Geophysical Research Abstracts
Vol. 18, EGU2016-16733-5, 2016
EGU General Assembly 2016

EGI-EUDAT integration activity - Pair data and high-throughput computing resources together

Diego Scardaci (1), Matthew Viljoen (1), Dejan Vitlacil (2), Giuseppe Fiameni (3), Yin Chen (1), Gergely sipos (1), and Tiziana Ferrari (1)

(1) EGI.eu Netherlands (diego.scardaci@egi.eu), (2) KTH Royal Institute of Technology Sweden, (3) CINECA Italy

EGI (www.egi.eu) is a publicly funded e-infrastructure put together to give scientists access to more than 530,000 logical CPUs, 200 PB of disk capacity and 300 PB of tape storage to drive research and innovation in Europe. The infrastructure provides both high throughput computing and cloud compute/storage capabilities. Resources are provided by about 350 resource centres which are distributed across 56 countries in Europe, the Asia-Pacific region, Canada and Latin America.

EUDAT (www.eudat.eu) is a collaborative Pan-European infrastructure providing research data services, training and consultancy for researchers, research communities, research infrastructures and data centres. EUDAT's vision is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure (CDI) conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres.

EGI and EUDAT, in the context of their flagship projects, EGI-Engage and EUDAT2020, started in March 2015 a collaboration to harmonise the two infrastructures, including technical interoperability, authentication, authorisation and identity management, policy and operations. The main objective of this work is to provide end-users with a seamless access to an integrated infrastructure offering both EGI and EUDAT services and, then, pairing data and high-throughput computing resources together.

To define the roadmap of this collaboration, EGI and EUDAT selected a set of relevant user communities, already collaborating with both infrastructures, which could bring requirements and help to assign the right priorities to each of them. In this way, from the beginning, this activity has been really driven by the end users. The identified user communities are relevant European Research infrastructure in the field of Earth Science (EPOS and ICOS), Bioinformatics (BBMRI and ELIXIR) and Space Physics (EISCAT-3D).

The first outcome of this activity has been the definition of a generic use case that captures the typical user scenario with respect the integrated use of the EGI and EUDAT infrastructures. This generic use case allows a user to instantiate a set of Virtual Machine images on the EGI Federated Cloud to perform computational jobs that analyse data previously stored on EUDAT long-term storage systems. The results of such analysis can be staged back to EUDAT storages, and if needed, allocated with Permanent identifyers (PIDs) for future use. The implementation of this generic use case requires the following integration activities between EGI and EUDAT: (1) harmonisation of the user authentication and authorisation models, (2) implementing interface connectors between the relevant EGI and EUDAT services, particularly EGI Cloud compute facilities and EUDAT long-term storage and PID systems.

In the presentation, the collected user requirements and the implementation status of the universal use case will be showed. Furthermore, how the universal use case is currently applied to satisfy EPOS and ICOS needs will be described.