EUDAT Service Road Map

Common Data Services

EUDAT’s mission is to design, develop, implement and offer “Common Data Services” as they have been introduced in the “Riding the Wave report [1]” to all interested researchers and research communities. According to this report these services will be offered through the Collaborative Data Infrastructure (CDI) which is being identified by many different initiatives at community, research organisation and cross-border level (disciplines and countries). Common data services obviously must be relevant to several communities and be available at European level and they need to be characterised by a high degree of openness: (1) Open Access should be the default principle; (2) Independent of specific technologies since these will change frequently and (3) Flexible to allow new communities to be integrated which is not a trivial requirement given the heterogeneity and fragmentation of the data landscape.

Following two years intensive work on a first set of services EUDAT understands the importance of certain aspects to offer such common data services:

- User communities must be in the driving seat to select high-priority services and their major functional characteristics.
- Close collaboration with senior technologists from different backgrounds is essential to guarantee the openness, technology independence and flexibility required.
- The establishment of trust is key for acceptance of data services and trust has many dimensions including irrational ones. Furthermore establishing trust require considerable time investment.
- Persistence of data and sustainability of the service framework are key aspects of trust establishment.
- The existence of data management plans that cover the various aspects of trust, quality, availability, etc. are non-negotiable.

In most cases common services can only be developed and shaped in close collaboration with interested research communities. And even then, EUDAT realised that areas identified as important for developing new services require the establishment of working groups to define the nucleus of what the “common service” should be.

Characteristics of Data

Data comes in many different characteristics and common services need to address them all. In general, so-called big data is produced by sensors and simulations and has a fairly regular structure. Long-tail data is often the result of scientific analysis, is highly irregular and complex and comes in many different formats where the end of dynamics is not so clear. The middle area which is characterized by a large amount of what is commonly called collections also needs to be addressed. Large amounts of organized data as a result of scientific activities are also

EUDAT receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 654065.
highly complex in nature.

Increasingly often it becomes clear that data streams are associated with a variety of logical layer information such as persistent identifiers and its attributes, metadata and its many different aspects, access permission information and more. However, for organizing this logical layer and its many relations there are no agreements yet. Common services in the CDI need to address not only the specific nature of the data in focus, but also deal with the heterogeneity of the data organization. It can only be hoped that initiatives such as RDA [2] will lead to more harmonization, but it will be a stepwise process.

Service Roadmap

The roadmap includes three overlapping phases of brainstorming, selecting and defining the services to be offered. In any case, development and improvement of services is a continuous process.

Phase 1 Services (October 2011 - October 2013)

EUDAT started intensive discussions two years ago with some user communities about high priority data services to be tackled which resulted in four phase 1 services: (1) B2SHARE addressing the long tail data and is open to all researchers. (2) B2SAFE offering safe replication services in different flavours addressing big and collection type data. (3) B2FIND offering a metadata repository for both EUDAT and other community data. (4) B2STAGE offering the possibility to move replicated data to the workspace of high performance computers and carry out scientific calculations. In addition a number of enabling services have been implemented such as a service to register and resolve persistent identifiers and methods for authentication and authorization in distributed infrastructures.
Phase 2 Services (December 2011 - February 2014)

While working on the first set of services, EUDAT started working on two more services: (1) B2DROP to synchronize file systems with a central store and (2) B2NOTE offering semantic referencing and annotation.

Phase 3 Services (October 2013 - ...)

Currently, EUDAT is brainstorming the third phase of services and is inviting all interested communities to participate in this discussion. A recent survey indicated that topics such as data lifecycle support, support for dynamic data, workflow execution environments and additional semantic services are urgently required. EUDAT is circulating a call for collaborations to all interested communities which will be turned into action this year and also in H2020. The results of the survey and the results of the call for collaborations will widely determine the priorities for the coming months and such calls will be repeated regularly.

Naturally, all new activities must fit into the general strategy and architecture of EUDAT. The EUDAT data domain is a domain of registered and described data.

[1] Research Data Alliance
[2] Riding the wave report from EC’s High Level Expert Group on Scientific Data

Author(s):
EUDAT

Read more