Wissenschaftlicher Einsatz von Grid-Middleware am Beispiel Erdsystemwissenschaft

Scientific Application of Grid Middleware in Earth System Science

M. Stockhause, MPI-M / IFM-GEOMAR
S. Kindermann, DKRZ
S. Makedanz, AWI

D-Grid Symposium, Informatik 2007, 25.09.2007, Bremen

Grids in
Earth System Sciences (ESS)

Grids in Earth System Science

Earth System Modelling:

- **HPC applications of coupled models**
- **Compute intensive**

Earth System Research / Data Analyzation:

- **Platform independent applications**
- **Data intensive**

Frameworks for Earth System Modelling: ESMF, COSMOS

- partly automatic coupled model instancing
- not grid enabled
- based on standardisation initiatives: ESMF, PRISM

Earth System Grid (ESG):

- data access of homogeneous data
- file archives
- central admin.

NERC DataGrid (NDG):

- data discovery
- data access
- federated data bases and file archives

Collaborative Climate Community Grid (C3Grid):

- data discovery
- uniform data access
- data processing
- federated data bases and file archives

Common Information Model (CIM), e.g. ESC, METAFOR

- software environment for assembling, running, and archiving information about earth system models
- early stages of development

C3Grid Consortium

Scientific Users

MPI-M (Max Planck Institute for Meteorology)
IFM-GEOMAR (Leibniz Institute of Marine Sciences)
University of Cologne
Freie Universität Berlin
PIK (Potsdam Institute for Climate Impact Research)
DLR (German Aerospace Center)
AWI (Alfred Wegener Institute for Polar and Marine Research)
GKSS

Partners in Information Sciences

University of Dortmund
Zuse Institute Berlin (ZIB)

Scientific Data Provider

World Data Centers
WDC Climate
WDC Mare
WDC RSAT

DWD (German Meteorological Service)

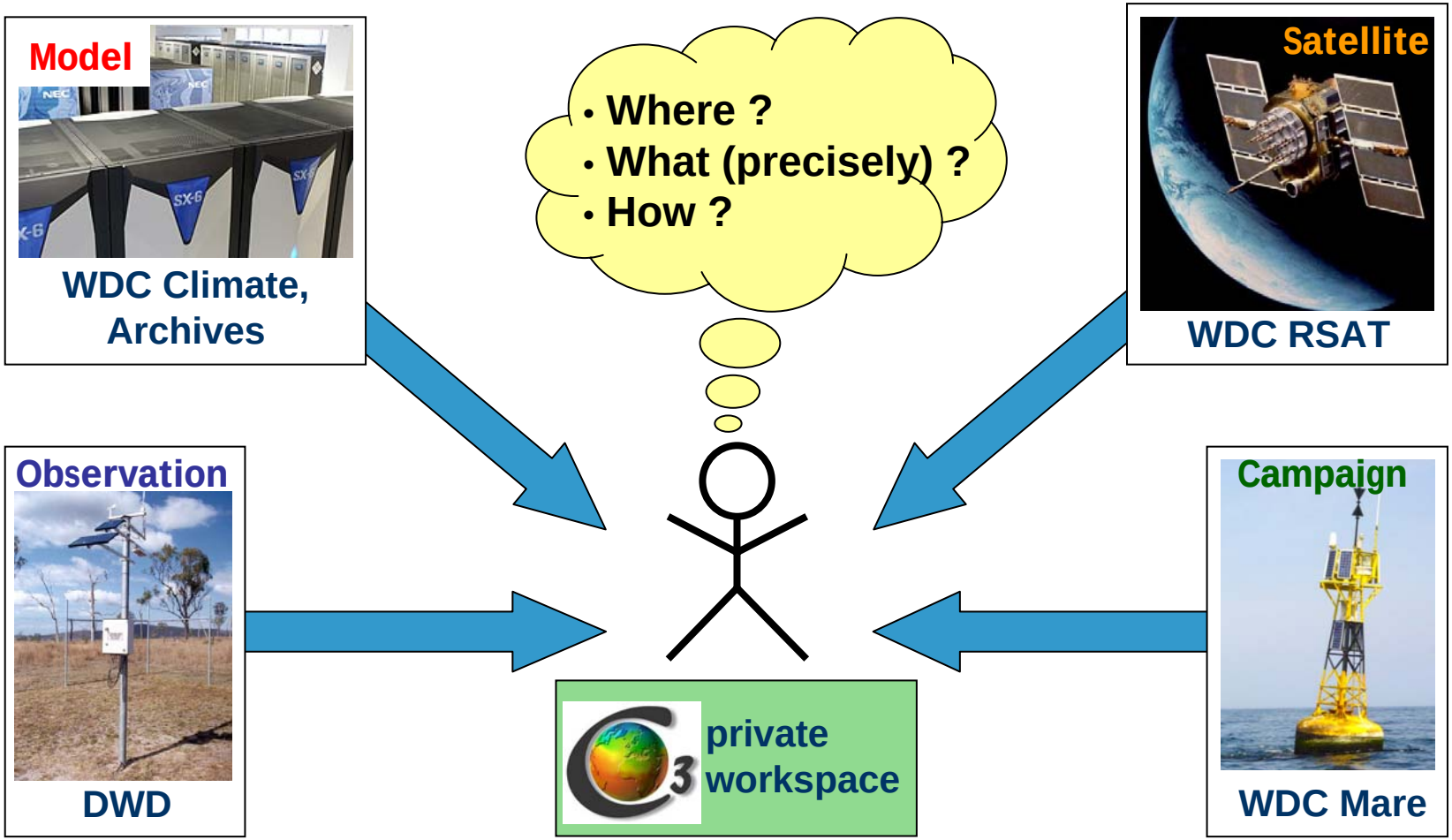
DKRZ (German Climate Computing Center)
PIK (Potsdam Institute for Climate Impact Research)
AWI (Alfred Wegener Institute for Polar and Marine Research)
IFM-GEOMAR (Leibniz Institute of Marine Sciences)
University of Cologne / ZAIK / FUB

Associate Partners

University of Hannover
University of Bonn
Science Center Karlsruhe

Sun Microsystems, Inc.
NEC Corporation
Brockmann Consult

Data for Scientific Workflows



C3Grid Middleware

Service Oriented Architecture - SOA

- Data Services
- Compute Services
- Metadata Services
- Security Services

Data Services – Data Management

SRM-dCache

Storage Resource Management (SRM) with dCache-protocol using GridFTP for file access

- Scientific Linux 3 (HEP)
- non-commercial
- for file archives

SRB / iRods

(Storage Resource Broker / i Rule Oriented Data Systems)

- SRB: cumbersome licencing process
- GT4 offers SRB-DSI (Data Storage Interface) using GridFTP
- iRods: new, some features missing (GSI)
- for file archives used

OGSA-DAI

- for data bases
- problems with blobs (DGI)
- asynchronous access necessary

C3Grid Data Request WS

Requirements:

- for file archives and data bases
- data amount reduction

Solution:

- C3Grid specific WSDL with data reduction functionality
- local data management solutions
- usable as wrapper on other solutions, e.g. OGSA-DAI at DWD

other solutions for data management

Data Services – Data Transfer

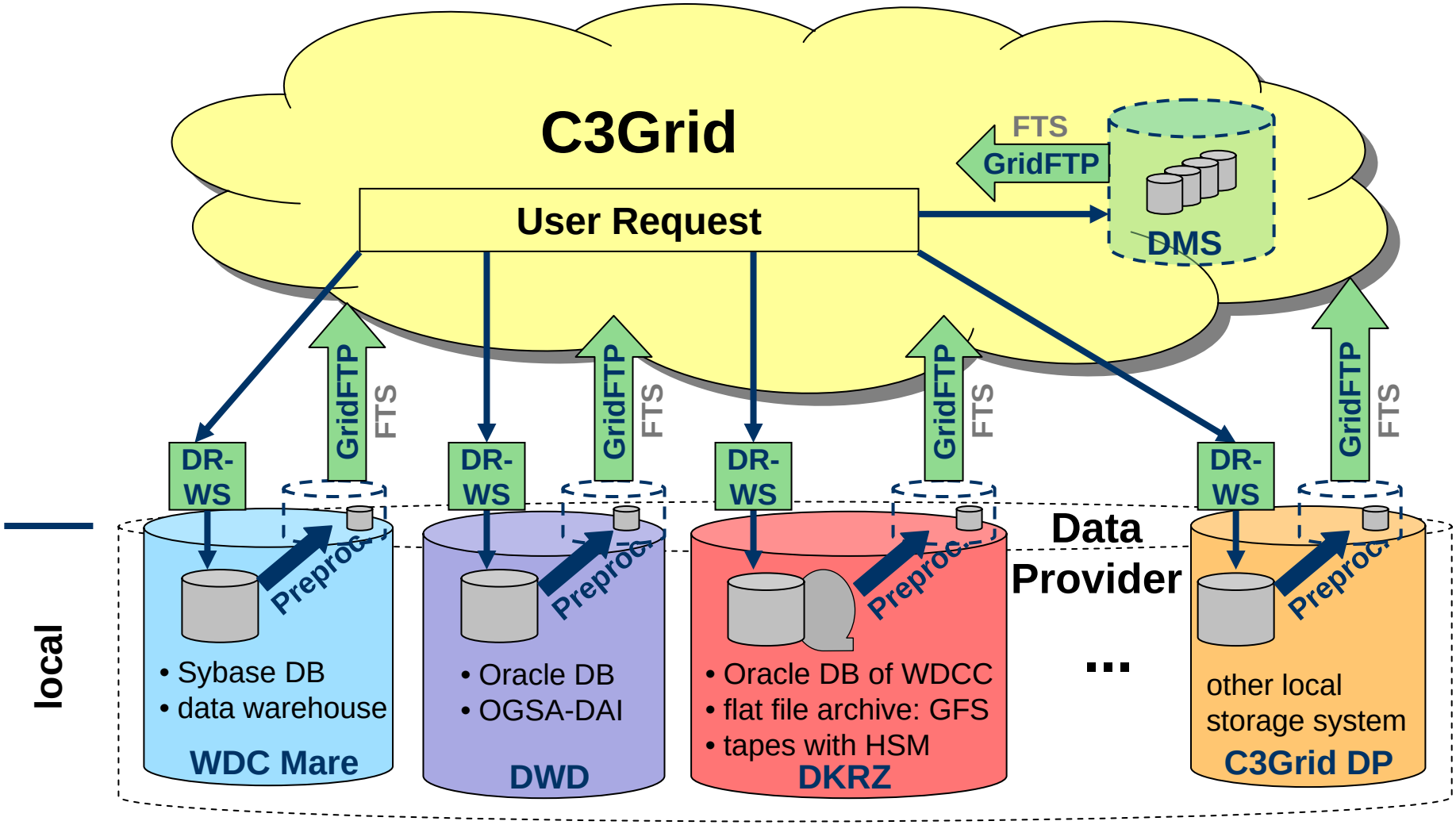
GridFTP
- low level service for secure transfer of files

RFT
(Reliable File Transfer)
- Management of multiple file transfer, success control by guaranteed finishing of transfer
- Webservice and database for protocolling
- status informations about data, transfer(s)

C3Grid Transfer Service
Requirements:
- estimation of needed transfer time for (large) files

Solution:
- **GridFTP** based service
- **C3Grid File Transfer Service**: estimation of transfer time and action triggered when exceeded (cancellation)

Data Service – C3Grid



Compute Services

UNICORE

- vertical integrated solution
- uniform interface for different HPCs
- parallel development to GT4

the globus toolkit

WS-GRAM

(Globus Resource Allocation Manager)

- WS interface for clients for job submission, monitoring and cancellation
- **MDS - WSRF** publishs information about local scheduler – queue
- **RSL** (Resource Specification Language) for job specification – job submit

gLite

- specific solution for: GT 2.4, Scientific Linux 3
- updates in progress: GT 4, Scientific Linux 4

C3Grid Compute Service

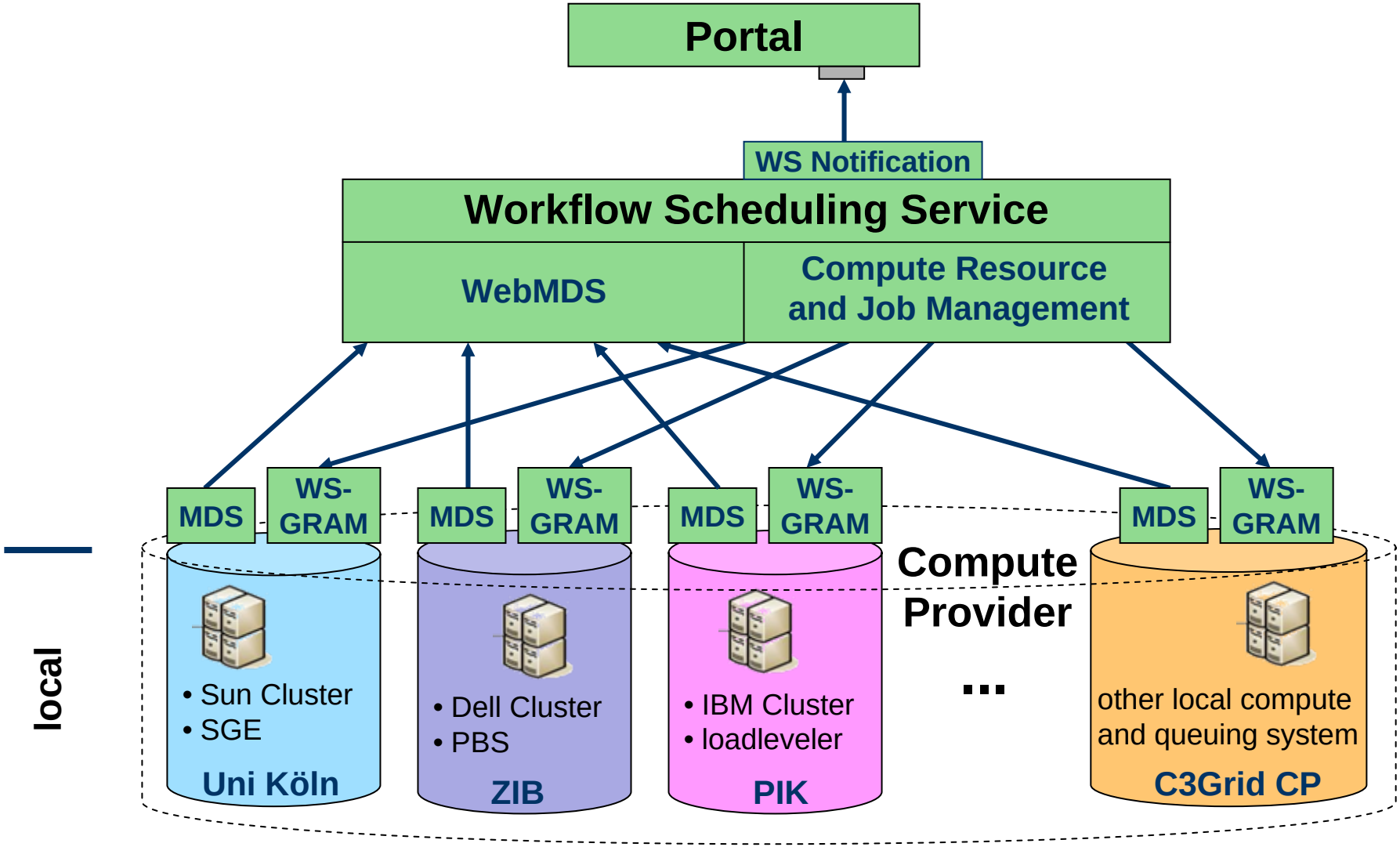
Requirements:

- Data-Job-Coscheduling

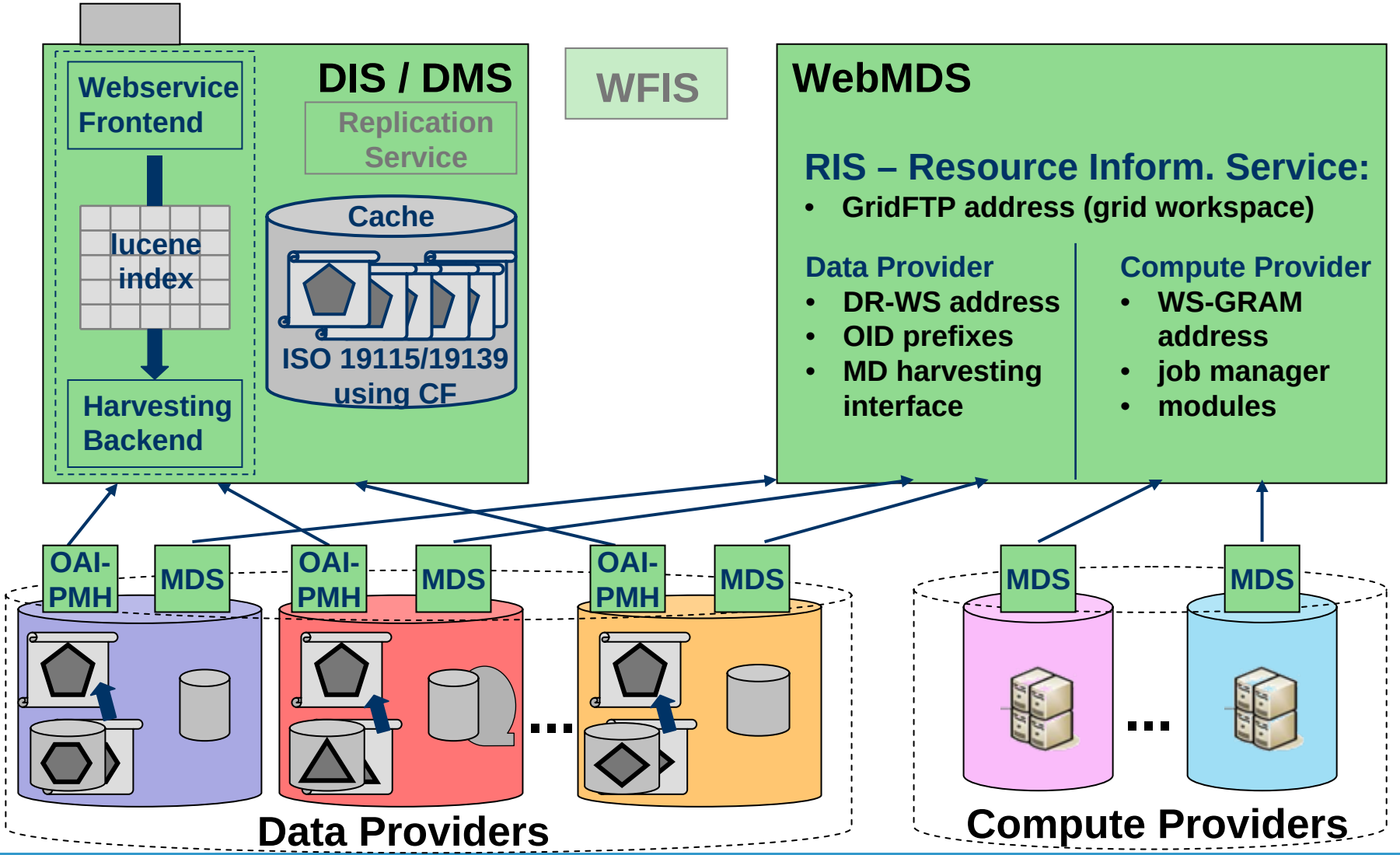
Solution:

- **WS-GRAM:** usage at local compute providers
- **WSS (C3Grid Workflow Scheduling Service):** central component for
  - collection of WS-GRAM data
  - C3Grid job management
  - C3Grid compute resource management
  - monitoring (and control) for user layer: WS Notification

Compute Service – C3Grid



Community Metadata Services C3Grid



Security Services

GSI (Grid Security Infrastructure)

- personalized secure communication
- based on X.509 personal certificate
- sigle-sign-on, delegation by proxy certificates

GridShib

- **Shibboleth:** federated attribute-based authZ
- **GSI:** personalized secure communication
- Delegation by proxy certificates with SAML-assertions

Shibboleth

- federated fine granular authorization based on attributes (for VOs)
- IdP give authorization information on request
- SAML based

C3Grid Security Service

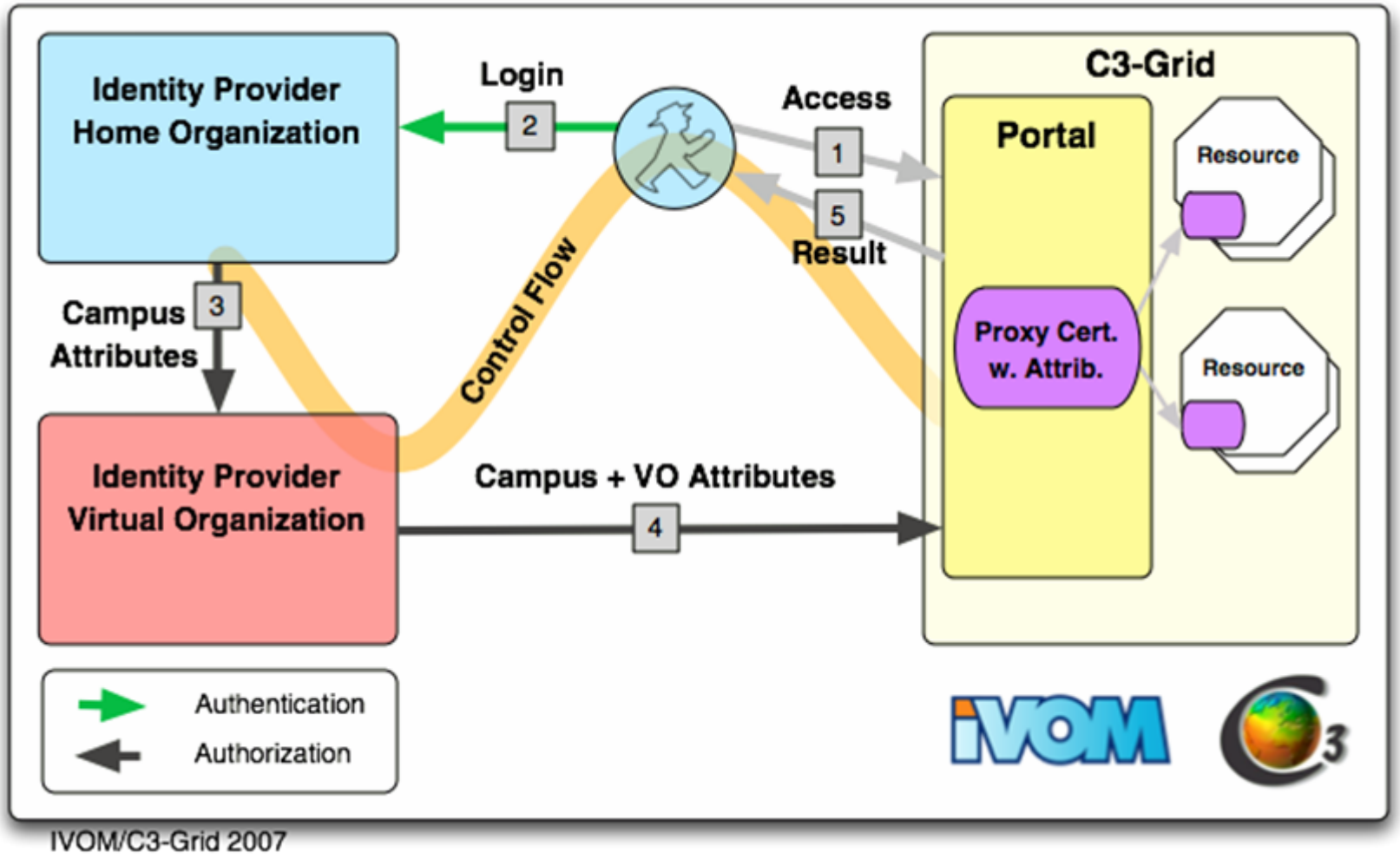
Requirements:

- description of fine granular access constraints
- changing user roles (sc. project)

Solution:

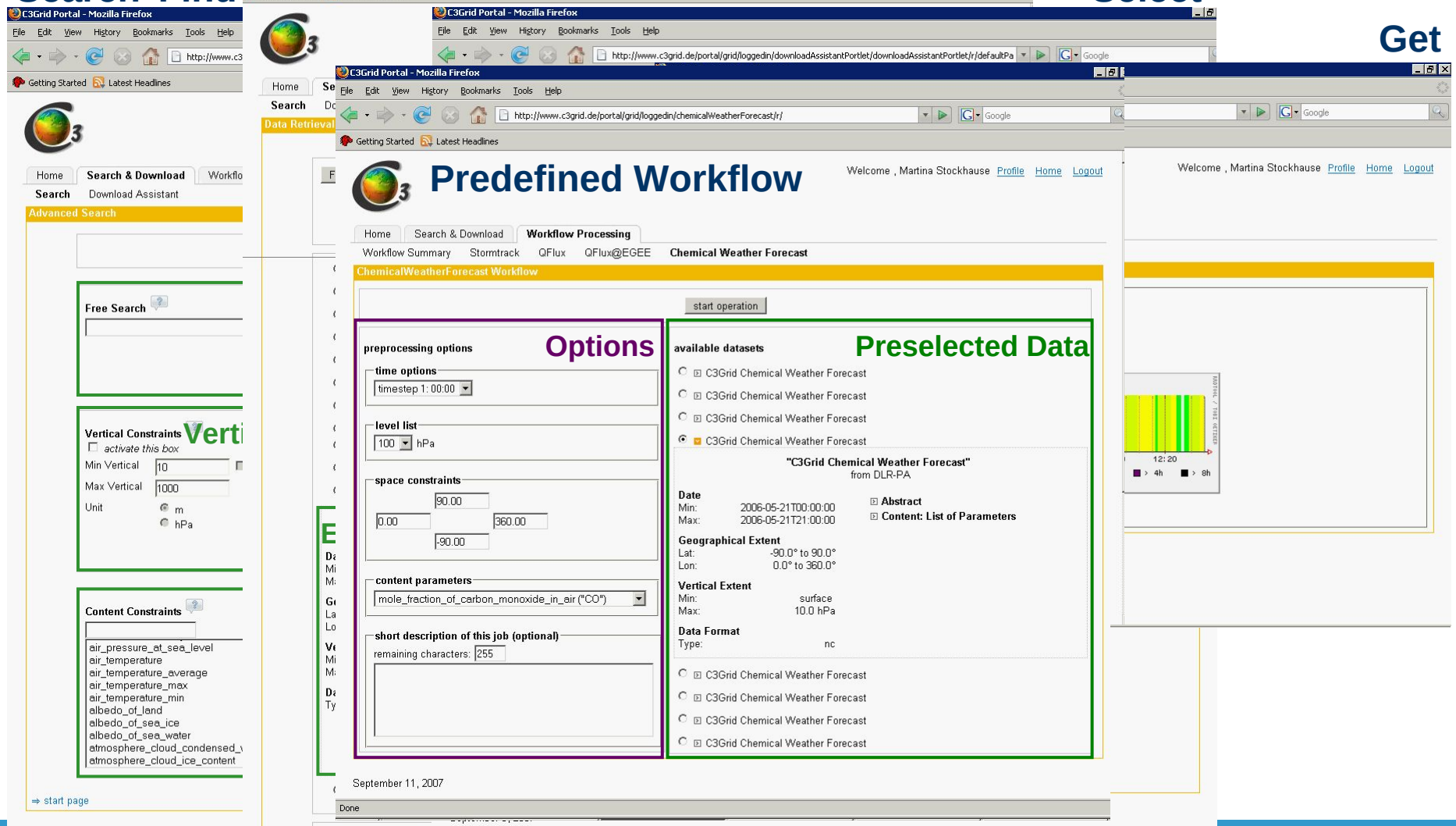
- **GridShib** with C3Grid specific Shibboleth attributes
- myVocs for VO Management planned

Security Services – C3Grid

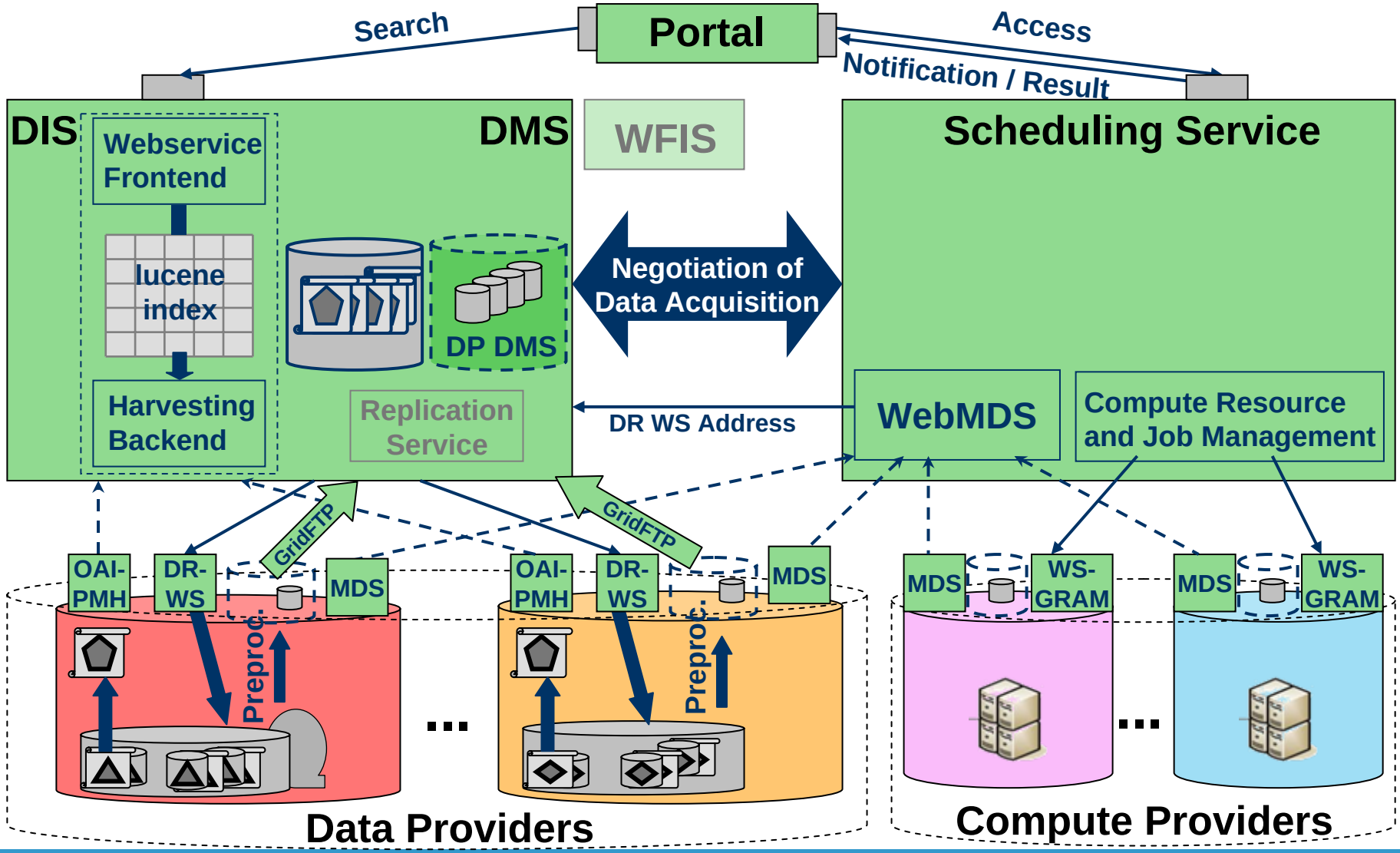


Conclusion

C3-Grid: Portal



C3-Grid Middleware



Open Issues

Open Issue	short	mid range	long
Security: GridShib			
Replica			
(Meta)data Republishing			
Data Archiving			
Data-Job-Coscheduling			
Workflow Information Service (WFIS)			
User Friendliness			

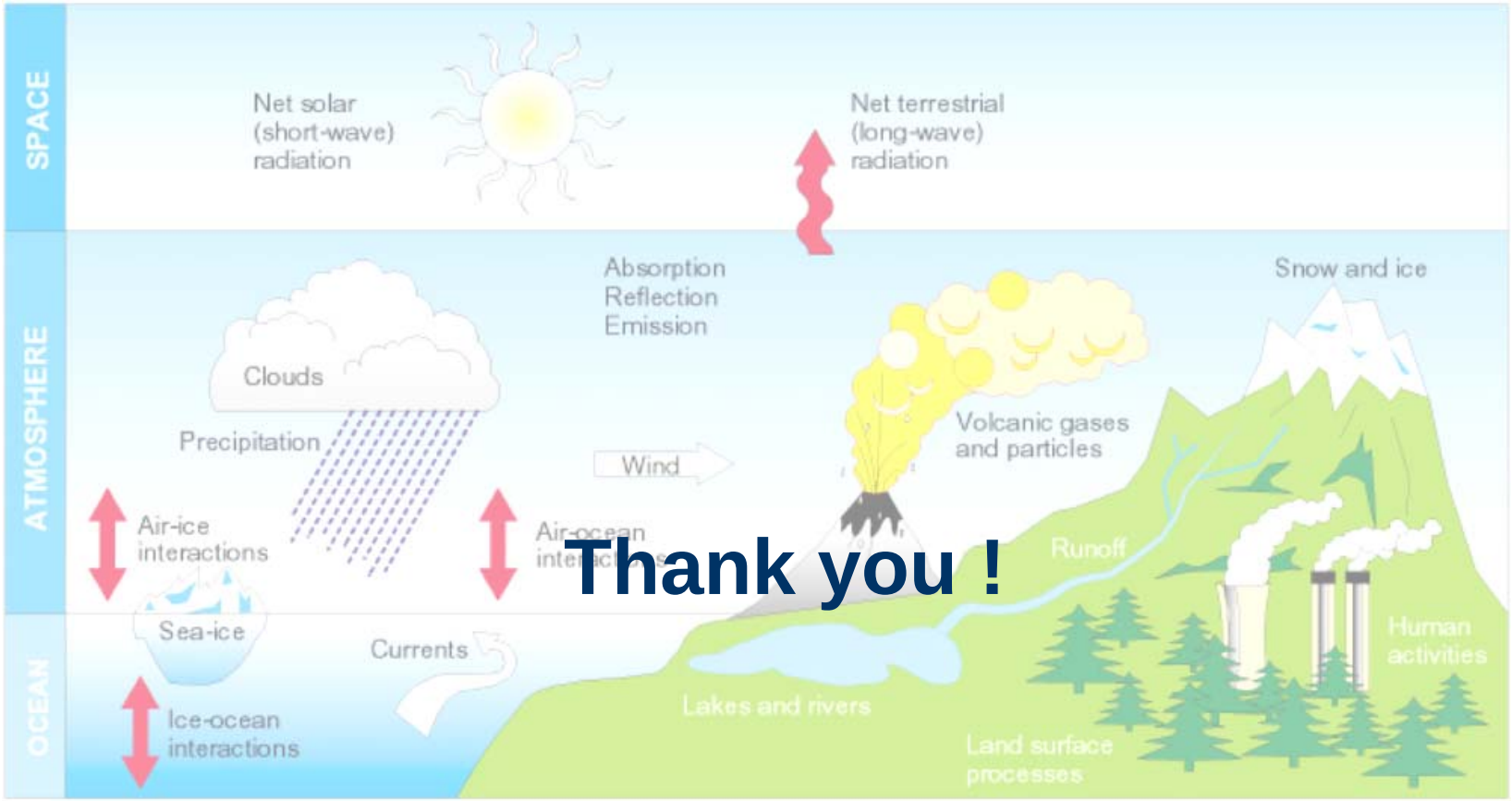
Outlook

IPCC-AR5:

- WDC Climate as one of federated data storage locations
- Provision of climate parameters ‘on demand‘
- Use of C3Grid Data Request functionality and Compute Resources in “griddified“ preprocessing of DKRZ / CERA DB

User groups with different applications and demands:

- working platform (private workspace) for
  - earth system scientists and
- information platform (provide derived data ‘on demand‘) for
  - scientists of other disciplines
  - non-scientists in economy and public authorities
  - education



Thank you !

www.c3grid.de

martina.stockhause@zmaw.de