

The GISELA Science Gateway

Valeria Ardizzone^a, Riccardo Bruno^b, Antonio Calanducci^{b,c},
Marco Fargetta^b, Elisa Ingra^a, Giuseppe La Rocca^c, Salvatore Monforte^c,
Fabrizio Pistagna^c, Rita Ricceri^c, Riccardo Rotondo^{a,c}, Diego Scardaci^c,
and Roberto Barbera^{b,c,d,1}

^a Consortium GARR, Via dei Tizii, 6, 00185 Rome – Italy
{valeria.ardizzone, elisa.ingra, riccardo.rotondo}@garr.it

^b Consorzio COMETA, via S. Sofia, 64, 95123 Catania – Italy
{riccardo.bruno, marco.fargetta, antonio.calanducci}@ct.infn.it

^c National Institute of Nuclear Physics, Division of Catania, via S. Sofia, 64,
95123 Catania – Italy
{giuseppe.larocca, salvatore.monforte, fabrizio.pistagna, rita.ricceri,
diego.scardaci}@ct.infn.it

^d Department of Physics and Astronomy of the University of Catania, Viale A. Doria, 6,
95125 Catania – Italy
roberto.barbera@ct.infn.it

Abstract. In this paper we present the architecture of a framework for building Science Gateways supporting official standards both for user authentication and authorization and for middleware-independent job and data management. The use case of the GISELA Science Gateway is also described.

Keywords: Science Gateway, Grid Computing, GISELA, Identity Federation.

1 Introduction

Grid infrastructures are being built in several areas of the world but, despite the huge investments made by the European Commission and by other funding agencies, both at national and international level, the total number of users is in the order of magnitude $O(10^4)$, much less than $O(10^7)$ which is the order of magnitude of the number of users of the international research and education networks (e. g., GÉANT in Europe). The reasons for this have been investigated through studies promoted by the European Commission itself [1] and mostly reside in the complexity for non-IT-expert users of the Grid security, based on a Public Key Infrastructure, in the little adoption of standards to let different middleware be interoperable among each other,

¹ Corresponding author.

and in the lack of general frameworks to easily build customizable high-level user interfaces.

In the recent past, interesting developments have been independently carried out by the Grid community with the Science Gateways and by the National Research and Education Networks with the Identity Federations to ease, from one side, the access and use of Grid infrastructures and, from the other side, to increase the number of users authorised to access network-based services.

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 (US Teragrid project).*”

An Identity Federation is made of “[...] *the agreements, standards, and technologies that make identity and entitlements portable across autonomous domains (Burton Group)*”. Identity Federations have the aim of setting up and supporting a common framework for different organisations to manage accesses to on-line resources. They are already established in many countries and currently gather a number of people which is in the order of $O(10^7)$.

To address the issue of the use of e-Infrastructures, the Italian National Institute of Nuclear Physics and the Consorzio COMETA are developing since almost two years, in the context of several EU funded projects such as DECIDE [2], EUMEDGRID-Support [3], GISELA [4] and INDICATE [5], a new type of Science Gateway that implements an authentication schema based on Identity Federations [6].

In this paper we report on the customization of this new type of Science Gateway for the GISELA project and its applications. The paper is organized as follows. The architecture of the Science Gateway is provided in Section 2 while Section 3 shows the implementation done for GISELA. Summary and conclusions are drawn in Section 4..

2 Architecture of the framework for Science Gateways

The primary requirements that drove the design and implementation of the Science Gateway framework were:

- Use of standards;
- Simplicity;
- Easiness of use;
- Re-usability.

Since the very beginning, the idea was not to build a “vertical” solution but rather to create a framework made of small pieces of software that, as LEGO® bricks, could have been customized and re-arranged in many ways in order to fulfill a large variety of applications and end-users. The success of LEGO® bricks resides in the fact that the “basic element” is simple and standard and it can be easily connected to other basic elements to create huge and very complicated constructions. For the development of the basic elements of the Science Gateway, the JSR 286 [7] standard (also known as “portlet 2.0”) was adopted. In this case, “our” LEGO® bricks are

standard portlets that can be easily arranged to create different, even complex, portals. As portlet container, the award winning Liferay [8] 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.

Figure 1 shows the reference model adopted to build the Science Gateway.

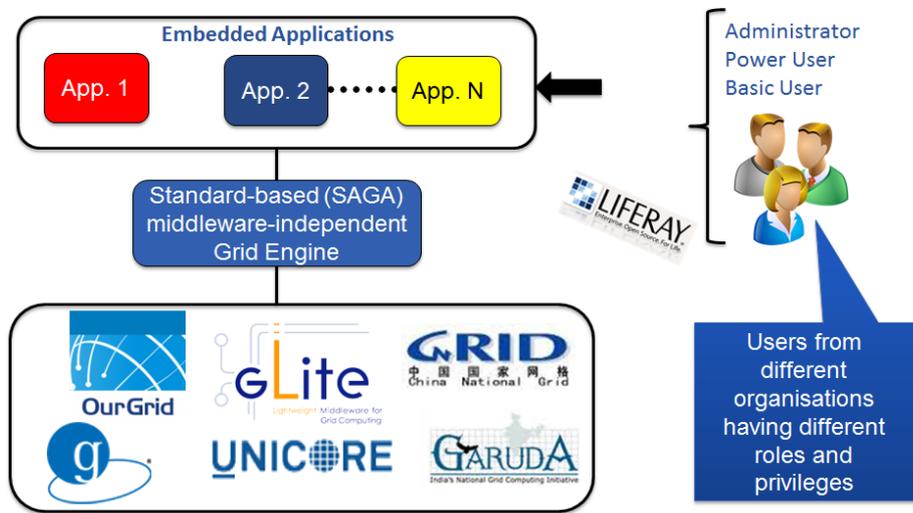


Fig. 1. Reference model of the Science Gateway.

Users belonging to different organizations may have different roles in the community the Science Gateway is developed for and different privileges on the applications and related data available in the gateway. They access the Liferay-based portal and, according to their role and privileges, they are allowed to run some applications embedded in the Science Gateway and exposed through its user interface. Applications are interfaced to the underlying Grid infrastructure through a library of software services which is based on standards and is middleware-independent.

One of the strengths of this Science Gateway is the decoupling of the authentication (AuthN) phase from the authorisation (AuthZ) one. In order to access the Science Gateway, a user must be both authenticated and authorized but we treat the two steps separately and with different technologies.

User authentication relies on Identity Providers (IdPs) that are members of one or more Identity Federations. We only support federations based on the SAML 2.0 [9] standard specifications and on its implementation done by Shibboleth [10] and SimpleSAMLphp [11]. We currently support several official Identity Federations and some of our Science Gateways are already registered as Service Providers of the eduGAIN [12] inter-federation service within the GÉANT project [13]. We also

support all the Identity Providers of the Grid Identity Pool (GrIDP) [14], a “catch-all” Identity Federation that we have expressly created to gather all the IdPs that do not already belong to any official federations and all the users of the Science Gateway who are not (already) registered in any IdPs. This is particularly important and useful in the contexts where it is necessary to authenticate the so-called “citizen scientist” (i.e., people belonging to the general public) and let him/her access the e-Infrastructure for dissemination and self-learning purposes. Inside the GrIDP Federation, we have also created a special IdP, the “Social Networks’ Bridge Identity Provider” [15], that allows people to get authenticated with the same credentials they already have with the most known and populated social networks.

Unlike authentication, user authorization is carried out at the level of the Science Gateway: users whose request to register is approved by the managers of the portal, are stored in a LDAP-based registry together with the roles they have and the privileges they are granted.

Users wanting to register to the Science Gateway have to fill a dedicate web form. Here, they can specify the Identity Federation she belongs or ask to be enrolled as a member of GrIDP. The request of registration, once it is confirmed by the user via email, is then forwarded to the administrators of the portal. If it is accepted, user information is stored on the LDAP registry and the user is notified that she can sign in. Otherwise, she is notified that her request has been denied. This procedure has been put in place in order to ensure that authorisations are not provided automatically to everybody and that a check be done on the requests by a human being.

Once a user has been authorised to access the Science Gateway, she can then sign in and run the applications she is allowed to from within the portal. The workflow of this phase is the following: when the user signs in, she is asked to select in a web page the Identity Federation and the Identity Provider she belongs to and then she is redirected to the login page of her Identity Provider where she can insert her credentials. If they are correctly verified, the control returns to the Science Gateway that checks if the user is inserted in the LDAP registry. If she is, the user is then presented with the web page(s) of the application(s) she has the privilege to run on the Grid infrastructure. At this time, the portal contacts an eToken server (see below) that returns a valid proxy to be used to perform the Grid transaction. Once done, the output of the application is presented back to the user who can thus download it on her computer.

It is worth mentioning that, in order to be compliant with the strict rules of the European Grid Infrastructure VO Portal [16] and Grid Security Traceability and Logging [17] policies, each operation done by the user inside the Science Gateway is stored on a User Tracking DB that can be inspected at any time by the administrator of the portal. This ensures the non-repudiability of Grid transactions which is one of the most important requirements of the Grid Security Infrastructure.

The authentication and authorisation mechanism described above has the big advantage of being based on standards and greatly simplifies the access to e-Infrastructure by non-IT-expert users avoiding the need for them to get and manage personal digital certificates. However, all Grid transactions must be signed with proxies generated by standard X.509 digital certificates so we have implemented in the Science Gateway a mechanism that creates proxies on the fly and on user request. This is done by a service called eToken server [18]. The eToken server generates proxies starting from robot certificates [19]. Robot certificates are special, yet

standard, digital certificates stored in USB Smart Card, referred to as etokens. It is possible to bind robot certificates with applications and allow people to run them without any personal credentials. According to this schema, when a user is authorised to access the Science Gateway and wants to run one of the applications she is allowed to, the portal retrieves on her behalf a valid proxy for the eToken server. The proxy generated on the fly contain the extensions specifying the role and privileges of the robot certificate inside the VO supported by the Science Gateway so different proxies can be created according to the different roles and privileges of the user in the LDAP registry. This ensures a fine grained authorisation and provides the portal manager with the complete control of deciding what a given user can see and do.

The core of the eToken server is a “lightweight” grid crypto library implemented according to the Service Oriented Architecture. The multi-threaded eToken server holds the web services to access the smart cards and interacts both with the Virtual Organisation and the automatic proxy renewal (MyProxy) server. A Java multi-platform client configured for inter-service communication via HTTPS completes the architecture. In order to improve the performances, the server is built on top of the Apache Tomcat Application Server and configured to accept requests only from a set of authorized “clients” (e.g., the Science Gateway). The adoption of the Apache Tomcat as Application Server ensures scalability and high performances especially when the server has to deal with huge numbers of requests. To further improve its performances and reduce the waiting time to get a proxy, the eToken server implements a mechanism for caching the proxies.

A RESTful API has also been implemented to access the eToken server. The currently available functions allow to:

- List all the robot certificates stored on the different smart cards physically connected to the server;
- Generate VO-specific proxies from a given robot certificate;
- Create and upload a long-term proxy on a MyProxy Server, allowing proxies to be automatically and securely renewed for much more than their default 12-hours lifetime.

Once generated, the proxy signed with the robot certificate is used to entrust the Grid transactions, such as job and data management ones.

In order to execute applications from within the Science Gateway in a middleware-independent way, the Simple API for Grid Applications (SAGA) standard specifications [20], defined by the Open Grid Forum [21], and its JSAGA [22] implementation, have been adopted.

As schematically shown in fig. 2, a software layer Grid Engine has been developed which contains a “job engine” and a “data engine” which, in turn, call the JSAGA API for job and data management. The Science GW Interface also contains the functions to interact with the User Tracking DB mentioned above.

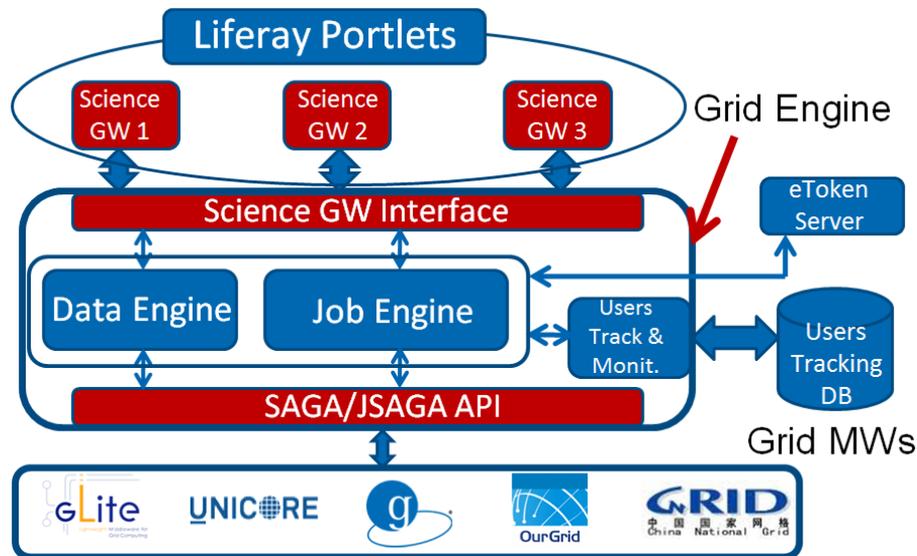


Fig. 2. Architectural overview of the Grid Engine.

As stated in the introductory section, the framework described above is already being used to create Science Gateways in the context of several EU funded projects. Since September 2011, almost 400 people have registered with the various portals with an exponential increase of this number as a function of time. The same trend is observed for the temporal distribution of the number of jobs submitted through the Science Gateways.

3 The implementation done for the GISELA project

GISELA (Grid Initiatives for e-Science virtual communities in Europe and Latin America) is a project co-funded by the European Union under its Seventh Framework Programme. GISELA aims at (i) implementing the Latin American Grid Initiative (LGI) sustainability model rooted on National Grid Initiatives (NGI) or Equivalent Domestic Grid Structures (EDGS), in association with CLARA, the Latin American NRENs and collaborating with the European Grid Initiative (EGI); (ii) Providing Virtual Research Communities (VRCs) with the e-Infrastructure and Application-related Services required to improve the effectiveness of their research.

In order to extend its user base and make Grid access and use easy for non IT-experts, GISELA has developed the VRC-driven Science Gateway [4] based on the architecture exposed in the previous Section. The VRC-driven GISELA Science Gateway is a web 2.0 environment which allows users to fully exploit the e-Infrastructure computing (jobs) and storage (data) services through a normal web browser exploiting the model described in the previous sections. Users can obtain access to the Science Gateway using the username/password pair (provided by their

Identity Provider) and submit applications simply filling web forms.

The GISELA Science Gateway is considered the key tool to reach the sustainability of the e-Infrastructure in Latin America. Indeed, it has the potentiality to attract a big number of Latin American scientific communities to the Grid world. The number of Identity Federations users is indeed very big (several millions) then, thanks to the authentication procedure we implemented, we could give access to the Science Gateway capabilities to a huge number of potential users. At present, all users belonging to the Identity Federation of Brazil (CAFe [23]), Italy (IDEM [24]), Portugal (RCTSaai [25]), Spain (SIR [26]) and to the eduGAIN inter-federation can access the GISELA Science Gateway. Those who do not belong to any identity federation, are enrolled in the GridP “catch-all” federation discussed in Section 2. Moreover, in collaboration with RedCLARA [27] and the Latin America NRENs, we are promoting and encouraging the creation of Identity Federations in other Latin American countries and their eventual participation in eduGAIN.

After this simple authentication mechanism, users are able to run selected high-impact applications in a very easy way. Then, the distance, in term of competences, between the discovery of the Grid services and the ability to submit jobs will be substantially reduced thus opening the Grid world to a huge community of users.

We decided to organize the GISELA Science Gateway as VRC-driven tool: different sections, specific for each supported VRCs (Life Sciences, Earth Sciences, Industry, etc.), are available inside the portal and can be selected by the users according to the roles and privileges they are granted inside the Science Gateway.

Currently, we created 4 VRC-specific Science Gateways and integrated 7 applications. They are all listed below:

1. Cultural Heritage Science Gateway:
 - a. ASTRA [28]: The ASTRA projects aims to reconstruct the sound or timbre of ancient instruments (not existing anymore) using archaeological data as fragments from excavations, written descriptions, pictures, etc.;
2. Industry Sciences Gateway:
 - a. Industry@Grid: The objective of this collaboration is to gridfy JobShop Scheduling and Supply Chain models using techniques based on heuristics and branch and bound methods. The achievement of this specific goal can create an important building block of a new industrial grid community, focusing on SMEs that will be able to use the grid as a leading-edge tool for their internal planning process;
3. Life Sciences Gateway:
 - a. ClustalW [29]: it is a tool for aligning multiple protein or nucleotide sequences. The alignment is achieved via three steps: pairwise alignment, guide-tree generation and progressive alignment;
 - b. GATE [30]: GATE (Simulations of Preclinical and Clinical Scans in Emission Tomography) is an advanced opensource software developed by the international OpenGATE collaboration and dedicated to the numerical simulations in medical imaging. It currently supports simulations of Emission Tomography (Positron Emission Tomography - PET and Single Photon Emission

- Computed Tomography - SPECT), and Computed Tomography (CT);
- c. Phylogenetics (MrBayes [31]): it is a program for the Bayesian estimation of phylogeny. Bayesian inference of phylogeny is based upon a quantity called the posterior probability distribution of trees, which is the probability of a tree conditioned on the observations. The conditioning is accomplished using Bayes's theorem;
4. Mathematics Science Gateway:
 - a. Octave [32]: GNU Octave is a high-level interpreted language, primarily intended for numerical computations. It provides capabilities for the numerical solution of linear and nonlinear problems, and for performing other numerical experiments. The Octave language is quite similar to Matlab so that most programs are easily portable;
 - b. R [33]: R is a free software environment for statistical computing and graphics.

4 Summary and conclusions

A framework for building new-generation Science Gateway with hundreds of millions of potential users has been presented. The authorisation and authentication mechanism, together with the middleware-independent Grid Engine, based on standards, have been described. The implementation done for the GISELA project and its applications has also been shown.

References

1. Final report *eResearch2020 - The Role of e-Infrastructures in the Creation of Global Virtual Research Communities* (www.eresearch2020.eu/eResearch%20Brochure%20EN.pdf).
2. V. Ardiszone, R. Barbera, A. Calanducci, M. Fargetta, G. La Rocca, S. Monforte, F. Pistagna, R. Rotondo, and D. Scardaci, The DECIDE Science Gateway, *Proceedings of IWSG-Life 2011*, <http://ceur-ws.org/Vol-819/>. The home page of the DECIDE project is www.eu-decide.eu while the home page of the DECIDE Science Gateway is applications.eu-decide.eu.
3. The home page of the EUMEDGRID-Support project is www.eumedgrid.eu while the home page of the EUMEDGRID Science Gateway is: applications.eumedgrid.eu/science-gateway.
4. The home page of the GISELA project is www.gisela-grid.eu while the home page of the GISELA Science Gateways is gisela-gw.ct.infn.it/science-gateways.
5. The home page of the INDICATE project is www.indicate-project.eu while the home page of the INDICATE e-Culture Science Gateway is indicate-gw.consortio-cometa.it/.
6. R. Barbera, M. Fargetta, and R. Rotondo, A simplified Access to Grid Resource by Science Gateways, *Proceedings of The International Symposium on Grids and Clouds, Taipei (2011), Proceedings of Science (ISGC 2011 & OGF 31)*, p. 23.

7. download.oracle.com/otndocs/jcp/portlet-2.0-fr-oth-JSpec.
8. www.liferay.com.
9. saml.xml.org.
10. shibboleth.internet2.edu.
11. simplesamlphp.org.
12. www.edugain.org.
13. www.geant.net.
14. gridp.ct.infn.it.
15. idpsocial.ct.infn.it.
16. <https://documents.egi.eu/public/ShowDocument?docid=80>.
17. <https://documents.egi.eu/public/ShowDocument?docid=81>.
18. G. La Rocca, R. Barbera, V. Ciaschini, A. Falzone, S. Monforte, A new “lightweight” Crypto Library for supporting a new Advanced Grid Authentication Process with Smart Cards, *Proceedings of The International Symposium on Grids and Clouds, Taipei (2011), Proceedings of Science (ISGC 2011 & OGF 31)*, p. 29.
19. R. Barbera, G. Andronico, G. Donvito, A. Falzone, J. J. Keijsers, G. La Rocca, L. Milanesi, G. P. Maggi, and S. Vicario, A grid portal with robot certificates for bioinformatics phylogenetic analyses, *Concurrency and Computation: Practice and Experience 23 (2011)*, no. 3, 246–255.
20. www.gridforum.org/documents/GDF.90.pdf.
21. www.gridforum.org.
22. grid.in2p3.fr/jsaga.
23. www.rnp.br/en/services/cafe.html
24. www.idem.garr.it
25. <https://rctsaai.fccn.pt/>
26. www.rediris.es/sir/
27. www.redclara.net
28. www.astraproject.org
29. www.clustal.org/
30. www.opengatecollaboration.org/
31. mrbayes.sourceforge.net/
32. www.gnu.org/software/octave/
33. www.r-project.org/