



SeaDataCloud

Upgrading CDI Data Discovery and Access service

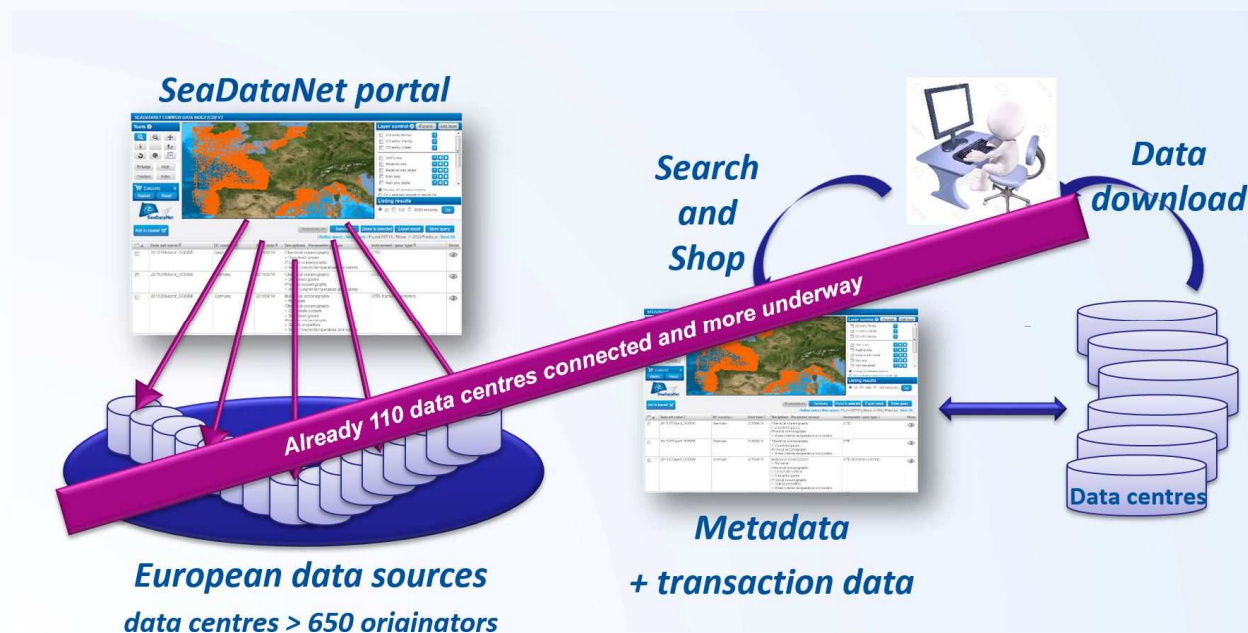
Dick M.A. Schaap (MARIS, NL, Technical coordinator)

SDC Plenary Meeting, 8 November 2018, Barcelona - Spain

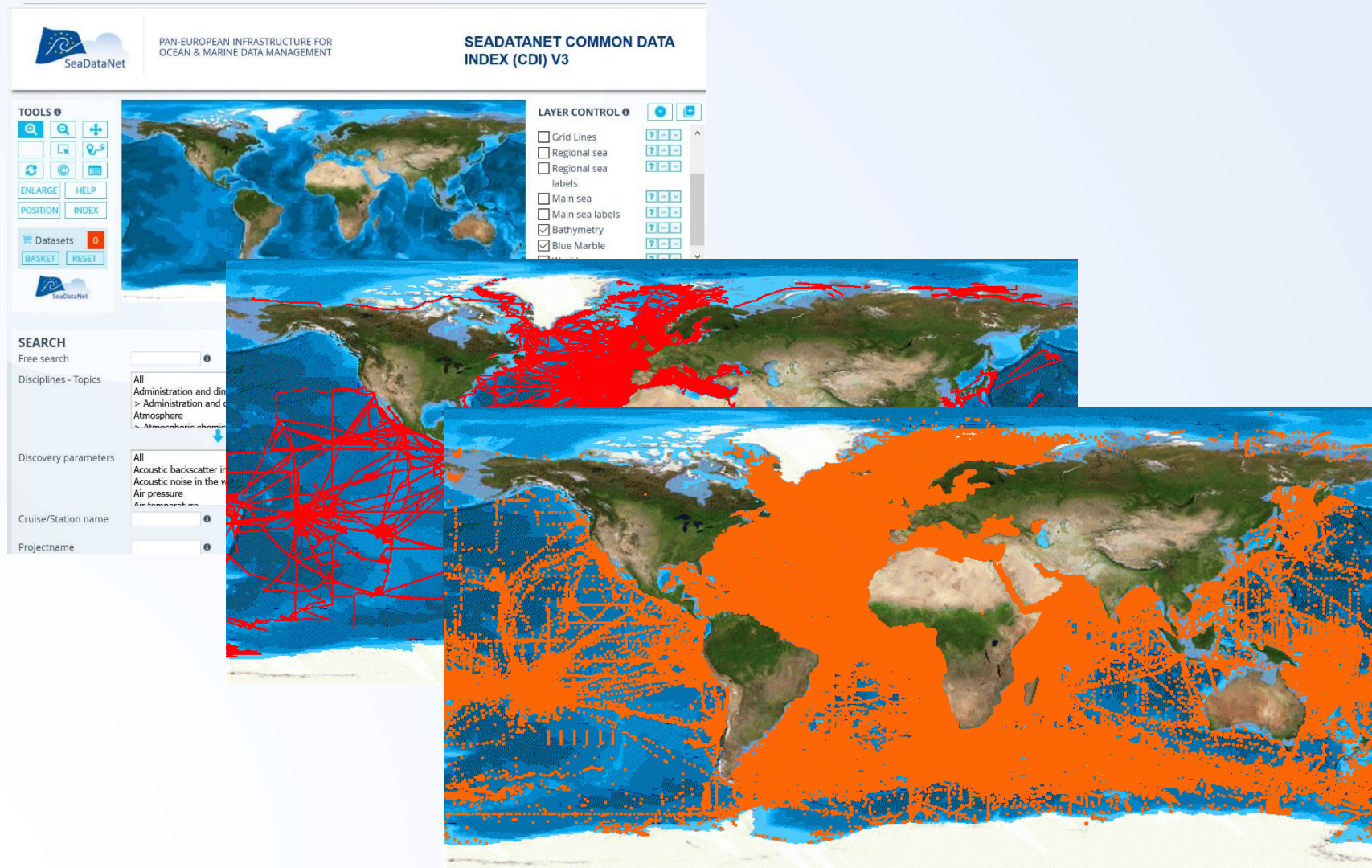
sdn-userdesk@seadatanet.org – www.seadatanet.org

CDI Data Discovery and Access service

- It is one of the core services of the SeaDataNet infrastructure
- It provides a highly detailed insight and unified access to the large volumes of marine and oceanographic data sets managed by the distributed data centres
- It is a fine-grained index (ISO 19115 – ISO 19139) to individual data measurements (such as a CTD cast or moored instrument record)

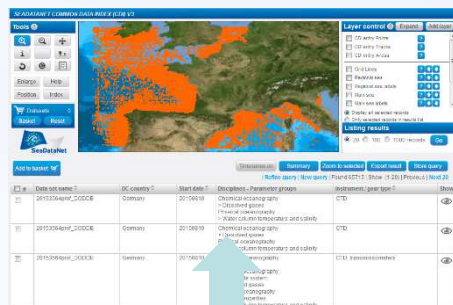


CDI service with global coverage



> 2.25 Million CDI entries for physics, chemistry, biology, geology and geophysics

CDI service contributes to many portals



Data discovery and access



> 110 data centres



NODCs; HO; GEOs; BIOS; ICES; PANGAEA

≈ 650 European data originators

CDI Data Discovery and Access service



GEOSS portal



IODE ODP portal

Aggregated collection



Black Sea portal



Caspian portal

Regional subsets

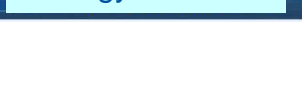
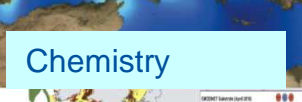


Geo-Seas portal

Thematic portals



Bathymetry



Installed base of CDI nodes (> 110)





Issues with current CDI service

- **performance for users:** CDI data access service interacts with the distributed data collections and databases at the connected data centres.
 - user can submit a shopping basket with requests for data from multiple data centres.
 - user must await the automatic data preparation by each of these data centres
 - user must download resulting data sets through the RSM as packages directly from each data centre, which implicates multiple download transactions
- **performance for users:** data centres are not always online, operational and have different machine capacities which might give extra delays
- **quality issues:** concerning formats of data files (ODV + NetCDF) and their consistency with CDI metadata.
- **installation and configuration** of the Download Manager software can be challenging due to different configurations, firewalls etc., which in practice results in having different versions installed



Upgrading the CDI service using the cloud

- To configure and maintain a **CLOUD buffer** to host **copies of unrestricted data resources**
- Exchange by dynamic **replication** from the individual data centres, following their updating of the CDI catalogue service
- In the cloud buffer:
 - checking overall quality of metadata and data, as extra check on top of local QA-QC by data centres
 - checking integrity of data files and metadata relations.
 - results of checks to be reported back to data centres for corrections
- Include transformation services for converting data sets to SeaDataNet ODV and NetCDF formats and relevant INSPIRE data models.
- Introduce versioning of metadata and data as part of provenance



New CDI service components

- **Local software tools** at data centres to prepare ingestions
- **Replication Manager** (RM) at data centres for exchanging to Import Manager and EUDAT cloud
- **Import Manager dashboard** to steer import and validation process
- **EUDAT cloud** with adapted EUDAT services
- Upgraded **CDI User Interface**, ordering and downloading facility

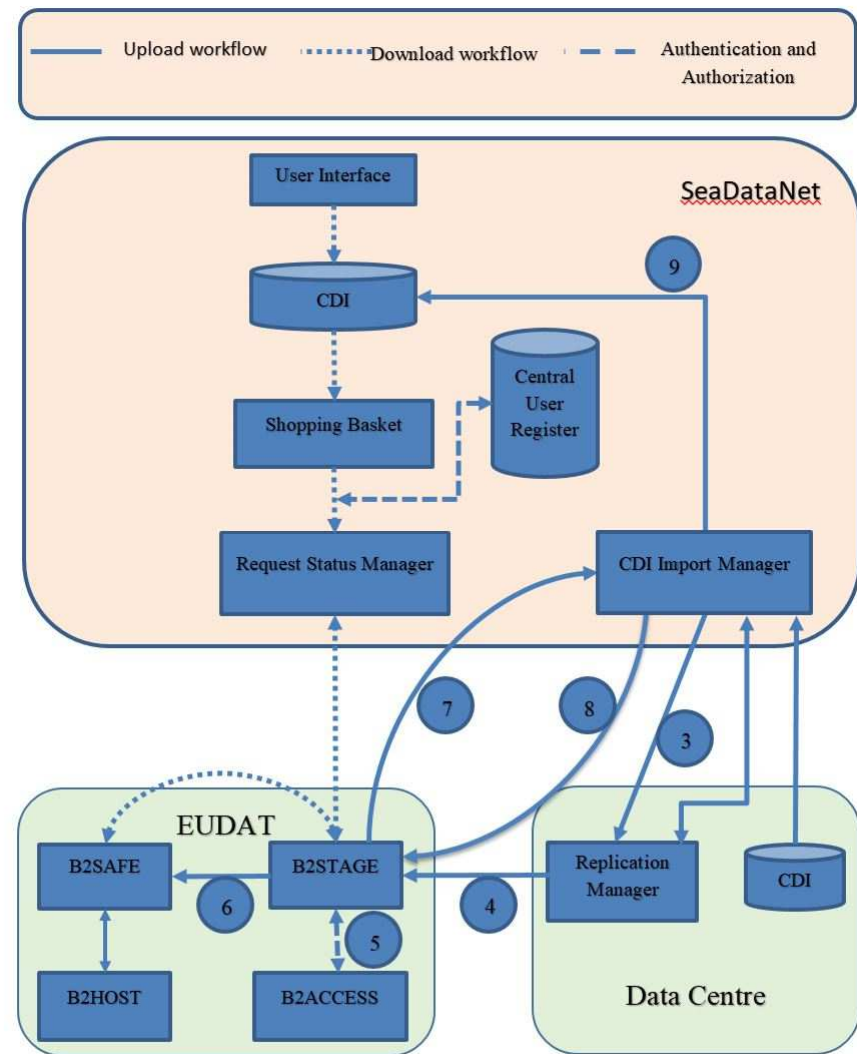
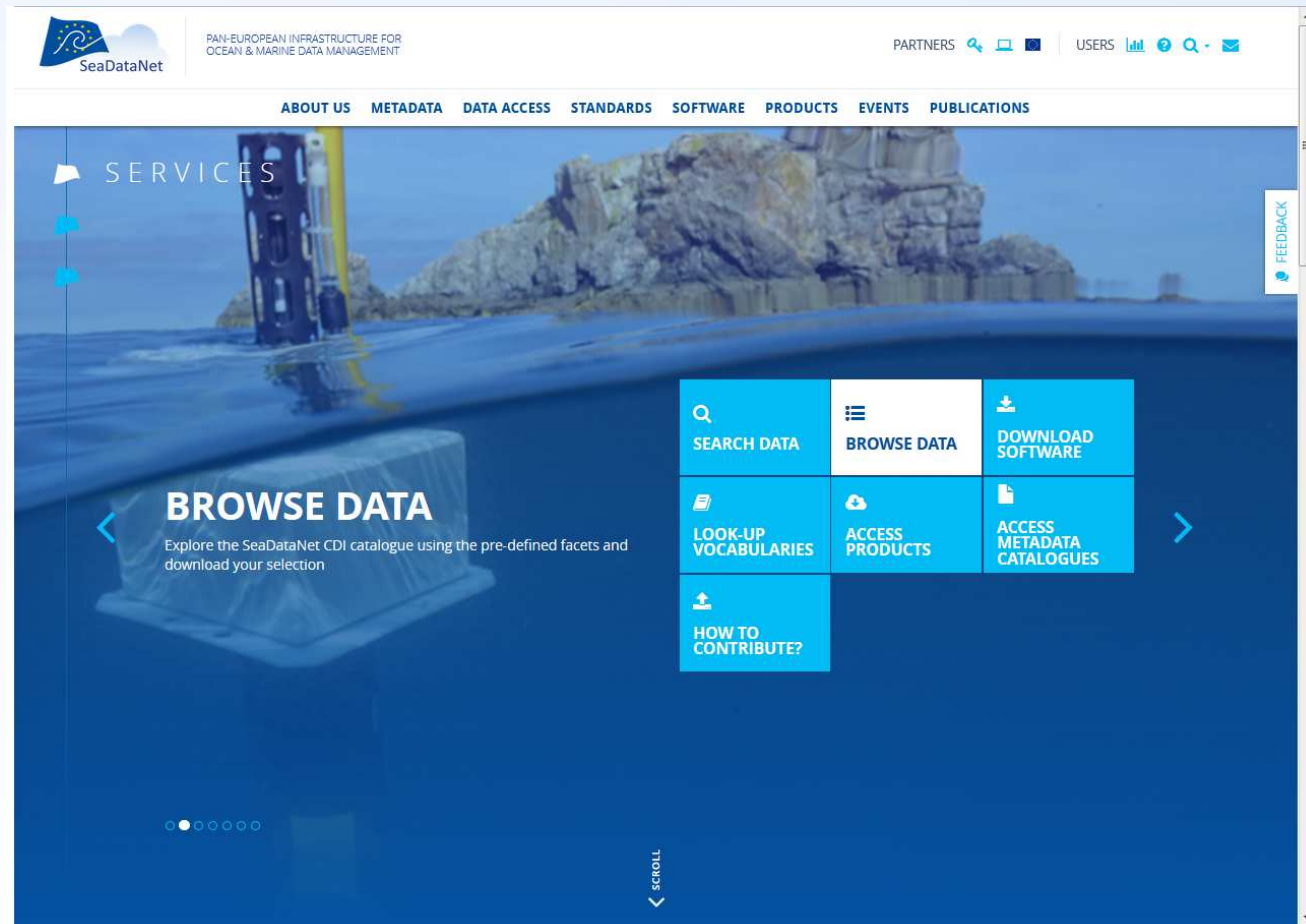


Figure High level architecture



www.seadatanet.org

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