

The Climate-G testbed towards a large scale data sharing environment for climate change

G. Aloisio

Scientific Computing and Operations Division Head, CMCC & University of Salento, Italy

S. Fiore

Scientific Computing and Operations Division, CMCC, Italy

EGEE09 Conference

Sept 21, 2009



Scenario, issues and needs

- **Climate community produces huge amount of data** at an international level
- There is a strong need to **share** and **integrate** data among several centres
- **Petabytes** of data related to:
 - **Climate Change data -> century types simulations**
 - **Seasonal to Decadal data -> decennial types simulations**
- **Next generation climate change infrastructures** must provide a seamless environment
- **Open, distributed** and **service-based** approach
- Issues: **data distribution, data format heterogeneity, metadata management, security, transparent access to the system, scalable approach, ...**

Grid and Climate Change: Climate-G

The main goal of Climate-G is to create an open and unified environment for climate change enabling geographical and cross-institutional data discovery, access, analysis, visualization and sharing.



This effort has been conceived as a proof of concept for the involved grid technologies (in particular GReIC grid metadata service) and it is supported by the Earth Science Cluster Community (EGEE Project).

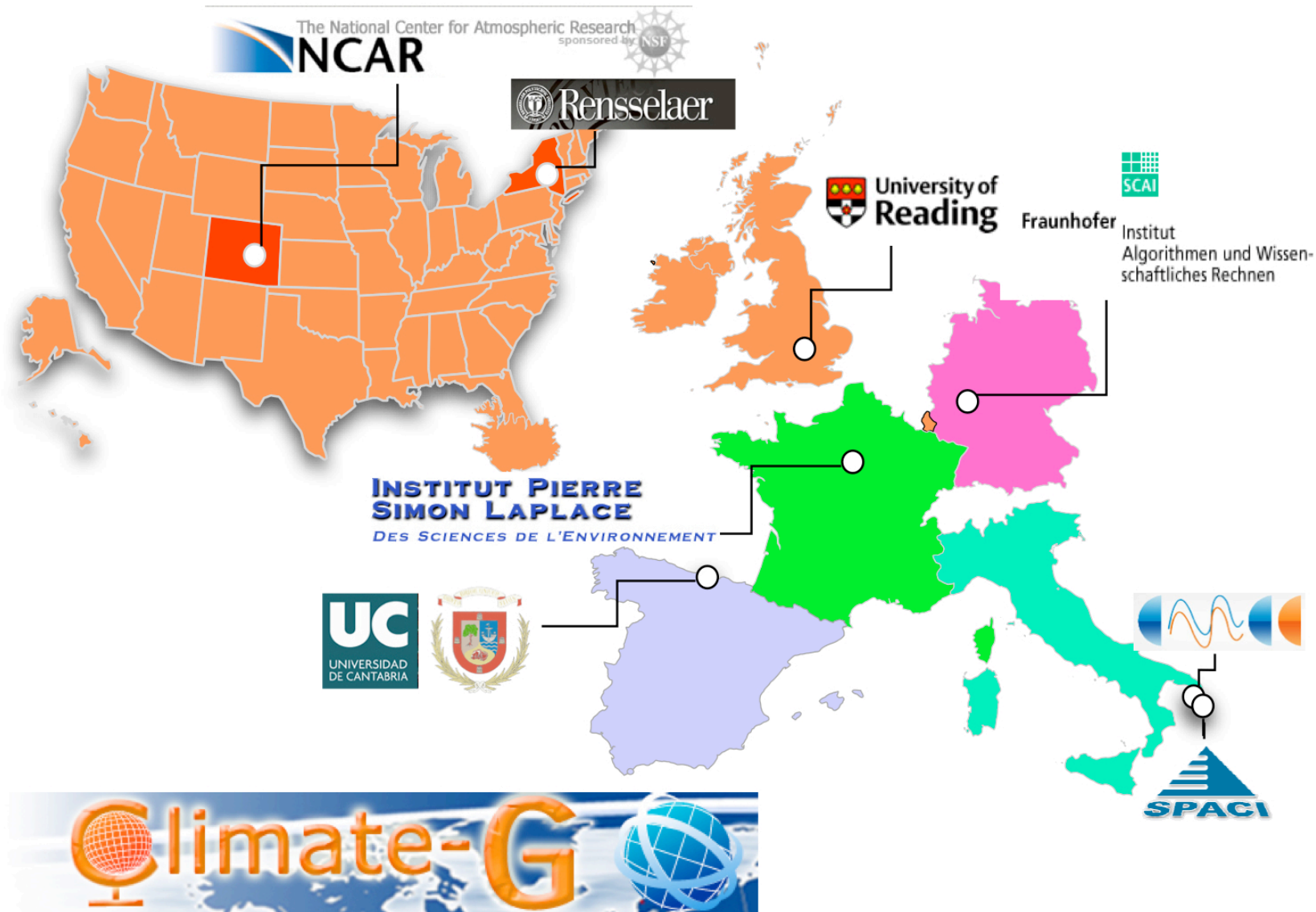


A virtual laboratory involving partners both in Europe and US

*Interdisciplinary effort: both **Climate Change** and **Computational Scientists***



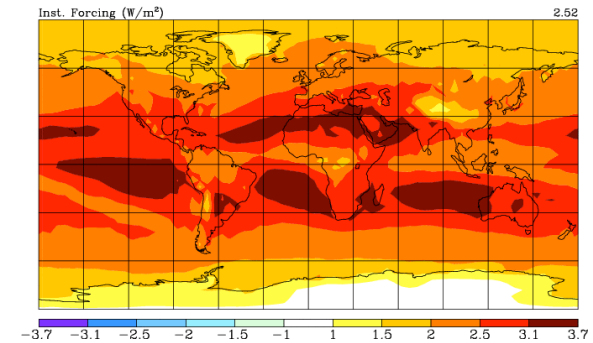
Climate-G partnership



Climate-G: main challenges and requirements

Management of PBs of distributed data

- performance
- scalability
- fault tolerance
- autonomy
- security
- transparency
- interoperability



Data Distribution Centre

- pervasive
- easy
- ubiquitous



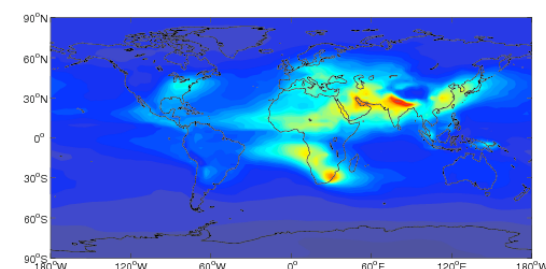
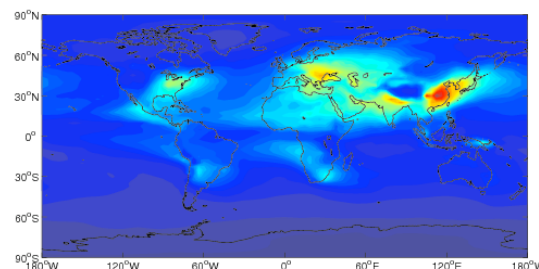
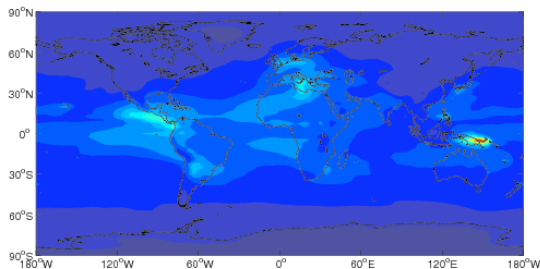
Integrated Environment

- Several tools integrated in the same
- web context. Modular approach
- Easily extensible



Why Grid?

- **A Grid based approach** provides the proper basis at an **infrastructural level**
- It ensures the right level of **flexibility**, **scalability** and **manageability**
- **Data virtualization** is a key point to build a transparent environment for **the climate community**
- **Grid metadata management** gives an efficient answer to **climate metadata** access, management, sharing and integration
- **Computational tasks** related to **post-processing** and **data analysis** can take advantage of a grid infrastructure



Climate-G and EGEE: Grid Services

- **GReIC Data Access and Integration Service (GReIC DAIS) - EGEE RESPECT**
 - Grid Metadata Management
 - Convergence between Grid & P2P systems
 - Full Metadata capabilities integrated into the Climate-G Portal
- **LHC File Catalog (LFC) - EGEE gLite**
 - Grid filesystem for the distributed climate data production
 - User-defined data collections
 - LFC access for replica management already integrated into the Climate-G Portal
- **Virtual Organization Membership Service (VOMS) - EGEE gLite**
 - Flexible and scalable role-based management
 - VOMS proxy creation already integrated into the Climate-G Portal
- **EGEE Farms, WMS and User Interface**
 - Computational services for post-processing and analysis tasks
 - Already available as CLI (next slide will show the involved EGEE environment)
 - Integration into the Climate-G Portal is not yet available. Ongoing activity

Climate-G and EGEE Middleware



Lightweight Middleware for Grid Computing

EGEE SERVICES



Workload Management System



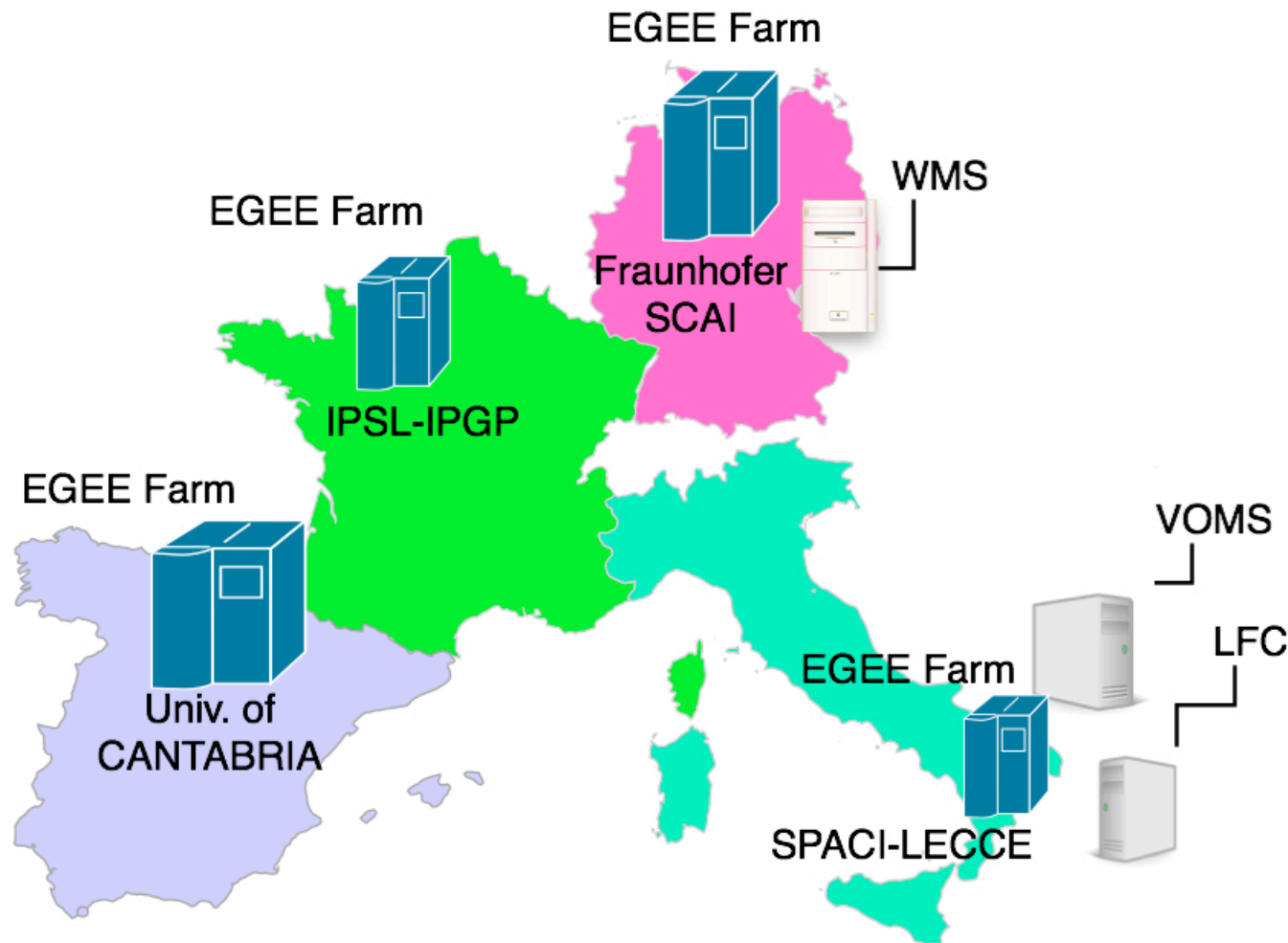
VOMS Service



LFC (LHC File Catalog)



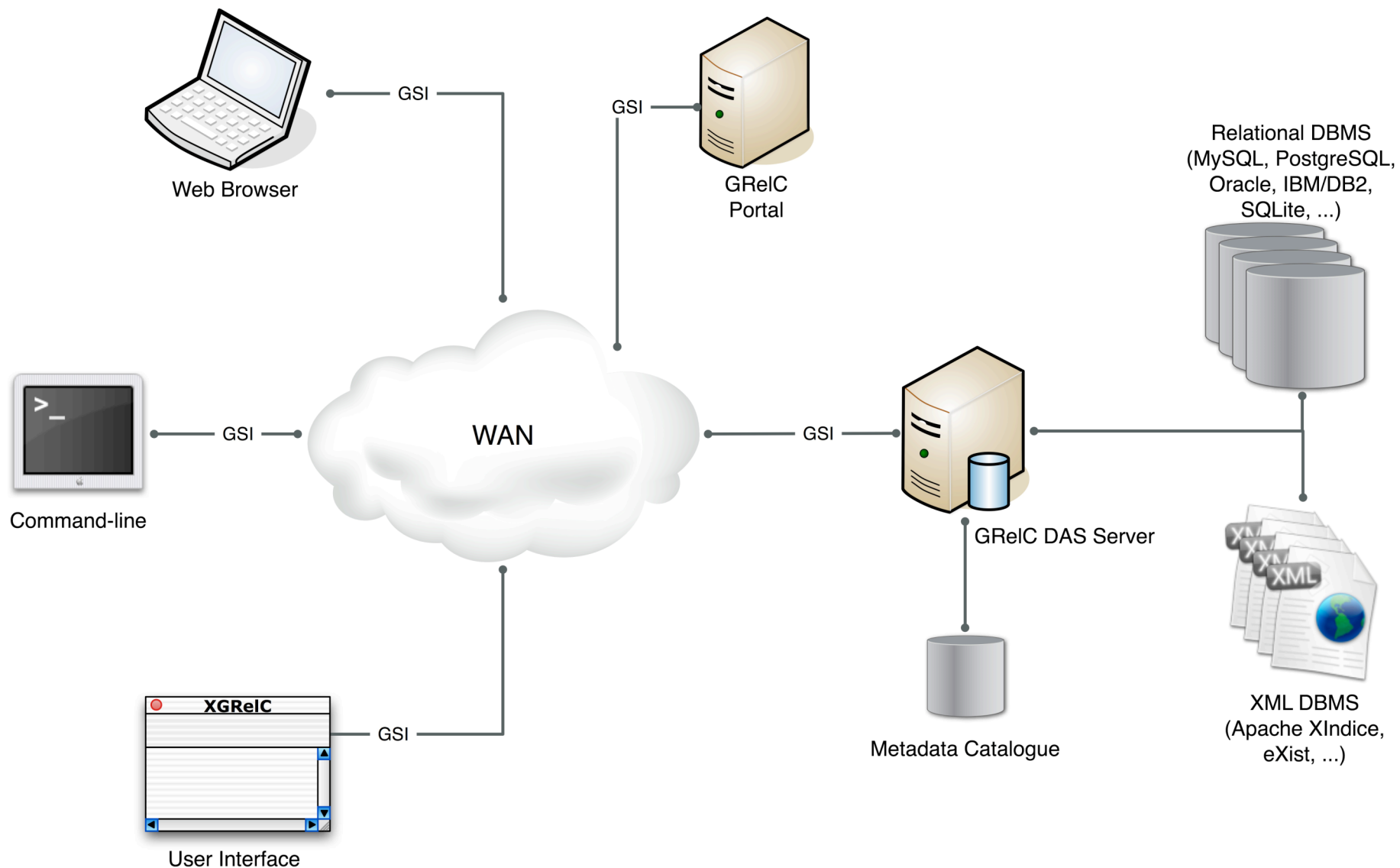
EGEE FARM SE+CE+WNs



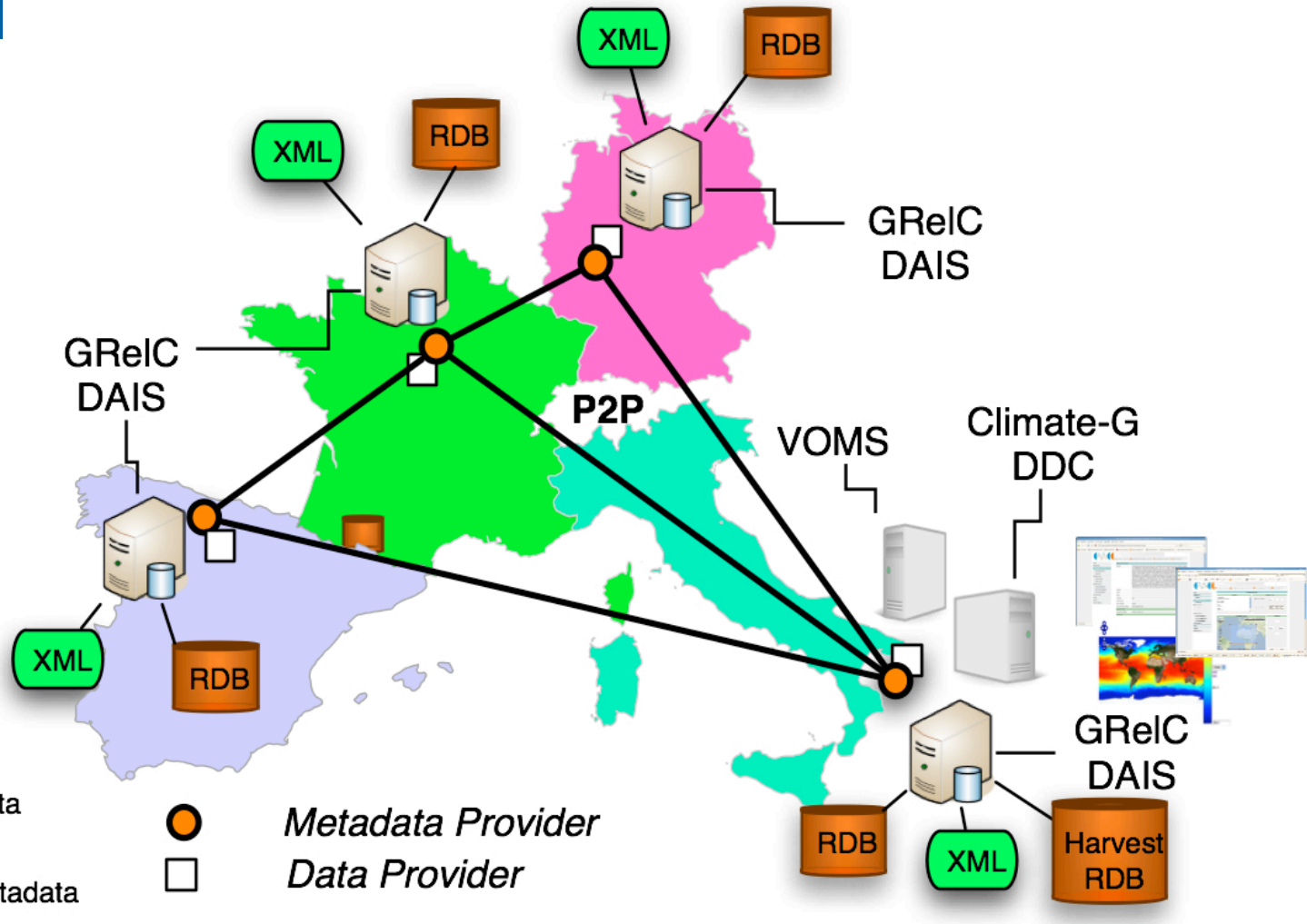
Grid Metadata Service: GReIC (EGEE RESPECT)

- **Metadata are information about data**
- Fundamental to perform **search** and **discovery** of climate datasets
- **A centralized approach** is not suitable (problem dimension, site autonomy requirement, scalability of the approach, etc.)
- **Distributed and grid-based metadata management** provides the basis for an efficient, transparent and effective climate metadata management
- **GReIC** provides a **Grid & P2P** based approach to metadata management
 - Fully compliant with gLite
 - Part of the EGEE RESPECT Program
 - Expand the functionality of the grid infrastructure for users
 - Manages several data sources
 - *XML and Relational*
 - General purpose, that is domain independent solution

Grid Metadata Service: GReIC (EGEE RESPECT)



Climate-G: Grid Metadata P2P System



Monitoring Climate-G Metadata System

Climate-G Data Distribution Centre

https://grelc04.unisalento.it:8443/ClimateG-DDC-PPS/batch_monitoring.jsp

Most Visited Start Headlines Thredds GODIVA2 EVO Grid-it EGEE - Accounting Dictionary.com Climate-G DDC Climate-G PPS CMCC DDC Voms 2.0 DDC User Mng DAS DAIS DAIS PPS BBC BBC2

Climate-G

GRelC P2P System

Filter by a specific VO: P2P System

Conncted Servers:

Server name	Istitution	City	Mail administrator	Status
scalmeta.scai.fraunhofer.de	Fraunhofer SCAI	Sankt Augustin	andre.gemuend@scal.fraunhofer.de	UP
vespucci.datagrid.jussieu.fr	IPSL-IPGP	Paris	david.weissenbach@upmc.fr	UP
grelc04.unisalento.it	CMCC & SPACI-LECCE	Lecce	grid-prod@sara.unisalento.it	UP
mon01.macc.unican.es	Universidad de Cantabria	Santander	grid-prod@unican.es	UP

Server Detail

Address: scalmeta.scai.fraunhofer.de
 Port: 18500
 Number of Local User: 3

GRelC Instances:

Name	Type	Host
monitoring	monitoring	localhost
batch_monitoring	batch_monitoring	localhost

Grid DBs:

Grid DB Name	Instance Name
cmcc-index-xml	xindice_on_localhost
cmcc-index	mysql_on_localhost
monitoring	monitoring
batch_monitoring	batch_monitoring

VOs:

VO Name	Grid DB Number
climate-g.vo.eu-egee.org	1

<< Back to admin interface

Fraunhofer SCAI - Sankt Augustin
 Contact the administrator of this GRelC DAS at andre.gemuend@scal.fraunhofer.de

Done

Climate-G: domain based services/tools

Climate-G includes domain-based services & tools into the infrastructure

- User community requirement: coexistence of grid and domain-based services
- Provides domain specific tasks. Well known, tested and widely adopted.
- Legacy systems already available and accessible

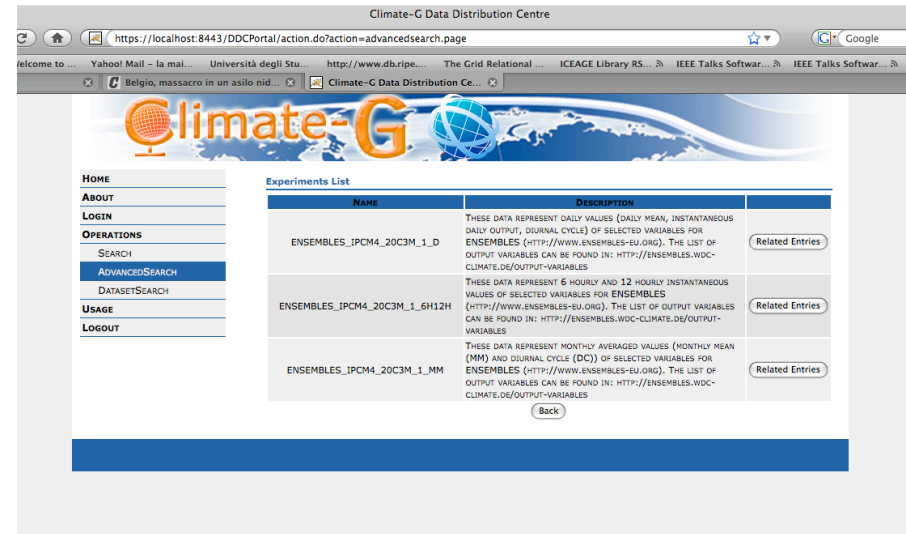
Some examples:

- **OPeNDAP (OPeNDAP Consortium)**
 - Provides access to climate data sources
 - Widely adopted in the Climate community
- **nc Web Map Service (Univ. of Reading)**
 - HTTP interface for requesting geo-registered map images from geospatial databases
- **Integrated Data Viewer (UNIDATA,UCAR) and Godiva2 (Univ. of Reading)**
 - Data visualization tools widely adopted by the Climate community

- In April, Climate-G has been recognized as a **new VO** by the EGEE Resource Allocation Group (*climate-g.vo.eu-egee.org*)
 - First VO devoted to climate change community!
 - Wide climate community in Europe potentially interested in Climate-G
 - Several Climate-G presentations in the Geoscience community (EGU09, ESA Workshop, etc.)
 - About 60 users joined the VO since April
 - 30 new users from CMCC Divisions (agricultural, soil & coast and economic impacts people) will soon join the VO
 - Most of them (more than 85%) comes from the climate context and are using a grid infrastructure for the first time -> new users
 - Interesting level of feedback from our users in terms of:
 - suggestions to improve the portal
 - new data sources and new tools to be included into the portal
 - application-level requirements (=> good for EGEE computational infrastructure)
- Several **EGEE sites have been configured** to support the “Climate-G VO” (*Fraunhofer SCAI, SPACI-LECCE, IPSL/CNRS IPGP, UniCantabria*)
 - More than 300 CPUs are now available for preliminary tests
- **Seed Resources** will be exploited by the Climate-G testbed/users
 - Thanks to the EGEE NA4 VO Support Group for their support
- The whole Climate-G EGEE infrastructure (data and computational) must be accessible through the Climate-G Portal, our scientific gateway

Climate-G Data Distribution Centre

- **Main Functionalities**
 - Search & Discovery
 - Data access & viz
 - Metadata browsing
 - Users and roles mng
 -



- **Features**
 - Easy to use interfaces
 - **Platform independent**
 - Secured by design
 - **No additional software is required**
 - It entirely replaces the Command Line Interface



Climate-G DDC: Snapshots

Climate-G Data Distribution Centre

https://localhost:8443/DDCPortal/action.do?action=advancedsearch.page

Climate-G

HOME
ABOUT
LOGIN
OPERATIONS
SEARCH
ADVANCEDSEARCH
DATASETSEARCH
USAGE
LOGOUT

KEYWORDS AND EXTENT

Keywords: IPCC-DDC, climate simulation, IPCC-AR4, IPSL-CM4, IPSL, 20C3M, GFA, ORCA05, ENSEMBLES Stream1

Temporal Extent: FROM (YYYY/MM/DD) yyyy mm dd, To (YYYY/MM/DD) yyyy mm dd

GEOGRAPHICAL EXTENT

MAP | SATELLITE | TERRAIN

North, West, South

Filter Projects & Experiments

PROJECTS

ENSEMBLES

Clear Projects | Filter Experiments | Project Info

EXPERIMENTS

ENSEMBLES_IPCM4_20C3M_1_D
ENSEMBLES_IPCM4_20C3M_1_6H12H
ENSEMBLES_IPCM4_20C3M_1_MM

Experiment Metadata | Experiment List | Related Entries

Climate-G Data Distribution Centre

https://grelc.unile.it:8443/ClimateG-DDC/action.do?action=basearch.page

Climate-G

HOME
ABOUT
LOGIN
OPERATIONS
SEARCH
ADVANCEDSEARCH
DATASETSEARCH
USAGE
LOGOUT

IDENTIFICATION INFO

Abstract: This dataset represents instantaneous 6 hourly (6H) or 12 hourly (12H) values of the selected variable for ENSEMBLES. The model output was prepared for the IPCC Fourth Assessment 20C3M experiment. For specific scenario details see experiment summary. These data are in netCDF format.

Purpose:
Credit:
Status:
Language: eng
Character Set:
Topic Category: climatologyMeteorologyAtmosphere

Points of Contact

Principal Investigator: Dufresne, Jean-Louis

Citation Info

Title: IPCM4_20C3M_1_6H_ps
Alternate Title:
Edition:
Edition Date:
Presentation Form:
Other Citation Details:
Collective Title:
ISBN:
ISSN:
Resource Format:
Keywords:
Geographical Extent:
West Bound Longitude: 0
East Bound Longitude: 360

Data Access - Complete OPeNDAP Support

The image displays two overlapping browser windows from Mozilla Firefox. The top window shows the 'OPeNDAP Dataset Query Form' for the dataset 'ENSEMBLES/PCM4_20C3M_1_DM_hfls_1-50760.nc'. It includes a title bar, navigation menu, and a main content area with the following sections:

- Action:** Buttons for 'Get ASCII', 'Get Binary', and 'Show Help'.
- Data URL:** A text field containing the URL: `http://gandalf.unile.it:8080/thredds/dodsC/ENSEMBLES/PCM4_20C3M_1_DM_hfls_1-50760.nc`
- Global Attributes:** A section with a title: "IPSL model output prepared for IPCC Fourth Assessment climate of the 20th Century experiment (20C3M)".

The bottom window displays the 'Dataset Information' page for the same dataset. It provides detailed metadata:

- title:** "IPSL model output prepared for IPCC Fourth Assessment climate of the 20th Century experiment (20C3M)"
- institution:** "IPSL (Institut Pierre Simon Laplace, Paris, France)"
- source:** "IPSL-CM4_v1 (2003) : atmosphere : LMDZ (IPSL-CM4_IPCC, 96x72x19) ; ocean ORCA2 (ipsl_cm4_v1_8_2x12L31) ; sea ice LIM (ipsl_cm4_v1_8) ; land ORCHIDEE (orchidee_1_3)"
- contact:** "Sebastien Denvil, sebastien.denvil@ipsl.jussieu.fr"
- project_id:** "IPCC Fourth Assessment"
- table_id:** "Table DF (17 November 2005)"
- experiment_id:** "climate of the 20th Century experiment (20C3M)"
- realization:** 1
- cmor_version:** 1.1
- Conventions:** "CF-1.0"
- history:** "YYYY/MM/DD : data generated At 08:03:51 on 07/22/2007, CMOR rewrote data to comply with CF standards and IPCC Fourth Assessment requirements"
- references:** "For coming informations : http://www.ipsl.jussieu.fr/poles/Modelisation/NotesSciences.htm"
- comment:** "For coming informations : http://www.ipsl.jussieu.fr/poles/Modelisation/NotesSciences.htm"
- Unlimited_Dimension:** time
- Variables in this Dataset:** lon: Array of 64 bit Reals [lon = 0.95]

The bottom window also shows the 'Attributes' section of the dataset, listing metadata for 'lon', 'lat', 'time', and 'hfls'.

```
Attributes {
  lon {
    String standard_name "longitude";
    String long_name "longitude";
    String units "degrees_east";
    String axis "X";
    String bounds "lon_bnds";
  }
  lon_bnds {
  }
  lat {
    String standard_name "latitude";
    String long_name "latitude";
    String units "degrees_north";
    String axis "Y";
    String bounds "lat_bnds";
  }
  lat_bnds {
  }
  time {
    String standard_name "time";
    String long_name "time";
    String units "days since 1860-1-1";
    String axis "T";
    String calendar "360_day";
    String bounds "time_bnds";
  }
  time_bnds {
  }
  hfls {
    String standard_name "surface_upward_latent_heat_flux";
    String long_name "Surface Latent Heat Flux";
    String units "W m-2";
    String cell_methods "time: mean (interval: 30 minutes)";
    Float32 _FillValue 1.0E20;
    Float32 missing_value 1.0E20;
    String original_name "hfls";
    String history "14 - 08:03:51 on 07/22/2007, CMOR altered the data in the following ways: multiplied by 1.00000500 to avoid overflow";
  }
}
```



Data Visualization (IDV support)

The screenshot displays the Unidata Integrated Data Viewer (IDV) software interface. The main window is titled "Unidata IDV - Map View - One Pane" and shows two map views. The top map is a 3D surface view of the United States, showing topography and weather data. The bottom map is a polar projection map of the Arctic region, showing weather data and range rings. The interface includes a menu bar (File, Edit, Displays, Data, Tools, Help), a toolbar, a legend, and a data table at the bottom.

Legend

- Maps
- Default Background...
- World Coastlines
- Hi-Res US
- Plan Views
- Z - Contour Plan...
- Level: 500 hectopascals
- Pm_msl - Color-S...
- 952
- 16
- 3D Surface
- windspeed - 3D L...
- Value: 45 m/s
- 0
- Cross sections
- T - Color-Shaded...
- 90

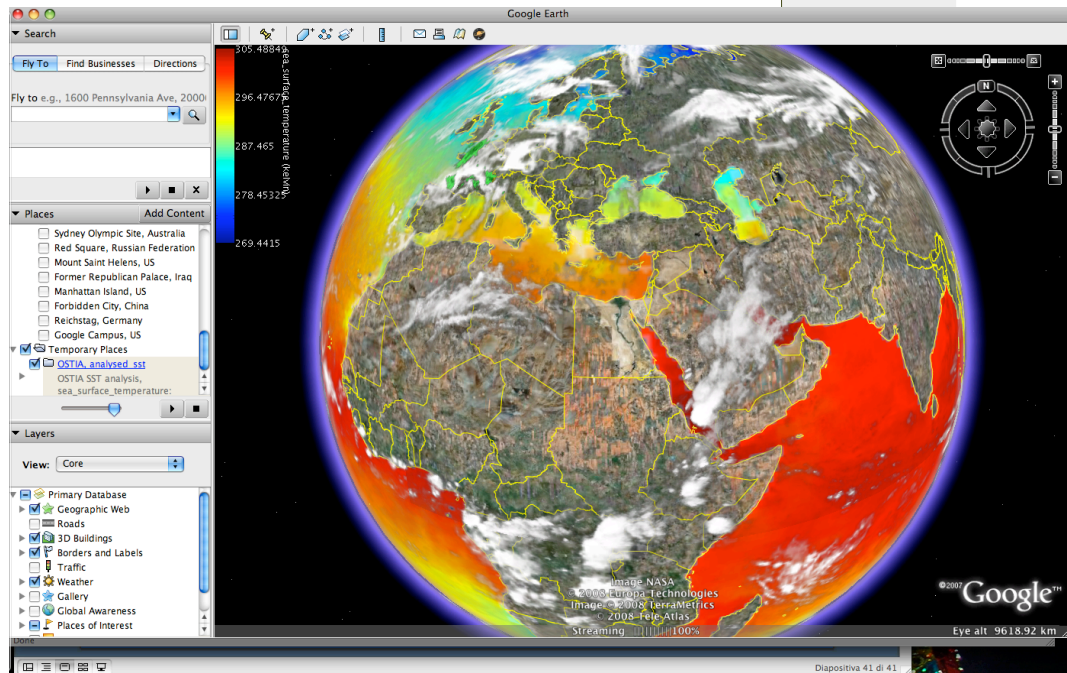
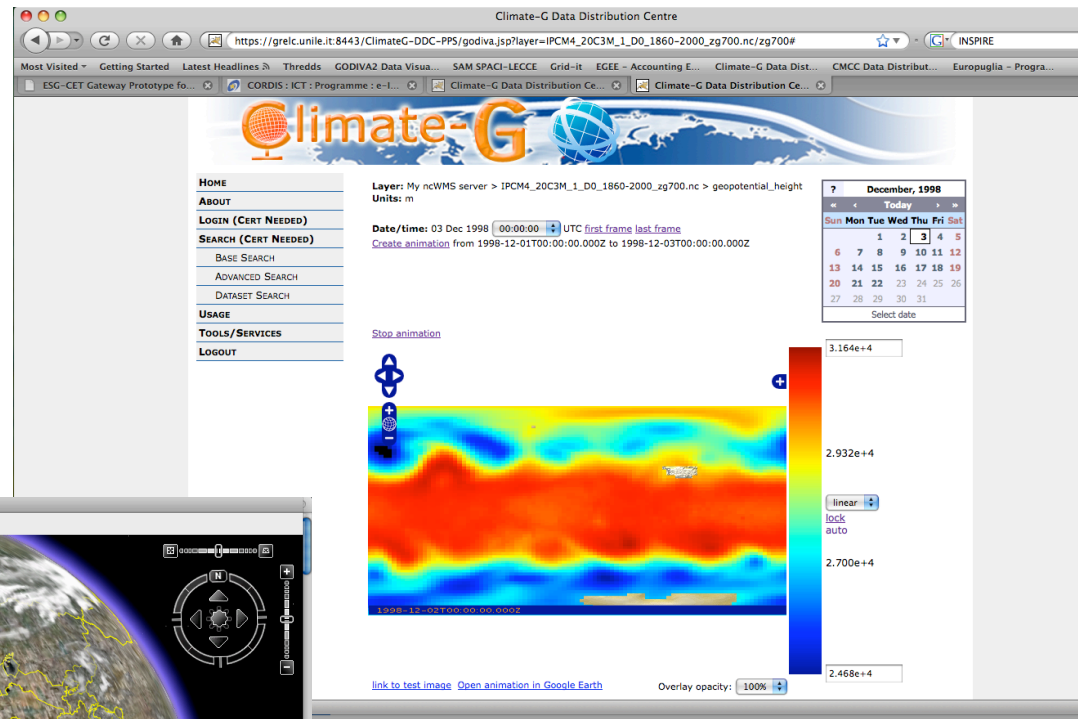
Data Table

File Name	DAF	DAS	DDS	INF	XML	BIN	IDV
ICE (GRELC04.UNILE.IT)							
(GRID009.CT.INFN.IT)	DAF	DAS	DDS	INF	XML	BIN	IDV
RODIGUER.IPSL.JUSSIEU.FR)	DAF	DAS	DDS	INF	XML	BIN	IDV
(PRODIGUER.IPSL.JUSSIEU.FR)							
GRELC DAIS					XML		
IMETA.SCAI.FRAUNHOFER.DE)	DAF	DAS	DDS	INF	XML	BIN	IDV
IP (GRELC04.UNILE.IT)	DAF	DAS	DDS	INF	XML	BIN	IDV
ICE (GRELC04.UNILE.IT)							
(GRID009.CT.INFN.IT)	DAF	DAS	DDS	INF	XML	BIN	IDV
OPENDAP (PRODIGUER.IPSL.JUSSIEU.FR)	DAF	DAS	DDS	INF	XML	BIN	IDV

Godiva2 Integration

Two-dimensional
Data visualization tool

Google Earth



Limitations and future work

- Presently Climate-G manages online data
 - In the next future access to deep storage will be managed via SRM interfaces
- Today Climate-G Portal manages the entire data infrastructure. Access to the computational part is now carried out via CLI
 - In the next future access to the computational part will be performed via the Climate-G Portal
- Climate-G Portal now manages Atmospheric and Oceanographic data
 - Climate-G will manage both climate and economic data. Economic impacts of climate change on health, coasts, soil, agriculture, etc. represent an important goal for our community
- Analysis & visualization tools currently supported: IDV and Godiva2
 - Climate-G will soon integrate into the portal support for the Grid Analysis and Display System (GrADS)
- This work could continue and evolve in Earth Science SSC (ES SSC)
 - The strong experience of the testbed could represent a solid basis for future works in the ES SSC context (grid metadata mng and Portal)

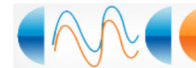
Conclusions

- Climate-G has a strong relationship with the **EGEE Project**
- A new **EGEE VO** for the Climate-G testbed has been created (April 2009)
- GReIC DAIS provides a grid based **distributed metadata management as well as harvesting solution**
- **Data oriented** EGEE services already integrated into the portal, **computational** ones soon available for *analysis* and *post-processing*
- **Climate-G Portal** to ease Metadata management via Web Interface
- Visualization tools have been integrated (**IDV, Godiva2**)
- Climate-G is conceived as a **Virtual Laboratory** for the involved people and technologies
- Seed resources will be exploited very soon

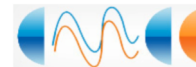
Acknowledgments

Many thanks to all of the involved people in the Climate-G testbed

Giovanni Aloisio (CMCC)



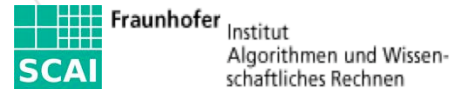
Sandro Fiore (CMCC)



Monique Petitdidier (CNRS/IPSL)



Horst Schwichtenberg (Fraunhofer-SCAI)



Sébastien Denvil (IPSL)



Peter Fox (RPI, NCAR)



Jon Blower (Univ. Reading)



Antonio Cofino (Univ. of Cantabria)

