SeaDataCloud technical challenges

Dick M.A. Schaap, MARIS, SDC Technical Coordinator

SDC Review, Brussels – Belgium, 6 December 2018
sdn-userdesk@seadatanet.org – www.seadatanet.org
European Open Science Cloud (EOSC)

- **6 May 2015** the EU adopted the Digital Single Markets strategy and announced the launch of a **cloud for research data**
- A **High Level Expert Group** was established to advise on scientific services to be provided on the cloud and its governance structure
- **16 April 2016** the EU published a package of measures for digitising European industry, including a communication on the **European Cloud Initiative**
- **12 June 2017** the **European Open Science Cloud Summit** took place bringing together key players from across Europe, to make the ‘**EOSC a reality by 2020**’
- **14 March 2018** the EU adopted the **Implementation Roadmap for the European Science Cloud**

**EOSC** should give the EU a global lead in research data management and ensure that European scientists reap the full benefits of data-driven science. It also foresees setting up a European Data Infrastructure, with high-capacity cloud solutions with super-computing capacity.
In anticipation of these EOSC planning activities, SeaDataNet submitted in March 2016 the SeaDataCloud proposal with the following aims:

- Further **developing skills, standards, tools and services** for dealing with marine data, such as handling data from new instruments, INSPIRE compliance, interoperability with other data infrastructures, vocabularies governance, and adopting new technical approaches such as ‘Linked Data’, and ‘Sensor Web Enablement’,

- Exploring the trend towards cloud storage and cloud computing, also taking into account **‘big data’ challenges**, for instance by analysing an architecture for a Virtual Research Environment (VRE) and developing a VRE pilot

- Improving and expanding services and tools for **data providers** for connecting and ingesting data AND for **users** to make it easier and more performing to find, access, and use datasets

- **Better joining in an early stage and riding the EOSC wave than losing ourselves in the undertow**
Cooperation with EUDAT

5 EUDAT members are partners of SeaDataCloud:
CINECA, CSC, DKRZ, GRNET and STFC
Cooperation with EUDAT
SeaDataCloud technical challenges

- SDC is about **improving and innovating services** such as:
  - upgrading the CDI Data Discovery & Access service by adopting the cloud
  - developing MySeaDataCloud custom services
New CDI service architecture

- User Interface
- CDI
- Shopping Basket
- Request Status Manager
- Data Cloud
- CDI Import Manager
- Replication Manager
- Download Manager
SeaDataCloud technical challenges

- SDC is about **improving and innovating services** such as:
  - upgrading the monitoring of the SeaDataNet system
  - Deploying a DOI minting service in combination with EMODnet Ingestion
SeaDataCloud technical challenges

- SDC is about **updating and further developing standards**, such as:
  - expanding vocabularies
  - linked data principle for richer metadata and searches
  - handling different data types such as HF Radar, Gliders, Flow cytometer data
SeaDataCloud – INSPIRE compliance

Relevant
Integration of multiple data streams

- Further expansion of the number of data sets populated in the **CDI service** by an increasing number of connected data centres, in particular through SeaDataCloud, EMODnet, AtlantOS, and other projects.

- Integrating and operating a **Brokerage Service** following ODIPII (Ocean Data Interoperability Platform) for connecting other major international marine data discovery and access services such as US NCEI, Australian AODN, and others. This is done through seeking interoperability. SDN users will be offered an extra facility for discovering available data sets in those other systems and given their links and services to access associated data sets;
Brokering Prototype – Semantic Discovery

- ODIP Test Portal
- JS API
- ODIP Broker
- SPARQL
- Rosetta Stone Translation Service
- HTTP (ISO CDI)
- CSW (ISO NCEI)
- CSW (ISO MCP)
- USA NCEI (NODC)
- SeaDataNet CDI
- IMOS AODN

sdn-userdesk@seadatanet.org
ODIP Brokerage Service

http://odip-prototype.essi-lab.eu/broker/search

sdn-userdesk@seadatanet.org – www.seadatanet.org
Integration of multiple data streams

- Using **EMODnet Ingestion portal** where third parties can ingest data collections. Submissions are assigned to expert data centres.
Integration of multiple data streams

- EMODnet Ingestion Pathways for working up the ingested data sets.
Integration of multiple data streams

Integrating the **online Sensor Web Enablement (SWE) Ingestion service (52N)** for ingesting near real time data sets from operational oceanography sensor networks:

- Closing the gap between operational oceanography and delayed mode validation and archiving. The SWE ingestion service will facilitate streamlining the data and metadata transfer from operational networks to a database buffer from which assigned SDN data centres can pick up the data timeseries for further elaboration and later population into the CDI service.

- Enriching and expanding the CDI service for its users with access to historical long timeseries AND the latest NRT data timeseries from the operational sensors linking to the 52°North “Helgoland” viewer. As part of SeaDataCloud, Helgoland will be further extended in close cooperation with users. This comprises especially the support of further data types as well as usability improvements.
Integration of multiple data streams

- a backend implementation to perform the collection of data from different sources, the interpretation and conversion into an internal data model (based on the ISO/OGC Observation and Measurements (O&M) standard, as well as the publication of the collected data into the database of an interoperable OGC Sensor Observation Service (SOS) instance;
SeaDataCloud technical challenges

- SDC is about **adopting and elaborating new technologies**, such as:
  - developing extra tools for data managers, such as OCTOPUS for SeaDataNet file format compliance checking and format conversions
  - developing a **Virtual Research Environment (VRE)** for researchers with dedicated cloud-based tools for processing and analyzing data
Leading concept => towards a Blue Cloud

- Cloud platform with common services for data pre-processing, analyses, visualizations, publishing, DOIs...
- Applying common standards and interoperability solutions for providing harmonised data and metadata
- Providing harmonised discovery and access to data output from multiple sources, European and international
Virtual Research Environment (VRE)

- Reviewing of existing VRE developments in Australia, USA and Europe, also as part of the Ocean Data Interoperability Platform (ODIP) activities
- Formulating a ToR for the VRE overall specifications
- Analysing and defining an overall architecture for the SDC VRE, considering EUDAT’s generic services and embedding dedicated SDN services, custom and generic
- Analysing the workflow and specific functions required for supporting the SDC pilot: generating a SeaDataNet Temperature & Salinity climatology
- Developing VRE components and integrating these into the SDC VRE pilot
Overall architecture for the SDC VRE