EUDAT - Towards a pan-European Collaborative Data Infrastructure

Objectives:

While research communities from different disciplines have different ambitions and approaches – particularly with respect to data organization and content – they do share many basic service requirements. This commonality makes it possible for EUDAT (European Data Infrastructure) to establish common data services, designed to support multiple research communities, as part of a Collaborative Data Infrastructure (CDI). By providing generic services to existing scientific communities, EUDAT enables research communities to focus a greater part of their effort and investment on services that are discipline-specific. EUDAT also provides individual researchers, smaller communities, and projects lacking tailored data management solutions with access to sophisticated shared services, thus removing the need for large-scale capital investment in infrastructure development. Lastly, by providing opportunities for disciplines from across the spectrum to share data and cross-fertilize ideas, EUDAT encourages progress towards the vision of open and participatory data-intensive science. EUDAT is a research data infrastructure initiative funded by the European Commission's FP7 programme (2011-2014) to build a sustainable cross-disciplinary and cross-national infrastructure that provides a set of shared services for accessing and preserving research data. The project is coordinated by CSC and brings together a consortium of 25 partners, including national data and high performance computing (HPC) centers, technology providers, research communities, and research councils from 13 countries.

On-going Activities:

EUDAT has reviewed the approaches and requirements of a first subset of communities coming from linguistics, solid earth sciences, climate sciences, environmental sciences, and biological and medical sciences, regarding the deployment and use of a cross-disciplinary and persistent data e-Infrastructure. Four generic services that have been identified by these communities as priorities and have been designed as part of multi-disciplinary task forces. The first prototypes are currently being tested before being handed over to the operations team for deployment onto the infrastructure. The services are: **data replication from site to site, data staging to computer facilities, metadata, and easy storage**. Training programmes are also being conducted to help researchers make the best out of the current and future services offered by EUDAT. Another important strand of activity concerns planning the evolution of EUDAT from a 'project-centric' collaboration of organizations to a sustainable infrastructure. This implies early definition of future partnership and business models for adopting, supporting and sustaining common services developed for, and partly operated by, the different research communities.

Results:

To date, four core services have been chosen as early candidate services and will soon be released for production. The Safe Replication service offers a robust, safe and highly available replication service allowing community and departmental repositories to replicate their data to guard against data loss, optimize access for users from different regions, and/or bring data closer to powerful computers for compute-intensive analysis, through the **Data Staging** service. The **Simple Store service** addresses the problem of many researchers of finding a simple, convenient and durable way of storing and sharing their data. It allows registered users to upload typical "long tail" data objects into the EUDAT store, and to share such objects and collections with other researchers. The **Metadata service** allows users to see what kind of data is stored through the Safe Replication and Simple Store services. As the project's initial data services all approach delivery, EUDAT is moving ahead to look at the next round of services to be developed. A number of new service candidates are currently being discussed, in particular generic services to handle real-time data streams and crowd-sourced data, and to support semantic annotation.

CONTACT DETAILS:
NAME: NAGHAM SALMAN
ORGANISATION: BARCELONA SUPERCOMPUTING CENTRE
PROJECT NAME: EUDAT
EMAIL: NAGHAM.SALMAN@BSC.ES
URL: WWW.EUDAT.EU