EGI-Engage

Platforms for the long tail of science

D5.2

Date	09 March 2016
Activity	SA1
Lead Partner	EGI.eu
Document Status	FINAL
Document Link	https://documents.egi.eu/document/2671

Abstract

This document presents the status of the long tail of science (LTOS) processes, policies and tools, and it includes a description of the architecture, the processes, and the associated services (the "platforms") to be integrated or under integration. The deliverable also reports on a survey involving NGIs and service providers aiming at gathering information about the experience gathered at national level when engaging with the long tail of science, and about the services for end-users that provide the greatest value proposition. Purpose of this survey is to collect information on new services to be accessed via the EGI LTOS and to formulate a business plan.

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	Partner/Activity	Date
From:	Peter Solagna	EGI.eu/SA1	9/03/2016
Moderated by:	Małgorzata Krakowian	EGI.eu/NA1	19/02/2016
Reviewed by	Tiziana Ferrari Kostas Koumantaros	EGI.eu/NA1, GRNET/JRA1	03/03/2016 29/02/2016
Approved by:	AMB and PMB		9/03/2016

DOCUMENT LOG

Issue	Date	Comment	Author/Partner
v.1	14/02/2016	Version ready for external review	Peter Solagna/EGI.eu
FINAL	9/03/2016	Final version after external review	Peter Solagna/EGI.eu

TERMINOLOGY

A complete project glossary is provided at the following page:
http://www.egi.eu/about/glossary/

Contents

Executive summary

The long tail of science platform (also referred to as "LTOS Platform" or "the platform" in the rest of the document) is a set of security policies, processes and tools for lightweight user authentication, authorization and resource allocation that allow small research groups and single users to benefit from granted access to a limited quota of advanced computing and data services in EGI, while operating in full compliance to the EGI security policies and trust model. The platform significantly advances the trust model of EGI (originally designed for long-term research collaborations with sustained long-term need of substantial amounts of storage and computing resources), to address the needs of short-term research projects. The objective is to provide alternative processes for authentication and authorization and GUIs that improve the accessibility and usability of various services of the EGI service catalogue.

The LTOS platform is designed for occasional short-term research projects for which creating or joining a Virtual Organization and getting resource allocations are processes with excessive overhead. The platform is also meant for potential users who initially need to a small allocation of a service quota to try the capabilities offered by EGI.

The LTOS platform is composed of a user registration portal, an e-token server, a pool of dedicate resources supporting a catch-all VO, and a set of science gateways providing data analytics and other community-specific services. The Platform is not a service per se for end-users, but rather a component service allowing access to a portfolio of capabilities, which can be arbitrarily expanded to address the needs of generic users. The Platform will be also an enabling service to expand the EGI service portfolio by including Software as a Service (currently EGI does not provide SaaS directly, but supports external providers by offering a hosting environment with the EGI Federated Cloud). So far the Platform has been demonstrated by integrating it with one science gateway – the Catania Science Gateway.

The document provides an overview of the services that are mostly wanted by the users. This information is needed to develop a business plan that suits the needs of the LTOS user groups.

Surveying users of the long tail of science at the current status of the platform is a difficult task, since by definition they are not organized as members of long-term scientific collaborations. The problem was overcome in collaboration with the national partners, who work closely with university and national research institutes. The result of this investigation is that generic services like cluster computing, cloud IaaS, and storage are the most popular, and should be higher priority in the LTOS business plan to become LTOS services.

DIRAC, iRODS and cloud brokering with Slipstream are some of the solutions that will be evaluated for high priority integration in the LOTS platform.

1 Introduction

EGI identifies Long Tail of Science (LTOS) users as the individual researchers and small research groups who - unlike large, organized collaborations - do not have access to computational resources and online services to manage and analyse large amount of data.

The LTOS platform addresses on demand consumers of EGI services, for example those who need to accommodate short-term bursts of computing workloads.

EGI has well-established processes to allocate resources for organized user communities, these involve a procedure for service level agreement negotiation engaging the user community manager and the EGI service providers, and resulting in medium-term service allocations (the average duration of a service level agreement is 2 years). These processes are not suitable for individual researchers and small research teams as they may involve:

- Obtaining an IGTF[1] personal certificate;
- Set-up a virtual organization (VO) or join an existing one after approval;
- Have the VO enabled in the distributed infrastructure, and use the resources.

Although EGI has processes and solutions to overcome these barriers, to some users the overhead associated is enough to discourage and make them look for alternatives, in most of the cases these new users are looking for quick access to resources to cope with their deadlines.

Recognising the need for simpler and more harmonised access for users of the LTOS, a prototype for a new platform was designed including ad-hoc processes and policies, to support these users and to reduce, if not eliminate, the abovementioned barriers. The platform main features are:

- User authentication via home username/password.
- Use of a pre-existing ad-hoc catch-all Virtual Organization, the LTOS VO, to support all the users of the platform.
- Pre-allocation of a centrally managed resource pool, which can be allocated to users via grants, and which is pre-configured to support the LTOS VO.

This LTOS platform allows individual researchers and small research teams to perform compute and data-intensive simulations on large, distributed networks of computers in a user friendly way.

The platform, besides the technical services, aims to create communication channels between EGI and the NGIs who have already implemented processes and services for the national LTOS users, and who could better support their users. Through the LTOS

---

[1] Interoperable Global Trust Federation: https://www.igtf.net/

platform, users will be able to get in contact with the user support teams of the respective NGI, in case the use case could be better supported locally.

This document provides an overview of the architecture of the LTOS platform and its integrated services, the components and the interactions, the processes that support the workflows and how its modular nature can evolve to expand the services offered to the users.

In summary the LTOS platforms is designed to be an almost zero-barrier access to the EGI services.

2 Architecture of the Long Tail of science Platform

The LTOS platform is composed by the following main technical components, as shown in Figure 1:

- A user registration portal (access.egi.eu);
- A catch-all VO;
- A pool of resources where the catch-all VO is pre-enabled;
- A set of science gateways and portal;
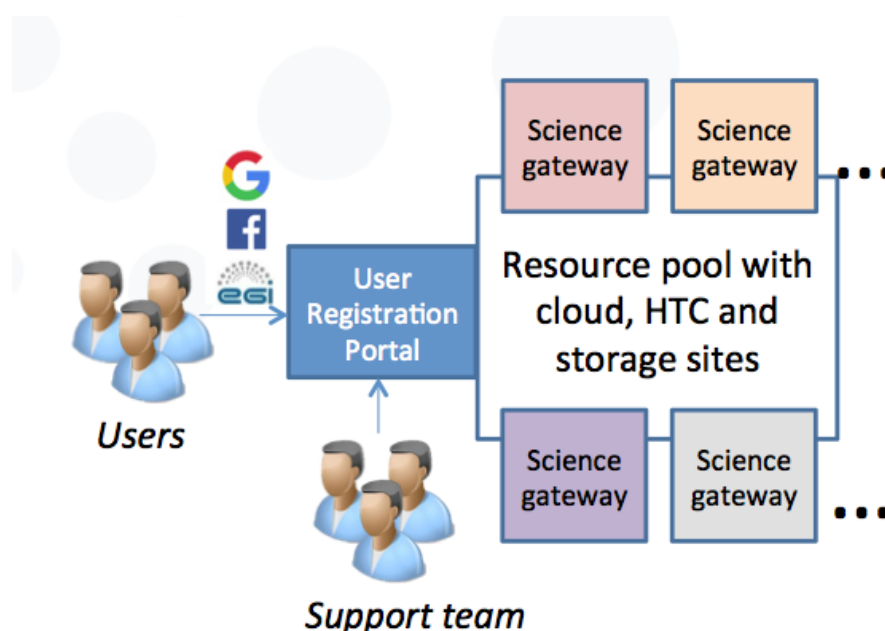- A X509 credentials factory.

Figure 1. High-level architecture of the LTOS platform

The **user registration portal** is composed by a website (access.egi.eu) which allows users to authenticate using either social credentials or EGI SSO IdP[2] and to provide information about their contacts, institutions, and research topic. The website is the initial user-facing interface for both the end users and the support team who will approve and manage the registrations and requests.

The user registration portal has also an interface for the services, such as the science gateways, in the form of an identity provider (IdP) that is used by the science gateways to get user authorization data and attributes. The user interface has been designed and

---

[2] EGI SSO identity provider: https://www.egi.eu/sso/

developed specifically for the LTOS use case, the forms used by the user to describe the type of research and the expected usage of resources use the interfaces developed for e-grant (the EGI resource allocation tool)[3]. The IdP interface is using Unity[4], a pre-existing authentication and authorization management solution using OpenID Connect as standard interface, allowing the use of social accounts as valid user credentials.

The **catch-all VO** is the Virtual Organization (vo.access.egi.eu) under which all the activities of the LTOS users are performed and accounted for. For security reasons, this VO can be used only by users owning credentials authorized by the LTOS platform.

The **pre-allocated pool of resources** is operated by various geographically distributed EGI Resource Centres, both cloud and HTC that accepted to enable the vo.access.egi.eu VO on their resources. The services on these sites are configured to recognize the X.509 per-user sub-proxies, and handle the credentials according to the LTOS procedures and policies.

The **Science Gateways** (SG) are user interfaces, usually web portals, which provide a set of high-level tools. Since command-line interface may be too complicated for new users, SGs provide a user-friendly web based interface. Integration of Science Gateways with the LTOS platform is needed to consume authorization information from the User registration Portal and to implement the use of the per-user sub proxies for traceability of user identifies through the credential factory. These integration developments have proven to be straightforward in terms of development effort.

The **X.509 credentials factory** (or e-token service) is directly contacted by the SGs to generate a per-user-sub-proxy (PUSP)[5] from a robot certificate[6]. Permission to request robot certificate proxies is granted only to the science gateways integrated in the LTOS platform and only when the user accessing the SG is authorized. The purpose of a per-user sub-proxy (PUSP) is to allow identification of the individual users that operate using a common robot certificate. This is achieved by creating a proxy credential from the robot credential with the proxy certificate containing user-identifying information in its additional proxy CN field. The user-identifying information may be anonymised where only the portal knows the actual mapping.

The LTOS platform integrated as X509 credential factory the solution developed by the Italian National Institute of Nuclear Physics (INFN), division of Catania, the e-Tokens server which allows central management of robot certificates and provides credential proxies to get seamless and secure access to computing e-Infrastructures, based on local, Grid and Cloud middleware supporting the X.509 standard for authorization. This solution enables the enrichment of the user identity with additional information

---

[3] http://e-grant.egi.eu

[4] http://www.unity-idm.eu

[5] https://wiki.egi.eu/wiki/Fedcloud-tf:WorkGroups:Federated_AAI:per-user_sub-proxy

[6] Robot certificates are X509 credentials that can be used programmatically, for example to generate proxy credentials, without the need for the user to perform direct actions, e.g. inserting a password.

during the proxy generation process to allow the identification of the individual users that operates using a shared robot certificate. This is key for accounting and traceability of the usage of e-Infrastructures.

The interaction between the components is described in Figure 2, the steps represented are the following:

1. User registers in the access.egi.eu portal, a user ID is associated to the user and kept in the user database of the portal;
2. The user, once approved, can log into the SG, to access EGI resources;
3. The SG confirms that the user is authorized with the IdP provided by the access.egi.eu portal and retrieves the UID of the user through the IdP interface.
4. The SG generates a PUSP from the credential factory (not in the picture), including the UID information in the sub-proxy, and uses it to submit the payloads of the users to the HTC or cloud resources in the LTOS pool.
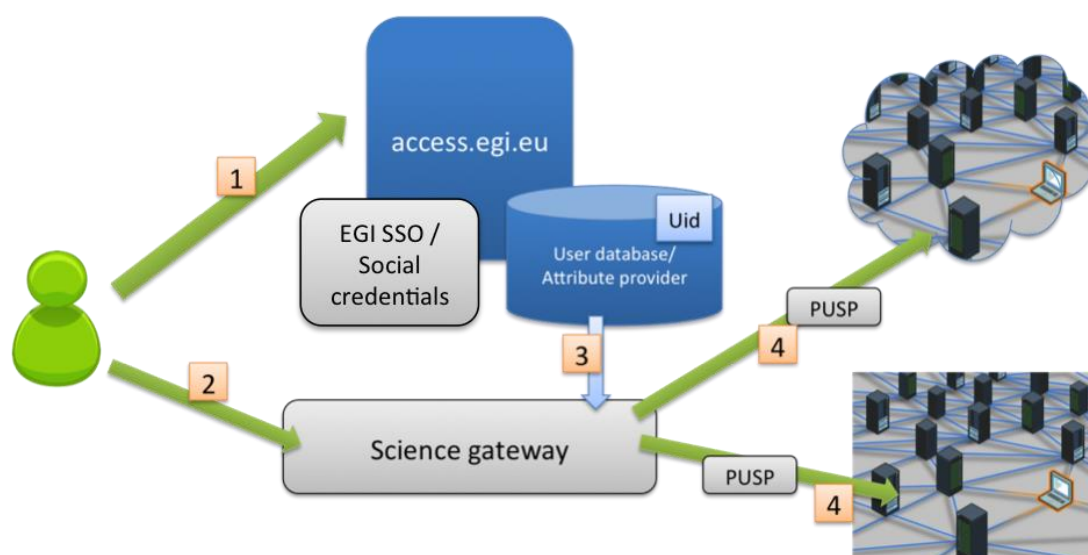
Figure 2, Interactions between the components of the LTOS

The advantages of the centralised user management are as follow:

*   EGI can control centrally the users' activities
    *   Users are associated to a persistent unique identifier that is used in all the services accessed through the platform.
    *   Thanks to the central management, users can be quickly suspended in case of a security issue.
*   Science Gateways do not have to keep a users database or manage authorization, since these information are hosted centrally

3 Operational processes and business model

This section analyses the policies and processed that have been developed to support the LTOS, providing an overview of the non-technical components of the platform.

3.1 User validation and user management

New users register in the user management portal (access.egi.eu), as described in the previous sections, providing information about their institution, referee and topic of the research for which they would like to use EGI resources.

New users registration required validation, and this is a human process that must be performed by either the EGI user community support team, in collaboration with the NGI user support representatives. The process of validation can exclude users who are not deemed eligible, and provides an additional level of assurance to the service providers who enable the LTOS VO trusting EGI in authorizing the right users. In the current pilot stage, no limitations in use are applied. The definition of the Terms and Conditions is part of the general LTOS platform business plan and is the roadmap. The only limitations are the ones described in the acceptable usage policy, summarized in section 3.2.

During this first pilot the validation of the users will be managed by the EGI user support teams at EGI.eu and the NGIs (UCST and NILs), the NGI NILs will be also involved in the process to manage requests coming from their national users.

The process for validating a user must be limited to the essential steps to reduce any additional overhead and to reduce the operational costs of the platform. The necessary checks to be implemented are:

- The user is affiliated to a research institution, or is collaborating with someone affiliated to a research institution.
- The user has no other active accounts in the LTOS platform.
- The user is accessing the LTOS platform for research purposes.

The checks that the user support team member will verify are the following:

1. Is the institution reported by the user a research institution, or another legal entity with research purposes?
   a. NILs can provide information about the research institutions in their countries, where this is not possible, the team can refer to official EC documentation[7].
   b. If not, the request will be suspended.
2. Has a user with the same name, or with the same email another account on the LTOS platform?

---

[7] One example of source of information for research institutions is: http://ec.europa.eu/eurostat/documents/203647/771732/Recognised-research-entities.pdf

   a. In case the users already has another account the new one will not be approved.

3. Has the user registered with an institutional email, and/or has the user provided the link to a contact page on the website of their institution containing the user's contacts?

   a. Using an institutional email, verified during the registration process, and reported on the contacts page will be usually enough to provide the minimum level of assurance to access the LTOS platform. If a phone number is available on the institution site of the user the user support team will use it for a phone confirmation of the data.

4. If the user suggests a referee in a research institution, this reference must be checked by phone.

5. Has the user previously used the LTOS, has they properly acknowledged EGI and reported results to be associated to their use of EGI services (e.g. publications)?

   a. User must provide an explanation for not following this policy, before being approved again.

6. Is the research topic relevant?

   a. This step is not a review of the quality of the research, but only checks that the resources devoted to the user will be used for research and non for profit goals.

Some simple tagging mechanisms are being implemented in the LTOS in order to make easier to apply filters to the notifications of new registrations sent via email, to make easier for the NGI representatives to filter out all the notifications that are not relevant to their NGIs.

The future integration with eduGAIN[8] will streamline the authorization process: for the users using eduGAIN credentials the LTOS platform will be able to automatically retrieve information about the users' association to the research and education community. After an initial pilot period, the eduGAIN users will be automatically approved, if their identity providers will release enough information about the user.

Once the user is approved, access of EGI resources will be possible until the approved status is valid. Currently the reasons for which a user can be suspended are:

- The user has used the maximum amount of resources allocated to given user
  - ACTION. Temporary suspension in order to understand if the allocation of additional resources is needed, or if the user should join another production VO in EGI. Alternatively, a larger amount of resources can be requested through the e-grant[9] tool.
- The user was approved 12 or more months ago

---

[8] http://services.geant.net/edugain/Pages/Home.aspx
[9] http://e-grant.egi.eu

  o ACTION. Temporary suspension in order to confirm the validity of the data reported by the users 12 months ago, before re-enable the user.
- The user behaviour does not comply with the EGI policies
  o ACTION. Depending on the severity of the non-conformity suspension can be permanent.

The Terms and Conditions drafted in this section will be tested during the LTOS platform pilot stage and will be refined as part of the business plan preparation.

3.2 Security policy and AUP

The deployment of the LTOS platform required policy work to support the technical solutions. The existing EGI VO portal policy[10] limits science gateways in the actions that can be performed on behalf of the user with a robot certificate.

A LTOS security policy[11] has been developed as part of the activities of the project to define what services integrated with the LTOS are allowed to do on behalf of the user, and what the EGI requirements are. The policy also describes the risk envelope of the LTOS platform, and it suggests what mitigations the resource providers should implement.

The main points in the security policy are:

1. User access should be subject to explicit approval.
2. User affiliation must be verified at the moment of the approval, and verified every 13 months
3. User information must be kept in the platform for at least 13 months after the user's last access
4. The resource providers shall accept the risks introduced by supporting the LTOS users, and shall protect the rest of the infrastructure from that additional risk

The security policy also suggests a list of implementation guidelines for the resource providers, to be compliant with the policy and minimize any risk associated to the LTOS to an acceptable level.

The security policy allows users of the LTOS to access a set of features that are similar to the capabilities offered to users owning a personal certificate, for example submitting arbitrary binaries. The verification of the user identity, although acceptable, is not as strong as it is performed by a certification authority who releases certificates. This introduces the additional risk that is evaluated in the policy and that should be mitigated by implementing the recommendations.

The second policy document produced to support the LTOS activities is the Acceptable Use Policy and Conditions of Use of the EGI Platform for the Long-tail of Science[12]. It is an

---

[10] https://documents.egi.eu/document/80
[11] https://documents.egi.eu/document/2734
[12] https://documents.egi.eu/document/2635

extension of the general AUP for the EGI users[13], which implements the specifics requirements and limitations of the platforms from a user's perspective. Every user is required to accept the AUP during registration in the portal.

The main points in the LTOS VO AUP are:

1. The user is authorized to use the EGI resources through the LTOS platform only for the purposes described during the registration.
2. Access is suspended after the usage of resources exceeded the assigned allocation (the same for every user), or the time limits.
3. User must provide appropriate acknowledgement of the EGI support in the outputs of their work
4. Contact information must be kept up-to-date in the access.egi.eu portal
5. The LTOS platform, and the resource providers, are allowed to track user's activities to be kept confidential and used only for operational and security purposes
6. User accepts the risk that their access to the LTOS platform can be suspended and is provided without any guarantee

Both policies have been approved by the EGI bodies and are in production.

3.3 Business model

The content of this section summarizes the information and features of the LTOS platform relevant to develop a business model to support it.

The customers of the long tail of science are individual users or small group of researchers who need to access computing and storage resources that are not available in their home institutions. These services must be provided through an easy-to-use platform designed to reduce to a minimum the overhead for the users, and possibly to reduce to a minimum the support that users need.

Existing services provided by the NGIs or by the Virtual Organizations can benefit from the LTOS as well. In case of a service that needs to be open and easy to access, for example to provide access to open data, the service provider can delegate user management to EGI and to the LTOS platform, since it is already providing the necessary features.

In particular, platforms or Virtual Research Environments that are using EGI resources would highly benefit from the out of the box provision of X.509 credentials and a pre-allocated pool of resources.

VOs or communities who are developing tools for a specific research discipline may want to make the tool available to as many users as possible, to increase the use base of their software. Integrating it with the LTOS, will both achieve the results of creating a shortcut for new users to join, and advertise the tool in the LTOS platform portal.

---

[13] https://documents.egi.eu/document/2623

The LTOS platform will provide easy access to EGI general-purpose compute and data management services. The LTOS platform implements a scalable process, requiring minimum effort and minimum bureaucracy, to allow individual users to access EGI resources. The platform will support users and small collaborations who have as main requirements computing and storage capabilities, while the provisioning of services that require dedicated support at resource centres level, is for the moment out of scope.

The platform will support the user registration, and the validation process that are required to enable the new users on the platform. The request validation process is a continuous activity that checks that new registrants are fulfilling the pre-requisites defined by the acceptable usage policy. The LTOS platform can be extended by integrating new services as described in the previous section, this activity is part of the maintenance and operation of the platform.

For NGIs who do not have existing processes to support their LTOS users, enabling the LTOS platform is as easy as enabling a new VO in their Resource Centres. The national LTOS user group could be redirected to the EGI LTOS portal allowing access to the platform.

The LTOS platform is an access mode to EGI services, not a service by itself. Therefore the LTOS platform will not be directly published in the EGI Marketplace, but it will be advertised as component enabling one access mode to other services. For example, the marketplace will list the LTOS access as a feature of the EGI services integrated with the LTOS platform.

Funding for the LTOS platform can have several sources. Users will not pay for the use of the resources, but they may access pay-per-use services if they need more capacity than what is provided by the LTOS pool of resources.

The LTOS platform can support more use cases, for example, it can be used as an early access, trial version, of new services, or to support citizen science use cases. Considering the potential applications of the tool, it could be supported by EGI members' fees and co-funding through EC projects.

This is expressed in the following table using the business model canvas[14] as a means for understanding the building blocks of the LTOS platform.

---

[14] https://en.m.wikipedia.org/wiki/Business_Model_Canvas

Table 1, LTOS Business Model Canvas

Key Partners	Key Activities	Value Proposition	Customer Relationships	Customer Segments
Cyfronet for hosting, operating and developing the platform Individual sites participating to make resources available for access by customers NGI NILs, who validate the request from the users	Customer validation ands support to ensure fulfilment of policies and terms of reference Establishing and ensuring applicable service management processes Operation and maintenance of the platform Promotion through communication team as well as UCST	An easy-to-use platform for individual researchers and small research teams researchers to access compute, storage and application services and perform compute and data-intensive simulations on large, distributed networks of resources	Minimum individual interaction mainly during verification process Minimum individual support Majority should be automatized Knowledge of the long tail of science is fed through established relationships with large research groups	Individual researchers and small research groups who aren't able to contribute resources to the infrastructure but still require services to perform complex research activities
	Key Resources Computing and storage dedicated to the activity made available through the participating sites Human effort for development/operation of the platform as well as verification of users		**Channels** The long tail cannot be handled on an individual basis, but through available and promoted online platform. This will mainly be done through the LTOS Portal and eventual EGI Marketplace. The EGI.eu Comms and UCST team will be the main sources of dissemination support.	

Cost Structure	Revenue Streams
At least 1000 CPU dedicated to the activity made available through the participating sites ~3PM for development/operation of the platform ~1PM per year for verification of users Development support through EGI-Engage project; future funding will be done through bidding process once added officially in EGI Service Catalogue	Indirect: all users through the LTOS will access the services/resources free of charge; however, is part of the sales strategy to get an increased number of customers using EGI, which can move to other revenue generating models as usage increases. For operation, maintenance and development of the platform will depend on whether the platform because a core service and is supported through the participant fees with shared costs through the bidding process. Co-funding will also be sought through future projects where individual users or small research groups need to test, access resources in a federated environment.

3.4 Integration of new services

As described in previous sections, the LTOS platform has developed a user registration and management process and associated a pre-allocated pool of resources. The link between the users and the resources are the science gateways or other services that can help users in easily use the resources.

The requirements for the science gateways and other services (called SG for simplicity) are the following:

- The SG must authenticate the users through the LTOS platform, using access.egi.eu

- The SG must allow the users to act on the EGI resources only if they are authorized in the access.egi.eu portal
- SG must use the UID provided by access.egi.eu to log user activity
- If X.509 credentials are required for some actions or to access other services, the SG must use robot certificates and per-user sub-proxies. The PUSP must contain the UID provided by access.egi.eu

To fulfil these requirements the development work for a SG is relatively limited.

- The SG should implement OpenID Connect authentication, since the unity service releases user information through the OIDC interface. Many portal frameworks already have OIDC plugins that can be used out of the box.
- The SG should retrieve the X.509 credentials from the credential factory. This can be done as easily as with a HTTPS call through a REST interface. The credential factory (token service) can provide robot certificate proxies already in the per-user sub-proxy format (UID is provided by the SG in the REST call, and added as an extension to the X509 proxy), and already with the VO extensions for the LTOS VO.

The steps required for the integration on the LTOS side are very simple as well:

- Authorize the SG to retrieve authentication and authorization information from access.egi.eu
- Add a link of the list of SG in the pages of access.egi.eu

The integration process is tracked through EGI helpdesk tickets[15] by the "Operations" support unit. Before the technical implementation, the operations and the user support team evaluate the new SG in terms of maturity and effectiveness to support LTOS users.

---

[15] http://helpdesk.egi.eu

4 Piloting of the LTOS platform

4.1 Catania Science Gateway

The Catania Science Gateway was selected as first service to be offered by the LTOS platform in its pilot stage. Additional services are being integrated and the business plan will define the future integration roadmap, also taking into account the input received from the NGIs through a dedicated survey.

The Catania Science Gateway Framework[16] (CSGF) is an open source toolkit developed by the Italian National Institute for Nuclear Physics (INFN), division of Catania, to build, in a fast and easy way, standard-based web 2.0 Science Gateways to exponentially increase the number of potential users of Distributed Computing Infrastructure (DCIs) worldwide.

The framework, originally conceived in the context of the CHAIN project, has been used in the last couple of years to develop Science Gateways for several EU co-funded projects such as: DECIDE[17], EarthServer[18], EUMEDGRID-Support[19], GISELA[20], DCH-RP[21], INDICATE[22] and CHAIN-REDS[23] just to name a few. The framework is being completely re-engineered in the context of the INDIGO-DataCloud[24] project to include additional cloud and data management functionality and to improve its performance.

The CSGF was selected as first pilot service as it is based on well-established and widely adopted standards making its integration with component services simple, these are: SAGA[25], SAML[26], OCCI[27] and it is fully compliant with both the EGI.eu VO Portal and User Traceability policies[28].

In more details, user Authentication and Authorization framework is based on the SAML standard for Identity Federations, including Social Networks' credential providers, while application interfaces are based on the OGF SAGA standard and are thus middleware-independent.

---

[16] http://www.catania-science-gateways.it/
[17] https://www.eu-decide.eu/
[18] http://www.earthserver.eu/
[19] http://www.eumedgrid.eu/
[20] http://www.gisela-grid.eu/
[21] http://www.dch-rp.eu/
[22] http://www.indicate-project.org/
[23] https://www.chain-project.eu
[24] https://www.indigo-datacloud.eu/
[25] https://www.ogf.org/documents/GFD.90.pdf
[26] http://saml.xml.org/
[27] http://occi-wg.org/
[28] https://documents.egi.eu/document/80

In the context of the EGI-Engage project, the CSGF has been further extended developing a new plug-in[29] to support the Authentication schema based on the OpenID Connect[30] protocol.

The CSGF allows users to execute applications on the EGI Federated Cloud and HTC resources, through web portals/SGs. The Science Gateways based on CSGF provide users with an intuitive web interface to execute applications on the Cloud as jobs and to manage these jobs during their running (check the status and download the output). The SG takes care of starting the VMs on the EGI Federated Cloud, of transferring the needed files (e.g. executable, input files, etc.), stopping the VMs and downloading the output in behalf of the user.

The LTOS users can exploit the CSGF features to:

- Execute applications already integrated in one of the existing Science Gateways;
- Add their applications in one of the existing Science Gateways;
- Deploy a new Science Gateway.

4.2 Services under integration

The following science gateways are in the pipeline for the integration with the LTOS platform:

WS-PGRADE[31]. WS-PGRADE is an open source science gateway framework that enables users to access HTC and cloud infrastructures. gUSE is developed by the Laboratory of Parallel and Distributed Systems (LPDS) at Institute for Computer Science and Control (SZTAKI) of the Hungarian Academy of Sciences.

WS-PGRADE is a Web portal hosted in a standard portal framework, using the client APIs of gUSE services to turn user requests into sequences of gUSE specific Web service calls. A graph editor component can be downloaded from WS-PGRADE and can be used to define the static skeleton of workflows, while the HTML pages of WS-PGRADE provide interfaces to add content to graphs, to generate complete Grid/Web service applications. The members of gUSE community can access WS-PGRADE via ordinary web browsers. Application developers can access to all the advanced workflow features (graph, abstract workflow, template, application and project) to develop new workflow applications and upload them to the gUSE repository. For scientific end-users WS-PGRADE gives full access to the parameterization and execution of applications downloaded from the gUSE repository.

WS-PGRADE is in a very advanced status of integration, at the moment of writing.

---

[29] https://github.com/csgf/OpenIdConnectLiferay
[30] http://openid.net/connect/
[31] http://guse.hu/about/architecture/ws-pgrade

QCG[32]. The QCG middleware is an integrated system offering advanced job and resource management capabilities to deliver to end-users supercomputer-like performance and structure. By connecting many distributed computing resources together, QCG offers highly efficient mapping, execution and monitoring capabilities for variety of applications, such as parameter sweep, workflows, MPI or hybrid MPI-OpenMP. Thanks to QCG, large-scale applications, multi-scale or complex computing models written in Fortran, C, C++ or Java can be automatically distributed over a network of computing resources with guaranteed QoS. The middleware provides also a set of unique features, such as advance reservation and co-allocation of distributed computing resources. The integration of QCG has started.

---

[32] http://www.qoscosgrid.org/trac/qcg

5 Assessment of the requirements

The LTOS platform needs to evolve by offering a rich set of SGs to increase the capabilities and tools available, which will make the LTOS service and related platform easy to market.

Identifying the services that would be needed by potential users of the LTOS is one of the major challenges of this activity. By definition LTOS users are extremely diverse and not organized, it is therefore very difficult to identify interlocutors to gather requirements.

Instead, the requirements for the LTOS platform have been gathered from the services providers, from the NGIs, who have extensive experience in supporting the LTOS users. The questionnaire targeting the NGIs contained few questions about the services that the NGIs consider important for the LTOS users, both general purpose and discipline specific. The survey was sent to NGIs Operations managers and NGI International Liaisons who are responsible for the national user engagement plan.

Those NGIs, who have been more active in the past in structuring the support for the LTOS in their countries, have been addressed directly to discuss their experience in greater detail.

The following sub-sections summarize the findings.

5.1 Data management services

Data services provided by NGIs are at the moment associated with HTC or Cloud computing. For example France Grilles provides data storage facilities via the iRODS[33] product. iRODS is accessible through the NGI France national catch-all VO, and allows users to manage their data in a virtualised storage, and to manage datasets, apply access control and implement rules for replication and high availability of the data. iRODS provides a native GUI.

5.2 Computing services

HTC Computing

The majority of the user support teams highlighted the importance of providing easy to access generic computing services. LTOS users often are interested in computing capacity to run their own scientific code, using their submission frameworks (e.g. scripts) which they were using on local computing resources until they needed to scale up the available capacity and applied for resources in EGI. To achieve this result, users could be given either access to a command-line interface or to a sufficiently generic user interface accessible using username/password-like credentials.

---

[33] http://irods.org/

NGI France has been providing DIRAC[34] for their long tail of science users, with good feedbacks. DIRAC is a workload management system with pilot jobs[35] introduced by the DIRAC project is now widely used by various communities as main tool for computing task submission. Through the web interface of DIRAC users can submit one or multiple computing tasks, monitor their status and retrieve outputs. DIRAC offers also command-line clients.

Other NGIs, such as NGI CZ, allow users to access with username/password machines where batch submission is enabled to the local clusters. On demand cluster computing and various different solutions will be evaluated for inclusion.

Cloud computing

Based on the answers provided to the survey, LTOS users are using GUIs to manage cloud resources. Users are both accessing directly the site cloud management framework GUI, for example Open Nebula, or using a cloud broker to use resources from multiple sites.

NGI France reports the use of slipstream[36], a cloud broker, to allow LTOS users to instantiate virtual machines across multiple cloud sites.

In terms of virtual machines images, users have access to general-purpose virtual machine images, where they can install their applications and use the virtual machine, or multiple virtual machines, to run it.

EGI-Engage will provide VM management capabilities via its virtual appliances store and its GUI (developments are planned in the context of WP4). These features will be integrated with the LTOS platform.

5.3 Discipline-specific services

None of the NGIs reported to provide discipline-specific or specialized tools to their LTOS users. Although NGI CZ reported that they are supporting their users with pre-installed software, such as R, or Matlab. In particular commercial software are highly appreciated by the users, who cannot access to licenses locally. A business development strategy will be discussed with innovators like CloudSME[37] for the offering of their simulation software via the LTOS platform.

The offering of licensed software will be at lower priority, as this is a service already offered locally by some NGIs. However, the offering of non-commercial software (R, for example) can be integrated and made available also at EGI level. It is already available in the Catania SG, and it could be used by users through virtual machines with pre-installed software. The need for R and similar software packages will be assessed during the pilot phase.

---

[34] http://diracgrid.org/

[35] https://en.wikipedia.org/wiki/Pilot_job

[36] http://sixsq.com/products/slipstream/

[37] http://cloudsme.eu/content/simulation-solutions

6 Conclusions and Future Work

The start of the piloting phase of the LTOS platform and the related services requires the completion of a number of activities:

1. The definition of the LTOS portal Terms and Conditions [timeline: April 2016];
2. The setup of the structures (team, processes) needed to support the LTOS platform and service operations including SLAs and OLAs [timeline: April 2016];
3. The integration of the LTOS portal with the operational structures of EGI like monitoring.
4. The improvement of the user experience through the LTOS portal according to the customer review outcomes reported in M5.1[38]; a second review of the ergonomics of the GUI will be conducted subject to the successful completion of the pilot phase. [timeline: May 2016];
5. The finalization of the LTOS business model defining the suppliers and the related underpinning agreements/operational level agreements [timeline: May 2016];
6. The definition of the LTOS service business plan to provide a rich set of LTOS services, building on the experience of the NGIs supporting their national LTOS users. The following initial priorities have been identified for the provisioning of new LTOS services in addition to the Science Gateways currently under integration. Products for cluster computing on demand and VM management are needed and will be evaluated [timeline: June 2016.

   Direct access to HTC computing. This is the highest priority based on the use cases listed by the NGIs through the survey. HTC computing is the most common service in EGI, and therefore it is the service for which the pool of resources could be expanded more easily, having more resource providers who could potentially contribute to it.

   The Catania SG, WS-PGRADE and QCG can all use HTC resources, but the requirements suggest that simpler access modes are necessary. In the context of the EGI LTOS platform, command line access is not possible – at the moment – since users should not be allowed to manipulate the X509 credentials generated by the token server. DIRAC is a workload management system that can support basic job submission, and its integration can improve the experience of the users. At the moment of writing EGI is evaluating with the DIRAC team the feasibility of its technical integration as described in section 3.4. Technical developments have been estimated to be limited [timeline: June 2016..

   Direct access to cloud computing. No cloud access from the LTOS platform is currently possible, but extensions already planned for the LTOS science gateways will enable this type of resource as well.

---

[38] https://documents.egi.eu/document/2697

Slipstream broker could be one solution to be adopted to allow users to span multiple sites with the VMs instantiations. Slipstream is being evaluated with the federated cloud with other EGI-related activities, with particular focus on the support for the federating interfaces promoted by the EGI federated cloud.

Based on the results of the testing, and the feasibility of the technical integration of slipstream, EGI will evaluate the offering of cloud resources to the LTOS through this tool.

As an alternative to the use of external brokers, the VM management capabilities of the Application Database will be integrated in September 2016.

Storage. iRODS is the solution offered by NGI France, and it seems to satisfy the needs of the LTOS users. Different solutions will be evaluated, including iRODS and the EGI Open Data Platform being developed in WP4 [timeline: June 2016]

The current status of the platform itself is ready to start its pilot phase during which the platform will be provided to collect users' feedback. The next step forward is the integration of more SGs, in order to offer to the LTOS users a attractive catalogue of services. What is missing is a simple interface to submit computational tasks, and instantiate virtual machines, this is only partially provided by the SG currently integrated.

There are open actions to complete the full operational integration of the LTOS platform in the EGI production infrastructure, for examples the integration of the LTOS portal with the monitoring infrastructure, the definition of OLA and SLA supporting the service. These actions are planned in the coming month.

The technical review of the document, described in M5.1[39], highlighted some technical issues. The most urgent issues require small developments to be implemented, and the development timeline of such requirements is being defined at the moment of writing.

Once the number of user will get to a minimum critical mass, another study or ergonomic review of the tool will take place. With active users registered in the portal, a feedback-based review will be possible, and this is critical to identify if any gaps are still missing to be filled to reach the 'zero barriers' goal.

---

[39] https://documents.egi.eu/document/2697