

International Conference on Computational Science, ICCS 2011

The Climate-G testbed: towards large scale distributed data management for climate change

Sandro Fiore^{1,2*}, Giovanni Aloisio^{1,2}, Peter Fox³, Monique Petitdidier⁴, Horst Schwichtenberg⁵, Sebastien Denvil⁴, Jonathan D. Blower⁶, Antonio Cofino⁷

¹ Euro-Mediterranean Centre for Climate Change (CMCC), Italy

² University of Salento, Italy

³ Rensselaer Polytechnic Institute (RPI) and National Center for Atmospheric Research (NCAR), USA

⁴ Institut Pierre-Simon Laplace (IPSL), France

⁵ Fraunhofer-SCAI, Germany

⁶ University of Reading, UK

⁷ University of Cantabria, Spain

Abstract

Climate-G is a large scale distributed testbed devoted to climate change research. It is an unfunded effort started in 2008 and involving a wide community both in Europe and US. The testbed is an interdisciplinary effort involving partners from several institutions and joining expertise in the field of climate change and computational science. Its main goal is to allow scientists carrying out geographical and cross-institutional data discovery, access, analysis, visualization and sharing of climate data. It represents an attempt to address, in a real environment, challenging data and metadata management issues. This paper presents a complete overview about the Climate-G testbed highlighting the most important results that have been achieved since the beginning of this project.

Keywords: Grid; Data Management; Metadata Management; Testbed; Climate change; Scientific Gateways.

1. Introduction

In the Geosciences contexts there are several research efforts addressing data management issues at different levels. One of them is Climate-G, a research effort (started in 2008) devoted to the climate change community. It is a distributed testbed for climate change acting as a virtual laboratory and addressing challenging data and metadata management issues at a very large scale.

Related work in this area are: (i) at international level the *Earth System Grid Federation* [1], [7] in the US and the

* Corresponding author. Tel.: +39 0832 297371; fax: +39 0832 298173.

E-mail address: sandro.fiore@unisalento.it.

EU funded *IS-ENES* project; and (ii) at national level the *C3Grid* [2] initiative in Germany, the *NERC DataGrid* [3],[16] in UK and the *CMCC* [4],[17] in Italy. All of these projects basically address and share important challenges such as the distributed data management, the transparent access to the data via scientific gateways, the metadata management (both at schema and service level), the data analysis and visualization, the large scale data replication, etc. which are crucial for this community.

The Climate-G testbed is an interdisciplinary effort involving partners from several institutions and joining expertise in the field of climate change and computational science. Its main goal *is to allow scientists carrying out geographical and cross-institutional data discovery, access, analysis, visualization and sharing of climate data.*

The Climate-G partners involved into the testbed are: the Centro Euro-Mediterraneo per i Cambiamenti Climatici (CMCC, Italy), the Institut Pierre-Simon Laplace (IPSL/CNRS, France), the Fraunhofer Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI, Germany), the National Center for Atmospheric Research (NCAR, USA) and Rensselaer Polytechnic Institute (RPI, USA), the University of Reading (Reading, UK), the University of Cantabria (UC, Spain) and the University of Salento (UniSalento, Italy).

This work provides a complete understanding about the testbed, presenting in particular the main challenges, a couple of use cases, the data and metadata infrastructure, the monitoring system, etc. The remainder of this work is organized as it follows: Section 2 highlights the main challenges, issues and requirements related to the testbed. Section 3 presents two relevant use cases, whereas Section 4 describes the path from the use cases to the architectural design towards the infrastructural implementation. Section 5 presents the Climate-G monitoring system. Finally the conclusions are given in Section 6.

2. Main Challenges, issues and requirements

Since its inception in 2008, the Climate-G testbed has been conceived as a large scale data sharing environment focusing on the following crucial challenges: *distributed data management, grid metadata management, data heterogeneity, scientific gateway* and *coexistence of grid and domain-based services*. The motivations behind each of them are presented and discussed in the remainder of this Section.

- *Distributed data infrastructure*: it represents the basic layer to access datasets, make them available to the community, provide fundamental primitives for data subsetting, aggregation and download. It is a key challenge for large scale data infrastructures and at the same time, it can also be considered an enabling factor for these environments since it allows:

- (i) to avoid centralized solutions which are no longer suitable to manage large volume of data (single point of failure and performance bottleneck are just two inherent drawbacks);
- (ii) to preserve sites autonomy implementing proper (loosely coupled) data federation strategies;
- (iii) to implement replication schemas which increase data availability, service performance and system scalability;
- (iv) each site to participate in several and different projects/contexts, without moving or copying the data several times.

Data distribution has been addressed in the Climate-G testbed giving each partner the possibility to join and contribute with new data just adding new data services to the infrastructure and binding them to the closer metadata service. A loosely coupled approach like the one described in Section 4.1 makes easy and transparent the join/leave of the data nodes in the scene without affecting the availability and the stability of the whole system.

- *Grid metadata management*: the metadata management enables data search and discovery, data sharing and a complete description of datasets and experiments. A metadata service represents a fundamental building block for the entire infrastructure, due to the key information it manages. Our attention has been focused on grid-based solutions able to address the following requirements:

- (i) strong support for Virtual Organizations;
- (ii) full compatibility and straightforward integration with the existing production grids such as EGI;
- (iii) uniform access interface and complete support for heterogeneous back-ends where metadata are supposed to be stored;
- (iv) grid-enabled queries addressing performance on a wide area network;
- (v) authorization schemas preserving local autonomy and addressing scalability.

A technological implementation addressing all of the aforementioned points is the GReIC service [5], [9], the

grid metadata solution adopted in Climate-G (see Section 4.1.2 for additional information).

- *Data Heterogeneity*: the data managed in the Climate-G testbed basically fall in two main classes: coarse (see use case 1) and fine grain (see use case 2). In the former case we refer to *files* (basically NetCDF, GRIB and HDF data formats), whereas in the latter to *tuples* (records) stored into relational databases and related to the impacts of climate change on economy. As it can be argued, they need a different support in terms of data management services, APIs and user interfaces. For this reason, in the context of the testbed, two different use cases were defined and implemented (see Section 4).

- *Scientific Gateway*: a pervasive, ubiquitous and seamless access to the entire infrastructure is a critical challenge for large scale environments such as Climate-G. A Scientific Gateway (web-based application) can enable end-users to exploit all of the available functionalities with easy to use high level interfaces. In the Climate-G testbed this challenge has been addressed through a *data-centric* portal in a *portal-centric* infrastructure.

- *Coexistence of grid and domain-based services*: from an infrastructural point of view the Climate-G testbed tries to put together both grid and domain-based services. While the former are basically exploited for metadata management and user credentials at VO level, the latter ones focus more on domain-oriented functionalities like data subsetting and aggregation or download of maps and grids. The coexistence of these heterogeneous services can lead to several kinds of mismatches (e.g. on security) which need to be properly managed. While the *coexistence* of grid and domain-based services is a short-term goal, the *convergence* between Grid [19] and OGC services/specifications represents a longer-term challenge. The Climate-G testbed addresses only the first one (coexistence).

3. Two use cases

The *use case driven design* is the design approach adopted in the testbed. Two significant use cases were identified to drive the design of the system. The first one deals with NetCDF-CF climate change datasets, whereas the second one with climate change impacts on the economy. They address different issues (such as, for example, the management of coarse versus fine grain data or data versus a computational oriented infrastructures) in a complementary way and strongly contribute to the resulting architectural design described in Section 4.

The two use cases, main actors and requirements are presented in the following subsections.

3.1. Use case 1: distributed data and metadata management

The first use case refers to a distributed **data infrastructure** where:

- (a) data consumers must be able to transparently and securely carry out data search & discovery, access and visualization of climate datasets through a high level web interface like a Scientific Gateway;
- (b) data providers must be able to publish their own climate change datasets;
- (c) according to point (b) data providers must be able to edit, modify and validate the related metadata information;
- (d) site administrators must be able to monitor and control the underlying infrastructure, as well as to setup, restart, turn on/off the services they manage in their site.

Important requirements are:

- each participant must be able to participate to the federation managing on site its own datasets;
- according to the previous requirement, each participant must also be able to manage on site the related metadata information;
- metadata information must be compliant with well known standards and directives;
- join and leave steps of a site, must be as light as possible;
- local faults on data and metadata services must not affect the overall system availability;
- community-based tools and domain-oriented services must be integrated in the overall infrastructure;
- the Scientific Gateway must be the only entry point to the infrastructure;
- the Scientific Gateway must integrate data, metadata, tools and services.

The second use case refers to a distributed **computational infrastructure** where:

- (e) users must be able to submit to the grid infrastructure several high throughput jobs related to a specific software application used to model the impacts of climate change on economy;
- (f) users must be also able to monitor, abort or delete their jobs;
- (g) according to point (e), users must be able to publish the output of their jobs into a multidimensional data repository for further data analysis;
- (h) users must be able to transparently and securely carry out data analysis activities through advanced interfaces integrated into the Scientific Gateway.

Important requirements are:

- users belonging to the testbed must be able to run computational jobs on the grid infrastructure;
- the data related to the impacts of climate change on economy must be available for On-Line Analytical Processing (OLAP).
- the Scientific Gateway must provide high level user interface to do data analysis.

4. From the use cases to the architectural design towards the infrastructural implementation

As stated before, the two use cases address different issues in a complementary way, so the resulting “Climate-G picture” is the merge of the data part (use case 1) with the computational one (use case 2). For the sake of clarity, the architectural design and the infrastructural implementation of the two use cases are separately discussed in the following two subsections.

4.1. Use case 1: distributed data and metadata management

4.1.1. Architectural Design

To address the requirements related to the use case 1 and preserve autonomy, locality and scalability, the following model related the generic “Climate-G site” has been proposed: *the Climate-G federation consists of several sites which must own at least one data service and may own at most one metadata service.*

Consequently, at the site level, in terms of data and metadata management, there are two possible scenarios:

- a site owns one or more data services and only one metadata service (full autonomous configuration).
- a site owns one or more data services and it is virtually bound to a metadata service deployed on another “close” site (partial autonomous configuration).

Obviously, for the metadata part, such a distributed architecture needs a specific component, named *harvester*, acting both as a collector and as a high level index with regard to the distributed metadata information available across the network of metadata services. From the end-user point of view (see use case 1), the access to the distributed architecture must be done in terms of search and discovery functionalities, data subsetting and download, maps delivery and visualization, metadata access and editing, service monitoring, that is via several high level interfaces provided by a Scientific Gateway. So, this web application must represent the entry point of the testbed and must integrate all of the operational requirements of the use case 1, coming from the testbed community.

An orthogonal layer, that must be considered into the whole design is the security part. This means that the access to the data, metadata, portal, tools and services must be carried out taking into account at least both authentication and authorization (the need for a set of data policies can be clearly argued by the statements (a), (b), (c), (d)).

4.1.2. Infrastructural implementation

Starting from the architectural design, it was quite straightforward to move towards the infrastructural implementation of the testbed. First of all, the *data services* adopted in the testbed were basically OPeNDAP [6] and THREDDS. An OPeNDAP service is easy to install and manage, providing efficient data access, subsetting and aggregation. On the other hand, THREDDS is able to provide datasets and collections browsing and it includes Web

Coverage Service (WCS) and Web Map Service (WMS) capabilities. OPeNDAP and THREDDS are well known in the community, they address community-oriented needs and requirements and they are widely adopted and deployed in the climate change context; for all of these reasons they represent a natural and in some way a mandatory choice for the data layer of the testbed.

Regarding the *metadata schema* adopted in the testbed, Climate-G did not provide its own schema and did not invest much effort on that point. The participants have agreed to adopt the ISO19115 schema and its related XML implementation (ISO19139) indeed, keeping just the needed (from the testbed point of view) and mandatory (from the ISO19115 point of view) parts of the standard. The ISO standards adopted in the testbed are also compliant with the INSPIRE directives.

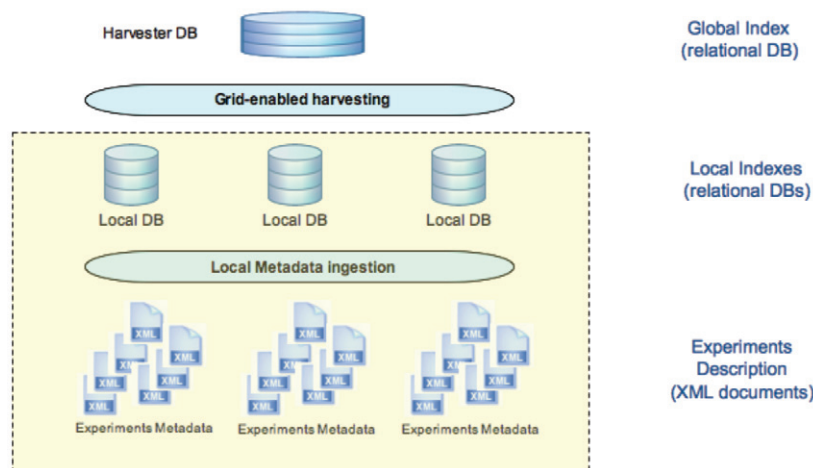


Fig. 1. The three-layer metadata hierarchy

Concerning the *metadata service*, as stated in Section 2, since the beginning our attention has been focused on grid-based solutions able to address performance, compatibility, scalability and transparency requirements and to support Virtual Organizations and more flexible authorization schemas. Definitively, due to its features, flexibility and efficiency, the grid metadata service adopted in the Climate-G testbed was the GReIC service [14].

The GReIC service is a web service GSI [8] and VOMS [10] enabled. It allows users to manage relational and non-relational databases (i.e. XML-DB) in a grid environments. If the target database stores metadata information, the GReIC service acts as a grid metadata service, as in the case of the Climate-G testbed. It is important to remind that the GReIC service is fully compliant with both gLite and Globus and it implements several authorization schemas both at VO and at site levels, addressing high scalability (needed in large scale environments) while preserving local autonomy in terms of policy management.

The metadata management infrastructure implements a metadata hierarchy composed of the following three layers:

- **Level 1** - the complete experiments description stored into XML-based documents and locally managed by a GReIC service;
- **Level 2** - key metadata information about the local experiments (geographical and temporal extents, project name and keywords) managed at site level. These metadata information (stored into a relational database - a local index - and accessible in grid through a GReIC service interface) are useful for local search and discovery purposes;
- **Level 3** - key metadata information (a subset of those ones described in the previous level) about all of the experiments available in the testbed, retrieved from the geographically spread GReIC services and collected into a central repository (a global index) by a harvester. The harvester DB is available through a GReIC

service interface too. In this case the metadata information are useful for distributed search and discovery purposes.

It is worth mentioning that a specific command line interface has been developed to address, at site level, the automatic extraction, transformation and loading into the local relational database of the key metadata information stored into the XML documents (from Level 1 to Level step 2). In the same way, the harvester represents the key point into the infrastructure that automatically retrieves the key metadata information from the distributed local indexes (Level 2) and collects them into the harvester DB (Level 3). Local ingestion and harvester are two GReC clients. Is it important to remind that the proposed harvesting solution is grid-enabled but not OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) compliant.

A key point of the infrastructure is the Climate-G Portal [15] (see Fig 2), which represents the scientific gateway of this collaboration. It is intended for scientists and researchers that want to carry out search and discovery activities on the available large scale digital library. It provides a ubiquitous and pervasive way to ease data publishing, metadata search & discovery, metadata annotation and validation, data access, etc. The Climate-G data portal security model includes the use of HTTPS protocol for secure communication with the client (based on X509v3 certificates that must be loaded into the browser), secure cookies to establish and maintain user sessions as well as a complete role-based authorization system. Login to the system is performed uploading a valid proxy to the portal. The proxy can be created and uploaded directly through the portal running a JNLP application (GReC Proxy Init) embedded into the system, which avoids users having grid-proxy-init or voms-proxy-init command line interface installed on the client side. The GReC Proxy Init application entirely hides the interaction with the VOMS service.

Climate-G

HOME

ABOUT

LOGIN (CERT NEEDED)

SEARCH (CERT NEEDED)

BASE SEARCH

ADVANCED SEARCH

DATASET SEARCH

DASHBOARDS

DASHBOARD

MONITORING

TOOLS/SERVICES

USAGE

LOGOUT

IDENTIFICATION INFO

Abstract

These data represent 6 hourly and 12 hourly instantaneous values of selected variables for ENSEMBLES (<http://www.ensembles-eu.org>). The list of output variables can be found in: <http://ensembles.wdc-climate.de/output-variables> The 20th century simulation(included year 2000) was initiated from the control preindustrial simulation when equilibrium was reached (corresponds to nominal year 1860 of 20C3M). Forcing agents included: CO2,CH4,N2O,O3,CFC11(including other CFCs and HFCs),CFC12; sulfate(Boucher). This is followed by a commitment experiment for the 21th century (year 2001-2100) with all concentrations fixed at their levels of year 2000. These datasets are available in netCDF format. The dataset names are composed of - centre/model acronym (e.g. IPCM4: Institut Pierre Simon Laplace) - scenario acronym (e.g. SRA2: SRES A2) - run number (e.g. 1: run 1) - time interval (MM:monthly mean, DM:daily mean, DC:diurnal cycle, 6H:6 hourly, 12h:12hourly) - variable acronym [with level value] (e.g. hur850: relative humidity, 850 hPa) --> example: IPCM4_SRA2_1_MM_hur850 For this experiment 1 ensemble run was carried out. Technical data to this experiment: IPSL-CM4 (2003): atmosphere: LMDZ (IPSL-CM4_IPCC, 96x72x19); ocean: OPAB.2 (ipsl_cm4_v1_8, 2x2L31); sea ice LIM (ipsl_cm4_v1_8); land ORCHIDEE (orchidee_1_3)

Purpose

Credit

Status

Language

Character Set

Topic Category

Points of Contact

Principal Investigator

Organization: Institut Pierre Simon Laplace

Address

E-Mail: jean-louis.dufraigne@lmd.jussieu.fr

OnLine Resource

URL: <http://ipsl.jussieu.fr/>

Function: information

Citation Info

Title

ENSEMBLES_IPCM4_20C3M_1_6H12H

Alternate Title

Contrasts in the effects on climate of anthropogenic sulfate aerosols between the 20th and the 21st century

Fig. 2. The Climate-G Portal – Metadata visualization

The Climate-G Portal offers the following functionalities:

- distributed and map-enabled data search and discovery;
- grid metadata access, browsing, editing and download;
- support for ncWMS, which implements the OGC WMS specification (provided by the University of Reading and integrated into the infrastructure);
- dataset download via HTTP;
- dataset access and subsetting via OPeNDAP;
- dataset visualization via Integrated Data Viewer (provided by UNIDATA) and Godiva2 [11] web interface (provided by University of Reading and completely integrated into the portal). Support for Google™ Earth is also available through Godiva2.

Currently, the Climate-G Portal publishes about 2TB of data related to the ENSEMBLES [12] project (also including distributed replicas of data) as well as to the IPCC AR4 [13]. Since April 2009, about 100 users (85% coming from the climate change community) got access to the portal. The opportunity to have distributed data (accessible via web in a transparent way), rich metadata descriptions (compliant with well known ISO standards), several data analysis and visualization tools (widely used and acknowledged by the community) harmonically integrated and concentrated into the same environment (the Climate-G Portal) was considered by the Climate-G users as the main achievement concerning the use case 1.

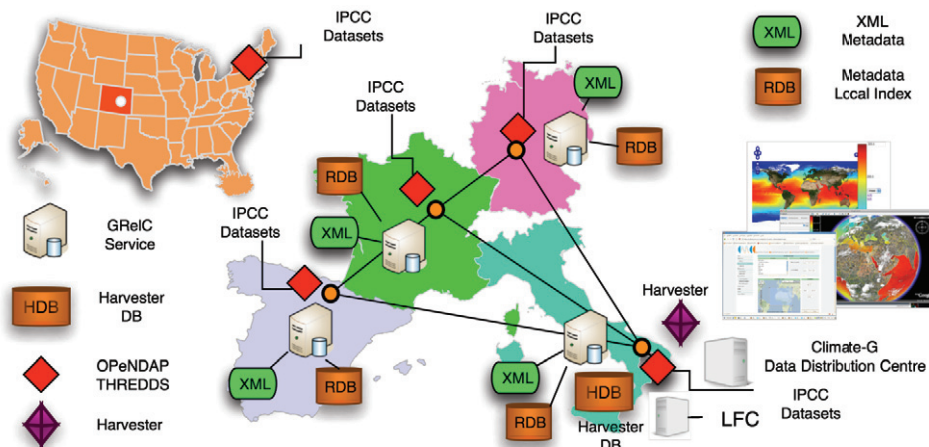


Fig. 3. The Climate-G Infrastructure – use case 1

Figure 3 provides a complete picture about the Climate-G data and metadata management infrastructure related to the use case 1. As it can be argued, the *co-existence* of grid (such as GRelC) and domain-related services (in particular OPeNDAP and ncWMS [11]) was successfully addressed at the tested level. Conversely, the *convergence* between OGC and Grid services still represents an open research issue.

4.2. Use case 2: distributed computational environment for high throughput analysis jobs

4.2.1. Architectural Design

To address the requirements of use case 2 and basically related to high throughput computational jobs it was decided to exploit the **grid paradigm**. A Computational grid allows the concurrent running of multiple independent jobs across a distributed infrastructure. The results of these computational jobs (output data of a CMCC application

used to model the impacts of climate change on economy) are stored into a data warehouse for post analysis and visualization purposes. The grid choice solved almost all of the architectural issues of this second use case.

As in the use case 1, the security part is a key point to be addressed. This means that again, the access to grid the infrastructure must consider at least authentication and authorization (the need for a set of data policies, someone shared with use case 1, can be argued by the statements (e), (f), (g), (g)).

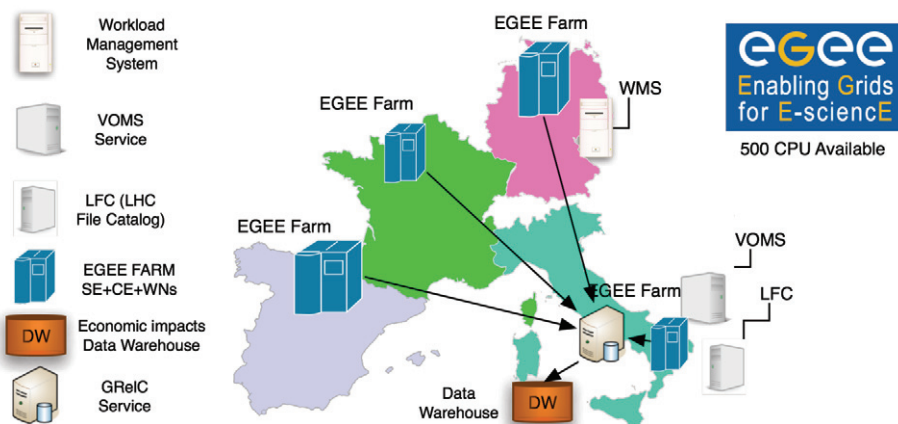


Fig. 4. The Climate-G Infrastructure – use case 2

4.2.2. Infrastructural implementation

To address the second use case, since 2009 a stronger interaction was established with the EGEE Project (now with EGI). First of all, a new EGEE Virtual Organization (VO), named *climate-g.vo.eu-egee.org*, was created for the Climate-G community. This way, by joining the VO, the VO users were allowed to exploit computational resources coming from the EGEE partners taking part into the testbed. Moreover, 500 CPUs (seed resources) were additionally allocated by the EGEE project to the new born VO and exploited to support the Climate-G testbed. A VOMS service as well as a LCG File Catalog were also set up to support VO level authorization schemas and grid file system capabilities. The grid infrastructure provided a strong support to run computational jobs (see Fig. 4). From the testbed participants point of view, there were just some VO-based configuration steps to map the Climate-G VO on the physical resources (farms, GRelC, User Interface, Workload Management System, LHC File Catalog).

The computational jobs ended their run, by storing the impacts data into a data warehouse exposed in grid through the GRelC service running in Lecce. The OLAP primitives available to the end users were basically related to roll-up and drill-down operations across the three analysis dimensions: time, experiment and geographical region. The data warehouse design was driven by the requirements coming from the impacts community involved into the experiment. Anyway, the design and implementation of the data warehouse are out of the scope of this paper and will be presented in a future work.

From a Scientific Gateway point of view, the user interface to do data analysis and visualization was also integrated into the Climate-G Portal.

5. Climate-G Monitoring System

The Climate-G Monitoring System is the Climate-G Portal section devoted to the testbed monitoring purposes. It provides different monitoring views: project, host and service level. The project view (see Fig. 5) provides aggregated metrics like the global availability and failure at the testbed level. The host view provides several metrics at the host level (host availability, failure, round trip time from a central common location) and the complete list of

services running on it. Finally, the service view consists of a complete dashboard with several charts and reports about the service availability and other useful service level statistics and data summaries.

Views, charts and reports can be customized by the user from different points of view. According to the Web2.0 [20] approach, each view can be easily exported into other web-applications, just adding a permalink into the target web page. The monitoring system is devoted to the system administrators of the testbed and provides them with real time monitoring capabilities about the involved resources. Additional information about the Climate-G Monitoring System are out of the scope of this paper and will be provided in a future work.

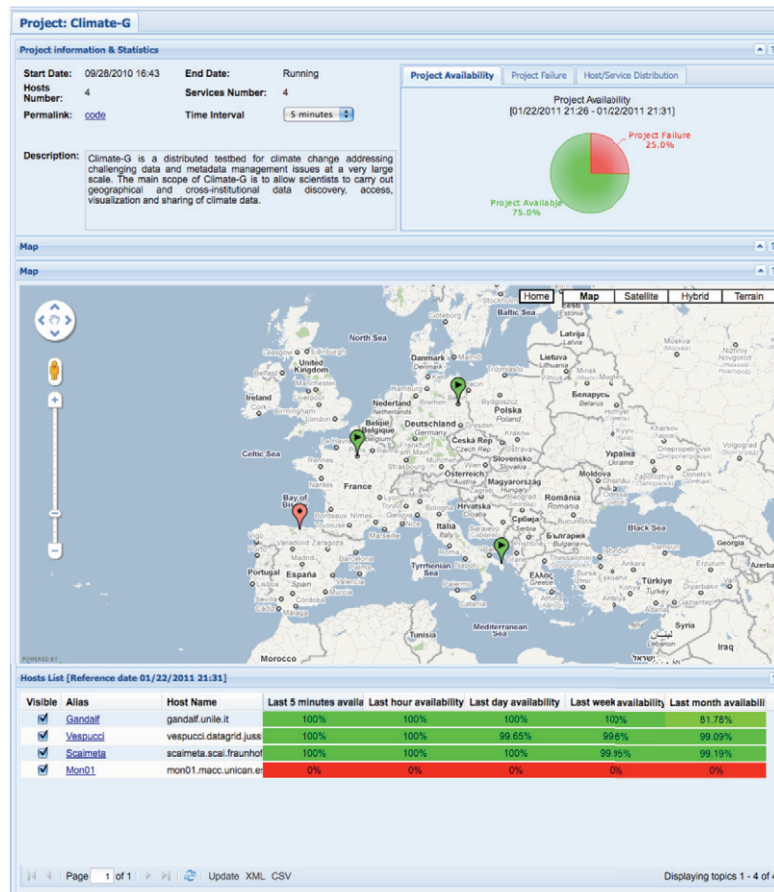


Fig. 5. The Climate-G Monitoring system

6. Conclusions

Climate-G is a distributed environment for studying climate change, acting as a virtual laboratory among several partners both in Europe and US. As described in the paper, the main goal of this testbed is to address large scale data sharing, access, analysis and visualization for the climate change community.

Two use cases have been presented and discussed in detail showing how complementary aspects like the management of *coarse* and *fine* grain data or the implementation of *data* and *computational* oriented infrastructures have been successfully addressed and solved.

A major and long-term challenge addressed by the testbed is the convergence among *Grids* and *OGC services*.

In June 2009, during the first year review of the EGEE-III Project (CERN, Geneva) this work was acknowledged by the EGEE Network Activity 4 Steering Committee, the EGEE Activity Management Board and the European

References

1. D. Bernholdt, S. Bharathi, D. Brown et al “The Earth System Grid: Supporting the Next Generation of Climate Modeling Research”. <http://arxiv.org/pdf/0712.2262v1>, 2007.
2. S. Kindermann, “Climate Data Analysis and Grid Infrastructures: Experiences and Perspectives”, GELA Workshop. IEEE, Paris, France, 2006.
3. K. O'Neill, R. Cramer, M. Gutierrez, K. Kleese van Dam, S. Kondapalli, S. Latham, B. N. Lawrence, R. Lowry, A. Woolf. “A specialised metadata approach to discovery and use of data in the NERC DataGrid”, Proceedings of the U.K. e-science All Hands Meeting. Ed. S.J.Cox, ISBN 1-904425-21-6, 2004.
4. S. Fiore, S. Vadacca, A. Negro, G. Aloisio, “Data Issues at the Euro-Mediterranean Centre for Climate Change”, *Journal of Earth Science Informatics*, pp. 23-35, DOI: 10.1007/s12145-009-0023-x, Ed. Springer, 2009.
5. S. Fiore, A. Negro, G. Aloisio, “The Data Access Layer in the GRelC System Architecture”, accepted for publication in *Future Generation Computer System*, Special Section Data Management for eScience. <http://dx.doi.org/10.1016/j.future.2010.07.006>.
6. J. Gallagher, N. Potter, P. West, J. Garcia, P. Fox, “OPeNDAP's Server4: Building a High Performance Data Server for the DAP Using Existing Software”, AGU Meeting in San Francisco, 2006.
7. D. Middleton, D. Bernholdt, D. Brown et al “Enabling worldwide access to climate simulation data: the earth system grid (ESG). Scientific Discovery Through Advanced Computing”. *J Phys* 46(2006):510–514. 2006.
8. I. Foster, C. Kesselmann, G. Tsudik, S. Tuecke: A security architecture for computational grids. In: Proceedings of 5th ACM Conference on Computer and Communications Security Conference, pp. 83–92 (1998).
9. S. Fiore, A. Negro, G. Aloisio, "The GRelC Project from 2001 to 2011, ten years working on Grid-DBMSs", to appear in book of "Grid and Cloud Database Management", Springer, 2011.
10. R. Alfieri, R. Cecchini, V. Ciaschini, et al, “VOMS, an authorization system for virtual organizations”, European Across Grids Conference 2003:33–40. 2003.
11. J. Blower, K. Haines, A. Santokhee, C. Liu, “Godiva2:Interactive visualization of environmental data on the web”, *Phil.Trans. Roy. Soc. A*, 367, 1035-9, 2009 doi:10.1098/rsta.2008.0180.
12. The ENSEMBLES Project: <http://ensembles-eu.metoffice.com/index.html>.
13. IPCC: Intergovernmental Panel on Climate Change (<http://www.ipcc.ch/>).
14. S. Fiore, A. Negro, S. Vadacca, M. Cafaro, G. Aloisio, R. Barbera and E. Giorgio, “An Architectural Overview of the GRelC Data Access Service”, in *Handbook of Research on Grid Technologies and Utility Computing: Concepts for Managing Large-Scale Applications*, Emmanuel Udoh (Editor), Frank Wang (Co-Editor), pp. 98-108, IGI Global, 2009.
15. The Climate-G Portal - <http://ddc.cmcc.it:8080/ClimateG-DDC-v2.0/>
16. K. O'Neill, R. Cramer, M. Gutierrez, K. Kleese van Dam, S. Kondapalli, S. Latham, B. N. Lawrence, R. Lowry, A. Woolf , “The Metadata Model of the NERC DataGrid”, Proceedings of the U.K. 2003. e-science All Hands Meeting. Ed. S.J.Cox ISBN 1-904425-11-9.
17. S. Fiore, S. Vadacca, A. Negro, G. Aloisio, “Euro-Mediterranean Centre for Climate Change Data Grid”, Proceedings of DAPSYS 2008, LNCS - September 3-5, 2008 - Debrecen, Hungary - pp. 61-74.
18. I. Foster, C. Kesselman, S. Tuecke, “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”, *International J. Supercomputer Applications*, 15(3). 2001.
19. F. Berman, G. Fox, T. Hey, “The grid: past, present, future”, in: F. Berman, G. Fox, T. Hey. *AJG (eds) Grid computing*. Wiley, New York, pp 51–63. 2003.
20. T. O'Reilly, What is Web 2.0 – design patterns and business models for the next generation of software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.