EUDAT - Towards a Collaborative Data Infrastructure for Science

Kimmo Koski and Damien Lecarpentier
CSC-IT Center for Science

EUDAT is a new pan-European data initiative, bringing together a unique consortium of 25 partners, including research communities, national data and high performance computing (HPC) centers, technology providers, and funding agencies from 13 countries research communities, aiming to contribute to the production of a collaborative data infrastructure (CDI) to support Europe's scientific and research data requirements. Although research communities from different disciplines have different ambitions and approaches, particularly with respect to data organization and content, they also share basic service requirements. This commonality makes it possible for EUDAT to establish shared pan-European data services, designed to support multiple research communities. Since its start on 1st October 2011, EUDAT has been reviewing the approaches and requirements of a first subset of communities from linguistics (CLARIN), earth sciences (EPOS), climate sciences (ENES), environmental sciences (LIFEWATCH), and biological and medical sciences (VPH) and is now extended this analysis to other communities. EUDAT has shortlisted six service cases identified by user communities as priorities: safe replication of data from site to site, data staging to HPC facilities, easy storage, metadata, AAI, and persistent identifiers. In each case, a multi-disciplinary taskforce involving representatives from communities and data centers has been set up to plan for the deployment of these services on the EUDAT infrastructure. EUDAT aims to deliver pilot services in 2012, with full services available to all research communities by the end of 2014.