Contribution ID: **26** Type: **oral presentation**

Intelligent, distributed climate data management

Thursday 10 May 2007 09:00 (20 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

Climate research is generally data-intensive. Observation, analysis and output data of climate simulations are traditionally stored in large archives and central databases. They vary highly in quality and in accessibility. Thus searching, finding and retrieving the data is often highly inefficient. To enable and ease effective data retrieval as well as collaborative work in the often interdisciplinary projects, an intelligent and sustainable (meta-)data management infrastructure is needed.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

As a first step, we used adapted tools and concepts developed in the German community driven Grid initiative C3-Grid (http://www.c3-grid.de/) to port the prototype of a typical climate data analysis workflow to the EGEE infrastructure (demonstrated at EGEE'06 conference).
Based on these experiences we now improve the structure and tighten the integration of the two systems. Of the C3-Grid, the central Web portal and uniform access interfaces, implemented at German Earthsystem science data centers, are used to initially search, find and retrieve the data and also to publish resulting metadata of already processed data. Processing can be done on EGEE. The Amga catalog is used as a central instance to receive and update the necessary runtime information of data produced or altered in the processing jobs. An interface is implemented to the Amga catalog to enable the automatic harvesting and hence republishing of the resulting metadata into the central C3 Web portal.

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)

The outlined approach does not only advance the interoperability of the two grid projects, but also stimulates synergy effects by combining the strengths of the two systems. Using standards along with international agreements allows for interoperability with international partners, such as the British NERC DataGrid and the US-American Earth System Grid, and with other communities, such as the GIS community. The next issue that needs to be solved is the match of the different security systems.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

Currently, to find, retrieve, and process climate data mostly complex individual solutions are used. Processed data is commonly stored locally and undocumented. Thus, identical analysis are redone by various scientists.
Enabling searching and browsing of the various data in a central catalog according to content, quality and processing history would ease the discovery of data. An intelligent, transparent data access would simplify the data retrieval. Selectable basic processing options and an automatic republishing of the processed data would support the daily workflows of climate scientists and facilitate further processing or usage of the results.
To realize this, besides conceptual agreements also common protocols and logging facilities, common authorization and authentication standards and finally also common resources to effectively share tools and data are required. EGEE offers solutions for most of these challenges and is thus a logical choice as basis for such an infrastructure

Primary authors: Dr RONNEBERGER, Kerstin (DKRZ); Dr STOCKHAUSE, Martina (MPI-M, Hamburg); Dr KINDERMANN, Stephan (DKRZ, Hamburg)

Presenter: Dr KINDERMANN, Stephan (DKRZ, Hamburg)

Session Classification: Data Management

Track Classification: Data Management