

The EarthServer project and its Lighthouse Applications: Exploiting Science Gateways and Mobile Clients for Big Earth Data Analysis

R. Bruno¹

INFN Division of Catania and Dpt of Physics and Astronomy of the University of Catania

Via S. Sofia, 64, 95123 Catania, Italy,

E-mail: roberto.barbera@ct.infn.it

A. Calanducci

INFN Division of Catania

Via S. Sofia, 64, 95123 Catania, Italy,

E-mail: antonio.calanducci@ct.infn.it

A. Messina

Software Engineering Italia s.r.l.

Via Ingegnere 91, 95125 Catania, Italy

E-mail: antonio.messina@softwareengineering.it

M. Pappalardo

Software Engineering Italia s.r.l.

Via Ingegnere 91, 95125 Catania, Italy

E-mail: marco.pappalardo@softwareengineering.it

G. Passaro

INFN Division of Catania

Via S. Sofia, 64, 95123 Catania, Italy,

E-mail: gianluca.passaro@ct.infn.it

The EarthServer project (www.earthserver.eu), funded by the European Commission under its Seventh Framework Program, aims at establishing open access and ad-hoc analytics on extreme-size Earth Science data, based on and extending leading-edge Array Database technology.

The project addresses the "Big Earth Data Analytics", basing its work on the open standards of the Open Geospatial Consortium: Web Coverage Processing Service (OGC WCPS) and the W3C XQuery. EarthServer combines both, thereby achieving a tight data/metadata integration. Further, the rasdaman Array Database System (www.rasdaman.com) is extended with further space-time coverage data types. On server side, highly effective optimizations - such as parallel and distributed query processing - ensure scalability to Exabyte volumes.

Six Lighthouse Applications are being established in EarthServer, each of which poses distinct challenges on Earth Data Analytics: Cryospheric Science, Airborne Science, Atmospheric

1

Speaker

Science, Geology, Oceanography, and Planetary Science. Altogether, they cover all Earth Science domains; the Planetary Science use case has been added to challenge concepts and standards in non-standard environments.

In this contribution we will report on the first applications integrated in the EarthServer' Science Gateway framework being developed by the INFN Dept. of Catania, integrated with a mobile client as well. We will also show how federated identity services can allow Big Earth Data Providers to expose their data in a distributed environment keeping a strict and fine-grained control on user authentication and authorisation.

Opportunities for the Asia-Pacific Earth Science research communities to profit from the project or contribute data and services to it will be discussed as well.

The International Symposium on Grids and Clouds (ISGC) 2013
March 17-22, 2013
Academia Sinica, Taipei, Taiwan

1. Introduction

The EarthServer is a co-financed project by the 7th Framework programme of the European commission. This project aims at open access and ad-hoc analytics on Earth Science (ES) data, based on the OGC geo service standards Web Coverage Service (WCS) and Web Coverage Processing Service (WCPS). The WCS model defines "coverages" as unifying paradigm for multi-dimensional raster data, point clouds, meshes, etc., thereby addressing most of Earth Science data. WCPS as aka "XQuery for raster data" allows declarative, SQL-style queries on coverages.

Integration of WCPS with XQuery allows mixed data/metadata queries. The unified service supports: navigation, extraction, aggregation, and ad-hoc analysis on massive n-D ES data through queries of open-ended complexity, achieving flexibility on coverage data, as it is known from SQL. Clients are ranging from mobile devices over Web tools to high-end immersive virtual reality.

A key component of EarthServer Implementation is the existing rasdaman server technology [16] developed by the Jacob University [24], which is one of the 10 EarthServer project partners.

The WCS and WCPS standards are tested on several partners super-scale archives of data centres serving: atmospheric, oceanography, geology, and general earth observation communities, for this reason six lighthouse applications have been identified each covering a particular Earth Science domain.

This paper will show how the Catania Science Gateway Framework has been integrated to serve some of the project lighthouse applications. It will be also shown mobile applications accessing the ES data.

1.1 Lighthouse applications

The six project lighthouse applications developed in the project are providing widely visible, convincing showcases. These applications are being established in EarthServer, each of which poses distinct challenges on Earth Data Analytics: Cryospheric Science Airborne Science, Atmospheric Science, Geology, Oceanography, and Planetary Science. Altogether, they cover all Earth Science domains; the Planetary Science use case has been added to challenge concepts and standards in non-standard environments.

All the lighthouse applications are based on the OGC' WCPS standard and the RASDAMAN [1] raster analytics server. Installations have been established at all project partner sites, prepared for sustained operation beyond the project. Alongside with the service as such, software, best practices, and trainings are being developed as elements of the project's outreach activities.

A brief description of all six lighthouse applications will be given:

- **Cryospheric Science:** The Cryospheric Data Service is developed by the EOX [17] partner and will, in the final stage, support the snow and land ice community by providing an online data archive and processing facility for relevant products. Currently, the service holds a selection of Snow Cover Products created by CryoLand, a Collaborative EU-FP7 Project (2011-2015) [18].
- **Airborne:** Airborne data with High Altitude Long Endurance (HALE) Uninhabited Aerial Systems (NASA) [22]. The Airborne Science Program within the

Earth Science Division is responsible for providing aircraft systems that further science and advance the use of satellite data.

□ **Atmospheric Science:** The Climate Data Service is developed by MEE0 [19] with the aim of supporting the Climate and Atmosphere communities in loading, visualizing and analysing 3D and 4D datasets at European and global scales. Another application related to the atmospheric science is the MERIS data repository coming from the ESA' MERIS spectrometer on board of the ENVISAT satellite.

□ **Geology:** The Geology Data Service is developed by the British Geological Survey (BGS) [20] to support the solid earth science community. The service currently delivers a range of coverage data sets and illustrates the use of WCPS queries and a range of visualisations

□ **Oceanography:** The Ocean Data Service is developed by Plymouth Marine Laboratory [21] and is intended to provide an exemplar portal, demonstrate large scale EO data services for the Ocean Science community and introduce new capabilities from the use of WCPS queries.

□ **Planetary Science:** The Planetary Data Service offers orbital spacecraft data over Mars. PlanetServer [23] aims at providing access and analysis capabilities to Mars Remote Sensing data. Ingested datasets include panchromatic, hyperspectral and topography data, at multiple spatial scales.

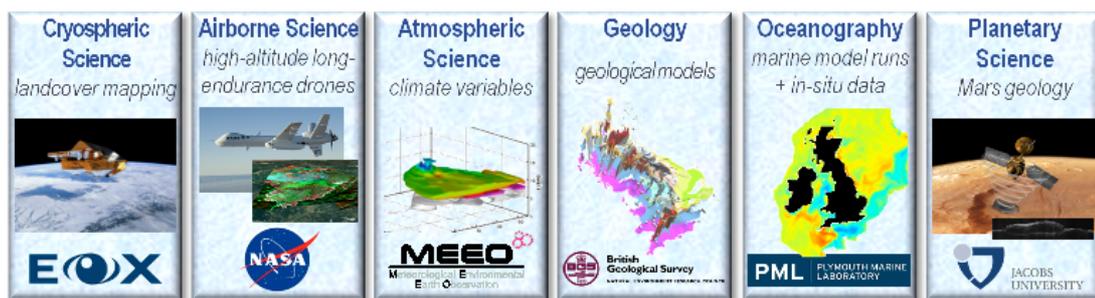


Figure 1-1: The six lighthouse applications banner

Some of the services above are already integrated in the EarthServer' Catania Science Gateway Framework.

2. Catania Science Gateway Framework

By definition a Science Gateway is: A community-developed set of tools, applications, and data that is integrated via a portal or a suite of applications, usually in a Graphical User Interface, that is further customized to meet the needs of a specific community [5]. The Catania Science Gateway Framework (CSGF) [4] complies with this definition offering a customisable environment, which tailors the need of many different user communities.

The CSGF has been designed to address large user communities, access different kind of distributed infrastructures and track any user activity in compliance with the EGI Portal traceability policy [6]. During the design phase of the CSGF architecture a particular attention has been devoted to building its components making use of standards and allowing a simplified access to the targeted distributed infrastructures. The adoption of widely known standards has many advantages, ranging from the software maintenance up to the personnel training; thus the

use of standards has been identified as a key element to reach the technological sustainability of the whole framework, with the idea that a reliable technological sustainability could also help to identify strategic routes towards the economical sustainability.

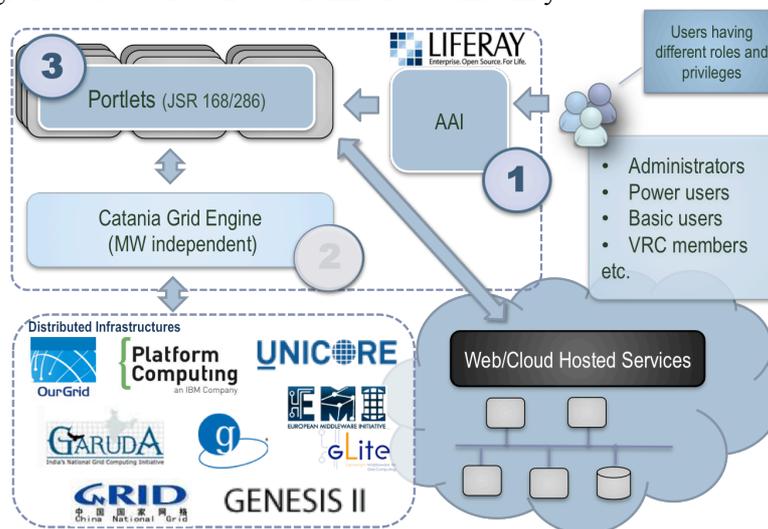


Figure 2-1 The Catania Science Gateway Components

To understand the standards adopted in the CSGF, starting from the top right side of the figure 2.1 the AAI module relying on SAML standard through its Shibboleth implementation allows users to be authenticated and guarantees SSO across trusted federated identities. The AAI module provides authorization registering user capabilities into an LDAP server interconnected with the Liferay portal. The portal itself adopts the Portlet standard JSR 286 and interacts through the GridEngine APIs to the underlying layer of distributed infrastructures exploiting the SAGA standard to access them. In this chapter the reader may appreciate how all those different standard components are strictly interconnected each other to build up a complex but stable and sustainable model that interacts with many different distributed environments such as Grid infrastructures and Clouds as well. On the other side a large and heterogeneous user community ranging from generic/anonymous users, up to site administrators and application developers can widely and securely access it via SSO across many Identity Federations (Idfs).

The adoption of standards in creating customized APIs and new software services offers many different advantages. A software solution made of standard components will be reliable and can be maintained during the service lifetime with minimal efforts. Due to the standard adoption it is also easy to find the necessary manpower that can maintain the services alive during the project lifetime and even beyond; for this reason this document dedicates an entire chapter that describes and locates all adopted standards.

2.1 Core components

The CTSGF consists of three core component that have been numbered in figure 2-1.

1. The AAI module manages its user Authentication and Authorisation with two separate modules each using different background technologies. CSGF authenticates users relying on Identity Providers (IdPs) that are members of one or more Identity Federations. The Catania Science Gateway Framework currently supports federations

based on the SAML 2.0 [10] and SimpleSAMLphp [11] to address the most important social network user identities.

2. The GridEngine is a generic software module able to interconnect the Scientific Gateway presentation layer with the underlying distributed infrastructures (Grid and Clouds) using standard technologies. It allows the quick creation of new Science Gateways providing their developers with a simple interface and avoiding worry about middleware specificities. This is possible thanks to the adoption of the SAGA [12] standard. SAGA stands for Simple API for Grid Applications. SAGA is a family of related standards specified by the Open Grid Forum (OGF) to define an application-programming interface (API) for common distributed computing functionality.

Web Applications: Is the remaining module of Catania Science Gateway Framework as shown in Figure 2-1 and consists of the user interface made of a set of web applications. For the development of these basic elements of the Science Gateway, the JSR 286 [7] standard (also known as "portlet 2.0") was adopted. These applications, normally in the form of portlets are used and they can be exchanged among different portals or many of them arranged together to build up complex and sophisticated scientific environments. As portlet container, the award winning Liferay [15] portal framework has been chosen which offers a rich, easy-to-use "web 2.0" interface using AJAX and other presentation layer technologies. It features effortless GUI-based personalization, drag-and-drop portlets, dynamic navigation, and an instant-add portlet library. The portal platform also integrates with most used packages such as YUI3 and jQuery and with the JavaScript library of the portal developers. Liferay is currently the most used framework to build Science Gateways in the Grid world.

2.2 gLibrary

gLibrary is the platform developed by INFN [8] to easily create, organise and access digital repositories on gLite-based Grid infrastructures, from Web, Mobile and Science Gateways. Using gLibrary users can navigate, search, download files stored on a gLite Grid exploiting a rich set of metadata customised for the specific application context (named repository). This service has been successfully integrated with the CSGF thanks to a rich set of REST API calls. The access to gLibrary metadata content will be securely protected by both the GSI for the physical storage of the data and the portal AAI module which handles the fine grained user access rights.

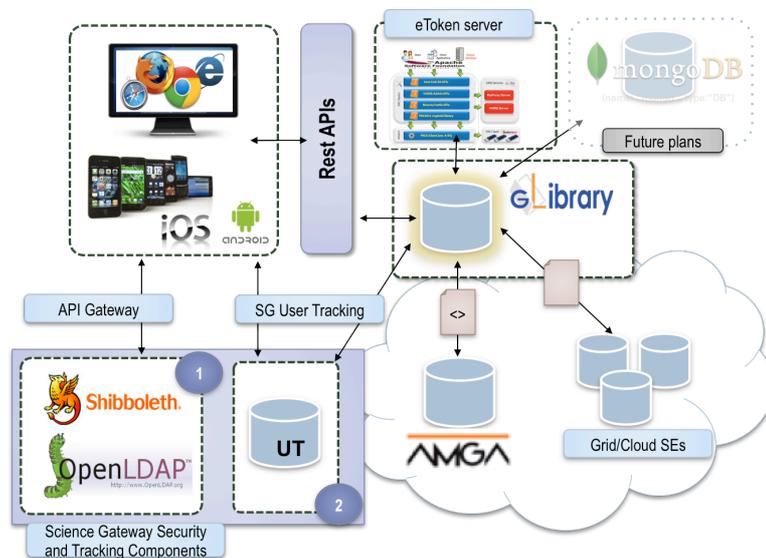


Figure 2-2 The gLibrary architecture

The gLibrary platform offers tools both to end-users and to service/component developers.

End-users could make use of the following:

1. gLibrary browser portlet
2. gLibrary annotator portlet
3. gLibrary browser/uploader Web App (Rich Internet Application)
4. gLibrary mobile browser for iOS (iPhone and iPad)

Service/Component Developers:

1. gLibrary Services RESTful API
2. OAI-PMH interface

Following is a description of the mentioned gLibrary components, one by one, trying to synthesize their functionality and possible audience:

1. the gLibrary browser portlet is used to browse repositories deployed in the EarthServer Science Gateway using pre-defined (by repository providers) filters, inspecting the full set of metadata per each entry, download the digital resource associated (if it will be stored on grid's storage) choosing from a set of replicas.

2. the gLibrary annotator/uploader portlet is used to add new entries to a repository and upload/replicate local digital resources to the grid, using the metadata schema set that has been defined (by repository administrators) for a given repository and its type.

3. the gLibrary annotator/uploader portlet is used to add new entries to a repository and upload/replicate local digital resources to the grid, using the metadata schema set that has been defined (by repository administrators) for a given repository and its type. Currently we have a simplified version of the annotator portlet, that allows only to annotate metadata using the Simple Dublin Core metadata set. Possible users are Data Curators/Annotators/Providers.

4. the gLibrary browser/uploader Web App, is a Rich Internet Application, installed on top of the gLibrary Server, that allows to navigate on all the repositories that are handled by a single instance of gLibrary server. It offer a browsing interface with filtering capabilities, metadata inspection/editing, replica download, upload of digital resources with metadata editing/annotation. While 1 & 2 require users to have an account to a Federation's Identity Provider, this web app, allow authentication via X.509 digital certificate, plain username/password, guest access.

5. This native iOS app, working both for iPhone and iPad, permits to access gLibrary repositories on the go, with filtered browsing, metadata inspection, replica download on the local storage for offline access. The iPad version allows to create graphical annotations on digital images, highlighting with polygons, and adding named comments. Comments and highlights could be shared among other users interested (subscribed to) in a given digital resources, using a subscription and real-time pushing notification mechanisms (using Apple Push Notification service).

6. Service/Component Developers could access gLibrary services through a set of HTTP(s) REST APIs that expose high level objects, like repositories, entries, filters, types, collections, hiding all the grid-related intricacies behind the scene. Currently we offer a set of APIs for querying repositories, retrieve the list of repository's entries, repository's filter and type's/collection's trees. We have also built a first draft API to upload and replicate local digital resources to Grid Storage Elements, and to add metadata to them. Administration (used by repository manager to create new repositories) APIs are still missing. All those APIs are authenticated via Shibboleth/SAML token or X.509 certificates.

The OAI-PMH interface can be enabled for any of the gLibrary hosted repository that need public access (no authentication is required). At the moment only the following verbs have been implemented: Identify, ListMetadataFormats, ListRecords.

2.3 Remote service access

The portal allows two different kind of user access to the remote services. A public mode, which normally provides a limited set of features, and a restricted mode normally available only to the portal' registered and authorised users. The figure below shows the architecture adopted to offer remote service accessibility from the Science Gateway.

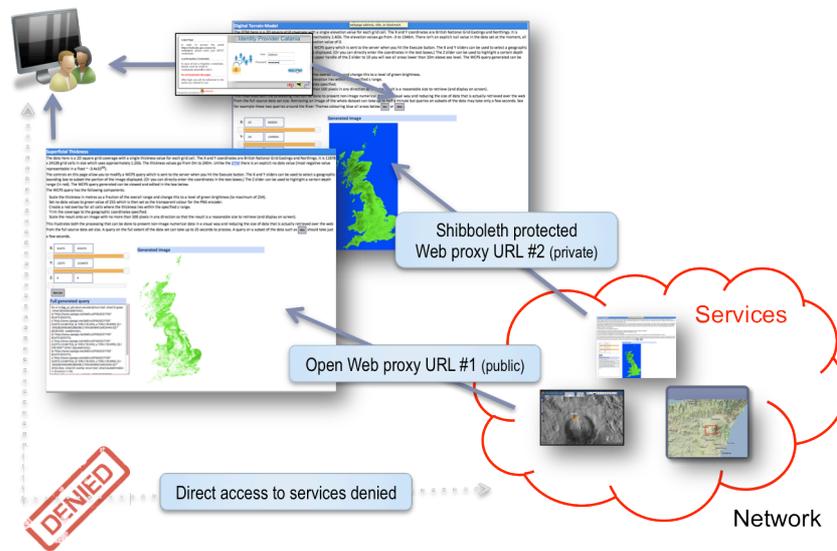


Figure 2-3 Remote service hosting architecture

Any user can only access the remote service only through the Science Gateway, so that any direct access will be denied by proper network/firewall configurations. Inside the Science Gateway it is possible to configure different URLs identifying those requiring user authentication and authorization through respectively Shibboleth and LDAP portal services. With this mechanism it is possible to define public accessible URLs normally used for demonstrative or anonymous access and private URLs accessible only by portal authorised users; normally used to grant the access only to a restricted group of people.

This model well suites with services hosted by Cloud node instances; for this reason the Catania Science Gateway Framework already provides a special portlet named myCloud which allows to manage Virtual Machines hosted by Cloud infrastructures through simple mouse clicks and drag-drop operations. This interface provides a more user-friendly approach to operations normally accomplished only via the use of the command line.

3. Integrated Light House Applications

In this chapter will be examined in detail the Catania Science Gateway Framework implementation built for the EarthServer project. As already described in the introduction, the main capabilities exploited by the EarthServer project is the embedding of external network resources hosted by the project partners and being part of the six lighthouse applications.

The next chapters will show a detailed view of each case.

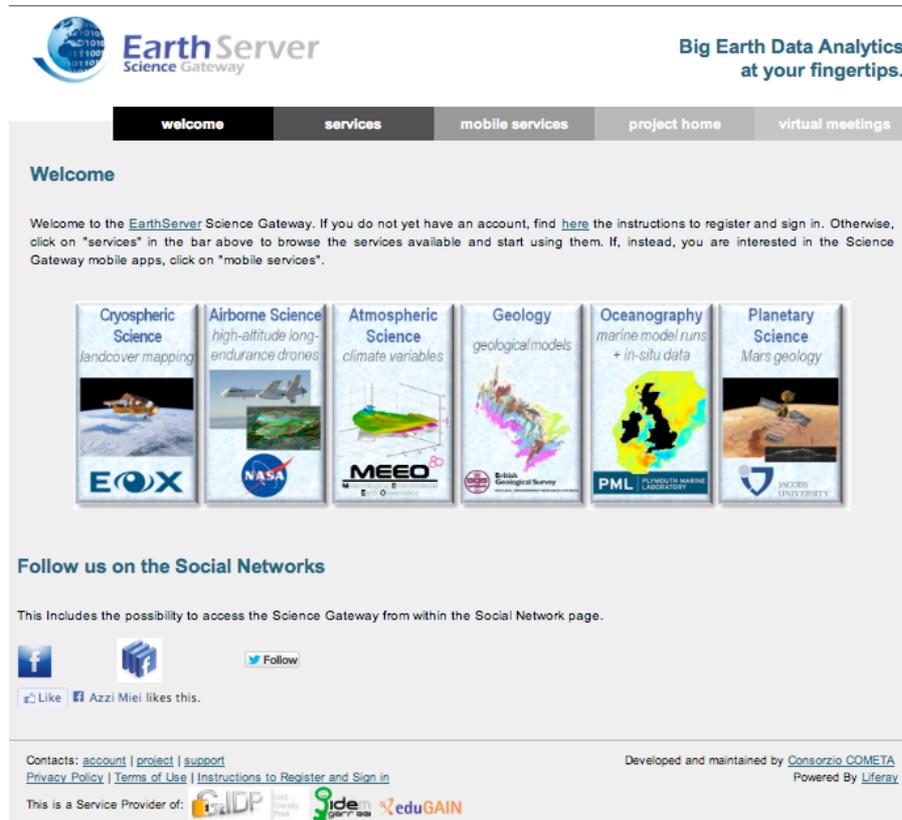


Figure 3-1 EarthServer science gateway home page

The overall portal look and feel recalls the main project web site and the only change consists of the different menu voices briefly described below:

- Welcome**, points to the EarthServer science gateway home page
- Services**, points to the list of all available lighthouse applications. This menu also allows the user to access a specific lighthouse application just clicking into the specific application name.
 - Mobile Services**, The EarthServer Science Gateway also offers a set of demonstrative mobile applications. By this page the user can access to the specific application market place and download the required mobile application.
- Project home**; redirects to the project main home page.
- Virtual Meetings**; this voice points to the project conferencing system.

3.1 Services

The EarthServer project is currently supporting some of the six available lighthouse applications which have been successfully integrated with the CSGF; in particular has been explored successfully the possibility to access third parties services from a single entry point providing protected and unprotected access with a fine grained user access rights. The next chapters will describe in detail each of these applications.

3.1.1 Planet Server

The Planetary Service, PlanetServer is dealing with orbital spacecraft data over Mars. PlanetServer aims at providing access and analysis capabilities to Mars Remote Sensing data. Ingested datasets include panchromatic, hyperspectral and topography data, at multiple spatial scales. This application was the first integrated service aiming to demonstrate the Catania Science Gateway capabilities among the project partners. The interface provides two buttons providing a restricted and a public/anonymous access.

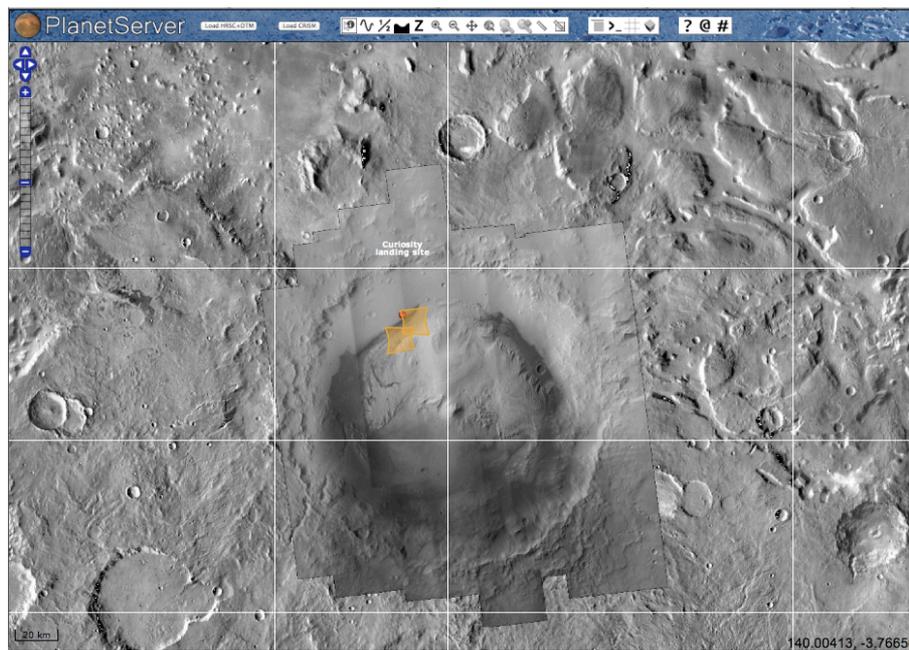


Figure 3-2 PlanetServer application

The Mars elevation data shown by this service have been elaborated by the COMETA Grid infrastructure [9] applying a particular algorithm to a huge set of satellite images. The image computation involved AMGA and LFC file catalog services while a dynamically generated set of parametric jobs were executed against the COMETA CPU cores that have been shared to the EarthServer project.

3.1.2 Meris Data

This application manages a repository of images exploiting the gLibrary Grid Service described above to associate metadata to existing image files. The application manages images coming from the MERIS spectrometer on board of the ESA' ENVISAT satellite. Each satellite image is stored into a Grid infrastructure and each file is enriched by a well-defined set of metadata fields. The Meris Data repository application allows the users to retrieve a satellite image querying on metadata contents. To access this application the user must be a registered member of the portal and granted to access this application.

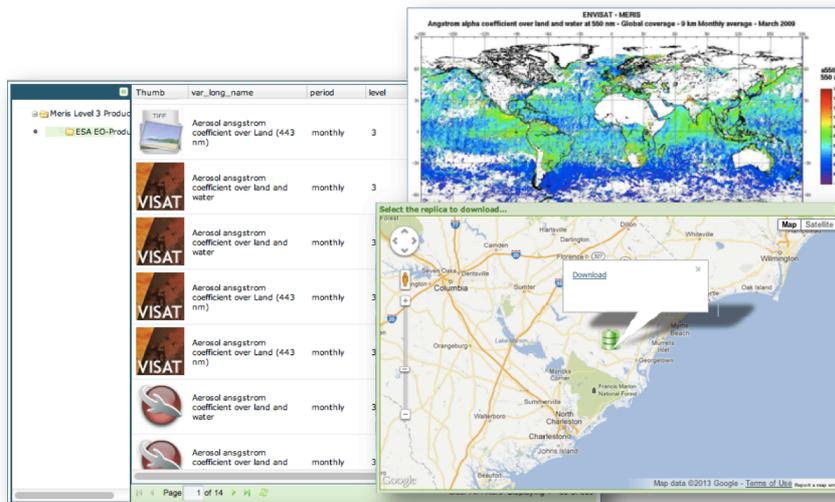


Figure 3-3 Meris data repository

At the beginning the interface provides a list of all available image files, then it is possible to refine the file list, properly filtering on the defined metadata content. Selecting a satellite image in the list first and then the application identifies the distributed infrastructure' storage elements containing the selected file. At the end the user can select one of these storage elements and then download the selected image file.

3.2 BGS Superficial Thickness and DTM clients

The BGS Geology lighthouse is experimenting with various kinds of data used in the British Geology Survey to demonstrate the potential of the EarthServer technologies for supporting users of these data. Currently there are examples of Superficial Thickness and DTM clients, which respectively provide a secured access to WCPS queries, and anonymous access to WCS queries.

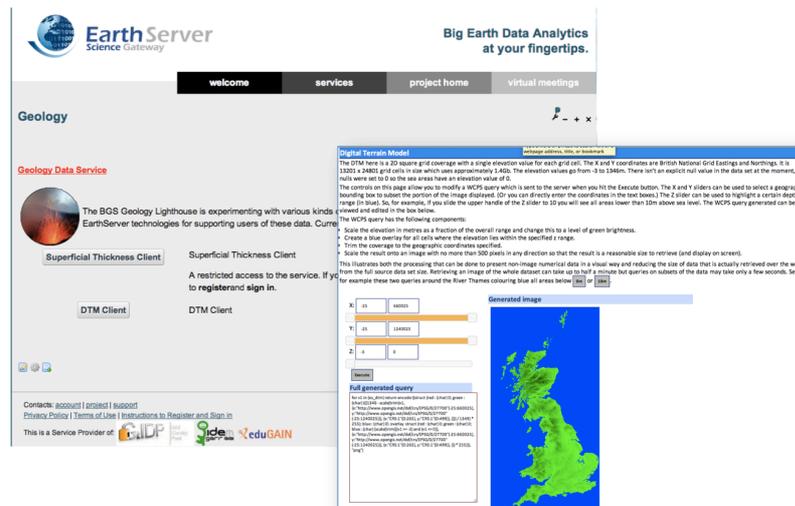


Figure 3-4 BGS Superficial Thickness and DTM clients

3.2.1 Mobile applications

The EarthServer project provides several mobile applications aiming to discover the features offered by the OSGC standards WCS and WCPS. Anyhow each mobile application makes use of the federated identities to grant access to the mobile applications.

IdF and IdP selection
Access to the resource



Figure 3-5 Mobile application IdF/IdP access

The authentication and authorisation procedure handled by the mobile application is the same used by the Catania Science Gateway Framework, which uses Shibboleth for the authentication, and LDAP for the authorisation.



Figure 3-6 EarthServer Meris data mobile application

Two demonstrative mobile applications have been developed, the first able to discover WCPS standard on top of coverages while the second provides the same capabilities offered by the Science Gateway application for MERIS data.

4. Conclusions

The EarthServer services and tools explained by the previous chapters offer to the Earth-Science user community a very powerful environment to easily, securely and seamlessly access the new features provided by OSGC' WCS and WCPS standards. The inclusion of the Catania Science Gateway paradigm to the project have an important role for the dissemination and achievement of the very interesting features provided by the OGC' WCS and WCPS standards in the field of Earth Science. The EarthServer Science Gateway portal is one of the main entry points for the project user community since it is able to hide the complexities behind the exploitation of distributed infrastructures such as Grids, Clouds, and any other kind of distributed data storage. The adoption of standard components in developing the Catania Science Gateway Framework is a key element to reach the technological sustainability of this new technology.

The EarthServer project represents an important opportunity to join the Earth-Science community to the world of distributed environments. Any user interested to access the project services and the related data, can join anytime to the EarthServer science gateway registering first into one of the supported Identity Federations such as the GrIdp (catch-all); then once registered, any user has to apply for the right credentials inside the project Science Gateway during the first log-in.

EarthServer project is also willing to exchange experiences, resources and services with other Earth-Science communities offering the possibility to sign a memorandum of understanding (MoU) as for example recently signed with another European project, ChainReds [25][26]. The EarthServer project is then offering a big opportunity to join the European earth science experiences with other research centres located in: Asia, Middle-East and any other region of the world.

References

- [1] P. Baumann: Array Databases and Raster Data Management. In: T. Özsu, L. Liu (eds.): Encyclopedia of Database Systems, Springer, 2009
- [2] RASDAMAN, <http://www.rasdaman.org>
- [3] Baumann, P.; Jucovski, C.; Stancu-Mara, S.: Efficient Map Portrayal Using a General-Purpose Query Language (A Case Study). DEXA 2009, August 31 - September 04, 2009, Vienna, Austria, Springer Berlin/Heidelberg, LNCS 5690, pp. 153-163
- [4] Valeria Ardizzone, Roberto Barbera, Antonio Calanducci, Marco Fargetta, E. Ingrà, Ivan Porro, Giuseppe La Rocca, Salvatore Monforte, R. Ricceri, R. Rotondo, Diego Scardaci, Andrea Schenone: “The DECIDE Science Gateway”. Journal of Grid Computing Vol. 10(4), pages 689-707, Editor Springer 2012.
- [5] Teragrid/Xede, <https://www.xsede.org>
- [6] EGI Portal traceability policy - <https://documents.egi.eu/public/ShowDocument?docid=80>
- [7] gLibrary: gLibrary/DRI:
- [8] Roberto B., Antonio.C. et al: A Grid-Based Platform to Host Multiple Repositories for Digital Content
- [9] Carmelo Marcello Iacono-Mann, Marco Fargetta, Roberto Barbera, Alberto Falzone, Giuseppe Andronico, Salvatore Monforte, Annamaria Muoio, Riccardo Bruno, Pietro Di Primo, Salvatore Orlando, Emanuele Leggio, Alessandro Lombardo, Gianluca Passaro, Gianmarco De Francisci-Morales and Simona Blandino: The Sicilian Grid Infrastructure for High Performance Computing
- [10] shibboleth.internet2.edu
- [11] simplesamlphp.org
- [12] saga-project.github.io
- [13] grid.in2p3.fr/jsaga
- [14] www.globus.org/security/overview.html
- [15] www.liferay.com
- [16] www.rasdaman.com
- [17] eox.at
- [18] www.cryoland.eu
- [19] www.meeo.it
- [20] www.bgs.ac.uk
- [21] www.pml.ac.uk

[22] airbornescience.nasa.gov

[23] planetserver.jacobs-university.de

[24] jacobs-university.de

[25] <http://documents.ct.infn.it/record/550/files/EARTHSERVER-CHAIN-REDS-MoU-Signed.pdf?version=1>

[26] www.chain-project.eu