Geophysical Research Abstracts, Vol. 9, 03184, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-03184

Data Networking in Earth System Sciences (C3-Grid)

M. Stockhause (1), S. Kindermann (2), and H. Ramthun (3)

(1) Max Planck Institute for Meteorology (MPI-M), Germany, (2) German Climate Computing Center (DKRZ), Germany, (3) Model and Data Group at MPI-M, WDC for Climate, Germany (martina.stockhause@zmaw.de)

Earth system science (ESS) research strongly depends on analysis and comparison of data. These data are distributed over many archives and databases (e.g. WDC Climate) and vary highly in quality, available description (metadata), and accessibility. In the German D-Grid initiative C3-Grid (Collaborative Climate Community Data and Processing Grid) data and metadata services are developed enabling transparent collaborative data management and access. The service implementation is based on international standards and protocols to create sustainable and in different international collaborations reusable components. We present the design and implementation of the interfaces as well as applications of these interfaces in different contexts: for the realization of C3-Grid workflows, for the interaction with international portals and the (meta-)data exchange with international e-science projects like EGEE.

The three key components of the described design are a standardized metadata description, a metadata harvesting approach for metadata exchange, and a common webservice interface for intelligent data access.

Standardized metadata: For a uniform data description C3-Grid adopts the recently standardized metadata schema for geographical data ISO 19115/19139, using CF (Climate and Forecast) standard names for variable specification. Therein, besides descriptions of content, quality, and processing history (providing reusability), also hierarchies of data sets can be reflected. In the international context C3-Grid is the first ESS grid project applying these ISO guidelines. A clear trend towards the support of ISO 19115 by international portals can be identified (Nerc DataGrid, Earth System Grid, GO-ESSP).

Metadata harvesting: ISO metadata is collected based on established interfaces and

protocols (OAI-PMH). The generation of the ISO metadata is performed automatically by transforming proprietary metadata formats (e.g. metadata of the CERA DB hosting the WDCC).

Intelligent data access and preprocessing: Typically only small parts of the available huge data amount (stored across multiple data sets) are required for scientific workflows: few variables, spatial or temporal parts, sometimes altered by statistical or coordinate transformations. The technical local implementation of such a data access interface is based on standard webservice technology. So, the intelligent data access including local preprocessing is completely hidden from the user. The requested data is finally provided in the grid workspace, accessible to the C3-Grid data management service as well as to other international portals or collaborative environments using grid protocols (e.g. GridFTP). Complex preprocessing tasks can be optionally distributed to free resources using grid technology.