The Climate-G Portal: The context, key features and a multi-dimensional analysis

Sandro Fiore [a,b,*], Alessandro Negro [b], Giovanni Aloisio [a,b]

[a] *Euro-Mediterranean Center for Climate Change, Via A. Imperatore 16, 73100 Lecce, Italy*
[b] *University of Salento, via per Monteroni, 73100 Lecce, Italy*

ARTICLE INFO

ABSTRACT

Grid portals are web gateways aiming at concealing the underlying infrastructure through a pervasive, transparent, user-friendly, ubiquitous and seamless access to heterogeneous and geographically spread resources (i.e. storage, computational facilities, services, sensors, network and databases). The Climate-G Portal is the web gateway of the Climate-G testbed (an interdisciplinary research effort involving scientists both in Europe and US) and it is devoted to climate change research studies. The main goal of this paper is to present the Climate-G Portal providing a complete understanding of the international context, discussing its main requirements, challenges, architecture and key functionalities, and finally carrying out and presenting a multi-dimensional analysis of the Climate-G Portal, starting from a general schema proposed and discussed in this work.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Grid portals [1,2] are web gateways aiming at concealing the underlying infrastructure through a pervasive, transparent, user-friendly, ubiquitous and seamless access to heterogeneous and geographically spread resources (i.e. storage, computational facilities, services, sensors, network and databases). Definitively they provide an enhanced problem-solving environment able to deal with modern, large scale scientific and engineering problems.

Scientific gateways are able to introduce a revolution in the way scientists and researchers organize and carry out their activities. Access to distributed resources, complex workflow capabilities, and community-oriented functionalities are just some of the features that can be provided by such a web-based environment.

In the context of the EGEE [3] NA4 Earth Science Cluster, Climate-G is a distributed testbed focusing on climate change research topics. The Euro-Mediterranean Center for Climate Change (CMCC) is actively participating in the testbed providing the scientific gateway (Climate-G Portal) to access to the entire infrastructure.

The main goal of this paper is to present the Climate-G Portal (i) providing a complete understanding of the international context (the Climate-G testbed and its distributed architecture), (ii) discussing requirements, challenges, architecture and main functionalities, and finally (iii) carrying out and presenting the

Climate-G multi-dimensional analysis starting from the general schema proposed and discussed in Section 2.

The outline of the paper is as follows. In Section 2, a multi-dimensional analysis of scientific gateways is presented. It defines a schema based on nine dimensions identifying important milestones referred to three historical phases about scientific gateways: Pioneeristic era, Service-Oriented Architecture (SOA) vision and Broad Grid revolution. Section 3 introduces the context, that is the Climate-G initiative, architecture and infrastructure whereas Section 4 addresses the Climate-G Portal presenting its requirements, architecture and main functionalities. Section 5 proposes the Climate-G multi-dimensional analysis starting from the general schema discussed in Section 2. Finally, Section 6 presents conclusions and future work.

2. A multi-dimensional analysis

Since the beginning of the *Grid* era, grid portals have played a key role in providing a seamless and ubiquitous way to access to the cyberinfrastructure.

A lot of things have been changed during the past decade according to several dimensions: *target, domain, class of users, technology, scale, environment, software re-usability, information architecture and simplicity*. In the remainder of this section, three main phases (Pioneeristic era, SOA vision and Broad Grid revolution) are described highlighting important milestones in our multi-dimensional analysis.

In the *first phase* (through 2001, *pioneeristic* era), grid portal projects were focused on general problems related to file transfer management, job submission, job monitoring, data management,

---

* Corresponding author at: Euro-Mediterranean Center for Climate Change, Via A. Imperatore 16, 73100 Lecce, Italy.

E-mail address: sandro.fiore@unile.it (S. Fiore).

etc. providing a web-based interface to middleware (basically Globus [4]) services. A complete analysis of the first portal efforts can be found in [5].

The main target was the *user* and such a class of portals basically provided the following functionalities:

- ability to specify the users *grid* (set of machines and more in general resources),
- job submission, check status, job delete,
- monitoring and information service,
- user profile management,
- file transfer from different grid resources exploiting GridFTP [6].

Such kind of functionalities can be classified as middleware level or *no-domain*. This is because at that time grid portals did not focus on a specific application domain, but rather on middleware services.

Consequently, *simplicity* was not addressed, because portals were oriented to *expert users* with a specific skill on grid computing especially at the middleware level. An immediate concern was the *scale* (less than thousand) of users that was obviously very low due to the technicalities and middleware implications and without community-based organizations.

The grid environments were usually *testbeds* allowing scientific research without production level quality standards in terms of resource availability, quality of service, etc. In this phase there was poor *re-usability* and *interoperability*, due to the total absence of portal programming interfaces, web service approach, standardized interfaces. This meant different projects starting from scratch, addressing the same goals, providing similar functionalities without collaborating, re-using, sharing code or making interoperable services and components.

In the *second phase* (after 2002, *SOA vision*), web service and portlets [7] changed the technological landscape basically providing a higher level of *interoperability* and *re-usability*. Exploiting web service approach [8,9] portal functionalities were completely decoupled from the presentation layer. The SOA standard vision (everything is a service) promoted service composition, service orchestration and more in general the complex workflow challenge [10,11].

Such a different methodology to build portals led to a higher level of *simplicity* in developing new portals, new functionalities, allowing people to concentrate on solving specific domain-oriented problems. The grid portal as a grid problem-solving environment (GPSE) started to be a more mature concept. Portals moved from *no-domain* to *single-domain*-oriented approach providing a strong support to the scientific research in several applications areas. Obviously, natural implications were the increased *scale* (thousands) of users (people from the application domain areas were attracted by the GPSE approach), the *simplicity* provided by the nature of the web, new opportunities coming from workflow mechanisms promising new frontiers in running complex jobs on distributed environments [12].

In the meanwhile stable production grid environments *with a particular application, technology or geographic focus* (defined in [13] as Narrow Grids) became a reality both in US and Europe (i.e. Teragrid and EGEE) giving groups of scientists belonging to Virtual Organizations, the opportunity to run millions of jobs on distributed and high performance computing machines. Consequently, huge volumes of data coming from scientific experiments, led to new exciting challenges centered around the word *data*: management, federation, mining, access, processing, visualization, etc. A lot of portals started providing a new set of data-oriented functionalities allowing users to search and discover distributed and geographically spread datasets, index large data collections, manage metadata, etc. The data deluge critically changed scenarios and needs. The transition from the first phase to the second one led to moving the *target* from *user* to *service* and

the *information architecture* from *data* to *information*, with a strong interest on metadata management and related services.

The *third phase* (after 2005, *Broad Grid revolution*) is characterized by a set of competing and different technologies such as grid middleware, web services, Web2.0 components, etc. Such a revolution impacts novel scientific portals [14] by (i) fostering the introduction of collaborative, cooperative and *people-oriented* functionalities, (ii) leveraging on client side Javascript libraries to build and compose grid services, (iii) building folksonomies of scientific user communities through social bookmarking, [15], etc. A community-based approach is definitely moving the *target* from *service* to *people* and, considering the *environment* dimension, from *Narrow grids* to the novel *Broad Grid* concept [13]. Web2.0 [16] success relies on several factors: *simplicity* and *re-usability* are two of the most notable ones.

The *simplicity* [17] of the approach is very different (and less standard) from the complex and huge SOA one. WS-* specification adoption failed for the complexity and quantity of interfaces. *Simplicity* of people-oriented approaches leads to a different *scale of users* (millions) as we can easily verify in well-known Web2.0 systems like Facebook, YouTube, Twitter. On the other hand, *re-usability* is consistently increased as it can be easily argued by (i) the higher number of re-usable components online available [18], (ii) the wide and fast adoption of components in mashups, like Google Maps, (iii) and its different (most important) role in new research projects (an example is myExperiment proposing use, re-use, repurposing and sharing of workflows in the Web2.0 perspective) [19–21].

In this phase, a wider and semantic integration of resources, a different target (from *service* to *people*) and an increasing need to infer new knowledge from the large volume of data, lead to a *knowledge-based information architecture*. Focusing on *people*, it becomes more and more important to understand (and foresee) users' desires and needs as well as users' behavioral patterns. Moreover, since large communities or groups of people are often very heterogeneous (talking about scientists we could focus on the interdisciplinary facet), the needed support should naturally follow a *multi-domain* approach (*everything is linked*). This should be reflected in the new scientific gateway generation that could enable a big revolution in eScience by exploiting Web2.0 approach. Scientific gateways could drastically change their role and impact, by providing new community-oriented functionalities, enabling a stronger collaboration, sharing and interaction among people and definitely considering *people as the big resource*.

A schematic representation of the three phases (according to our multi-dimensional analysis) is depicted in Fig. 1. In the remainder of this paper, the Climate-G initiative [22], related architecture and in particular, the Climate-G portal will be described. Following our schema, a multi-dimensional analysis of the Climate-G portal will be also carried out at the end of this work.

3. The Climate-G initiative

The Climate-G testbed is an interdisciplinary and unfunded research project joining expertise in the field of climate change and computational science. The primary goal of Climate-G is to create a Virtual Research Environment allowing scientists and policy makers to transparently and securely perform geographical and cross-institutional data discovery, access, visualization, analysis and sharing of climate change and impacts data through the Climate-G Portal.

Several partners belong to the collaboration; after the initial phase involving *Centro Euro Mediterraneo per i Cambiamenti Climatici, Institut Pierre-Simon Laplace and Fraunhofer Institut fr Algorithmen und Wissenschaftliches Rechnen other centers and institutions such as National Center for Atmospheric Research and Rensselaer Polytechnic Institute, University of Reading, University of Cantabria, and University of Salento* joined the testbed.
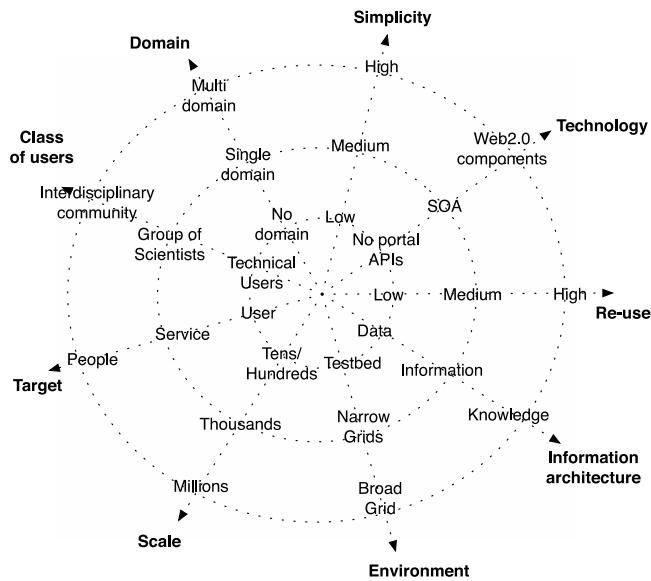
Fig. 1. Multi-dimensional analysis.

Fig. 2. Climate-G Infrastructure.

3.1. The Climate-G architectural design

The Climate-G testbed exploits the grid paradigm to manage widespread and heterogeneous resources distributed across several institutions and domains. From an architectural point of view, the design of the testbed reflects the following classical stack: Fabric, Middleware and User Interface.

– *Fabric*: it represents the lowest layer, which is the physical infrastructure.
– *Middleware*: a set of services acting as a glue. The adopted services fall in two main categories: *general purpose* and *domain-based*. The former addresses computational and data-oriented general issues and needs exploiting the distributed infrastructure. The latter has to provide domain-oriented functionalities exploiting well-known, widely deployed and adopted (in the climate change community) services.
– *User Interface*: a Scientific Gateway represents the access point to the entire infrastructure and it provides a set of user-friendly web-based interfaces to transparently, seamlessly and securely manage collections of data and experiments, storage and computational resources, users, services, and so forth.

Finally, a fundamental and orthogonal component of this architecture is *security*. The system has been secured by design paying special attention to authentication, authorization, delegation and communication protection.

3.2. The Climate-G Infrastructure

In this section, the infrastructural choices are described according to the architectural design. Main components (see Fig. 2) are *fabric, middleware, user interface* and *security infrastructure*.

3.2.1. Fabric

In our testbed, it is comprised of data, metadata, network, storages, computational resources.

3.2.2. Middleware

The adopted grid services come from two of the most notable middleware developments: Globus and gLite.

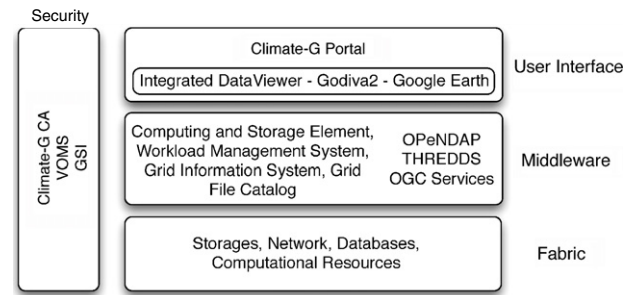In terms of *data management*, the data services exploited in the testbed are mainly:

– gLite Disk Pool Manager (DPM), OPeNDAP [23] and ncWMS services. While the first one is general purpose (it provides data-oriented functionalities regardless of data semantic) the other two can be classified as domain-oriented (delivery of maps, slicing and dicing of data, are just some examples of provided functionalities) and are adopted and deployed by the testbed community.
– gLite DPMs make datasets available through Globus Grid Security Infrastructure (GSI) [24] enabled interfaces, support GridFTP as data transport protocol, provide VO-oriented support and VO-based authorization. Replicas of data stored on several gLite DPMs can be managed through the LHC File Catalog (gLite LFC).
– OPeNDAP service moves data across the Internet using existing and well-understood technologies (based on the HTTP protocol).
– ncWMS is a Web Map Service for geospatial data that are stored in CF-compliant NetCDF files. It is designed to support interactive browsing of data over the Web and supports OGC WMS 1.3.0 specification.

Concerning *metadata management*, Climate-G exploits a distributed metadata solution (leveraging grid technologies) named GRelC DAIS [25]. The GRelC DAIS (included in the EGEE RESPECT Program) enables geographical data sharing, search and discovery and it is able to manage metadata information held in heterogeneous grid data sources.

It exposes a GSI and VOMS (Virtual Organization Membership Service) [26] enabled interface providing full security support (in terms of authentication, authorization, data integrity, data confidentiality and delegation). In the Climate-G testbed, metadata are stored both in relational databases (Index-DB) and XML ones (Metadata-DB) and are available through the same grid enabled web-service-based GRelC interface. While the Index-DBs contain just key information, the Metadata-DBs contain full metadata descriptions stored into XML files. Harvesting of metadata is performed through the GRelC Harvester component built on top of the GRelC service. Presently, the metadata schema managed by the GRelC metadata services is compliant with the ISO 19115/19139 standards. The GRelC metadata services (see Fig. 3) are P2P [27] connected in a self-monitored and self-controlled network. Monitoring facilities are also provided by this service giving the administrators full control of the underlying metadata system with real time monitoring capabilities, reports and statistics about the involved resources. Additional details about the internals of this service can be found in [28].

Finally, with regard to the *computational infrastructure*, the Climate-G testbed is now exploiting the computational and storage resources belonging to the EGEE NA4 partners involved into the EGEE Earth Science Cluster activities. Four sites (Fraunhofer-SCAI, IPSL-CNRS, University of Cantabria and SPACI) provided about 300 CPUs to run post-processing and analysis software. Each site participates in the testbed with a classical EGEE Farm hosting
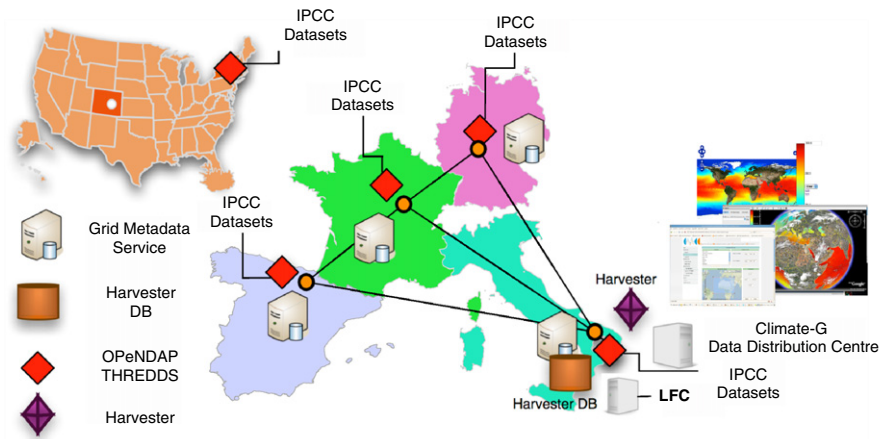
Fig. 3. Grid metadata management framework (GRelC system).

the following components: computing element, storage element and several worker nodes. The provided infrastructure is basically IA32-based. Additional resources (about 200 CPUs) are being provided by the EGEE Project itself (namely seed resources) due to the interest around the application domain and the scientific activity.

3.2.3. Security

The Globus Grid Security Infrastructure (GSI) [24] has been adopted in this testbed as it represents a standard de facto in the grid projects; X509v3 digital certificates (issued by trusted EUGridPMA certification authorities) are used to identify users; a VOMS service manages global authorization at the VO level whereas Access Control Lists on the service side manage local authorizations. A wide set of data access policies allows a variety of capabilities that can be granted or revoked to the users.

Starting from the Climate-G activity, since April 2009 in the context of the EGEE Project, a new Virtual Organization has been created to support the climate change community needs. About 100 users joined the VO and more than 85% come from the application domain.

3.2.4. User interface

The user interface in our testbed is the Climate-G Portal. It is web-based and Section 4 gives a complete description about it.

4. Climate-G Portal

The Climate-G Portal is the access point to the entire testbed infrastructure. It represents the scientific gateway of the testbed and it is intended for scientists and researchers that easily and transparently want to manage the available climate change experiments and datasets.

Basically, *all of the activities related to the testbed infrastructure must be carried out exploiting high level web interfaces available through the Climate-G Portal.* This means that all of the involved actors in the scene (system/portal administrators, scientists, guest users, data and metadata providers, etc.) have to perform most/all of their activities through the portal.

Fig. 4 shows the central role of the portal in the Climate-G infrastructure. Despite the previous general architecture, the figure highlights all of the interactions between the portal and the distributed testbed infrastructure. In the following subsections challenges and requirements, architectural design and key functionalities will be presented.

4.1. Main challenges and requirements

The Climate-G Portal has to face important and critical challenges and satisfy and address key requirements. In the following, the most relevant ones are presented and discussed.

Transparency: the portal has to provide transparent access to the underlying infrastructure preventing users from dealing with low level details and the complexity of a distributed grid environment.

Security: users must be authenticated and authorized on the portal to access and exploit portal functionalities. A wide set of roles is needed to clearly assign the proper one to each user. The access to the computational grid must be completely secured, since the target infrastructure to run jobs is a production grid environment. A security infrastructure (based on X509v3 digital certificates) is strongly needed.

Pervasivity and *ubiquity*: the access to the system must be pervasive and ubiquitous. This is easily true due to the nature of the needed web approach.

Usability and *simplicity*: the portal has to provide simple, high level and user-friendly interfaces to ease the access and exploitation of the entire system.

Coexistence of general purpose and domain-oriented services: along with general purpose services (file transfer, job submission, etc.), the portal has to provide domain-based services and functionalities.

Subsetting (slicing and dicing) of data, visualization of 2D maps around a virtual globe, delivery of maps through OGC-compliant interfaces (i.e. Web Map Service—WMS) are just some examples.

Integrated working environment: the Climate-G Portal must represent an integrated working environment where scientists can find huge amounts of data, complete metadata support, a wide set of data access services, data visualization and analysis tools, easy access to the underlying grid infrastructure (i.e. workflow capabilities), advanced monitoring interfaces, etc.

4.2. Portal architecture

In this section we will delve into details of the Climate-G architectural design, internal components and modules. The design of the Climate-G Portal follows a layered architecture comprised of the following levels: Java CoG, Axis, GRelCJProxy and Climate-G Portal Presentation Layer.

The Java CoG Kit [29] provides the implementation of Java-based GSI, gridFTP, myProxy, GRAM client implementations. Among these functionalities, the concept of grid abstractions
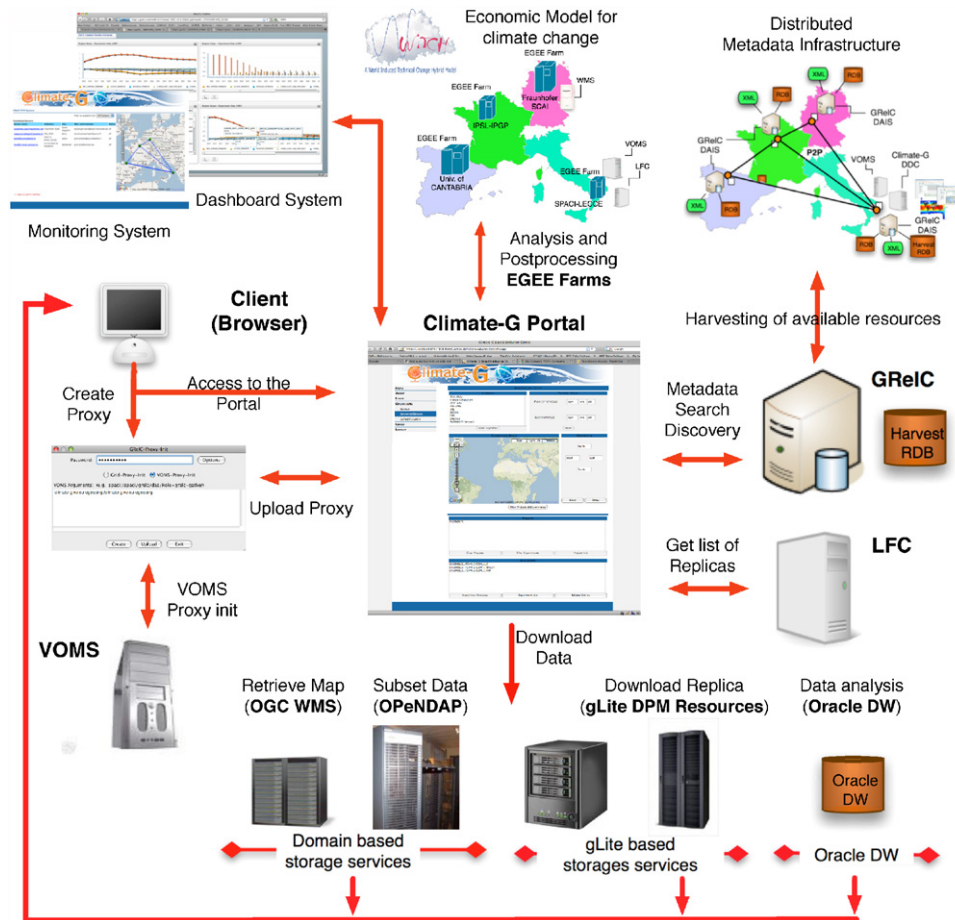
Fig. 4. A portal-centric view of the Climate-G Infrastructure.

and providers, workflows, and support for portal developers can be enumerated. However, in the proposed system, Java CoG is exploited just to provide GSI support.

The Axis level refers to the packages exploited to provide SOAP messages marshalling and unmarshalling, stubs and skeletons for the GRelC service and other web/grid services.

The GRelCJProxy relates to the GRelC client side software developed as Java APIs. It provides all of the functionalities needed to interact with and manage the GRelC service with a higher level of abstraction, hiding all the details related to the grid infrastructure, security and WS paradigm. Some examples of available methods concern query submission (both synchronous and asynchronous [30]), user and VO management, logging support, grid-database management (both relational and XML), P2P integration query, etc.

The Climate-G Presentation Layer represents the Climate-G Portal upper layer. Main functionalities, like dataset search, metadata access exploit the SOA paradigm through the available interfaces provided by the GRelC DAIS service. As stated before, in the Climate-G context the GRelC DAIS works as metadata management service, providing metadata harvesting capabilities and allowing user to search and discovery data as well as manage metadata in grid.

Moreover, to simplify user-portal interaction, parts of the functionalities provided by this layer are composed as a mashup of functionalities coming from different sources, according to the Web2.0 paradigm.

In more detail we integrate in our portal service, APIs, interfaces and tools like:

- Google Maps APIs: to allow user to select a specific geographical area of interest and to show geo-referenced data;
- Godiva2 interface [31]: it displays 2D maps (and animations) and also exports maps for display on the Google Earth virtual globe.
- THREDDS: to enable experiments and datasets browsing via web;
- Integrated Data Viewer (IDV): to visualize data through a Java Web Start Application available from Unidata/UCAR;
- OPeNDAP: to make data available via HTTP and to seamlessly give users access to data in a variety of different formats. Dataset access form, dataset data structure and dataset attribute structure web pages have been easily integrated into the portal;
- GRelC Proxy Init: it allows users to generate a valid grid proxy (both with and without VOMS extensions) by using a simple GUI available as Java Web Start Application. Consequently, no GSI Command Line Interface (CLI) needs to be installed on the client side to use and access to the portal. This component comes from the GRelC Portal [32].

From a technological point of view the Climate-G presentation layer exploits JSP/Servlet. Moreover we chose and adopted AJAX, as communication paradigm between the pages and the web server, and JQuery as library and programming model to introduce a higher level of dynamicity, simplify the portal usage and improve re-usability.

Benefits connected with this approach concern with the increased web page's interactivity, speed, functionality and usability. Cascading Style Sheets (CSS) to separate the stylistic elements of a page from the content and Javascript-based client side input validation are also widely adopted.

Finally, the entire system is managed by a Tomcat container version 6.0.20.

4.3. Portal functionalities

The main features currently available on the Climate-G Portal are mainly related to the search functionalities. Through the portal the users can perform data search and discovery of climate datasets, browsing and searching through the available climate change experiments and projects.

A key non-functional requirement is simplicity, that has been addressed in the Climate-G Portal by means of user-friendly and highly interactive search interfaces able to help people from different domains to easily and quickly search, find and access to the relevant data.

In more detail the user can:

- search data via three different search pages;
- filter available datasets by project, by experiment, by variable and by geographical and temporal extent;
- view metadata related to dataset, experiment or project;
- visualize datasets exploiting functionalities provided by Godiva2 Interface and IDV;
- download datasets from different data sources and via different protocols (i.e. HTTP).

To ease the data search and discovery, the Climate-G Portal provides three different types of search (base, advanced and direct dataset search) with different capabilities and filters.

The *base search* is the simplest and direct way to search for a dataset. In this case the user has to choose a set of projects, filter the related experiments, pick up a specific experiments and related datasets. Using *advanced search* the users can filter projects and experiments specifying keywords, temporal extent and geographical extent of the needed datasets.

Finally *direct dataset search* interface allows users to filter dataset specifying a particular variable (air pressure, sea surface temperature, etc.).

The user can specify the target of each search choosing a metadata server or the harvester. The first option allows to search across the datasets managed by a specific metadata server (a single GRelC DAIS); conversely the second one allows to search across all of the available datasets (managed by multiple GRelC DAIS servers) by querying just the harvester.

Once the user discovers a dataset, she can carry out (via the portal) several actions such as dataset download, subsetting, visualization, etc., leveraging on functionalities developed in the context of the project or available through external components like Godiva2 and IDV.

The Climate-G Portal provides another important functionality, that is the monitoring system. Since search queries can be distributed across several GRelC DAIS servers, it becomes strategic to monitor the status of the different servers involved in the P2P network of GRelC DAIS. For this reason, by means of a specific monitoring page administrators can know exactly the current status of each GRelC DAIS server in the P2P network, close relations of each server, their geographical location, the supported VOs, the grid-databases handled by each server, site specific information (site-name, contacts, etc.).

To get access to the portal the user has to login. The login procedure requires that the user has to upload on the portal a valid X509v3 proxy file and she must have a valid grid certificate loaded into the browser.

The certificate is used to authenticate the user, while the proxy will be used by the portal (on behalf of the user) to carry out search and discovery activities on the GRelC DAIS servers as well as on the harvester, to submit jobs on the EGEE grid, transfer files from grid storages, etc.

The portal access control system leverages on a Role-Based Access Control (RBAC) [33] mechanism that involves the use of filters to protect the different pages and functionalities. Thanks to this approach we can handle several types of user, each one with specific capabilities related to the pages that they can access to or actions that they can accomplish.

We can define at *web page, action* or *button* levels a minimum capabilities mask, such as a bitmap representing a set of permissions that a user must have to (i) gain access to a specific web page, (ii) perform a specific action or (iii) view a specific object (e.g. button).

Portal administrators can grant and revoke security roles assigning the right privileges to each user (*full admin, data provider, metadata provider, banned*, are just some examples). A role represents a set of capabilities. During the login phase the system reads the user role from the catalog and creates the related user capabilities mask. When a user tries to access a web page or to accomplish a specific action the associated filter verifies that the user capabilities mask matches the minimum capabilities mask needed to access to the page or to perform the action. Moreover, when a web page is created during a specific user session, the system verifies for each object loaded into the web page that the user role contains the minimum capabilities needed to see it. This means that objects are dynamically loaded (or hid) into the web pages according to user privileges.

5. Climate-G Portal: a multi-dimensional analysis

An analysis of the Climate-G Portal (see Fig. 5) can be performed according to the nine dimensions reported in Section 2.

Domain: the Climate-G Portal is devoted to several application domains. For instance available and registered data comes from atmospheric, oceanographic and impact (i.e. on economy) scientific groups.

Class of users: the Climate-G class of users is an interdisciplinary community. Economists, scientists in the atmospheric and oceanographic domains, numerical research team and computational scientists represent some of the most relevant groups of users.

Target: community-based functionalities are not yet integrated into our gateway. Anyway, future releases will be more targeted on people. Right now main target of the portal is the service.

Scale: the scale of users is today low. As stated before, since April 2009, about 100 users registered the portal. We expect to increase the number of users by (i) increasing the volume of data available through our system, (ii) adding new functionalities more targeted on people and enabling community-based activities (sharing of experiments, best practices, user workspace, etc.) centered around climate change topics, (iii) providing new and high quality tools directly accessible through the portal and strongly exploiting client side capabilities and (iv) working on different authentication mechanisms and security infrastructures solutions as grid certificates represent a strong barrier to attract new users [34].

Environment: the environment of the testbed is the EGEE Production grid, that is a notable example of Narrow Grid.

Information architecture: the information architecture is now centered on information. Metadata management is a key point addressed by the GRelC service and a lot of functionalities are already available through the portal.

Re-usability: this is a key point we aim at addressing with better results by strongly following a Web2.0 approach. The current re-usability level can be considered at a medium level (server side functionalities provided through SOAP-based services). In the future Climate-G APIs will be available to developers allowing
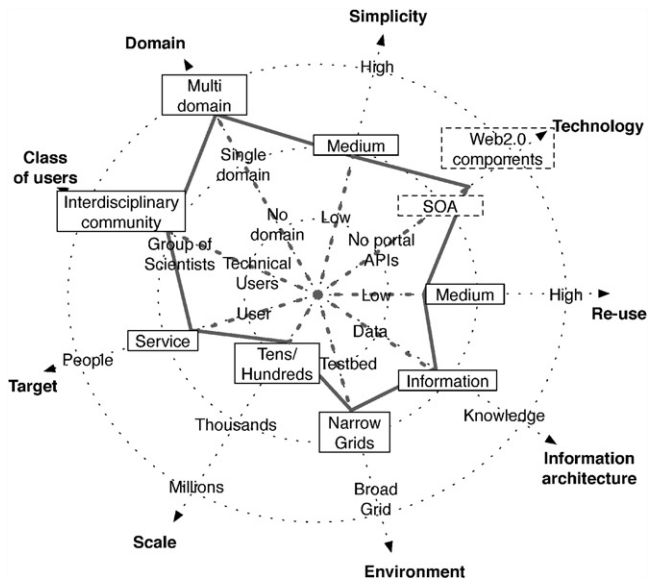
Fig. 5. Climate-G Portal multi-dimensional analysis.

them to integrate, e.g., climate search capabilities in their own portal.

Technology: the Climate-G implementation is characterized by Java servlets and fast adoption of components in mashups, like Google Maps. A stronger adoption of Web2.0 components is already part of our future developments.

Simplicity: the approach is simple from a scientist point of view, but simplicity cannot be considered so high if the final target is a wider user community.

6. Conclusions and future work

This paper focused on the Climate-G Portal, highlighting and discussing both the context (the Climate-G testbed) and the main requirements, challenges, architecture and key functionalities.

The paper also proposed a multi-dimensional schema based on nine dimensions (*target, domain, class of users, technology, scale, environment, software re-usability, information architecture* and *simplicity*) identifying important milestones related to three main phases discussed in this work: Pioneeristic era, SOA vision, Broad Grid revolution.

Finally, a multi-dimensional analysis of the Climate-G Portal has been presented according to the schema proposed in this work.

Future work will relate to some Climate-G Portal extensions in terms of workflow capabilities and community-oriented functionalities. Following the Web2.0 paradigm, particular attention will be paid to the Climate-G Portal APIs in order to release them in the near future. A refinement of the proposed multi-dimensional analysis will be part of our future work.

References

[1] D. Gannon, G. Fox, M. Pierce, B. Plale, G. von Laszewski, C. Severance, J. Hardin, J. Alameda, M. Thomas, J. Boisseau, Grid portals: a scientist's access point for grid services, GGF Community Practice Document, Working Draft 1, September 2003.

[2] M.P. Thomas, et al., Grid portal architectures for scientific applications, Journal of Physics: Conference Series 16 (2005) 596–600. doi:10.1088/1742-6596/16/1/083.

[3] EGEE Project. http://eu-egee.org/.

[4] I. Foster, C. Kesselman, Globus: a metacomputing infrastructure toolkit, International Journal of Supercomputer Applications 11 (2) (1997) 115–128.

[5] Geoffrey Fox, Dennis Gannon, Mary Thomas, Overview of Grid Computing Environments, in: Fox, Berman, Hey (Eds.), Grid Computing: Making the Global Infrastructure a Reality, Wiley, 2003, Chapter 20.

[6] J. Bresnahan, M. Link, G. Khanna, Z. Imani, R. Kettimuthu, I. Foster, Globus GridFTP: whats new in 2007, in: Proceedings of the First International Conference on Networks for Grid Applications, GridNets 2007, 2007, p. 15.

[7] J. Alameda, M. Christie, G. Fox, J. Futrelle, D. Gannon, M. Hategan, G. Kandaswamy, G. von Laszewski, M.A. Nacar, M. Pierce, E. Roberts, C. Severance, M. Thomas, The open grid computing environments collaboration: portlets and services for science gateways, Concurrency and Computation: Practice and Experience 19 (6) (2007) 921–942.

[8] D. Gannon, B. Plale, M. Christie, L. Fang, Y. Huang, S. Jensen, G. Kandaswamy, S. Marru, S.L. Pallickara, S. Shirasuna, A. Slominski, Y. Simmhan, Y. Sun, Service-oriented architectures for science gateways on grid systems, in: B. Benatallah, F. Casati, P. Traverso (Eds.), International Conference on Service-Oriented Computing, ICSOC, in: LNCS, vol. 3826, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 21–32.

[9] D. Gannon, B. Plale, D. Reed, Service architectures for e-science grid gateways: opportunities and challenges, in: Proceedings of Grid Computing, High Performance and Distributed Applications, GADA'07, Springer Verlag, Vilamoura, Algarve, Portugal, 2007, 2007.

[10] D. Gannon, G. Fox, in: Workflow in Grid Systems Concurrency and Computation: Practice and Experience, vol. 18, No. 10, August 2006, pp. 1009–1019. Editorial of special issue prepared from GGF10 Berlin.

[11] S. Perera, Suresh Marru, Chathura Herath, Workflow infrastructure for multi-scale science gateways, in: TeraGrid Conference, June, 2008.

[12] B. Plale, D. Gannon, J. Brotzge, K. Droegemeier, J. Kurose, D. McLaughlin, R. Wilhelmson, S. Graves, M. Ramamurthy, R.D. Clark, S. Yalda, D.A. Reed, E. Joseph, V. Chandrasekar, CASA and LEAD: Adaptive Cyberinfrastructure for Real-Time Multiscale Weather Forecasting, in: Computer special issue on System-Level Science, 39, IEEE Computer Science Press, 2006, pp. 56–63. http://doi.ieeecomputersociety.org/10.1109/MC.2006.375.

[13] G. Fox, M. Pierce, Grids challenged by a Web2.0 and multicore sandwich, in: Special Issue of Concurrency and Computation: Practice and Experience, on Seventh IEEE International Symposium on Cluster Computing and the Grid—CCGrid 2007.

[14] G.C. Fox, M.E. Pierce, Web2.0 for eScience: SC07 education program tutorial education program tutorial at SC07 November 12, Reno, Nevada, 2007.

[15] Marlon E. Pierce, Geoffrey C. Fox, Jong Y. Choi, Zhenhua Guo, Xiaoming Gao, Yu Ma, Using Web2.0 for scientific applications and scientific communities Concurr, Computing: Practice and Experience 21 (5) (2009) 583–603.

[16] T. OReilly, What is Web2.0 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.

[17] G. Fox, M. Pierce, Grids meet too much computing, too much data and never too much simplicity, White Paper, 2007.

[18] Programmable Web site. http://www.programmableweb.com.

[19] myexperiment Home Page. http://www.myexperiment.org/.

[20] D. De Roure, C. Goble, J. Bhagat, D. Cruickshank, A. Goderis, D. Michaelides, D. Newman, myExperiment: defining the social virtual research environment, in: 4th IEEE International Conference on e-Science, Indianapolis, Indiana, USA, 7–12 December 2008, p. 86. doi:10.1109/eScience.

[21] D. De Roure, C. Goble, R. Stevens, The design and realisation of the myexperiment virtual research environment for social sharing of workflows, Future Generation Computer Systems 25 (2009) 561–567. doi:10.1016/j.future.2008.06.010.

[22] Climate-G Testbed. http://grelc.unile.it:8080/ClimateG-DDC.

[23] J. Gallagher, N. Potter, P. West, J. Garcia, P. Fox, OPeNDAPs server4: building a high performance data server for the DAP using existing software, in: AGU Meeting in San Francisco, 2006.

[24] 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, 1998, p. 8392.

[25] G. Aloisio, M. Cafaro, S. Fiore, M. Mirto, The Grid Relational Catalog Project, in: L. Grandinetti (Ed.), Advances in Parallel Computing, in: Grid Computing: The New Frontiers of High Performance Computing, Elsevier, 2005, pp. 129–155.

[26] R. Alfieri, R. Cecchini, V. Ciaschini, L. Dell'Agnello, A. Frohner, A. Gianoli, K. Lorentey, F. Spataro, VOMS, An Authorization System for Virtual Organizations, in: European Across Grids Conference, 2003, pp. 33–40.

[27] G. Aloisio, M. Cafaro, S. Fiore, M. Mirto, S. Vadacca, GRelC data gather service: a step towards P2P production grids, in: Proceedings of ACM SAC 2007, Seoul, Korea, March 11–15, vol. 1, 2007, pp. 561–565.

[28] S. Fiore, A. Negro, G. Aloisio, The data access layer in the GRelC system architecture, Future Generation Computer Systems 27 (3) (2011) 334–340. http://dx.doi.org/10.1016/j.future.2010.07.006.

[29] G. von Laszewski, J. Gawor, S. Krishnan, K. Jackson, Grid computing: making the global infrastructure a reality, in: Commodity Grid Kits—Middleware for Building Grid Computing Environments, in: Communications Networking and Distributed Systems, Wiley, 2003, pp. 639–656, Chapter.

[30] S. Fiore, M. Cafaro, S. Vadacca, A. Negro, E. Verdesca, M. Mirto, G. Aloisio, Asynchronous query mechanisms within the GRelC data access service, in: Proceedings of IASTED PDCN 2008, Innsbruck, Austria, February 12–14, 2008, pp. 49–54.

[31] J. Blower, Keith Haines, Adit Santokhee, Chunlei Liu, Godiva2: interactive visualization of environmental data on the web, Philosophical Transactions of the Royal Society A 367 (2009) 1035–1039. doi:10.1098/rsta.2008.0180.
[32] S. Fiore, A. Negro, S. Vadacca, E. Verdesca, A. Leone, G. Aloisio, The GRelC portal: a seamless and ubiquitous way to manage grid databases, in: Proceedings of PDCAT 2008, Dunedin, New Zealand, December 01–04, 2008, pp. 413–418.
[33] M. Schumacher, E. Fernandez-Buglioni, D. Hybertson, F. Buschmann, P. Sommerlad, Security Patterns: Integrating Security and Systems Engineering, Wiley, ISBN: 978-0-470-85884-4, 2005.
[34] D. Krefting, et al., MediGRID: towards a user friendly secured grid infrastructure, Future Generation Computer Systems 25 (3) (2009) 326–333.

Alessandro Negro received a summa cum laude bachelor degree in Computer Engineering from the University of Lecce, Italy in February 2004 and a summa cum laude master degree in Computer Engineering from the same University in April 2006. His research interests include GUI Development, Data Management, Distributed, Grid Computing and Web Services. He is also interested in web and pattern-oriented design. Since 2003 he is a Team Member of the GRelC Project. From July 2008 to November 2009, he held a contract position at the Euro-Mediterranean Center for Climate Change.

Sandro Fiore received the Ph.D. degree in Computer Engineering from the Istituto Superiore Universitario di Formazione Interdisciplinare (ISUFI) at the University of Lecce in 2004. His research activities focus on High Performance Computing, Distributed and Grid Computing, in particular on distributed data management. Since 2001, he is the Principal Investigator of the Grid Relational Catalog (GRelC) project (www.grelc.unile.it). In 2006, he joined the Scientific Computing and Operations Division at the Euro-Mediterranean Center for Climate Change (CMCC). He was involved in the following EU Projects: EGEE-I, EGEE-II and EGEE-III and it is now actively involved in the EU FP7 IS-ENES and EGI-Inspire projects. He is author and co-author of more than 50 papers in refereed books/journals/proceedings on parallel and grid computing and holds a patent on advanced data management. He is a member of the ACM.

Giovanni Aloisio is Full Professor of Information Processing Systems at the Engineering Faculty of the University of Salento, Lecce, Italy. His research interests are in the area of High Performance Computing, Grid and Cloud Computing and are carried out at the Department of Innovation Engineering of the University of Salento. He has been a co-founder of the European Grid Forum (Egrid) which then merged into the Global Grid Forum (GGF), now Open Grid Forum (OGF). He has founded SPACI (Southern Partnership for Advanced Computational Infrastructures), a consortium on ICT and grid computing among the University of Salento, the University of Calabria and HP Italia. He is the Director of the "Scientific Computing and Operations" Division of the Euro-Mediterranean Center for Climate Change "(CMCC)". He is the author of more than 100 papers in refereed journals on parallel and grid computing.