

WRF4SG: A Scientific Gateway for climate experiment workflows

Carlos Blanco, Antonio S. Cofiño and Valvanuz Fernández-Quiruelas

Santander Meteorology Group

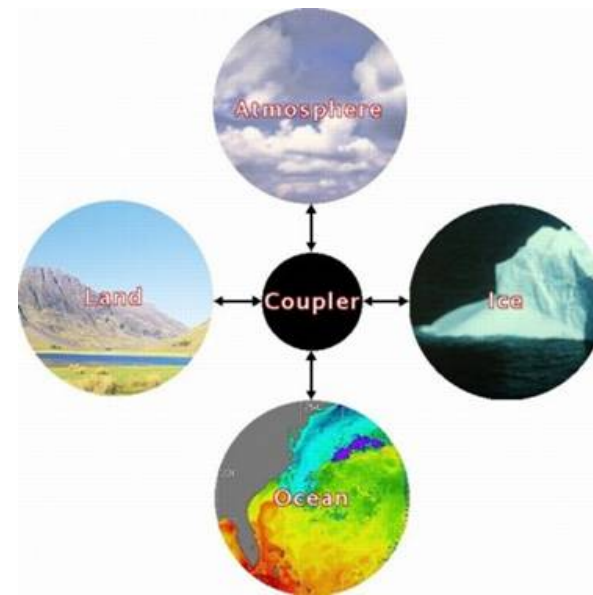
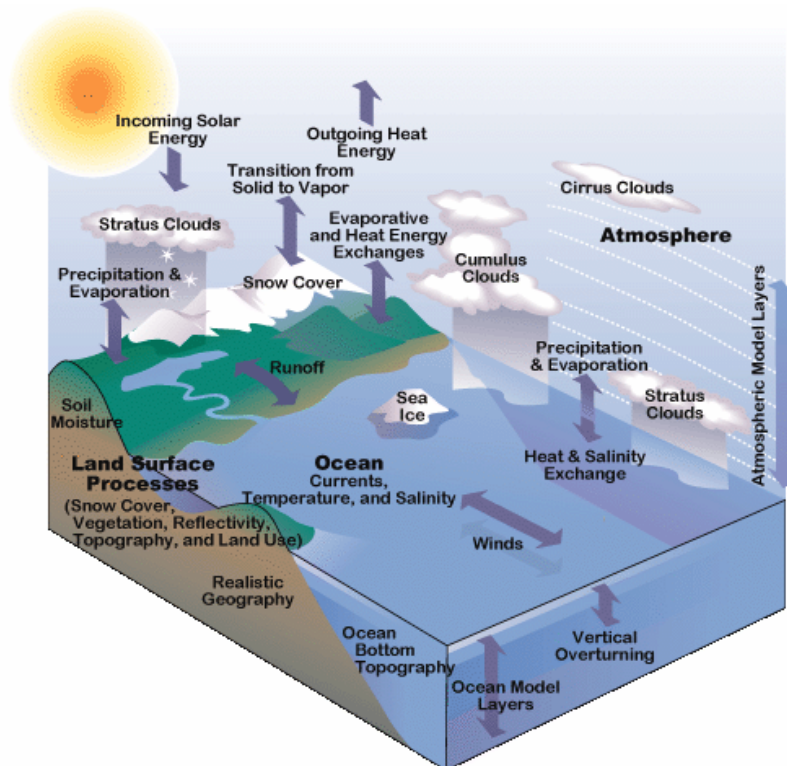
Department of Applied Mathematics and Computer Sciences

Universidad de Cantabria, Santander, Spain



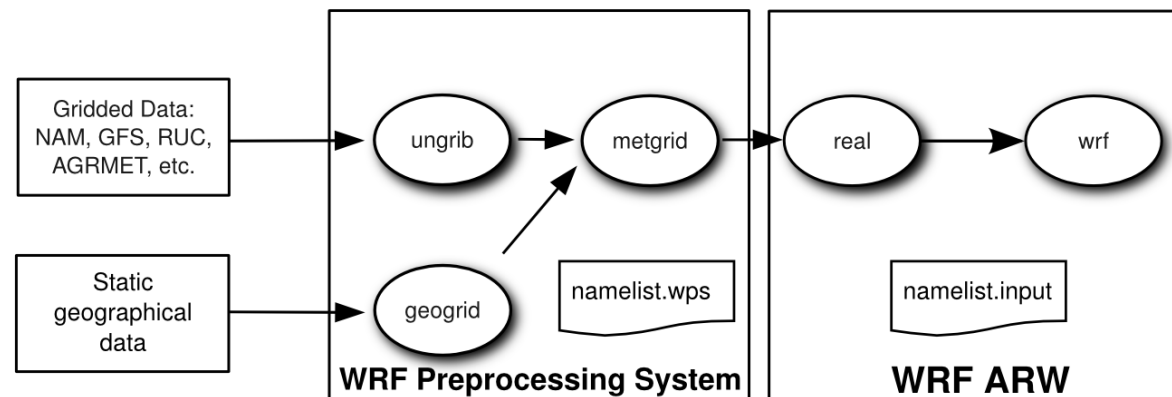
Aknowledgments: This work is funded by the Spanish PLAN NACIONAL de I+D+i 2008-2011 (Ref.# CGL2011-28864) and ERDF)

- Climate Models are numerical models that simulate the interactions of the atmosphere, oceans, land surface and cryosphere.
- Global Models (CAM) simulate the whole globe while Regional Models (WRF) provide a bigger resolution over a geographical region.



Working with Climate Models is not trivial...

- The compilation and execution are quite complex: WRF 400k lines F90.
- Models are composed by several components (executables) with dependencies among them and also among their data.
- Running a given experiment involves preparing data, preparing several configuration files and executing different programs.
- Large input and output data transfers.
- Intensive use of CPU and memory (one simulation months of Walltime, even in parallel!!!!).



- WRF provides a **flexible** and computationally-efficient framework which allows solving a **variety of problems** for different time-scales, from weather forecast to climate change projection
- WRF is also widely used as a research tool for cutting-edge advances in **modeling physics, dynamics, and data assimilation**
- **Large community**. Currently, 6000+ registered users

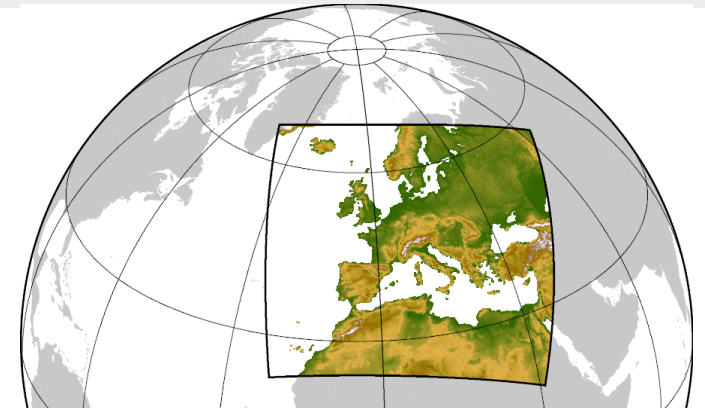


Reanalysis/Reforecasts/Hindcast

- High number ($\sim 10^4$) of independent jobs
- High volume of output-data (>TB)
- Requires **scalability**

Regional climate simulation

- Long, continuous simulations;
 - weeks of walltime of an MPI job
- High volume of output data (>TB)
- Recovering system for **simulation restart**

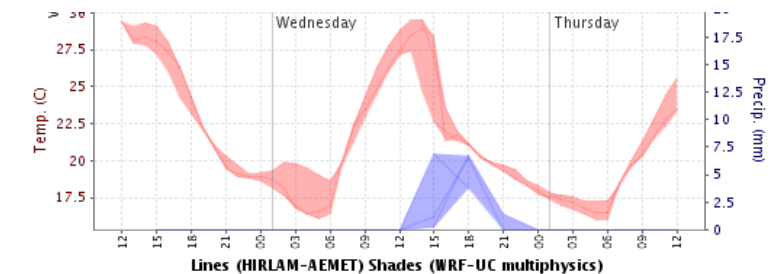


Weather Forecasting

- QoS and optimal resources: **deadline for delivery**

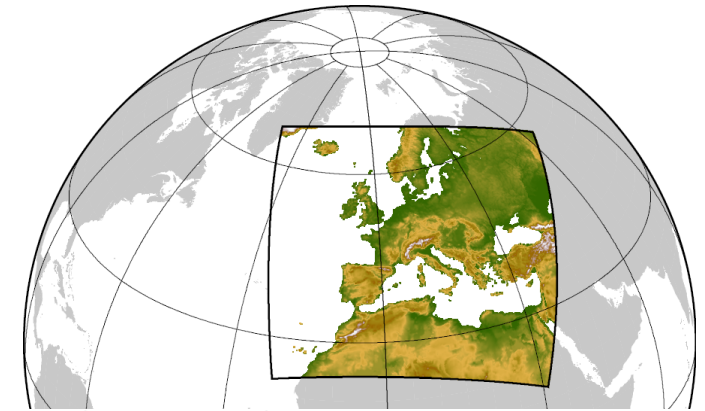
Sensitivity/ensemble studies

- Physical schemes, initial conditions and boundary conditions: uncertainty sampling
- Resource demanding: experiments composed by many **independent simulations**



SEAWIND experiment set: past reforecasts (1989-2009) of an improved wind field over Europe for off-shore wind farms.

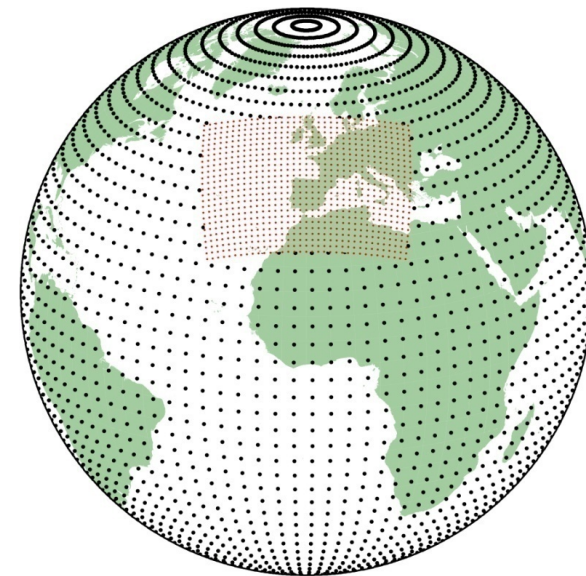
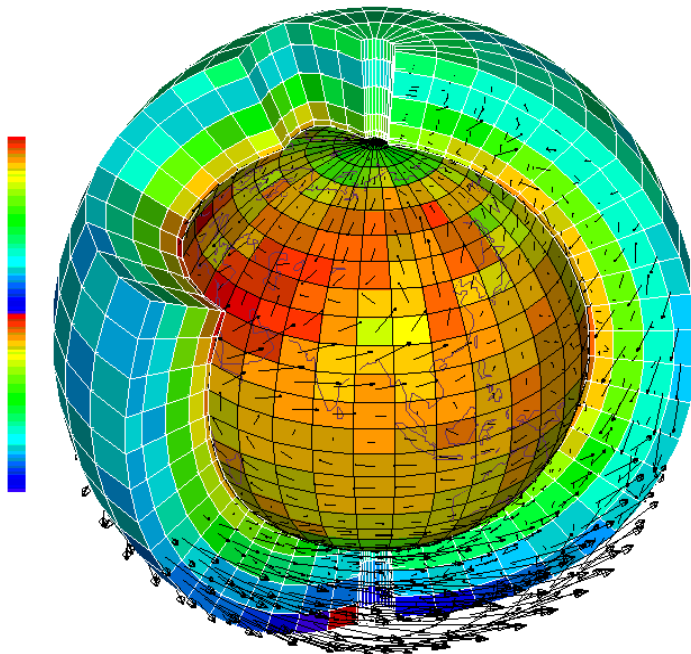
- Features of the experiment
 - 21 years of daily reforecasts (36h each)
 - 7,665 independent simulations
- Computing cost of the experiment
 - Working Node Architecture
 - CPU: Intel(R) Xeon(R) CPU E5620 @ 2.40GHz 8 Cores
 - RAM Memory: 16 GB
 - Result
 - Walltime (MPI job) = 21 x 365 x 70' ~ **1 year**
 - Output = 21 x 365 x 17 GB ~ **130 TB**



To test what is the best physical parametrization on the boundary layer a **sensitivity study** is required. This studies usually test tenths of parameters combinations just limited by computer resources available: **10 params → 10 years → 1.3PB**

WRF4G is a framework developed by the Santander Meteorology Group, provides:

- Flexible WRF experiment **design**, **execution** and **monitoring**, and ...
- ... the ability to run these experiments on different computing resources concurrently in a **transparent** way (DRM4G).



WRF4G is a framework developed by the Santander Meteorology Group, provides:

- Flexible WRF experiment **design, execution and monitoring**, and ...
- ... the ability to run these experiments on different computing resources concurrently in a **transparent** way (DRM4G).

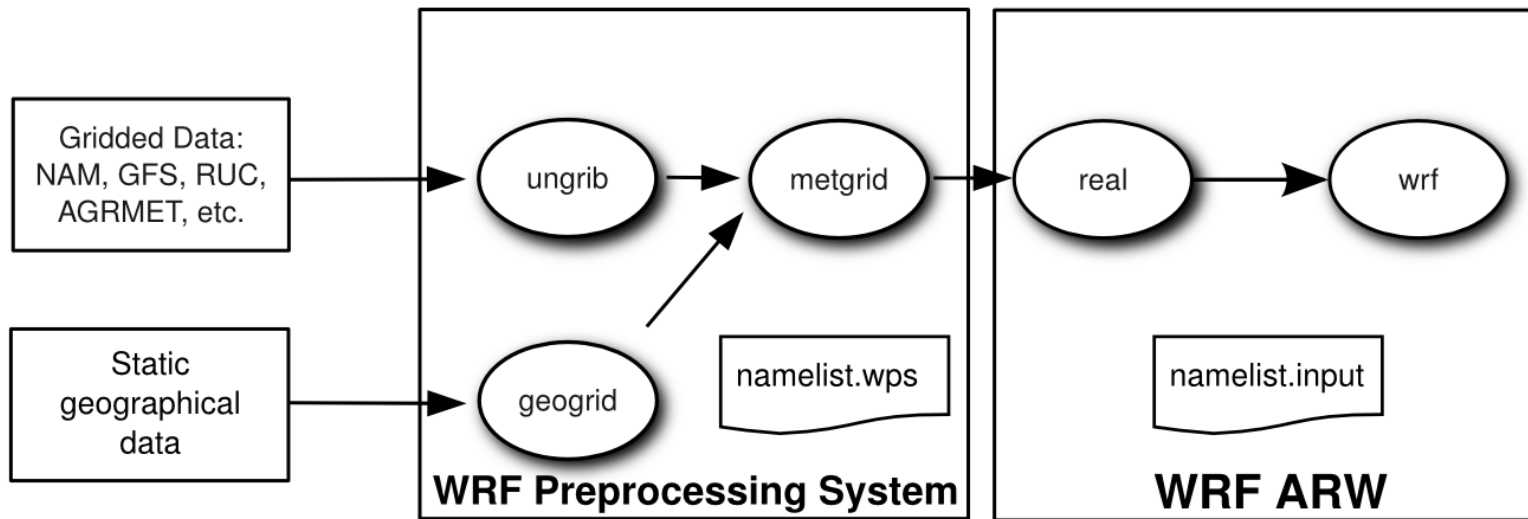
If you don't know what is WRF4G then you missed Monday's oral presentation at NH1.8 session:

WRF4G project: Adaptation of WRF Model to Distributed Computing Infrastructures

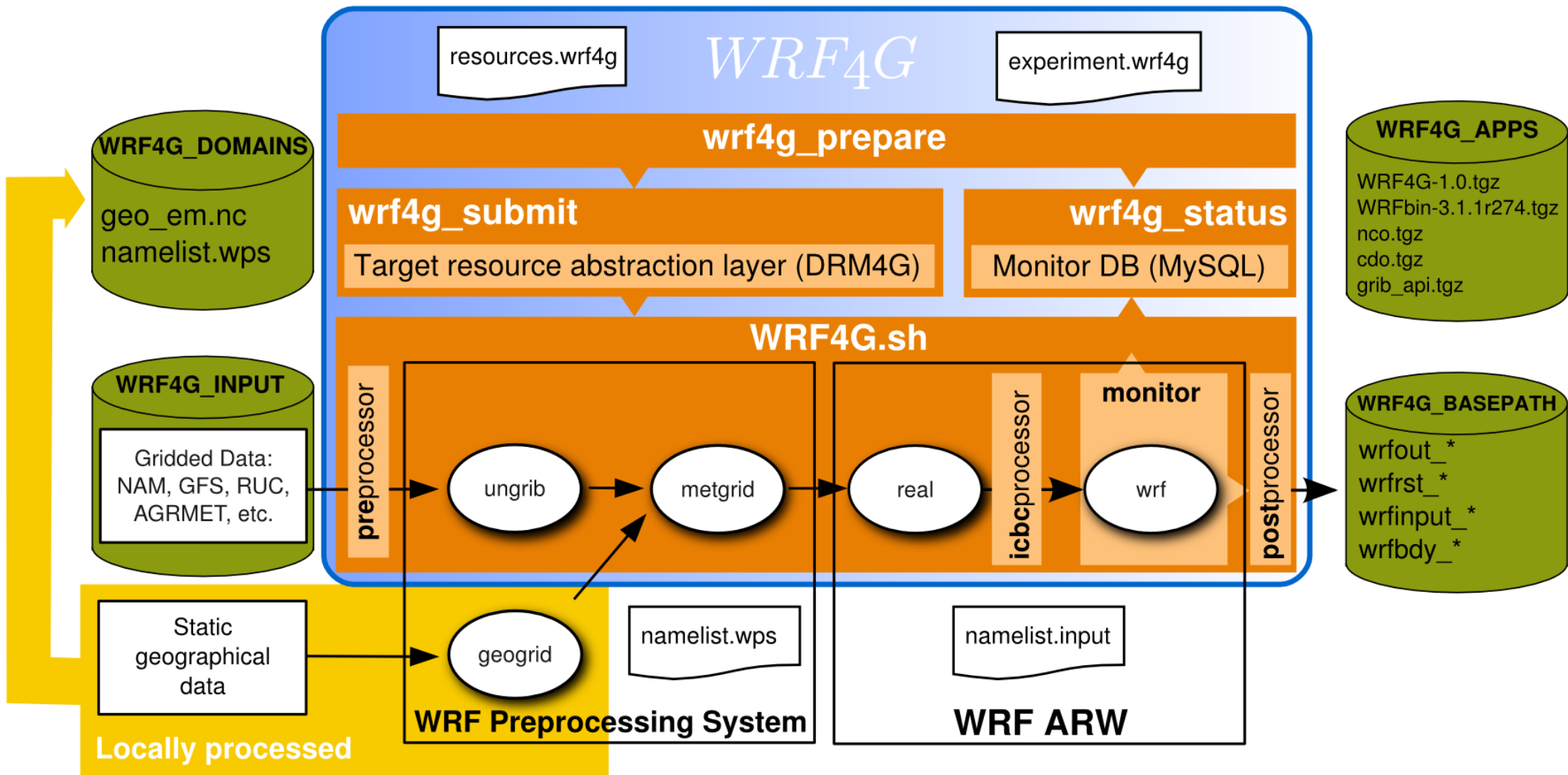
Santander Meteorology Group

A multidisciplinary approach for weather & climate

WRF Workflow



WRF4G Framework



- Needs to be deployed on a **Linux OS** such as Ubuntu, Debian, Centos, ...
- The user interface offered by WRF4G is a **Command Line Interface (CLI)**
- Access to new **cloud resources** such as IBM Cloud or Amazon Web Services
- No **multiuser** application. Therefore, each user has to deploy it

- **Framework of tools** that allows scientists to run applications with little concern for where the computation actually takes place
- It is similar to cloud computing in which applications run as **web services** on remote resources in a manner that is not visible to the end user
- Gateways often let users store, manage, catalog, and **share large data collections** or rapidly evolving **novel applications** they cannot find anywhere else
- **Training and education** are also a significant part of some science gateways

- A gateway packaged as a **web portal** with users in front and services in back
- A **grid-bridging gateway**. Often communities run their own resources, devoted to their specific areas of science. The gateway in these cases is a mechanism to extend the reach of their community grid
- A gateway that involves **application programs** running on users own personal computers or workstations like applications user interfaces **using CLIs or GUIs**

The objectives of the WRF4G Scientific Gateway has to be :

- Facilitate the **management and access of DCI** resources to the WRF community
- **Improve and fix the WRF4G** missing features
- Tackle **WRF** complex **climate experiments management** such as :
 - weather forecast
 - extreme weather case studies
 - future climate projections
 - re-forecasts



- **From scratch**
 - It takes long time to develop a complex gateway
 - High development cost
 - It is difficult to extend in a scalable way
- **From existing gateway technology**
 - It reduces development time and cost
 - Produces a robust service
 - Support to deploy and update

- From scratch
 - It takes long time to develop a complex gateway
 - High development cost
 - It is difficult to extend in a scalable way
- From existing gateway technology
 - It reduces development time and cost
 - Produces a robust service
 - Support to deploy and update

WS-PGRADE/gUSE

Collaboration with the SCI-BUS Project

- **Easy to extend** a generic gUSE Portal to a specific Scientific Gateway (Portlets)
- **Access** to different types of **DCIs**:
 - Clusters (PBS, LFS)
 - service grids (GT2, GT4, GT5, gLite , ARC, etc)
 - desktop grids (BOINC)
 - Clouds (Google App Engine cloud, IBM Cloud or Amazon Web Services)
- Advanced **data-flows and workflow**
- Simultaneous **execution of workflow** nodes on different resources

WRF4SG (WRF for Scientific Gateway) will provide services between WRF users and WRF4SG interface for different use cases:

- **User authentication.** When user wants to use the WRF4SG, they will be prompted for their credentials: **username** and **password**. After login, they will have to use **Security WS-PGRADE Portlet** (available now) in order to manage their Grid Certificates:
 - **Upload certificate:** The user uploads its certificate in order to create the proxy. Then, they indicate a MyProxy server in order to upload the created proxy to it.
 - **Download proxy:** The user downloads the proxy certificate from the MyProxy server into the WRF4SG.
 - **Associate to VO :** The user will subscribe to VO and will request the signed VO

WRF4SG (WRF for Scientific Gateway) will provide services between WRF users and WRF4SG interface for different use cases :

- **Data processing:** Climate experiments demand significant data storage and data management tools.
 - **Pre-processing:** The user will be able to **select data sources** and prepare the **domain** for the experiment. The **WRF Domain Wizard** write namelists automatically: <http://wrfportal.org> (TBD)
 - **Post-processing:** The output of the experiment will be **checked, filtered, and pushed to the location** specified by the user in the experiment configuration step. The user will be able to move and check the data by **LFC Portlet** (existing portlet)

WRF4SG (WRF for Scientific Gateway) will provide services between WRF users and WRF4SG interface for different use cases :

- **Design of Experiment workflow:** In order to configure the experiment, the user will be able to use :
 - **Experiment selector:** It indicates the **type of the experiment** such as weather forecast, re-forecasts or climate change. (TBD)
 - **Experiment features selector:** Experiment attributes will have to be configured such as the **fields of interest** (e. g. wind, precipitation, ...), the **start and date simulation**, **chunk size**, **components** (pre and post processors), etc. (TBD)
 - **Experiment requirements:** It is highly recommended that the user specifies the requirements for the experiment because climate experiments are very demanding applications. For instance, the requirements on hardware, **RAM memory** and **storage**.(TBD)

WRF4SG (WRF for Scientific Gateway) will provide services between WRF users and WRF4SG interface for different use cases :

- **Workflow monitoring:** In order to monitor the experiment, the user will subscribe to receive **notification messages**, and the gateway will display the **experiment progress** as soon as notifications arrive. WRF4SG displays a table with the user experiments with the following information (TBD):
 - Experiment identifier
 - Experiment status
 - Workflow components executed
 - Workflow components logs
 - The current date of the experiment

In order to develop **WRF4SG**, design of **no-existing portlets** has to be made for the WS-PGRADE/gUSE portlet repository :

- **Experiment design.** **WRF4SG** will provide a portlet with a **step-by-step process** to define a WRF experiments, based on WRF4G experiment definitions.
- **Experiment management and monitoring.** **WRF4SG** is going to provide elements to control all the available simulations in an experiment. The user will have a **set of option buttons** to submit, hold, stop, re-submit or cancel their experiments and simulations. It will provide information regarding the **simulations progress** as well.
- **Pre and Post experiment data management.** This portlet is going to allow user to **locate**, **select**, **check** and **visualize** both input and output experiment data.

The front-end prototype is now available on *ce02.macc.unican.es*

It has been configured gLite resources in order to use *esr*, *prod.vo.eu-eela.eu* and *earth.vo.ibergrid.eu* VOs

The screenshot shows a web browser window with the URL `ce02.macc.unican.es:8080/liferay-portal-6.1.0/`. The page features the LIFERAY logo and a navigation menu with 'Welcome', 'Statistics', 'Publications', and 'Help'. A breadcrumb trail shows 'WRF4SG > Welcome'. The main content area is titled 'Welcome to WS-PGRADE Portal!' and includes a paragraph describing the portal's purpose and supported resources. Below this, there are several service tiles: 'Graphical User Interface: WS-PGRADE' (orange), 'Presentation tier' (yellow), 'Workflow Storage', 'File Storage', 'Workflow Interpreter', 'Application Repository', and 'gUSE Information System' (blue). A 'NEW' badge is next to 'Job Submission Service: DCI-BRIDGE' (green), which lists supported resources like gLite, ARC, and Unicores. The 'Architectural tier Available middlewares' (red) tile lists gLite, ARC, Unicores, Globus, LSF, PBS, BOINC, GBAC, Gemica, Local, and Google App Engine. The 'Middle tier gUSE services' (blue) and 'Presentation tier' (yellow) tiles also feature the gUSE logo.

Santander Meteorology Group

A multidisciplinary approach for weather & climate

WRF4SG Prototype

The front-end prototype is now available on *ce02.macc.unican.es*

It has been configured gLite resources in order to use *esr*, *prod.vo.eu-eela.eu* and *earth.vo.ibergrid.eu* VOs

Climate workflows are available in order to run **WRF model** and pre and post process experiment datasets

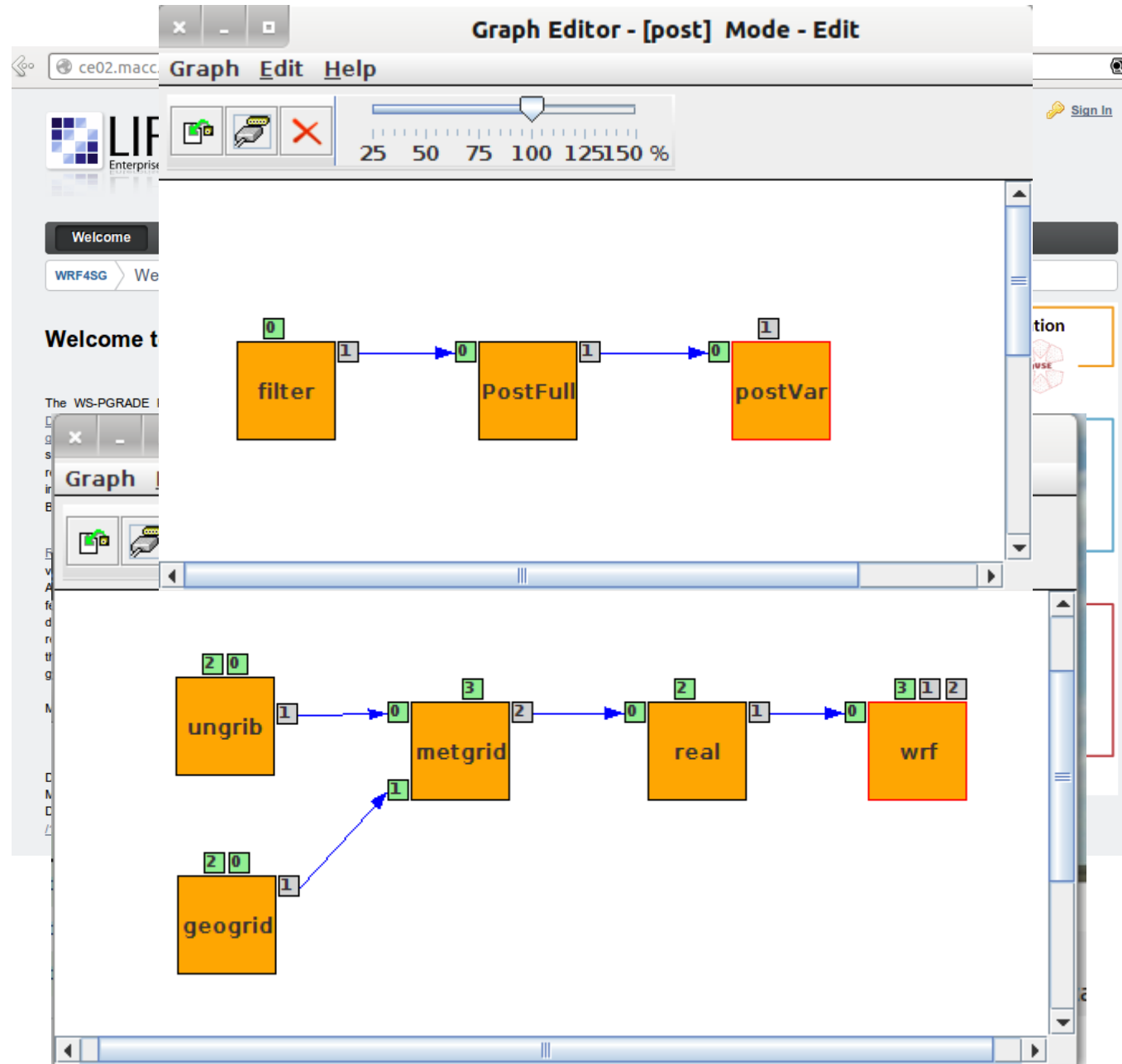
The screenshot displays a web browser window at `ce02.macc.unican.es:8080/liferay-portal-6.1.0/`. The page features the LIFERAY logo and navigation tabs for Welcome, Statistics, Publications, and Help. A breadcrumb trail shows 'WRF4SG > Welcome'. A prominent orange button reads 'Graphical User Interface: WS-PGRADE', and a 'Presentation tier' badge is visible. Below, a message states 'Welcome to WS-PGRADE Portal!' and notes the portal was developed by the 'Laboratory of Parallel and Distributed Systems'. The main content area is a 'Graph Editor - [wrf] Mode - Edit' window. This editor shows a workflow diagram with five orange boxes: 'ungrid', 'geogrid', 'metgrid', 'real', and 'wrf'. Arrows indicate the flow: 'ungrid' (with a '2 0' label) connects to 'metgrid' (with a '0' label); 'geogrid' (with a '2 0' label) connects to 'metgrid' (with a '1' label); 'metgrid' (with a '3' label) connects to 'real' (with a '2' label); and 'real' (with a '2' label) connects to 'wrf' (with a '0' label). The 'wrf' box also has '3 1 2' labels. A zoom slider at the top of the editor is set to 100%.

WRF4SG Prototype

The front-end prototype is now available on ce02.macc.unican.es

It has been configured gLite resources in order to use *esr*, *prod.vo.eu-eela.eu* and *earth.vo.ibergrid.eu* VOs

Climate workflows are available in order to run WRF model and pre and **post process** experiment datasets



WRF4SG is going to ...

- ... achieve more ambitious experiments to WRF community
- ... facilitate the access and management of the available DCIs
- ... ease the diversity of WRF experiment description and execution

Santander Meteorology Group

A multidisciplinary approach for weather & climate

Thank you!

Contact: blancojc@unican.es

More info: <http://www.meteo.unican.es/software/wrf4sg>

Aknowledgments: This work is funded by the Spanish PLAN NACIONAL de I+D+i 2008-2011 (Ref.# CGL2011-28864) and ERDF)

