxValue
:valueReference -> datum
:additionalProperty -> interval
om:observation
:result [a dcat:distribution]
:resultTime
:validTime
:resultQuality
:featureOfInterest [a dc:coverage]
:phenomenonTime [a dc:temporal]
:parameter
:observedProperty SDN:P01
:procedure
om:observedProperty
cpm:statisticalMeasure SDN:S07
cpm:property
cpm:objectOfInterest SDN:S27
cpm:matrix SDN:S26
cpm:unitOfMeasure SDN:P06
cpm="http://purl.org/voc/cpm#"
dcat="http://www.w3.org/ns/dcat#"
dct="http://purl.org/dc/terms/"
foaf="http://xmlns.com/foaf/0.1/"
locn="http://www.w3.org/ns/locn#"
om="http://def.seegrid.csiro.au/isotc211/iso19156/2011/observation#"
prov="http://www.w3.org/ns/prov#"
rdfs="http://www.w3.org/2000/01/rdf-schema#"
sam="http://def.seegrid.csiro.au/isotc211/iso19156/2011/sampling#"

First two SparQL endpoints ready, more to come.

Details presented in yesterdays session OS 2.4…

CDI: IMPROVED DATA DISCOVERY AND ACCESS

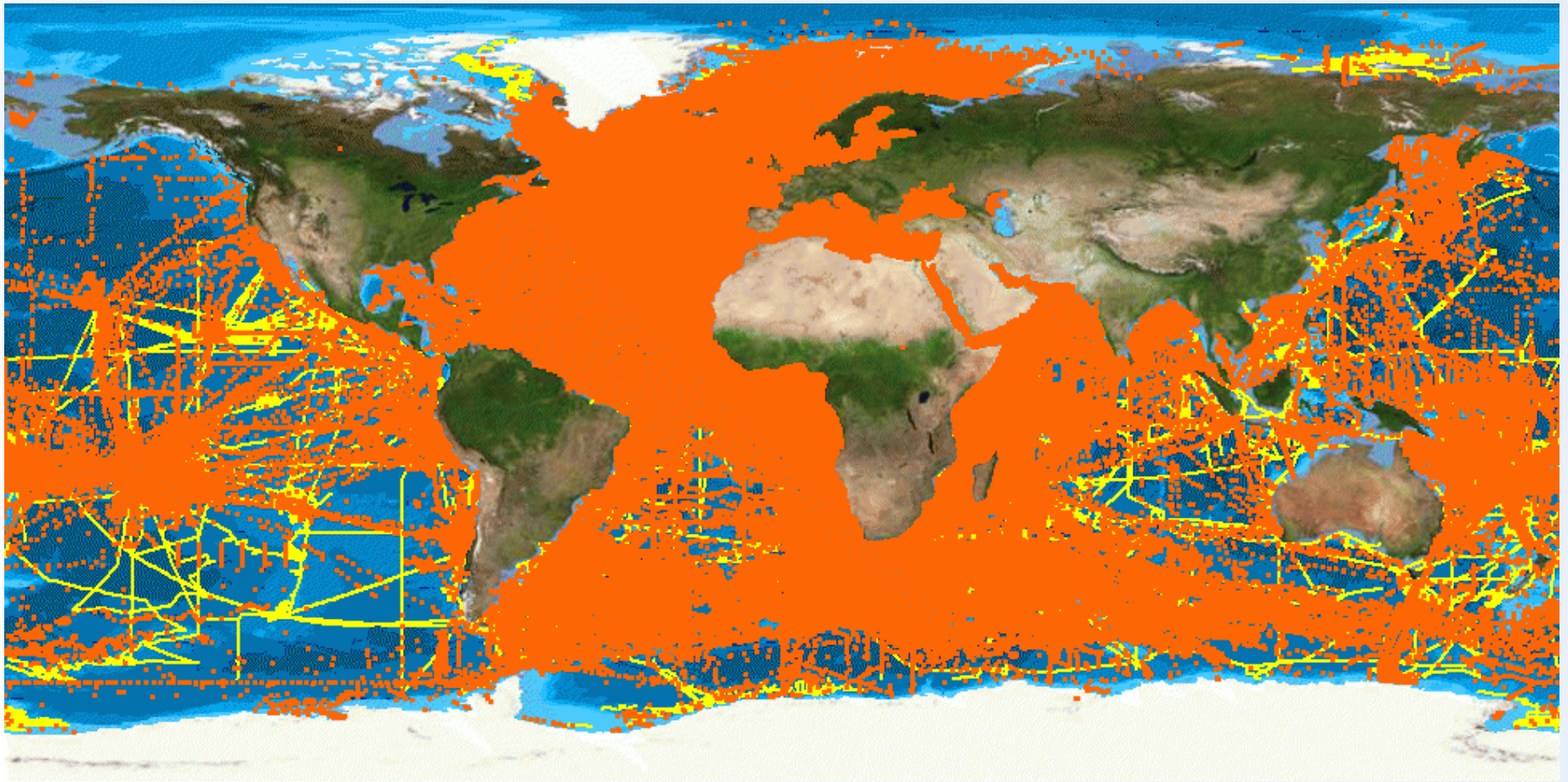
CDI Data Discovery and Access service



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

110 connected data centres

CDI data population



- **2 million** CDI entries from **34** countries, **110** data centres and > **650** originators
- physics, chemistry, geology, geophysics, bathymetry and biology;
- from **1805 to 2017**; **86%** unrestricted or under SDN License

Responding to user issues: Upgrading CDI service using the cloud

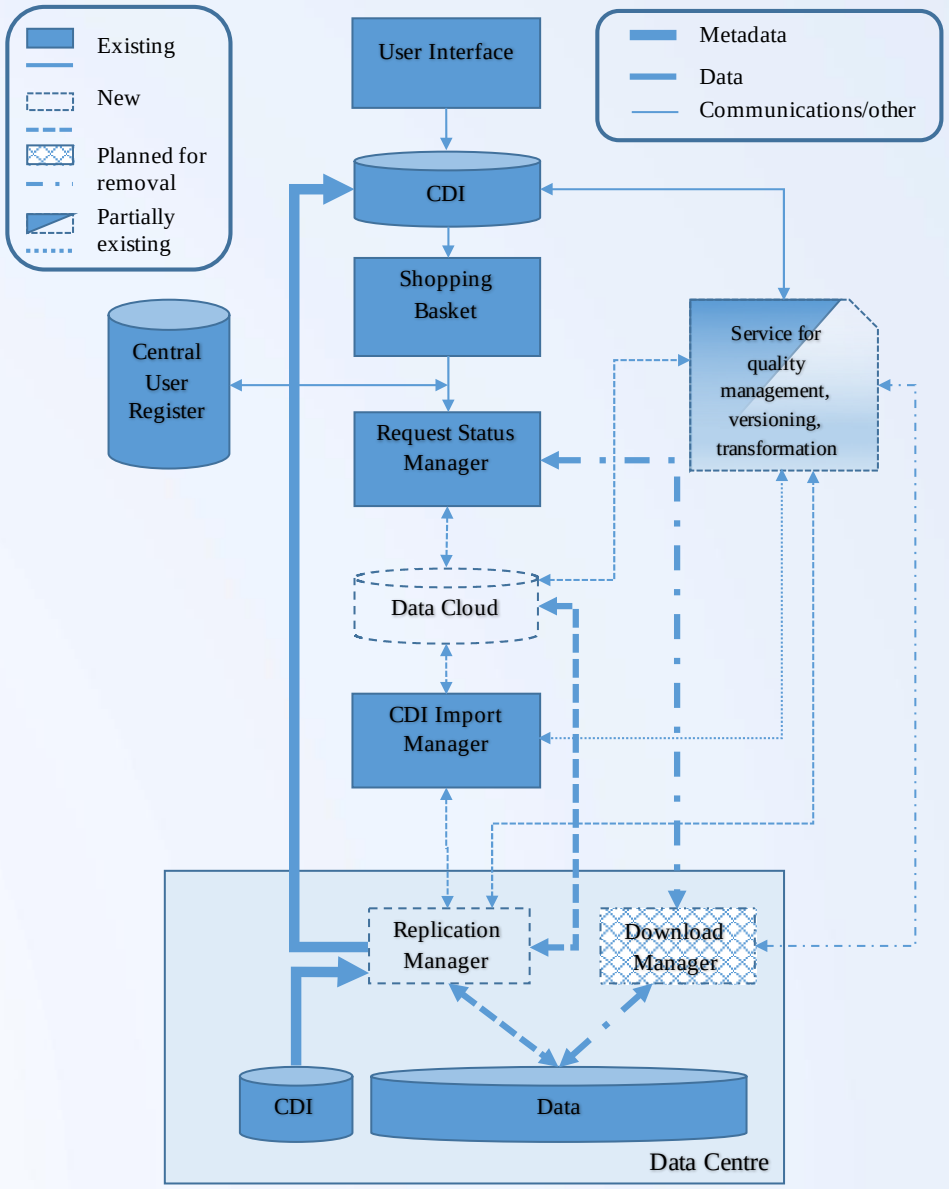
- Configure and **maintain a cloud environment as a 'cache' to host copies of all data resources** (from the distributed data centres)
- Exchange via **dynamic replication** from the individual data centres, following their updating of the CDI catalogue service
- In the cloud buffer **new QC functions**:
  - checking possible duplicates
  - Checking overall quality of formats
  - Checking integrity of data files and metadata relations.
  - Results of checks reported back to data centres for amendments of their submissions

- Develop a Virtual Research Environment (VRE) to facilitate collaborative and individual research by users
- Provide customised services (MySeaDataCloud) to let users have search profile, receive alerts on new available data, ingest and manage their own datasets

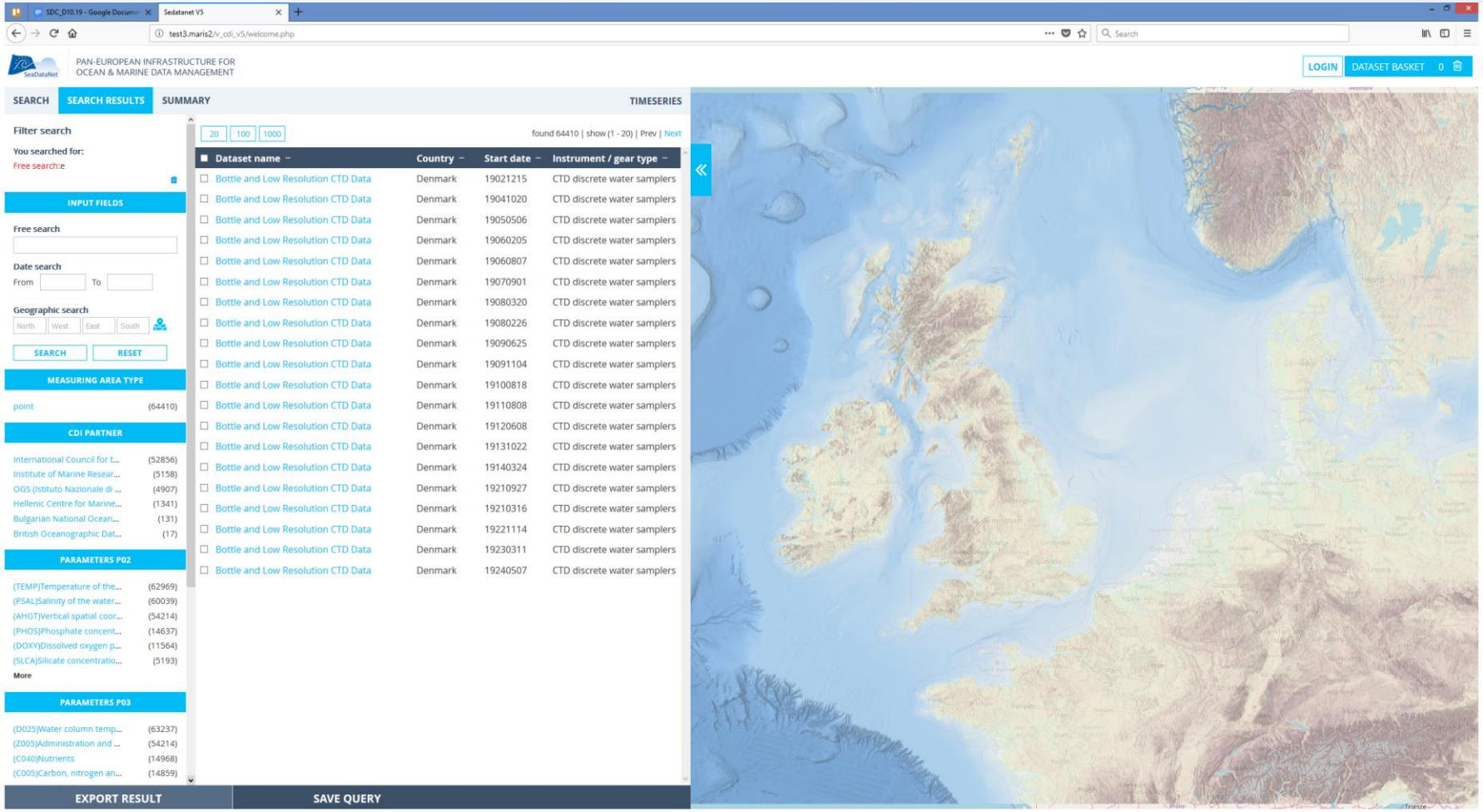
Upgrading CDI service using the cloud

SeaDataCloud

- New modern interfaces
- Transformation services
- Buffer quality control
- Data caching
- Import manager
- Replication Managers at Data nodes



And of course: New user interfaces



Benefits for CDI service and its users

- Increased download performance (less steps, faster, easier)
- Improved overall quality and coherence (data – metadata)
- Less complex replication module
- Data versioning introduced: Required a.o. in the context of the MSFD for scientific papers.

SeaDataCloud

THANK YOU!

MORE INFO:
CHECK WWW.SEADATANET.ORG
OR EMAIL PETER@MARIS.NL