



Collaborative Climate Community Grid C3-Grid



Grid Computing for the Climate Community

C3-Grid will improve the means of collaboration in the German climate and earth system research community.

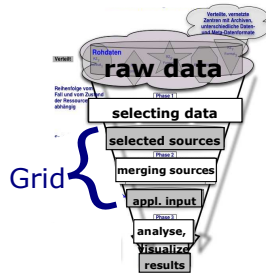
Therefore the following new features will be established:

- share local **high-volume data** with other researchers.
- remote computations allow **remote analysis of model and observation data**.
- **preprocessing** at data providers reduces transferred data.
- fast transfers via **high-bandwidth network** connections.
- **common metadata scheme** provides interoperability.
- cooperating **replica- and job-management** avoid unnecessary transfers.

Workflows

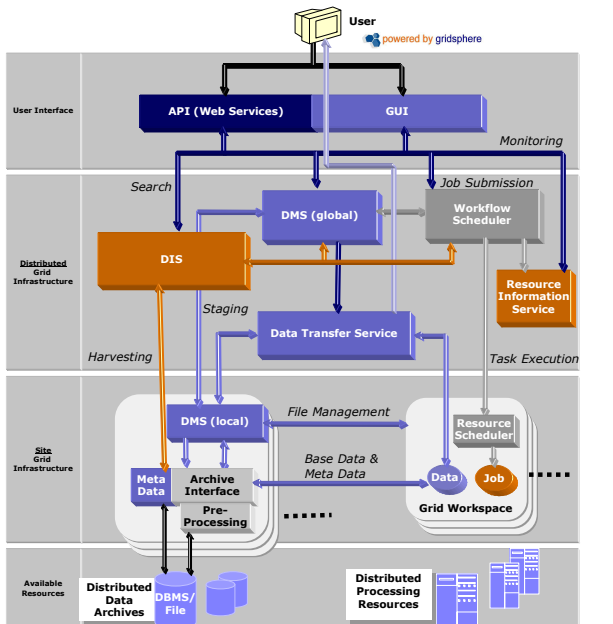
C3-Grid focuses on three use cases:

- **analysis** of different global simulations
- **regional climate** simulation using data from global models
- accompanying actual measurements



Existing applications are run on the grid. They are not modified but composed to grid workflows.

Services and Components

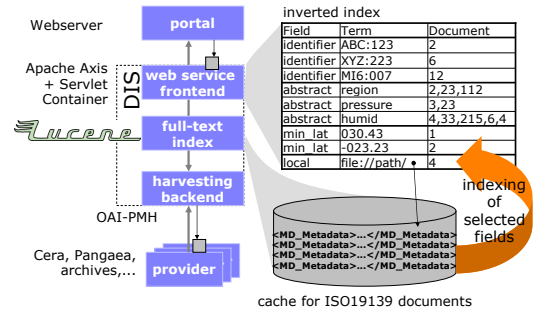


DIS: Data Information Service
DMS: Data Management Service

Distributed Grid Infrastructure

C3Grid is based on the Globus Toolkit 4, X.509 proxy certificates (GSI) and the **inter-institutional Shibboleth** authentication and authorization infrastructure.

Data Information System



The Data Information System (DIS) is aware of the community-specific data.

- 3D mesh of geographic variables in discrete time steps
- scientific description as **ISO 19115/139** metadata
- huge volume metadata (10s of GB)

Featured search capabilities:

- **ranges** (longitude, latitude, altitude, time)
- **full-text** (authors, abstract, ...)

Implementation:

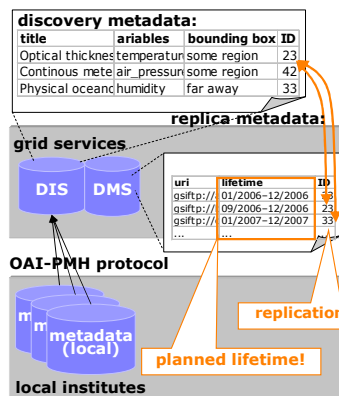
Apache Lucene, OAI-PMH, Apache Axis, ...

Data Management System

The Data Management System (DMS) stages data from primary sources, manages replicas and inter-workspace data transfers.

Challenges: Plan the future!

- at which sites will file A exist tomorrow at 11:00 am?
- copy file A from X to Y today between 6 pm and 10 pm
- How long will a transfer take?
- When can file X be restored from tape to disk earliest?



Co-scheduling

The Workflow Scheduler relies on the transfer plans of the DMS for its own scheduling.

Provenance tracking

Answers: How were these data produced? Which data were derived from this?

Implementation: based on GridFTP, Reliable File Transfer (RFT) and Apache Lucene

