

The aim of earth system sciences is a better understanding of the earth system to preserve the planet as the human habitat. Particularly climate changes caused by anthropogenic activities are in the focus. For the last decades the amount of data has increased enormously in this field. On the one hand, due to rapid rise in computing power scientists are now able to use models with higher resolution, to perform long term simulations and to couple several components in complex cumulative simulations producing petabytes of output. On the other hand, the monitoring of earth with satellites results in a second huge data stream for earth system science. The data are stored in distributed archives. Up to now, no uniform access to these data is available. The project „Collaborative Climate Community Data and Processing Grid – C3-Grid“ proposes to link distributed data archives in several German institutions. With the help of grid technologies we will build up an infrastructure for the scientists in climate research which provides tools for effective data discovery, data transfer and processing.

Distributed Grid Environment

The GridSphere Web portal provides a user-friendly interface for the metadata search and for the submission of grid processing jobs. The underlying middleware is based on the distributed Grid environment provided by *Globus Toolkit 4*. Security is assured by the X.509 certificate-based *Grid Security Infrastructure* (GSI) and the inter-institutional *Shibboleth* authentication and authorization infrastructure.

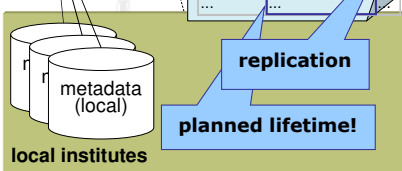
Handling of Data-specific Metadata

discovery metadata:			
title	variables	bounding box	ID
Optical thickness	temperature	some region	23
Continuous meteor	air_pressure	some region	42
Physical oceanog	humidity	far away	33

grid services

uri	lifetime	ID
gsiftp://a.c01/1987-12/200	23	23
gsiftp://b.c01/2987-12/300	23	23
gsiftp://c.c01/1987-12/200	33	33

OAI-PMH protocol



Main Middleware Components

Workflow Scheduler

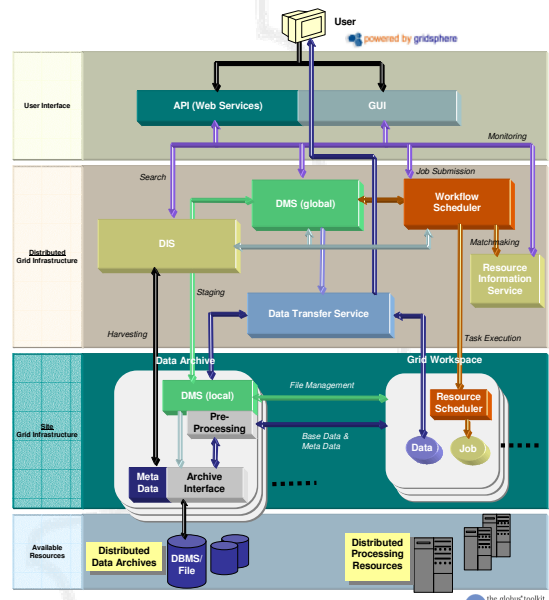
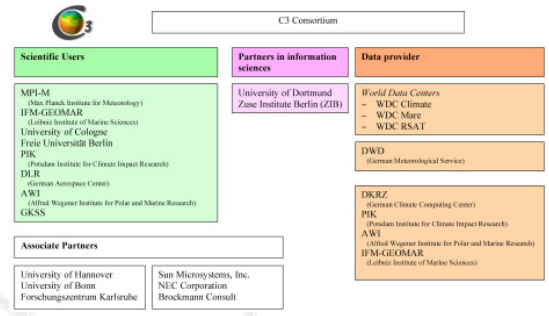
- workflow (DAG-based) scheduling; management of inter-dependent jobs
- Coordinated job and data scheduling; orchestration of staging and processing tasks (matchmaking and planning)
- technology: *Globus Toolkit 4*, *GRMS*, *JSDL*

Data Management Service (DMS)

- data staging from local institutes: tape archives, DBMS, file systems
- data reduction and pre-processing on site
- replication and caching for data locality
- co-scheduling, planning with prediction of data transfer and staging times
- technology: *GridFTP*, *Reliable File Transfer*

Data Information Service (DIS)

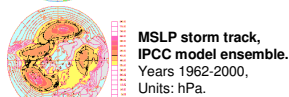
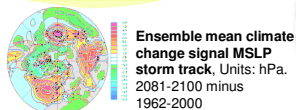
- community-specific metadata ISO19115
- Google-like full-text and range search on metadata
- technology: *Apache Lucene*, *OAI-PMH* meta-data harvesting



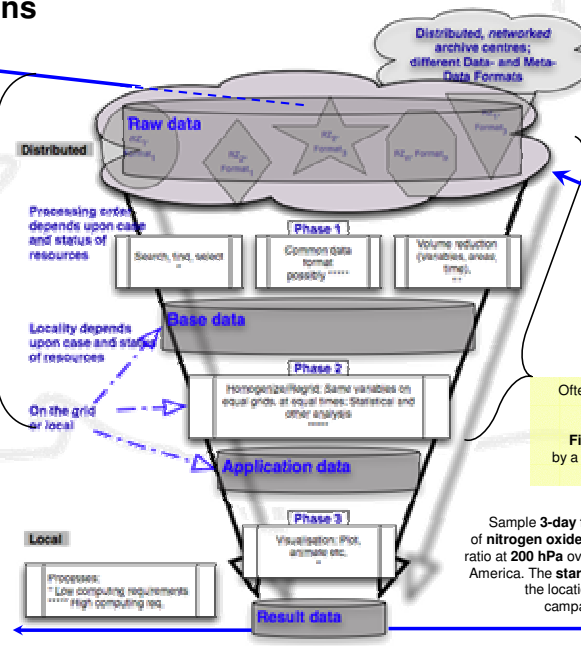
Different users applications

Data Near Processing: This processing is a part of numerous meteorological workflows by doing variable selection, time-space cut, format conversion, and simple processing like time averaging for data amount reduction. It processes the data sets from "Raw Data" to "Phase 2". The central processing tools are the CDOs (Climate Data Operators: <http://www.mpimet.mpg.de/~cdo>).

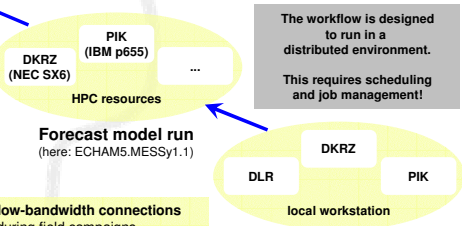
Diagnosis tools: Calculating important key-values by using different datasets and various values of huge data amounts (e.g.: storm track, cape, water fluxes, eady parameter). Processing from "Phase 2" to the visualization of the result data (Tools created by the Meteorological Institutes of the Universities of Cologne and Berlin).



Shifted and enhanced cyclone activity close to Europe



The workflow "chemical weather forecasting" will provide forecast charts of the atmospheric chemical composition. These forecasts are essential for planning flight routes during field measurement campaigns.



Often: low-bandwidth connections during field campaigns. Filtering and volume reduction by a "data near processing" workflow is essential!

