EUDAT & SeaDataCloud

SeaDataCloud Kick-off meeting

Damien Lecarpentier
CSC-IT Center for Science

www.eudat.eu

EUDAT receives funding from the European Union’s Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065
A consortium of high performance computing (HPC) / data centres, libraries, scientific communities, data scientists

EUDAT – European Data Infrastructure
Collaborative Data Infrastructure

- **Data Generators**: User functionalities, data capture, data transfer, virtual research environments
- **Community Support Services**: Data discovery & navigation, workflow generation, annotation, interpretability
- **Common Data Services**: Persistent storage, identification, authenticity, workflow execution, mining

---

**EUDAT**
What is the EUDAT Service offer?

Data
Metadata
Authentication

B2HANDLE
PID Management

Integrity, authenticity, locality

B2SAFE
Policy-based data management

B2STAGE
Dynamic replication for processing

B2FIND
Aggregated metadata inventory

B2DROP
Secure data exchange

B2SHARE
Reliable data storage and sharing

Network of trust

B2ACCESS
Identity & Authorisation

B2FIND
Find Research Data

B2SHARE
Store and Share Research Data

B2ACCESS
Identity & Authorisation

B2NOTE
Data Type Registry

B2HANDLE
Register your Research Data

B2STAGE
Get Data to Computation

B2SAFE
Replicate Research Data Safely

B2DROP
Sync and Exchange Research Data
Addressing the full lifecycle of research data

Regular webinar: “How to write a Data Management Plan”
eudat.eu/events/webinar/

eudat.eu/training/research-data-management
The CDI – A Service Infrastructure
Who are EUDAT Communities?

**PHYSICAL SCIENCES & ENGINEERING**

**ENVIRONMENTAL SCIENCES**

**BIOMEDICAL & MEDICAL SCIENCES**

**MATERIALS & ANALYTICAL FACILITIES**

**SOCIAL SCIENCES & HUMANITIES**

**CLARIN**

**EURO VO**

**pandata**
EUDAT ENV Core Communities

EUDAT ENV Data Pilots

Institutions and partners

For more information visit - https://eudat.eu/use-cases
EUDAT & SeaDataCloud

- **Strategic cooperation**

- **Main goal:**
  - Strengthening SeaDataNet underlying IT infrastructure
  - Improving access to and discovery of SeaDataNet data & products

- **Output:**
  - A cloud environment for storing datasets and providing cloud computing capabilities
  - A Virtual Research Environment (VRE) with various analysis and visualisation tools

---

**Implementation phases**

1) Replicate data from a subset of SeaDataNet centers to the EUDAT centers (technical cache)

2) Define relevant services for SeaDataNet (e.g. data subscription, etc.)

3) Develop a common cloud and computing environment integrated with the SeaDataNet portal and other supported services
EUDAT & SeaDataCloud
<table>
<thead>
<tr>
<th>WP</th>
<th>Title</th>
<th>EUDAT Partners</th>
<th>PMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP2</td>
<td><strong>Project Network Coordination</strong></td>
<td>CSC, CINECA, DKRZ, STFC, GRNET</td>
<td>6</td>
</tr>
<tr>
<td>WP3</td>
<td><strong>Training of Data Providers</strong></td>
<td>CSC, CINECA</td>
<td>3</td>
</tr>
<tr>
<td>WP5</td>
<td><strong>Expansion and governance of metadata and data content</strong></td>
<td>CSC, CINECA, DKRZ, STFC, GRNET</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>• WP5.4: Installing, configuring and taking into operation local components for the upgraded CDI data discovery and access service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP6</td>
<td><strong>Core and Advanced Services</strong></td>
<td>CSC, CINECA, DKRZ, STFC, GRNET</td>
<td>35</td>
</tr>
<tr>
<td>WP7</td>
<td><strong>Tuning of requirements and overall integration</strong></td>
<td>CSC</td>
<td>2</td>
</tr>
<tr>
<td>WP8</td>
<td><strong>Governance of standards and development of common services</strong></td>
<td>CSC, CINECA, DKRZ, STFC, GRNET</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>• ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• WP8.5: Upgrading the operational Monitoring system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP9</td>
<td><strong>Developments of upstream services</strong></td>
<td>CSC, CINECA, DKRZ, STFC, GRNET</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>• WP9.1: Upgrading the CDI service making use of the cloud</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• WP9.2 Developing and deploying additional services in the cloud for ensuring integrity and conformity of the CDI related cloud data resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• WP9.4: Developing integrated online services for ingesting autonomous observatory data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• WP9.8: Development of a solution for coordinated distributed DataCite DOI minting service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP10</td>
<td><strong>Developments of downstream services</strong></td>
<td>CSC, CINECA, DKRZ, STFC, GRNET</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>• WP10.1: Development of the Virtual Research Environment (VRE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• WP10.2: Developing advanced services for search, processing, analytics, quality, visualisation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• WP10.3: Developing MySeaDataCloud</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Deliverables

<table>
<thead>
<tr>
<th>Deliverable number</th>
<th>Title</th>
<th>Time</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>D9.1</td>
<td>Specification of the SeaDataNet European cloud environment</td>
<td>6</td>
<td>CSC</td>
</tr>
<tr>
<td>D9.2</td>
<td>Enhanced EUDAT B2SAFE replication software</td>
<td>12</td>
<td>CINECA</td>
</tr>
<tr>
<td>D9.3</td>
<td>Adapted EUDAT B2HOST service operational</td>
<td>12</td>
<td>DKRZ</td>
</tr>
<tr>
<td>D10.1</td>
<td>Specification of the Virtual Research Environment (VRE) in the cloud and development plan</td>
<td>12</td>
<td>DKRZ</td>
</tr>
<tr>
<td>D10.2</td>
<td>Virtual Research Environment (VRE) in the cloud operational</td>
<td>30</td>
<td>ALL</td>
</tr>
<tr>
<td>D.10.19</td>
<td>Specification of MySeaDataCloud and development plan</td>
<td>18</td>
<td>CSC/CINECA</td>
</tr>
<tr>
<td>D10.20</td>
<td>MySeaDataCloud operational</td>
<td>30</td>
<td>STFC (ALL)</td>
</tr>
</tbody>
</table>
EUDAT Partners in SDC

- **CSC – IT Center for Science** Ltd, established in 1971, is the Finnish national HPC centre providing high-performance computing, networking and data services to Finnish academia, research institutes, public sector and industry. CSC is a non-profit limited company whose shares are fully owned by Finnish state, and governed by the Finnish Ministry of Education and Culture.

- **CINECA**, established in 1969, is the Italian national supercomputing facility and one of the largest in Europe, providing high performance computing resources, data management and storage systems, tools, middleware services and expertise, aiming to develop and promote technical and scientific discoveries into the HPC field for the Italian and European research communities. CINECA is a non-profit consortium of 67 Italian Universities, the National Institute of Oceanography and Experimental Geophysics (OGS), the National Research Council (CNR), and the Ministry of Education, University and Research (MIUR).

- **DKRZ** is a national German facility, providing state-of-the-art supercomputing data service and other associated services to the German and international scientific community to conduct top of the line Earth System and Climate Modeling. DKRZ is partner in ENES (European Network for Earth System Modelling) and is one the representatives of the Earth system research community in the EUDAT project. DKRZ is operating with its ICSU World Data Center Climate (WDCC21) a community specific long-term data archive. Connected with its WDCC
EUDAT Partners in SDC

- **STFC** is the UK public sector research organization providing access to large scale scientific facilities. The Scientific Computing Department (SCD) — provides the STFC facilities with advanced IT infrastructure including massive data storage, high-end supercomputing, vast network bandwidth, and interoperability with other infrastructure in the UK and internationally. STFC is a major provider of resources (both compute and storage) for many large projects including WLCG, ENES (through CEDA where it also provides high performance computing and data archival services) and PanDATA providing multi-petabyte archival services for Diamond Light Source and ISIS.

- **GRNET** - the Greek Research and Technology Network - is a state-owned company, operating under the auspices of the Greek Ministry of Education. Its mission is to provide high-quality electronic Infrastructure services to the academic, research and educational community of Greece; and link these with global e-Infrastructures; GRNET is the main IT enabler for big science as well as long-tail science in Greece, and core player in international activities.
A landmark agreement sustaining the pan European Collaborative Data Infrastructure for the next 10 years

3rd October 2016

16 major European research organisations, data and computing centres signed an agreement to sustain the EUDAT – pan European collaborative data infrastructure for the next 10 years. The organisations stand together behind a long term sustainability plan and commit to develop, maintain and deploy pan-European research data services and to promote harmonization of research data management practices across centres.

What does this mean?

Service providers, both generic and thematic, and research communities have joined forces as part of a common framework for developing an interoperable layer of common data services. Known as the EUDAT Collaborative Data Infrastructure (CDI), this is essentially a European e-infrastructure of integrated data services and resources to support research. This infrastructure and its services have been developed in close collaboration with over 50 research communities spanning across many different scientific disciplines and involved at all stage of the design process. The establishment of the EUDAT CDI is timely with the imminent realization of the European Open Science Cloud[1] which aims to offer open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines.

What benefits will this bring to researchers?

Researchers can rely on innovative data services to support their research collaboration and data management. Additionally they benefit from a common service management framework delivered by CDI service providers and the connection between sites. These 16 organisations[2] from 11 countries with a further 4[3] in the process of joining, are European leaders in the deployment of data management and data infrastructure services. By joining the EUDAT CDI they are able to better serve their local customers across borders and support cross-national collaborations.
Thanks for your attention.