PROCEEDINGS OF SCIENCE

A Simplified Access to Grid Resources by Science Gateways

Roberto Barbera
Department of Physics and Astronomy of the University of Catania and INFN
Viale A. Doria 6, 95125 Catania, Italy
E-mail: roberto.barbera@ct.infn.it

Marco Fargetta[1]
Consorzio COMETA
Via S. Sofia 64, 95123 Catania, Italy
E-mail: marco.fargetta@ct.infn.it

Riccardo Rotondo
Italian National Institute of Nuclear Physics, Division of Catania
Via S. Sofia 64, 95123 Catania, Italy
E-mail: riccardo.rotondo@ct.infn.it

Science Gateways are playing an important role in e-Science research and their relevance will further increase with the development of more efficient network technologies and bigger Grid infrastructures. Through the highly collaborative environment of a Science Gateway, users spread around the world and belonging to different organisations can easily cooperate to reach common goals and exploit all the resources they need to accomplish their work. A major task of a Science Gateway is the connection of each user to the available services, denying the use of not authorised resources. This mapping between users and resources complies with the role of users inside the Virtual Research Community. Nevertheless, the security mechanisms inside the Science Gateway have to hide the security details of the technologies underneath and allow each institution to keep the control of their users. In this paper we present a new Science Gateway model supporting three levels of security mechanisms and allowing users to access Grid resources based on the credentials provided by the organisations they belong to. The idea is to combine Shibboleth2 System identities with X.509 "robot" certificates. The former enables the federation of organisations having different authentication policies while the latter allows access to Grid resources by individuals not owning any personal certificate. An LDAP server in the back-end allows this combination providing the map between the two different authentications layers. The proposed approach is very general and provides users with a single sign-on mechanism on Science Gateways built on top of computing and storage resources owned by different organisations and middleware.

The International Symposium on Grids and Clouds and the Open Grid Forum
Academia Sinica, Taipei, Taiwan
March 19 - 25, 2011

[1] Speaker

PoS(ISGC 2011 & OGF 31)023

1. Introduction

Nowadays, scientists access computing tools and software to conduct their studies and perform their investigations. Despite the heterogeneity of these tools, some similarities can be identified: for example, researchers access and use them by a computer (or similar devices like tablets, smart phones, etc.) and perform a direct or indirect access to the hardware of specific tools or sensors. The availability of high speed networks around the world has stimulated the adoption of remote tools while the demand for high performance computing moved the analysis and the execution of complex algorithms and large data-sets from a single workstation to cluster of servers located in different computer centres.

Scientists working together in a specific research or project do not need any more to operate side by side in a laboratory but the new tools allow them to interact and exchange data, information and results in a common virtual environment. The union of scientists, data and tools working remotely for a single goal is often referred to as a Virtual Research Community (VRC).

The Science Gateway [9] paradigm has been introduced to overcome the difficulties met by users of a VRC to access the provided tools. Its main goal is to hide the underlying complexity of a system, which gathers together heterogeneous resources, and prevent undesired access to non-authorised users. In more detail, a Science Gateway is defined as "*a framework of tools that allows scientists to run applications with little concern for where the computation actually takes place. This is similar to cloud computing in which applications run as Web services on remote resources in a manner that is not visible to the end user. However, a science gateway is usually more than a collection of applications. Gateways often let users store, manage, catalogue, and share large data collections or rapidly evolving novel applications they cannot find anywhere else. Training and education are also a significant part of some Science Gateways*".

One of the most used frameworks to build Science Gateways is Liferay [8] because of its easy to use interface based on modern web frameworks and the extensible architecture employing the most prominent technologies. An extensible security mechanism has a central role for the development of Science Gateways because it provides a way to integrate different Single Sign-On (SSO) technologies: once a user has accessed the portal, he/she has not to provide additional credentials to access other services or data (if stored on a secure storage) controlled by the same SSO infrastructure. This result is possible in Liferay thanks to the total compatibility with the modern authentication and authorization frameworks currently employed by the majority of organisations in both business and research area. However, only few authentication frameworks are available out of the box. Liferay does not provide any technologies to enable users to access grid services requiring Grid credentials to ensure the non-repudiability of transactions performed in the Grid environment. This problem has been successfully solved keeping in mind one of the main benefits taken from Science Gateway approach: simplicity.

In this paper we present how these technologies have been integrated in order to build a Science Gateway which can be used in various application contexts such as the medical research community, the digital cultural heritage and many others where easy access and different level of authorisations are key requirements. The paper is organised as follows. In Section 2 an overview of a Science Gateway and the frameworks used to build it is presented. Section 3 provides an overview of the reference scenario used for the design and implementation. In Section 4 we will present the authentication method proposed. Section 5 compares the proposed solution with other approaches. Finally, conclusions and some future works are drawn in the last section.

2. Building a Science Gateway

In the last few years, the use of web technologies in building and developing graphic user interfaces has steadily increased and in many contexts they have exceeded the other alternatives. Pages full of dynamic contents and information are commonly used by users to access complex services.

A portal able to offer different functionalities to a large number of users spread across different boundaries and enabling them to access e-Science services takes the name of Science Gateway [9][10]. Usually, a Science Gateway gathers a number of tools requested by a Virtual Research Community for its purpose. In addition, it should facilitate the intercommunication and collaboration among users spread around the world.

The main components deployed to build the Science Gateway, described in this paper, are Liferay and Shibboleth. These provide, respectively, the portal container for all the applications made available and the security infrastructure.

Liferay

The Liferay Portal Framework [8] is an open source tool for those who want to mash-up all their services within an single portal. Since the adoption of the standard Java Specification Request (JSR) 168 and 286 [4], developer can highly and widely make use of the “portlet” technology in order to offer an easy access to the services deployed in the portal, regardless their complexity.

The Liferay CE (Community Edition) comes with more than 60 out-of-the-box tools which allow user interaction, activities tracking, workflow management and documents sharing. Besides the existing tools, it is also possible to easily add new functionalities according to the tasks foreseen for the Science Gateway.

Shibboleth

Nowadays, many organisations have to deploy different web resources for their activities and systems for Single Sign-On (SSO) became important to create an integrated environment for end users. Additionally, when organisations have to co-operate the web resources should be shared so the SSO has to work across different organisation boundaries.

The Shibboleth System® [6] is a tool supporting cross-organisation SSO based on the OASIS Security Assertion Markup Language (SAML) [3]. SAML is a XML dialect developed to communicate users authentication, entitlement, and attribute information. Therefore, the Shibboleth System provides both clients and services for the exchange of SAML messages among web resources.

The main components of a Shibboleth-based infrastructure are the Identity Provider (IdP) and the Service Provider (SP). The IdP is the component entitled to provide users credentials. Each organisation can manage its own IdP so, in case of cross organisations SSO, the IdPs of all organisations can be grouped to create a federation. The SP represents the web resource that users can access. The system is very modular and it is provided with the full source code for an easier customisation and integration with the existing infrastructures. This was very important in order to be integrated with Liferay.

A Lifeary plug-in for Shibboleth system has been developed. Shibboleth provides a component for the Apache Web Server which can read the SAML messages coming from an IdP and move the information into the HTTP headers for the tools behind the server. The developed plug-in is able to retrieve the header information and use them to authenticate Liferay users.

3. The use case scenario

A Science Gateway aims at satisfying the requirements of a specific community.. Therefore, to develop a new model for Science Gateway it is important to have some reference scenarios as use cases.

Currently, the fields of medicine and bioinformatics provide the most interesting use cases because their demand for storage and computing is steadily increasing. Additionally, the scientists in these fields are not computer experts hence the possibility to access remote resources in a transparent way may have a very positive impact.

To better identify the problems of building Science Gateways several projects aiming to build infrastructures for such communities have been evaluated. Among these, particular the DECIDE (Diagnostic Enhancement of Confidence by an International Distributed Environment) project [7] was quite relevant because it aims at providing a dedicated Grid e-Infrastructure to the Neuroscientific and Medical communities, relying on the Pan-European backbone GEANT and the National Research & Education Networks (NRENs). Over this e-Infrastructure, a secure and user-friendly service has to be provided for the computer-aided extraction of diagnostic markers for the Alzheimer Disease and schizophrenia from medical images. Therefore, the goals of DECIDE project fit perfectly with the adoption of a Science Gateway and provide an invaluable amount of information about its reference community

The outcome of the DECIDE analysis is a Science Gateway that provide a service made of many applications which offer access to large distributed reference databases, high computation and storage resources and intensive image processing tools. Figure 1 shows the overall layout of the expected service and infrastructure.

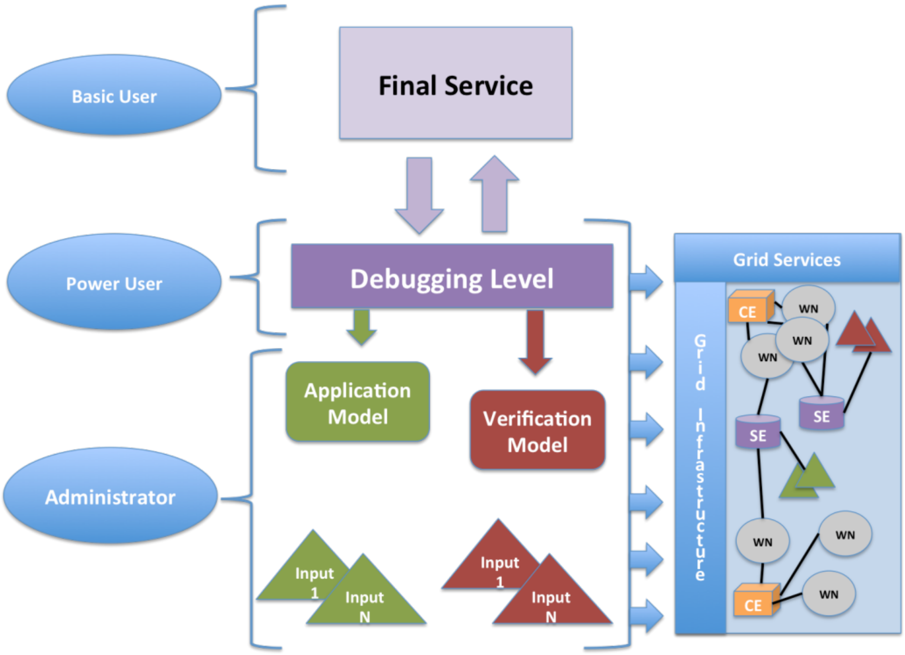

Fig. 1 –Service infrastructure.

Despite the strong orientation towards few communities, with few changes the infrastructure can be easily adapted/extended to other fields since the higher-level functionality (i.e., security, user interface, etc.) should be common to all users.

Nevertheless, every service for the end-user, have to be executed using a Grid middleware that provides: (i) authorisation and secure access to largely distributed databases for reference images, (ii) computationally intensive processing, (iii) image processing on patient images residing locally, compliant with the strict data-sharing policies of hospital.

4. Portal and Grid Authentication

The Science Gateway has to be a web portal providing access to the underneath Grid infrastructure and applications. Figure 2 shows the relations among the different layers representing the portal. A user can access the service portal and this allows in turn to access the applications making the service and these can access or be submitted to the Grid resources for execution.

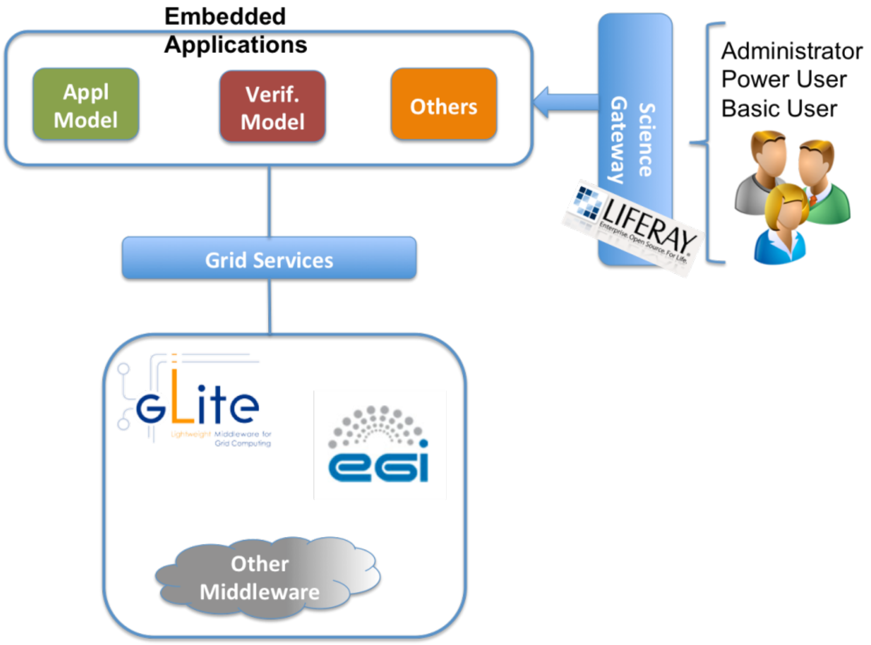

Fig. 2 – Infrastructure layers.

The whole system has to support the security requirements needed to work with medical data, which include a fine-grained control on the user authorisations in order to limit the access to the data according to different policies. Additionally, Grid infrastructures have their own security mechanisms for the authentication and authorisation of users, based on X.509 certificates, that is difficult to integrate in the portal. Therefore, the portal and the Grid resources have to use different approaches to identify users. Moreover, to keep the service easy to use for end-users, the portal has to provide the user credentials to the Grid resources so that the user has not to perform multiple log-ins depending to the activity and the layer to interact with.

In order to create this automatic log-in to lower layers and, at the same time, keep the total control of user activities, several strategies have been adopted. These involved the deployment of different tools and their integration in a common workflow providing a smooth access to the users. To better describe the overall workflow, two steps are identified: the access to the Science Gateway and the use of the Grid resources. Obviously, the first step is mandatory for any activity including the use of Grid resources.

4.1 Accessing the Science Gateway

The Science Gateway developed to support the above scenario is based on the Liferay portal framework described in Section 2. Liferay supports several authentication and authorisation mechanisms off the shelf. Additionally, it provides an interface to extend the security chain of the portal with the integration of new mechanisms. These can manage only some aspects of the authentication and authorisation process and it is possible to mix different tools to validate the user credentials.

To let users access the applications, the security workflow requires an explicit request of authorisations by the users. Therefore, independently of the user provenance, he/she has to fill a form up with some information about the activity to perform and apply to obtain an access to the portal (nevertheless, if the user works for an institution providing the Science Gateway or having a special agreement with it then the process could be faster). On the other hand, the authentication process has no particular requirements but the institutions may require to handle this step for a better implementation of the SSO with the internal infrastructures. In fact, managing the authentication makes easier to create the integrated environment requested by users because they would access all internal resources with the same credentials and a single log-in.

To support the different security requirements, separate tools, integrated in the security chain provide authentication and authorisation steps. These are an LDAP server for the authorisation and a Shibboleth federation (see section 2.2) for the authentication.

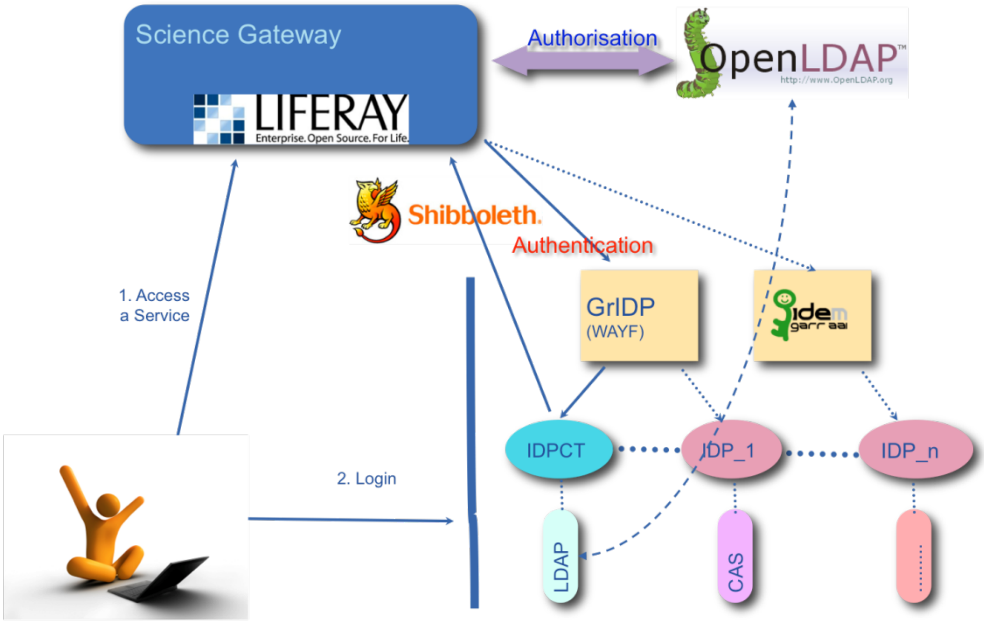

Fig. 3 – Access to the portal.

Figure 3 shows the interactions among all the components managing the authorisation/authentication steps. When a user tries to access a service in the restricted area of the portal (action 1), Liferay redirect the user to the authentication process. This is located behind a Shibboleth area defined in the web server so the user is redirected to a federation[1]. The federation page allows the user to select an IdP for the authentication and sends the user to the authentication page (action 2). After the authentication with the IdP, users can access the log-in page of the portal. The plug-in developed for Liferay, introduced in Section 2.2, takes effect and log-in the user with the credentials provided by the IdP. Therefore, users can operate regularly in the portal performing the log-in when requested and this is managed by the own institution. Although, the authentication is driven by the browser it is totally transparent to the users.

After the authentication with the IdP, in order to log-in, the user must exist in the LDAP server storing the authorisations. If the user is not present, or is not authorised to work in the

[1] Multiple federations should be supported by the portal before the end of the DECIDE project.

Science Gateway, then the access is denied even if his/her own IdP provides the authentication and an authorisation request page is shown.

After the authentication, Liferay retrieves from LDAP the list of groups the user is member of and performs the association. For each group defined in the LDAP, Liferay creates automatically a role, which is the element where authorisation can be specified inside the portal. Therefore, LDAP is responsible to manage user groups and, according to the group, the user gets the roles and the authorisation from Liferay.

The portal administrator is the responsible for the consistency between the group defined in the LDAP and the authorisation linked to the role associated with the group.

4.2 Interacting with the Grid

Obtaining the access to a service in the Science Gateway is only the first part of the authentication process. Users still need to be authenticated by the Grid security mechanism in order to access the available resources and/or execute the applications.

Grid authentication is entirely based on X.509 certificates so a user should use his/her own certificate to execute an application. This is not recommended in a Science Gateway because it usually creates complications for non-experts.

The proposed solution to interact with the Grid makes use of temporary proxies generated by robot certificates [1]. Robot certificates are special certificates stored in USB Smart Cards (referred to as e-tokens) that allow the creation of proxy certificates. Therefore, it is possible to bind robot certificates with applications and allow people to run them without any personal credentials. This approach has been adopted and Figure 4 shows the proposed architecture and workflow. According to the proposed schema, in order to use a Grid service, a user has to be authenticated and authorised on the portal, so the log-in described in section 4.1 comes in action. Then, when an operation to a Grid resource is requested the authorisations are verified and the portal retrieves a valid proxy for the service (action 3 and 4 of Figure 4) from a special proxy server managing the robot certificates [2]. The voms extensions bounded in the proxy depend on the service and the user authorisations so different users can obtain different proxies for the same service but providing different privileges. System managers and developers control the mapping between LDAP authorisations and voms extension.

In some scenarios, services do not access the Grid resources directly but their architecture requires extra layers between the portal and the Grid. If this is the case, the communication between the portal and the components in the middle carries the user information and authorisation so these components can retrieve the proxy accordingly.

Actually, proxies associated to the Grid transactions are not strictly bound to the user since the Distinguished Name (DN) in the proxy does not contain any information about the user. Hence, from the Grid point of view, a single user is performing all the different operations made by the portal. To monitor the user activity and ensure the non-repudiability feature of Grid transactions, all the services have to track the users requesting operations on Grid. Information are collected by an extra component to be deployed in the infrastructure, the User Tracking System, that combines the information coming from the services with the records in the Logging and Bookkeping service [5]. This component performs the association between users and operations in a non-repudiable way.

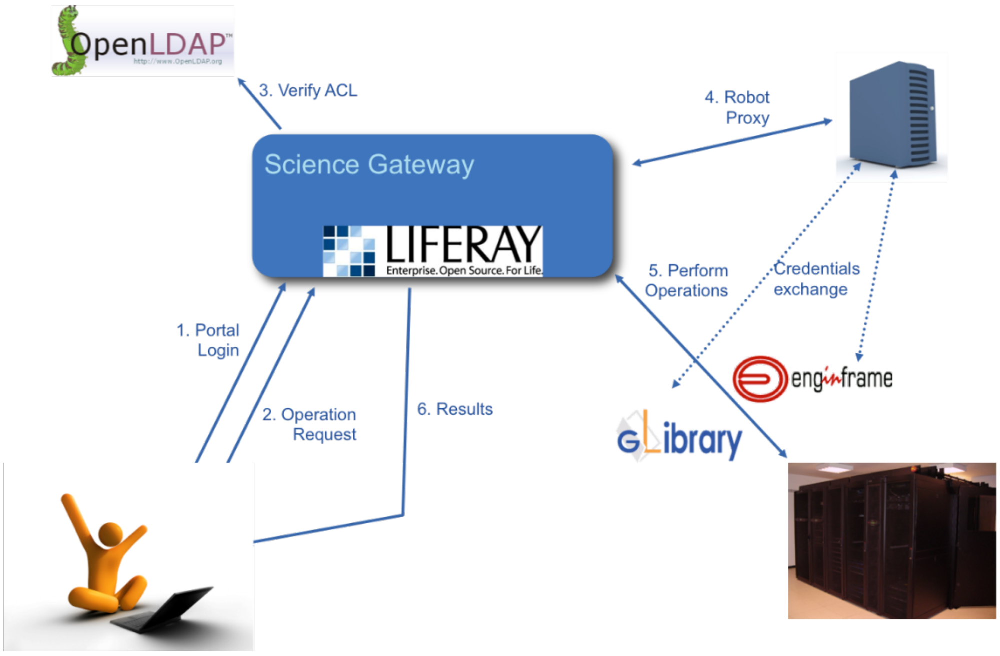

Fig. 4 – Interaction with Grid resources.

5. Related Works

Several research groups have identified in the web technologies and portal framework the best way to meet the needs of a virtual research community. These elements combined together allow building Grid Portal, like GENIUS [11][12], and Science Gateways as done by TeraGrid Community and others [9].

TeraGrid Science Gateway follows an approach similar to the one described, using Liferay and other similar technologies. However, the security mechanism is quite different because TeraGrid does not use robot certificates in their portal.

Differently, GENIUS Portal has been the first Grid Portal supporting an authentication method that grants an easy access respecting all the Grid policies in terms of security. The previous work developed for the GENIUS Portal has been updated and improved to power our Science Gateways with a brand new authentication method. In this way, INFN group has developed a single sign on method between the restricted resource offered by the Science Gateway and Grid Services. User has not to deal with complex procedure to get a personal certificate and with the usage of one password can access all the services the Science Gateway is created for. Robot certificates allowed this essential step thanks to their adoption by the INFN CA (Certification Authority); certificate designed for an easier access to grid functionalities and make Grid Services available to larger community.

Conclusions

This work has shown a way of building a Science Gateway meeting the requirements of a specific Virtual Research Community that takes in strong consideration all the matters related to the security level of the final portal.

Any Science Gateway needs a framework that provides developers and web designers with the essential tools to build the web portal. After a deep comparative analysis, thanks to its rich set of features, keywords of its success, we identified the Liferay portal framework as the component able to fully support this complex process of portal shaping.

One of the main aspects of Liferay responsible for its spread is its simplicity. Straightforwardness not only because of its user friendly interface or because users do not worry about where their applications run but also related to the fact that they can finally focus on their work, not caring about how the security of the system is actually provided.

The simple access to the resources behind a Liferay portal is based both on an internal management of users' data made by Liferay itself and, on the other side, done by the interaction of other authorisation and authentication systems which are used by those organisations that want to keep their services accessible by a selected community.

The adoption of the SAML standard is motivated by the needs of enabling access to different organisations working on the same project through the Science Gateway.

Our future work, after the prototype phase, will consist in testing the system on a production environment in order to refine the methodologies established in this study. The developed technology has been made available to the DECIDE project as a result of the support provided during the development. Actually, the project is developing a Science Gateway for its partners that should reach production quality by the end of the project

Other projects, working in different fields, are evaluating the use of the above model in their infrastructures in order to provide users with an easy access to the provided resources. The requirements coming from these projects will be integrated in the design and implementation to make the approach as general as possible.

References

[1] R. Barbera, G. Andronico, G. Donvito, A. Falzone, J. J. Keijser, 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.

[2] R. Barbera, V. Ciaschini, A. Falzone, and G. La Rocca, *A new "lightweight"crypto library for supporting an advanced grid authentication process with smart cards*, International Symposium on Grids and Clouds (ISGC), Taipei. Taiwan, 2011, Oral presentation.

[3] S. Cantor et al., *Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0*, Tech. report, OASIS SSTC, http://docs.oasis-open.org/security/saml/v2.0/saml-core-

2.0-os.pdf , 2005, Document ID: samlcore-2.0-os.

[4] *JSRs: Java Specification Requests*, http://www.jcp.org/en/jsr/all.

[5] L. Matyska, A. Křenek, M. Ruda, D. Kouřil, M. Voců, J. Pospíšil, M. Mulač, and Z. Salvet, *JobTracking on a Grid - the Logging and Bookkeeping and Job Provenance Services*, Tech. Report 9, CESNET, 2007.

[6] *The Shibboleth System*, http://shibboleth.internet2.edu/.

[7] DECIDE Project homepage available at: http://www.eu-decide.eu/.

[8] Liferay framework home page available at: http://www.liferay.com.

[9] N. Wilkins-Diehr, *Special issue: Science gateways - common community interfaces to grid resources.*, Concurrency and Computation: Practice and Experience **19** (2007), no. 6, 743–749.

[10] N. Wilkins-Diehr, D. Gannon, G. Klimeck, S. Oster, and S. Pamidighantam, *Teragrid science gateways and their impact on science*, IEEE Computer **41** (2008), no. 11, 32–41.

[11] Andronico G, Barbera R, Falzone A, Lo Re G, GENIUS: a web portal for grid. Available at: https://genius.ct.infn.it Nucl. Instrument and Methods in Phy., 2003.

[12] Barbera R, Falzone A, Ardizzone V, Scardaci D. The GENIUS Grid Portal: Its Architecture, Improvements of Features, and New Implelemtations about Authentication and Authorization. Enabling Technologies: Infrastructure for Collaborative Enterprises, 2007.