



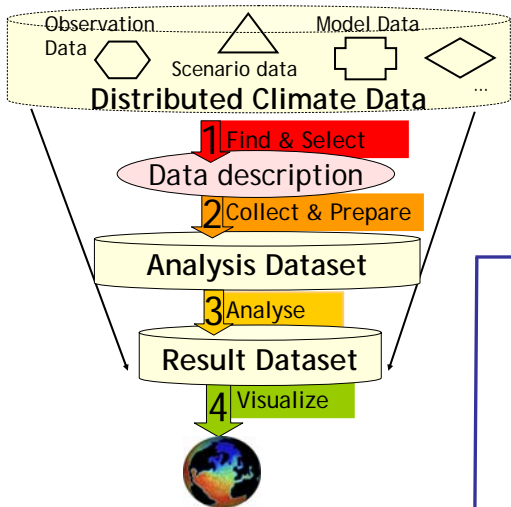
Grid Based Climate Data Analysis

Kerstin Ronneberger, Stephan Kindermann, Joachim Biercamp

Background

Climate research

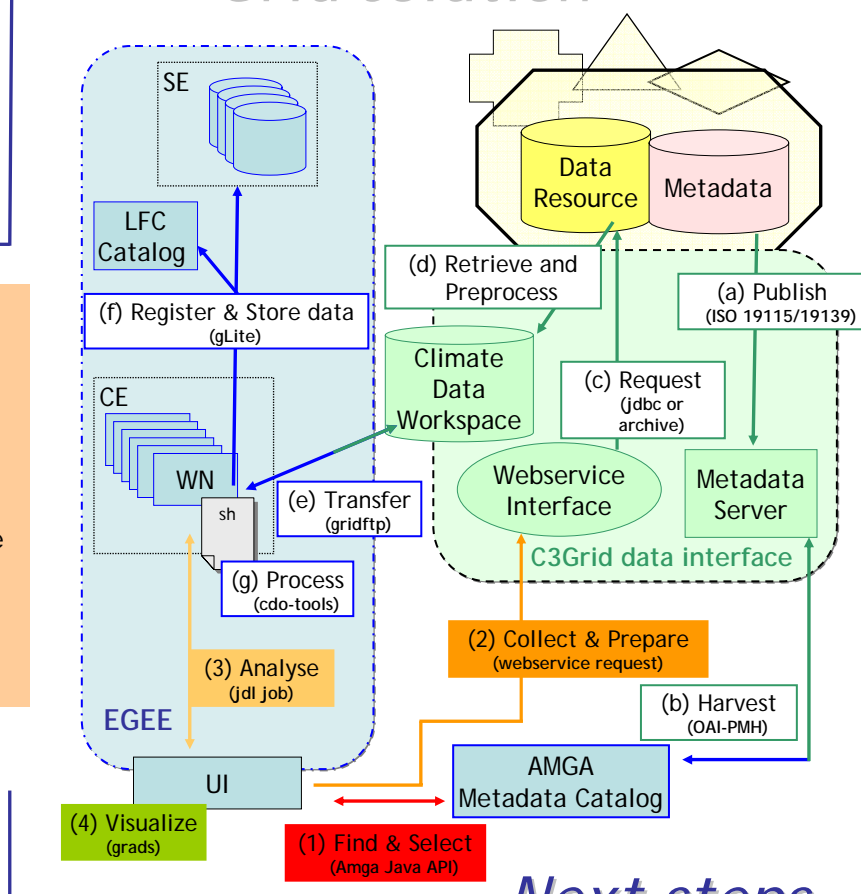
Is largely based on analysis of large datasets; A typical workflow basically consists of four steps:



Addressed problems

- The data is stored in distributed data centers (e.g. world data centers)
 - No central metadata catalog based on common metadata schema
 - Different data access interfaces with different AA policies
- A common platform with uniform access and standardized data description is needed

Grid solution



C3Grid & EGEE

- In the German C3Grid Project[1] a grid-enabled data interface has been developed to uniformly find and access data of the distributed climate data centres
- By using the C3-tools together with gLite software, climate analysis workflows can be ported to the EGEE infrastructure
- In this way EGEE can offer the common platform, needed for climate research to share, access and analyse climate data

Gridification-status of German climate-data providers

Data Centers	Current Volume	Grid enabled
DKRZ Archive (MPI-M/others)	~ 4 PB	~3 TB
WDCs (Climate/Mare)	~200 TB	~5 TB
IFM Geomar	~1 TB	~100 GB
DWD	~200 GB	The rest is coming soon...
FUB	~1 TB	
PIK	~700 GB	
AWI	~300 GB	
DLR	~60 GB	

Planned Extensions

- GUI/Portal components for AMGA interactions
- Tagging of EGEE-registered files in Amga
- Integrating the visualization in the EGEE workflow
- International cooperation with NDG[2] and ESG[3]

Next steps

Open issues

- Direct transfer of external files to, and registration in EGEE
- Interoperability of the AA infrastructures of EGEE and C3

References

- [1] Collaborative Climate Community Data and Processing Grid: <http://www.c3grid.de>
 [2] Nerc Data Grid: <http://ndg.badc.rl.ac.uk/>
 [3] Earth System Grid: <http://www.earthsystemgrid.org/>