

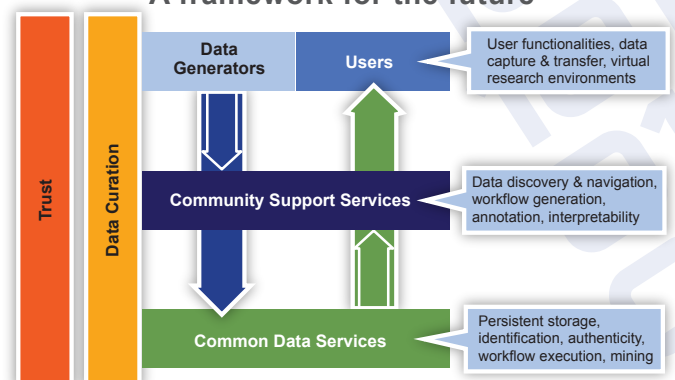
TOWARDS A COLLABORATIVE DATA INFRASTRUCTURE

On 1 October 2011, the EUDAT project was launched to target a pan-European solution to the challenge of data proliferation in Europe's scientific and research communities through the deployment of a Collaborative Data Infrastructure (CDI) driven by research communities needs. Building the CDI will require active collaboration between all actors, and EUDAT is calling for the contributions of all stakeholders interested in adapting their solutions or participating to the design of the CDI. EUDAT is currently carrying out a comprehensive review of research communities approaches and requirements to the deployment and use of a common and persistent data e-Infrastructure. With the help of the research communities, it will investigate and design appropriate services and technologies to match these requirements, which will be operated as part of its distributed infrastructure.

THE CASE FOR A GENERIC MULTI-DISCIPLINARY DATA INFRASTRUCTURE

Although research communities from different disciplines have different ambitions, particularly with respect to data organization and content, they also share basic service requirements. This commonality makes it possible to establish generic pan-European services designed to support multiple communities, as part of a Collaborative Data Infrastructure. For the CDI to succeed, an abstract architecture is required, facilitating integration of pre-existing data solutions from participating communities and data centers willing to support common data services. Reuse and recombination of data in such open scenarios is based on suitable integration and interoperability solutions.

The Collaborative Data Infrastructure: A framework for the future



Source: High Level Expert Group on Scientific Data, *Riding the wave*, 2010.

WHAT KIND OF SERVICES WILL EUDAT OFFER

Nine core service areas have been identified as essential for the different research communities, with corresponding community-oriented and enabling services that can be designed to support multiple communities:

COMMUNITY-ORIENTED SERVICES

- SIMPLE DATA ACCESS AND UPLOAD
- LONG-TERM PRESERVATION
- SHARED WORKSPACES
- EXECUTION AND WORKFLOW (DATA MINING ETC.)
- JOINT METADATA AND DATA VISIBILITY

ENABLING SERVICES (MAKING USE OF EXISTING SERVICES WHERE POSSIBLE)

- PERSISTENT IDENTIFIER SERVICE (EPIC, DataCite)
- FEDERATED AAI SERVICE
- NETWORK SERVICES
- MONITORING AND ACCOUNTING

EUDAT SERVICE DESIGN APPROACH

The EUDAT service building approach is based on three main strands of activity. The first strand involves capturing community requirements. The second activity strand concerns the appraisal of technologies and service candidates: this involves identifying, designing and constructing appropriate services, using existing solutions where possible. The third activity strand is operation of the collaborative infrastructure, particularly provisioning of secure, reliable (generic) services in a highly available production environment, with interfaces for cross-site and cross-community operation. The operation of the infrastructure will provide full life cycle data management services, ensuring the authenticity, integrity, retention and preservation of data, especially those marked for long-term archiving.

EUDAT aims to deploy a first set of services in 2012, as part of its distributed infrastructure. We anticipate that some services will support a broad range of communities, while others will only be of interest to a subset of communities or disciplines. This flexibility allows an incremental approach, where initial and associate communities are taken into account step by step.

WHO IS INVOLVED AND HOW TO JOIN THE INITIATIVE?

The EUDAT consortium comprises 25 European partners, including data centers, technology providers, research communities and funding agencies from 13 countries. It includes key representatives from research communities in linguistics (CLARIN), earth sciences (EPOS), climate sciences (ENES), environmental sciences (LIFEWATCH), and biological and medical sciences (VPH), all of which have been allocated project resources to help specify their requirements and co-design related services. Other communities have joined EUDAT as associate members, representing 15 research disciplines across all major fields of science. The EUDAT User Forums provide new communities interested in joining the initiative with a unique opportunity to learn about the project and contribute to the design of the CDI. Several task forces focusing on service cases and policy matters are also open to external stakeholders and allow an immediate integration of interested external partners into the concrete work carried out within the project.

