Enabling Grids for E-sciencE

Requirements of Climate modelling applications on Grid infrastructures

C3-Grid and EGEE

Joachim Biercamp,
Kerstin Ronneberger
Stefan Kindermann

www.eu-egee.org

Information Society and Media

Enabling Grids for E-sciencE

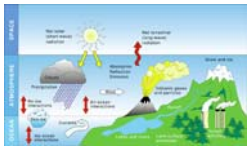
Overview

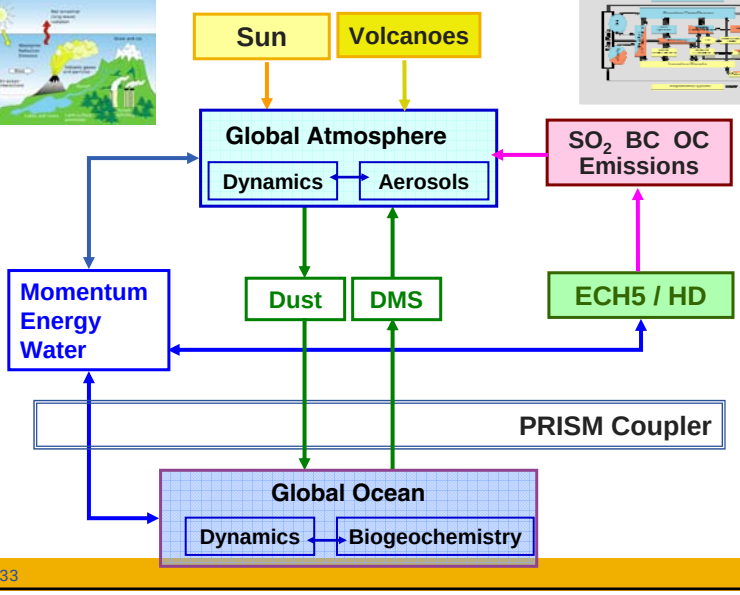
- **Climate Modeling Applications**
  - Characteristics
  - Potential Grid Added Value
  - Examples / (Possible) Case Studies
- **C3-Grid:**
  - Tackling the data challenge in collaborative climate modeling

INFSO-RI-508833

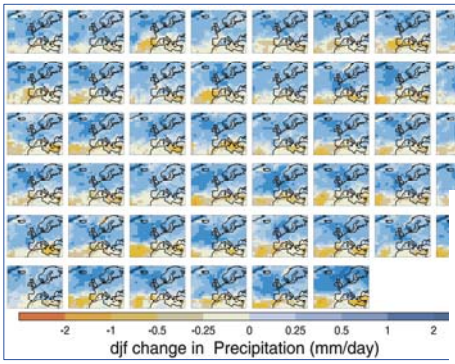
Models / Tools

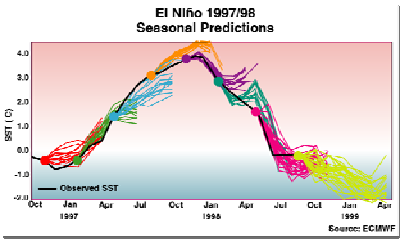
Modelling the Earth System

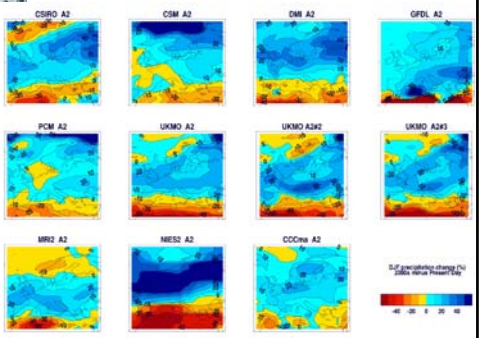




"Ensembles"







- **Climate Modeling …**

  …. is very compute intensive

  …. is very data intensive

  …. requires close international collaboration

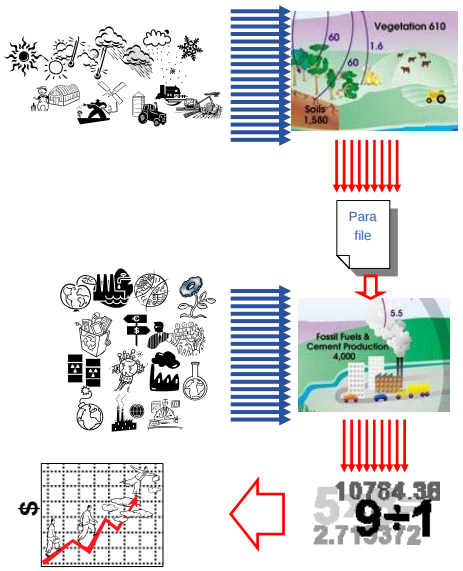
  …. has various degrees of high complexity

Grid Added Value ?!

(Case Studies)

Project example A : ENFA

- **Aim**
  - Rate cost-efficient mitigation and adaptation strategies for the land use sectors to react to climate change impacts
- **Strategy**
  - estimates of crop productivity for ensembles of different environmental and management scenarios by means of soil dynamic and crop growth models
  - Assessment of economic and emission implications by means of an economic land use choice model using the productivity estimates as input



http://www.uni-hamburg.de/Wiss/FB/15/Sustainability/ENFA/

Grid added value

Enabling Grids for E-sciencE

Benefits

- Solution time reduction by means of parallel computing of parametric jobs
- Shared platform for the 13 institutes from 9 different nations
- Common and consistent data management eases collaborative analysis

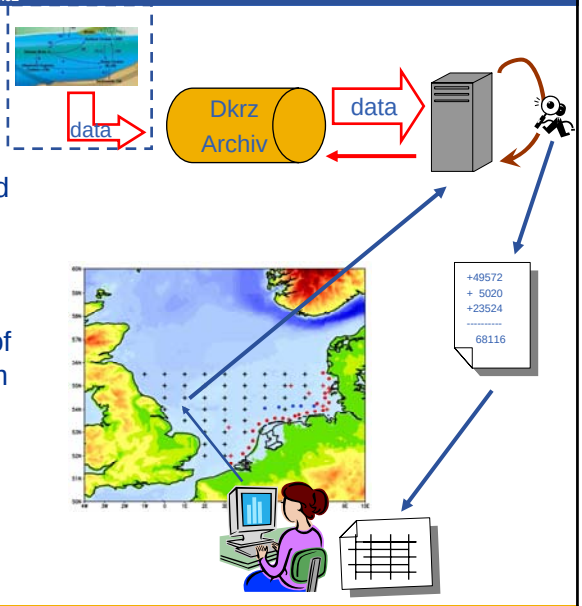
Requirements

- Port some of the models from windows to Linux and solve the respective licence problem (used software: GAMS)
- Define a secure shared area (voms?)

Project example B : NordseeAtlas

Enabling Grids for E-sciencE

- **Aim**
  - Provide statistical predictions of location and situation specific sea level variations and related parameters to the interested public
- **Strategy**
  - Statistical accounting of high resolved long term data, produced by hindcast-models of ocean currents, near-surface wind and sea level rise



Grid added value

Enabling Grids for E-sciencE

Benefits

- Distributed access for the public via standardized web-interfaces
- Distributed data storage enables parallel searching
- Direct processing (in contrast to pre-processed "answered") allows for flexible "questions"

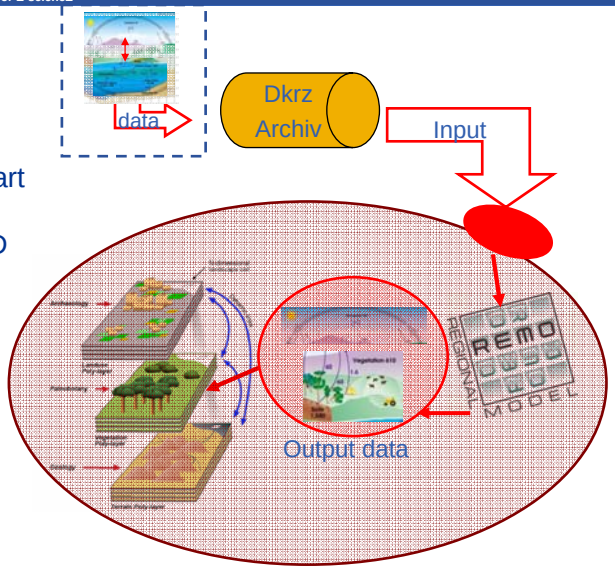
Requirements

- Transfer the data to - and store them on grid storage
- Provide a web-interface

Project example C : ArchaeoGRID

Enabling Grids for E-sciencE

- **Aim**
  - Dynamically simulating Paleoclimate as part of ArchaeoGRID (see ArchaeoGRID presentation)
- **Strategy**
  - Drive the regional climate model REMO with large scale climate predictions.



EGEE Infrastructure

Grid added value

Enabling Grids for E-sciencE

Benefits

- Common platform for interdisciplinary collaboration
- (see ArcheaoGRID slides)

Requirements

- Secure and direct access to external data archive
- Storage and management of the output data of REMO (many terabyte per hindcast)

Enabling Grids for E-sciencE

The data challenge

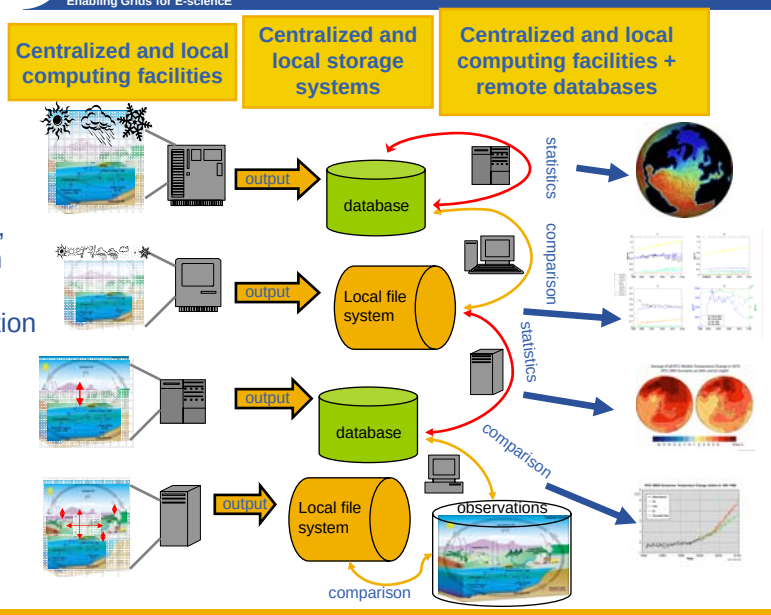
- Common to all climate modelling applications
- Solutions:
  - C3-Grid
  - EGEE

INFSO-RI-508833

Analysis

Enabling Grids for E-sciencE

- Validation, calibration and interpretation of model results

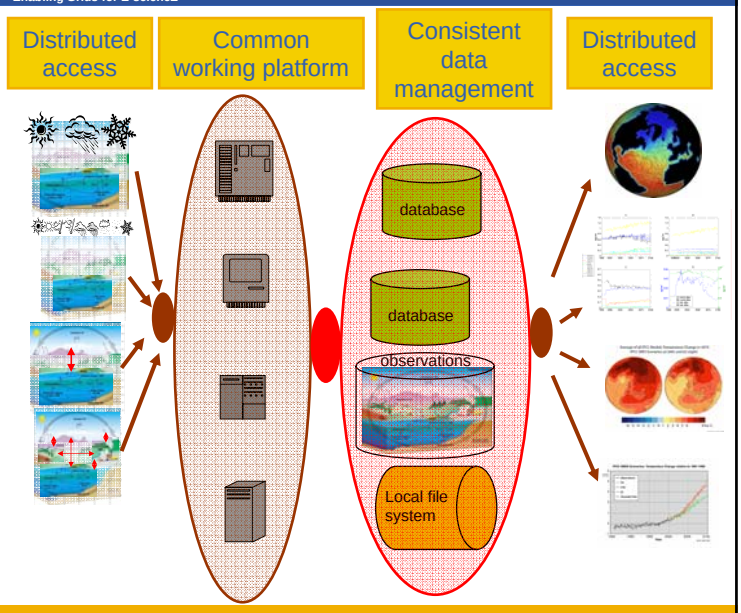


INFSO-RI-508833

Grid added value

Enabling Grids for E-sciencE

- Distributed access from different institutes to
- Common working platform
- With consistent data management



INFSO-RI-508833

Climate data analysis workflow

Enabling Grids for E-sciencE

Raw Data:
+ Metadata ( )

Databases

WDC Climate

...

Files (e.g. in HSMs)

DKRZ

...

Search / Selection based on Metadata

- Different Metadata schemata or no metadata available
- Heterogeneous search interfaces

Access

Preprocessing / Filtering

- Different interfaces and authorization mechanisms
- Large data sets
- Different filtering functionalities at data providers

- large data sets of gridded data (e.g. netCDF, GRIB)
- "implicit" metadata: file naming (besides file headers / search history)

- Workflows adapted to existing infra-structures
- interactive parts

Homogenization / Regridding

Variable(s) extraction / processing

statistical analysis

visualization / archiving / publishing (+ metadata)

EGEE and climate data analysis

Enabling Grids for E-sciencE

Climate data analysis resources

- Metadata databases
- Data databases
- HSMs
- Dedicated interfaces

...

WDC Climate

...

GFS/ PFS

- specific (parallel/global) file systems
- infrastructure adapted data analysis packages

? Resources ?

EGEE infrastructure:

? Metadata ?

? Data ?

- No community metadata catalog
- No standard interfaces to integrate external data providers (e.g. databases)
- Difficult to port to community ressources
- No support to integrate community authorization mechanisms

VOMS

RB

UI

EGEE file catalogs

RGMA

SE

SE

CE

→ Interfaces: C3Grid perspectives

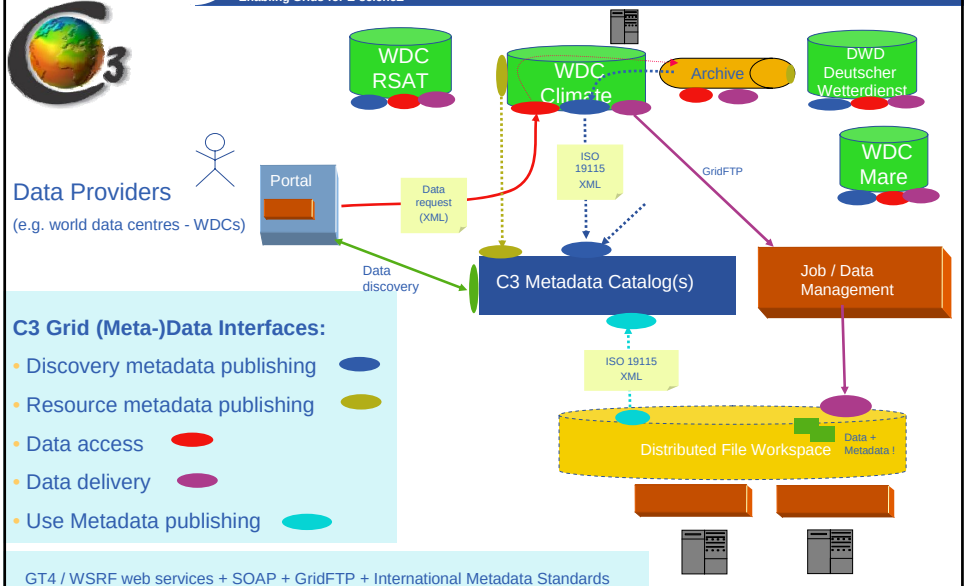
GEANT Networking Infrastructure

C3-Grid

Collaborative Climate Community Data and Analysis Grid

C3Grid Components and Interfaces

Enabling Grids for E-sciencE



C3 Grid (Meta-)Data Interfaces:

- Discovery metadata publishing
- Resource metadata publishing
- Data access
- Data delivery
- Use Metadata publishing

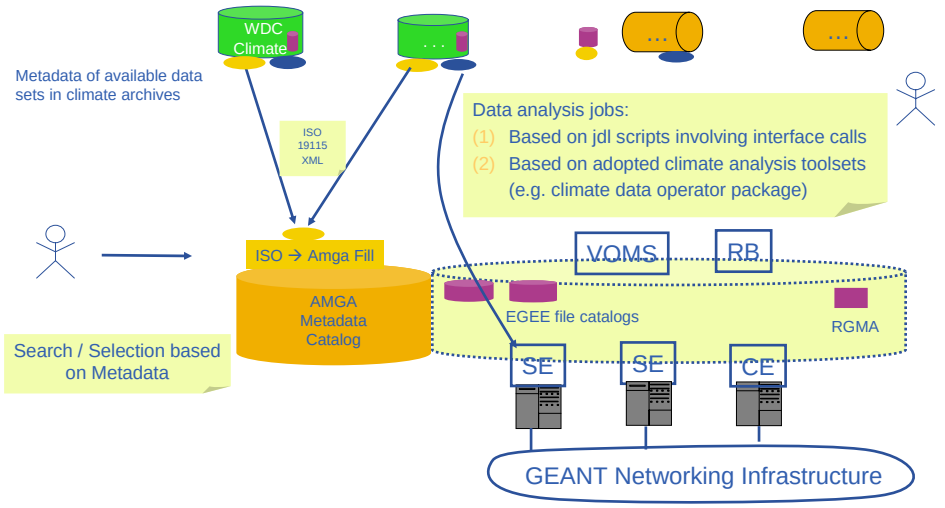
GT4 / WSRF web services + SOAP + GridFTP + International Metadata Standards

INFSO-RI-508833

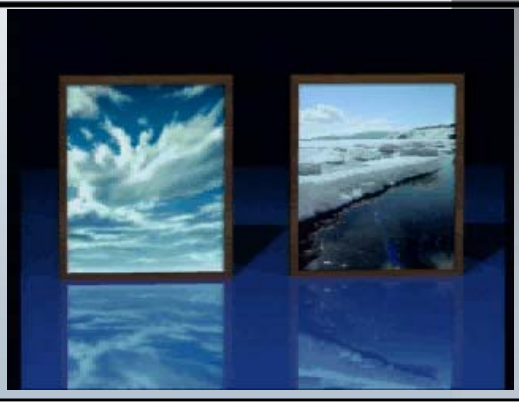
EGEE and climate data analysis

Enabling Grids for E-sciencE

- **EGEE + C3Grid interfaces → enabling climate data analysis workflows**



INFSO-RI-508833

Thank you

Joachim Biercamp (biercamp@dkrz.de)