Towards a pan-European Collaborative Data Infrastructure

On October 1, 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. Aiming to contribute to the production of a Collaborative Data Infrastructure driven by researchers' needs, the project is coordinated by CSC - IT Center for Science, Finland, and co-funded by the European Commission's Framework Programme 7.

EUDAT aims at providing Europe's scientific and research communities with a sustainable pan-European infrastructure for improved access to scientific data. Burgeoning volumes of valuable and complex data – newly available from powerful new scientific instruments, simulations and digitization of library resources – represents a fantastic opportunity for science, but has created new challenges related to data management, access and preservation.

The EUDAT consortium comprises 25 European partners, including data centers, technology providers, research communities and funding agencies from 13 countries, who will work together to deliver a Collaborative Data Infrastructure that can sustainably meet future researchers' needs.

"EUDAT will fill an important gap in the current European e-Infrastructure landscape," said Dr. Kimmo Koski, CSC Managing Director and EUDAT Project Coordinator. "We aim at developing a generic infrastructure for scientific

data management that can be used by a diversity of research communities and existing infrastructures. This can only be achieved through a systematic and focused approach covering the entire life cycle of data objects, and by encouraging collaboration between the various stakeholders, in particular between the communities involved in designing specific services and the data centers willing to provide generic solutions."

EUDAT addresses these challenges and exploits the opportunities using its vision of a Collaborative Data Infrastructure (CDI). The key idea is to acknowledge that research communities from different disciplines follow different approaches with respect to data organization and content, but they also share the need for common data services. Exactly this commonality makes it possible to offer generic services designed to support multiple communities complementary to their existing specific support services thus leading to a federated CDI. An abstract flexible framework facilitates the integration of existing data generators and solutions from users into the CDI and is supported by data centers offering common data services as a strong and sustainable data backbone.

Multi-disciplinary Collaboration and Data Sharing

The EUDAT partners include key representatives from research communities in linguistics (CLARIN), earth sciences (EPOS), climate sciences (ENES), environ-
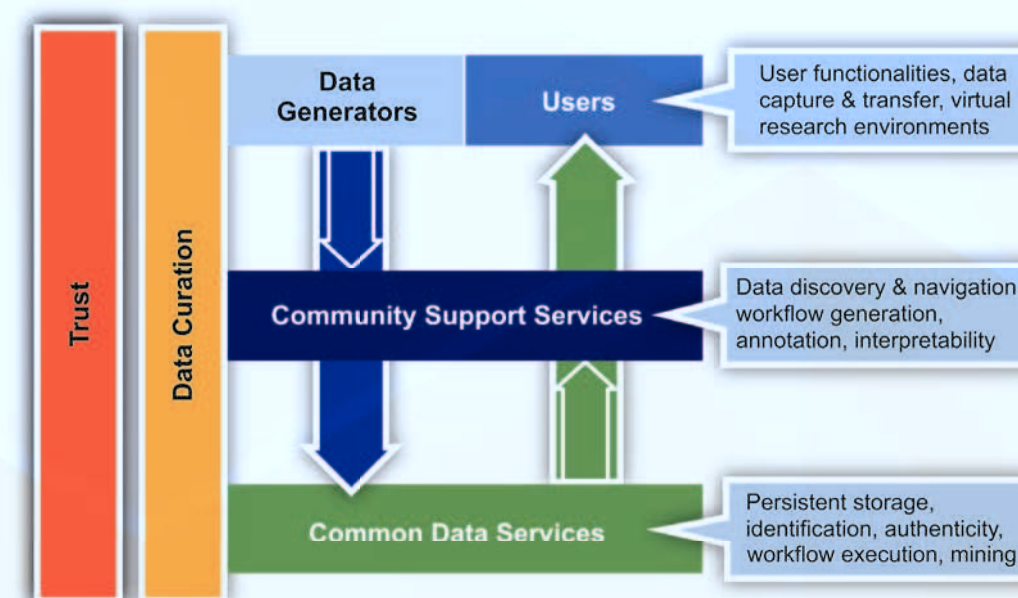

Figure 1: Collaborative Data Infrastructure

mental 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.

EUDAT Scientific Coordinator Peter Wittenburg, from the Max Planck Institute for Psycholinguistics at Nijmegen, the Netherlands, said EUDAT will open considerable new opportunities for research communities. "Beyond offering common services such as data hosting and preservation, EUDAT is paving the way towards integrated and interoperable access to data and will facilitate new science and allow efficient knowledge creation. It is this double opportunity that makes the EUDAT initiative so interesting for research communities and infrastructures."

The common services that have already been identified and that will be the

added value of the EUDAT CDI include but are not limited to federated AAI, data access and upload, long-term preservation, persistent identifier, workspaces, web execution and workflow, monitoring and accounting, network, and metadata. All these services will be established in a flexible manner to enable the adaptation of community solutions, which is a guiding principle for the design of the CDI. EUDAT aims at deploying its first production services in 2012.

First EUDAT Conference

The first EUDAT conference will take place in Barcelona from October 22 - 24, 2012. It is a high level international event where EUDAT's first results will be demonstrated and a forum to discuss the future of EUDAT and data infrastructures. The conference will host the second EUDAT User Forum and EUDAT training tutorials.

More information is available from www.eudat.eu

- Daniel Mallmann[1]
- Benedikt von St. Vieth[1]
- Morris Riedel[1]
- Jedrzej Rybicki[1]
- Kimmo Koski[2]
- Damien Lecarpentier[2]
- Peter Wittenburg[3]

[1] Jülich Supercomputing Centre

[2] CSC - IT Center for Science

[3] Max Planck Institute for Psycholinguistics