How Can Grid technologies Help in Earth System Sciences?

B. Fritzsch (1), W. Hiller (1) and R. Budich (2)

(1) Alfred Wegener Institute for Polar and Marine Research Bremerhaven, Germany, (2) Max Planck Institute for Meteorology, Hamburg, Germany,

(Bernadette.Fritzsch@awi.de)

Earth System Modelling strongly relies on a wide data base: Data serve as input in models and the model results have to been compared with other models or observational data. Unfortunately, the existing data are distributed over many archives and databases and vary highly in quality, available description (metadata), and accessibility. This encumbers collaborative work of scientists and makes routine workflows often tedious.

The project "Collaborative Climate Community Data and Processing Grid – C3Grid"meets this challenge and aims at linking distributed data archives in several German institutions. In order to ensure sustainability and reusability of C3Grid components in different international collaborations, the architecture is based on standard grid middleware and adheres to international standards wherever possible. Where required the architecture is completed by some specific components to supply the needs of the climate community.

A precondition for uniform access is the definition of a common metadata system. C3-Grid adopts the recently standardized metadata schema for geographical data ISO 19115/19139, using CF (Climate and Forecast) standard names for variable specification. An own C3 metadata profile was established, where 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 Earth System Sciences grid project applying these ISO guidelines.

The implemented infrastructure provides tools for effective data discovery and data transfer. A central catalogue of discovery metadata enables the user to seek after data in all connected data archives. The user can specify regional and time constraints, variable names and experiment labels for his data query. The provider specific implementation of the data access as well as of the data preparation is hidden behind a common WSDL defined web service interface. Thus the user has transparent and uniform access to all data archives within C3Grid.

Beside the data discovery and download functionality C3Grid offers a simple method to execute predefined workflows, which can be used for diagnostic purposes of model results. As first example the storm track and humidity flux analysis is implemented. The user submits the jobs in a portal and will get there the diagnosis results after job execution. In the background a scheduler distributes the incoming jobs in collaboration with the Data Management System to optimize data transfers within the grid.