How Can Grid Technologies Help in Earth System Sciences?

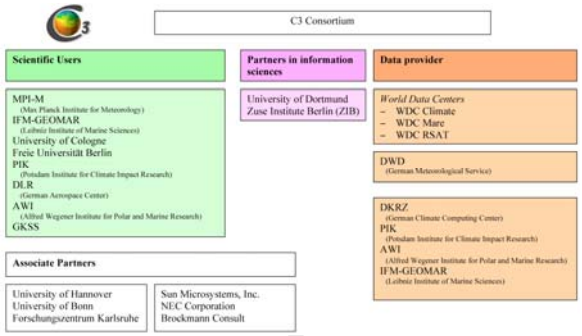
B. Fritzsch [1], W. Hiller [1] and R. Budich [2]

1 Wegener Institute for Polar and Marine Research Bremerhaven (Project Coordination)

2 Max Planck Institute for Meteorology, Hamburg

The "Collaborative Climate Community Data and Processing Grid – C3Grid", funded by the German Ministry for Research and Education (BMBF), is setting up a grid infrastructure for a seamless and fast access to the numerous data resources in the community of earth system research. C3Grid will ease model setup as well as data comparison and gives a broad scientific community access to model results and observational data.

The world data centres WDC Climate, WDC RSAT and WDC Mare as well as Germany's National Meteorological Service (DWD) and several other scientific institutes with specialised datasets provide a variety of data resources. Scientists from all major German earth science institutions are in the consortium and take part in the development and implementation of the C3Grid. They are supported by specialists from applied computer science from ZIB and University Dortmund.



How can C3Grid help in modeling?

Model setup

- preparing initial and boundary conditions, forcing data (find data, cut out the interesting spatial and temporal region, format conversion, regridding)

Model intercomparision, comparison of results with measurements

- find and prepare data, which fits the requirements (cut out the interesting spatial and temporal region, format conversion, regrinding)
- compute intensive analysis tools
- visualization

Current status in data providing

Provider	Data	in C3Grid
WDC Climate	Simulation results IPCC	~ 63 TB
WDC Mare	measurements (JGOFS, Southern Ocean)	~ 10 GB
WDC RSAT (DLR)	satellite data (Ozon profiles)	~ 60 GB
DKRZ Archive		
IFM-GEOMAR*	Simulations Nemo	~ 370 GB
GKSS*	Simulations Paleo	~ 1,1 TB
MPI-M*	Simulations IPCC	~ 1,3 TB
PIK	gridded meteorological & Carbon data	~ 9 GB
DWD	climatological data	~ 200 GB
AWI	Simulations OMIP	~ 300 GB
FUB/ Uni K	Simulations IPCC	~ 900 GB

Status	full data access	data searchable	Metadata exist	setup phase

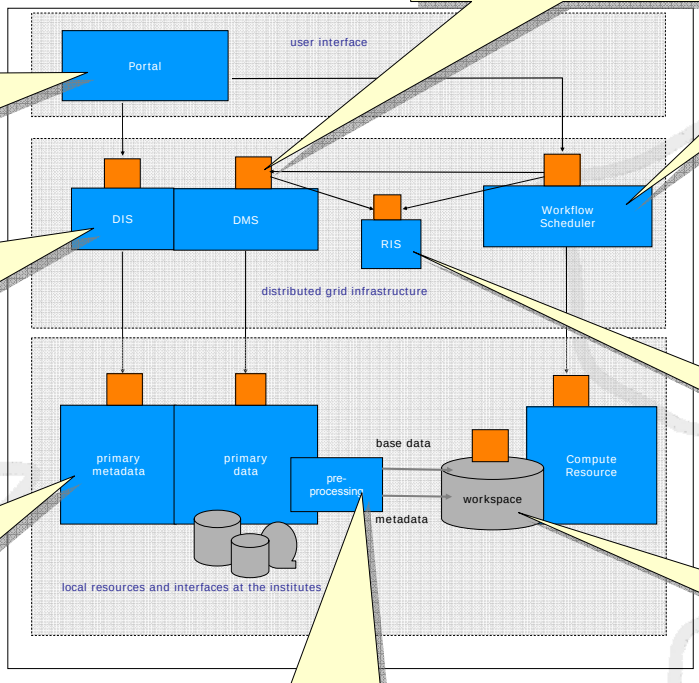
General implementation issues

- basic middleware Globus Toolkit 4.x with some C3 specific components
- components coupled via Web services
- C3 data publications standards
- workspace for logical/physical namespace mapping
- workflows consisting of several sequential but mutual dependent chains of elementary tasks → WSL

Portal
- based on the GridSphere open source portal framework
- modular and flexible portlet structure and the integration of the Java Commodity Grid Kit for direct use of Globus Toolkit features (e.g. GridFTP)
- single sign-on point to C3Grid (AAI)
- data queries
- workflow/job submission

Data Management Service DMS
- makes data available for grid applications
- maintains, exports and publishes local workspace data
- supports the workflow scheduler by performing its own planning of future transfers

Workflow Scheduler
- coupled with DMS via web services
- atomic tasks described in JSDL
- complex workflows described by C3Grid-specific Workflow Specification Language (WSL)



Data Information Service DIS
- based on the Open Archives Initiative Protocols (OAI-PMH) and Apache Lucene
- fast full text search engine
- efficient temporal and spatial range queries
- harvesting of C3 published metadata
- fully Java-based and directly integrated into the portal to avoid performance losses

AAI
- preliminary work on implementation of Shibboleth based C3 security infrastructure
- Identity Management in C3 member institutions

Ressource Information Service RIS
- based on Globus MDS4
- keeps information for resource discovery and monitoring

Metadata
- based on the international schema required for describing geographic information and services (ISO 19115 / 19139)
- C3Grid metadata profile was established with specific extensions
- Data providers have to map their individual metadata schemes to that profile.

local Workspace
- grid applications read/write on local workspaces
- grid data (from data providers as well as intermediate results of workflows) are stored temporarily

Data Archive Interface
- based on standard web service technology
- unifies data access by a common WSDL specification
- hides provider-specific storage infrastructure and preprocessing

Status:
Generation 0 implementation of C3Grid with high attention in development of infrastructure for data discovery and retrieval
- Gridification of two typical diagnosis workflows (stormtrack and humidity flux analysis)

Generation 1 (release in september 2007):
- with data from all data providers
- enhanced workflow scheduling

Further steps:
- integration of further data provider
- user support for gridification of new workflows by prototypic implementation of a Workflow Information Service (WFIS)

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung