EGI-InSPIRE

Services for the Life Science Community

EU MILESTONE: MS617

Document identifier:	EGI-MS617-1289-v1.2
Date:	**27/08/2012**
Activity:	**SA3**
Lead Partner:	**CNRS**
Document Status:	**FINAL**
Dissemination Level:	**PUBLIC**
Document Link:	https://documents.egi.eu/document/1289

Abstract

This report details the services offered to the Life Sciences community and how the intended users can access them.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/Activity	Date
From	Johan Montagnat	CNRS/SA3	9/08/2012
Reviewed by	**Moderator:** Karolis Eigelis **Reviewers:** Dean Flanders	EGI.eu	
Approved by	**AMB & PMB**		

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1.0	27/07/2012	Internal submission version	CNRS, SPACI
1.1	18/08/2012	Revised version	CNRS
1.2	24/08/2012	Revised section 2.1	SPACI

IV. APPLICATION AREA

This document is a milestone for the European Commission, applicable to all members of the EGI-InSPIRE project, beneficiaries and Joint Research Unit members, as well as its collaborating projects.

V. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors. The procedures documented in the EGI-InSPIRE "Document Management Procedure" will be followed:
https://wiki.egi.eu/wiki/Procedures

VI. TERMINOLOGY

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

VII. PROJECT SUMMARY

To support science and innovation, a lasting operational model for e-Science is needed – both for coordinating the infrastructure and for delivering integrated services that cross national borders.

The EGI-InSPIRE project will support the transition from a project-based system to a sustainable pan-European e-Infrastructure, by supporting 'grids' of high-performance computing (HPC) and high-throughput computing (HTC) resources. EGI-InSPIRE will also be ideally placed to integrate new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and desktop grids, to benefit user communities within the European Research Area.

EGI-InSPIRE will collect user requirements and provide support for the current and potential new user communities, for example within the ESFRI projects. Additional support will also be given to the current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences, as they move their critical services and tools from a centralised support model to one driven by their own individual communities.

The objectives of the project are:

1. The continued operation and expansion of today's production infrastructure by transitioning to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding.
2. The continued support of researchers within Europe and their international collaborators that are using the current production infrastructure.
3. The support for current heavy users of the infrastructure in earth science, astronomy and astrophysics, fusion, computational chemistry and materials science technology, life sciences and high energy physics as they move to sustainable support models for their own communities.
4. Interfaces that expand access to new user communities including new potential heavy users of the infrastructure from the ESFRI projects.
5. Mechanisms to integrate existing infrastructure providers in Europe and around the world into the production infrastructure, so as to provide transparent access to all authorised users.
6. Establish processes and procedures to allow the integration of new DCI technologies (e.g. clouds, volunteer desktop grids) and heterogeneous resources (e.g. HTC and HPC) into a seamless production infrastructure as they mature and demonstrate value to the EGI community.

The EGI community is a federation of independent national and community resource providers, whose resources support specific research communities and international collaborators both within Europe and worldwide. EGI.eu, coordinator of EGI-InSPIRE, brings together partner institutions established within the community to provide a set of essential human and technical services that enable secure integrated access to distributed resources on behalf of the community.

The production infrastructure supports Virtual Research Communities (VRCs) – structured international user communities – that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

VIII. EXECUTIVE SUMMARY

This milestone report lists the services delivered by the Life Science Heavy User Community, and describes how each of them can be accessed. All services have not been developed or deployed already and some sub-sections refer to work planed rather than actual instances of deployed services.

TABLE OF CONTENTS

1 INTRODUCTION

This document lists the services provided by the Life Sciences Heavy User Community from EGI-InSPIRE. These include (i) application services intended for the end-users and (ii) operation assistance services intended for VRC administrators.

The application services developed for the Life Sciences HUC are:

- GRelC (see section 2.1) is a generic database access interface that is useful for any grid application requiring data/metadata management. Along with user community support, the main goal in the context of the EGI-InSPIRE project is to deliver a new system (DashboardDB) to publish/share/discover the grid-database resources available in the EGI infrastructure, exploiting a social-oriented approach;
- Hydra (see section 2.2) is a file encryption/decryption service used to protect sensitive data stored on grid storage resources;

The GRelC service is delivered as a production-level service. The DashboardDB system has been recently released. To effectively reach the user community, a strong dissemination/communication plan is going to be implemented till the end of the project in close collaboration with NA2. The Hydra service is provisioned as a prototype for testing.

The operation assistance services include:

- The Life Sciences dashboard (see section 3.1), which aims at integrating community monitoring and administration tools into a single, comprehensive interface; and
- Tools for Life Sciences community (see section 3.1), that help managing a multi-Virtual Organization (VO) scale community.

The work reported falls in the tasks TSA3.2 ("Shared services and tools") and TSA3.4 ("Services for Life Sciences") of the EGI-InSPIRE DoW as follows:

- TSA3.2.1: Dashboards (section 3.1);
- TSA3.2.3: GRelC (section 2.1) and Hydra (section 2.2); and
- TSA3.4: Tools for Life Sciences community (section 3.1).

Beyond the EGI-InSPIRE project's own definition of a Heavy User Community, members of the Life Sciences area self-organized into a multi-grid users community named Life Sciences Grid Community (LSGC). The LSGC implements a Virtual Research Community (VRC). This document lists the services delivered or being developed by the Life Sciences VRC within EGI-InSPIRE. It does not provide an exhaustive list of all services developed by and available to the community outside of the EGI-InSPIRE project. The Life Sciences Grid Community (LSGC) wiki page [R4] is another important source of information for users from the LS community.

This document provides information on the development status of each service and the bootstrapping information needed for any user from the community who would like to make use of the services deployed.

2 SERVICES FOR END USERS

2.1 GRelC database access and integration service

The management of databases plays a crucial role in different scientific domains like Earth Science, High Energy Physics and Life Science. In the EGI context, the GRelC project has been supporting the Life Science community from several points of view through the analysis and implementation of new use cases, the user support, training and tutorial activities stemming from the biosciences domain. The key activity related to this task is the DashboardDB application [R18].

2.1.1 Service overview

This DashboardDB has been deployed and integrated into the official GRelC website [R15], so it can be easily accessed along with the other project information, software, news, rpms, etc. already available. The functionalities tested, validated and now available to the end users in the official release are: user registration, user profile management, grid-database registry, grid-database search and discovery, discussion groups (including group subscription and notification), grid-database rating and tagging.

Since the DashboardDB is now available online, the user has a direct way to test the functionalities offered by this system. The access can be performed through a web browser and no grid certificates are needed, since the system relies on a login/password based authentication. The user registration can be carried out online.

2.1.2 Service delivery status

The DashboardDB has been released at the end of Y2 [R18]. The system is now being updated and populated in terms both of grid-database services and grid-database resources. A strong dissemination plan is going to be implemented during Y3 to reach the user communities and make them aware about the features and opportunities offered by this new system. At the same time, the production release will be improved and fixed as new bugs will be discovered.

The delivered system consists of two parts:

- the *DashboardDB registry*;
- the *DashboardDB global monitoring*.

The **DashboardDB registry** aims at providing a web access interface to the database resources available in the EGI production grid. Such a "registry" complements the functionalities provided by the EGI Application Database [R10] and support the users using social network features, e.g. a message board on the DashboardDB platform (similar to shoutboxes found on public community websites), create discussion groups, rate existing resources, etc. The users can publish their own data sources and discover new ones already deployed through the search & discovery functionality. Since the registry has been deployed at the end of the Y2, it is now being updated with new data sources and grid database access services. The questionnaire prepared during Y1 will be sent again to the HUC to both (i) identify the available database resources (along with specific needs like the porting in grid of new databases) and to (ii) include them into the registry.

The **DashboardDB global monitoring** allows the monitoring of the GRelC instances deployed in EGI. Such a monitoring part complements the existing monitoring tools already available in the EGI environment providing a different perspective more focused on "grid-database management" services.

Since the the DashboardDB registry and global monitoring have been designed as two self-consistent components (they can be easily exported and embedded in other web contexts like a web gadget) they

are going to be publicized as two new EGI gadgets at the European level (www.egi.eu/user-support/gadgets).

Four permalinks are available to export each gadget both in a *secure* (login-password based) and *guest* (free access) way. They can be directly integrated into existing web application just using the following single HTML lines of codes.

1. Authenticated Permalink DashboardDB Registry:

```
<iframe
src="http://adm05.cmcc.it:8080/dashboardDB/monitoring/ProjectMini?request_locale=en&idProject=1&iframe=monitoring/ProjectMini%3Frequest_locale%3Den%26idProject%3D1" height="600" width="100%"></iframe>
```

2. Free Permalink DashboardDB Registry:

```
<iframe
src="http://adm05.cmcc.it:8080/dashboardDB/monitoring/ProjectMini?request_locale=en&idProject=1&guest=1&iframe=monitoring/ProjectMini%3Frequest_locale%3Den%26idProject%3D1%26guest%3D1" height="600" width="100%"></iframe>
```

3. Authenticated Permalink DashboardDB Global Monitoring:

```
<iframe
src="http://adm05.cmcc.it:8080/dashboardDB/grelc/ProjectRegistryMini?request_locale=en&idProject=1&iframe=grelc/ProjectRegistryMini%3Frequest_locale%3Den%26idProject%3D1" height="600" width="100%"></iframe>
```

4. Free Permalink DashboardDB Global Monitoring:

```
<iframe
src="http://adm05.cmcc.it:8080/dashboardDB/grelc/ProjectRegistryMini?request_locale=en&idProject=1&guest=1&iframe=grelc/ProjectRegistryMini%3Frequest_locale%3Den%26idProject%3D1%26guest%3D1" height="600" width="100%"></iframe>
```

The permalink feature has been exploited to integrate the DashboardDB application into the official GRelC website, under the tabs:

- "*Registry*" [R19] and
- "*Monitoring*" [R20].

From an end-user perspective, the GRelC project provides support through the official project website [R15] where users can find information about the status of the project, new releases, installation guides, software development kit, rpms, etc. The website represents the best way to bootstrap the activities related to the GRelC software, find news related to next events, available papers, tutorial sessions, new features, releases and so on. An integration plan of the gLite 3.2 compliant release of the GRelC software into the Italian Grid Initiative (IGI) release has been defined and started. A new repository at the IGI level has been setup for this task and will be available during the sumer.

Support is also provided through some tutorials available on the GILDA [R6] website where people can learn more about the GRelC service, the GRelC Portal, the Command Line Interface, and so on. Tutorials can be found on the GILDA Wiki [R7] and are organized to provide different skills according to different users' needs and requests. Basically, the GRelC service provides support to the HUC in terms of grid database management and in terms of grid metadata management. Most of the use cases already implemented in the past, fall in one of these two classes. In both the cases, it is possible to define common use cases that can act, at the same, time both as best practices and effective and tested architectural/usage patterns. Such knowledge can be easily exported in several contexts in a shortened time scale. In terms of grid database management and user support for the Life Science community, the solution adopted to access to the UNIPROT biological databank could be easily exploited to port in grid other data sources related to the Life Sciences domain. A major goal regarding the support for the LS community is to create in Lecce (SPACI node, Italy) "*a grid database node hosting several data banks addressing LS needs*". Concentrating at the same place many different biological databases (at this moment the UniProtKB/Swiss-Prot and GeneOntology have been deployed) is relevant for the LS community and could be also crucial to attract new users.

2.1.3 GRelC Service configuration, command line interface usage and DashboardDB access

The **GRelC service** is a stand-alone service. It relies on 1 rpm, which must be installed on a UI-like machine. The server keeps the whole status (grid-databases, users, etc.) into a system catalog (a relational database) running under the Postgresql RDBMS.

The server runs under the grelc UNIX user, using the host certificates of the machine. The target databases (both relational and XML-based), the system catalog and the GRelC server are usually on the same Local Area Network (it is not required to be on the same machine).

The GRelC server can be configured using one XML file containing two classes of information:

1) the system catalog configuration parameters (database connection string)
2) some GRelC server parameters (listening port, log-level, etc.)

```
<CONFIGURATION>
      <CATALOG>
            <DBMS>POSTGRES</DBMS>
            <DATABASE_NAME>dbaccess</DATABASE_NAME>
            <DBMS_HOST>127.0.0.1</DBMS_HOST>
            <LOGIN>postgres</LOGIN>
            <PASSWORD>postgres</PASSWORD>
            <DBMS_PORT>5432</DBMS_PORT>
      </CATALOG>
      <DAS_PARAMETERS>
            <PORT>18500</PORT>
            <SOCKET_TIMEOUT>2400</SOCKET_TIMEOUT>
            <ERROR_LOG>./error.log</ERROR_LOG>
            <LOG_LEVEL>3</LOG_LEVEL>
      </DAS_PARAMETERS>
</CONFIGURATION>
```

Default values (like the ones in the example above) are provided in the configuration file available in the rpm. More detailed instructions about how to setup a GRelC server are available in the grelc-dais-install-guide.pdf [R5].

The server can be managed through a rich Command Line Interface (CLI). The CLI provides a set of functionalities (secured through the adoption of the Grid Security Infrastructure) to interact with the GRelC server and to register a new grid-database, define user access policies, add/remove grid-users, etc. VOMS is supported but not mandatory due to an alternative local authorization mechanism available by default on the GRelC server and based on Access Control List (ACL). Additional information regarding the GRelC server administration are available in the grelc-dais-admin-guide.pdf guide [R5], whereas more detailed instructions about how to setup the VOMS-based authorization can be cound in the grelc-dais-authorization-guide.pdf [R5].

From an end-user point of view, a grid-database configured on a GRelC server can be queried by exploiting different grid-enabled queries (see grelc-dais-user-guide.pdf [R5]).

The grid-enabled queries are provided by the **GRelC client** rpm (available on the GRelC website) and can be easily installed on a gLite UI machine.

The most relevant clients to submit a query are:

- grelc-dais-query-memory

- grelc-dais-query-dime
- grelc-dais-query-stream

In the first case the recordset is retrieved "in memory", in the second one it is delivered as "a file" and finally in the third case it is sent in "streaming mode" (from a performance point of view the query stream if the most efficient one [R17]).

To give an example about the usage of these clients, here is the help of the grelc-dais-query-dime (the other 2 clients share almost the same options).

grelc-dais-query-dime [options]

-s, --server VAL	*Sets GRelC DAIS IP or hostname (default: localhost)*
-p, --port NUM	*Sets GRelC DAIS port (default: 18500)*
-D, --database VAL	*Sets grid database name*
-Q, --query VAL	*Sets query value*
-c, --chunk NUM	*Sets chunk dimension (default: 0 no chunk)*
-z, --gzip	*Compresses result file*
-t, --table	*Prints result in table format (default)*
-x, --xml	*Prints result in xml format*
-r, --raw	*Prints result with no format*
-f, --file VAL	*Print result in a file*
-H, --html	*Prints result in html format*
-h, --help	*Shows usage*

Some (non mandatory) options are related to the target GRelC server (-s, -p), other two (mandatory) are related to the target resource (-D, -Q); two more options concern the delivery of the resultset (-c, -z) and finally five non-mandatory options relate to the output format (-t, -x, -r, -f, -H).

In the following, few examples related to the grelc-dais-query-dime client submission are presented (the response of the query – the recordset - is not reported since not relevant in the examples):

1) query submission with compression
*grelc-dais-query-dime -s gandalf.unile.it -p 19000 -D grid_database -Q 'select * from table' -z*

2) query submission with chunking (chunk size = 50 tuples)
*grelc-dais-query-dime -s gandalf.unile.it -p 19000 -D grid_database -Q 'select * from table' -c 50*

3) query submission with chunking (chunk size = 100 tuples) and compression
*grelc-dais-query-dime -s gandalf.unile.it -p 19000 -D grid_database -Q 'select * from table' -c 100 -z*

Finally, the **DashboardDB** is a web-based application. It can be considered at a different (higher) level with regard to the GRelC server and the main goal is to create communities around the grid-database resources deployed in EGI. The target users are both grid and non-grid users. To reach a wider audience, the authentication is login/password based (no grid certificates – which could be potentially a barrier - are needed) and a user registration is also available online. Firefox and Safari are currently the supported browsers. The system has been designed and implemented taking into account social capabilities like rating or tagging a grid-database, joining a discussion group, post a comment.

The main goal is to create an environment for scientific databases where people can find/publish/share grid-database resources creating a community around them.

2.1.4 Future work

The future work includes two main activities: (1) extending the functionalities provided by the DashboardDB and (2) informing the users about this new tool publishing 'Scientific Databases'.

Concerning point (1) the DashboardDB registry will provide additional "social oriented" features (some of them now being tested) like:

> 1.1) creating multiple discussions associated to each grid-database (rather than just a single one),
>
> 1.2) including multimedia (image/video) content (rather than only text) into users' posts,
>
> 1.3) adding like/dislike and reply-to-post functionalities,
>
> 1.4) linking users.

whereas the DashboardDB monitoring one will include additional charts regarding:

> 1.5) a stronger Service-level view with monitoring capabilities;
>
> 1.6) the VOs authorized for each grid-DB service,
>
> 1.7) the number of published grid-DB resources for each grid-DB service over the time.

Regarding point (2) a strong dissemination/communication plan has been already defined in June (in close collaboration with NA2) and it is going to be definitely addressed through the following actions:

> 2.1) publicize the 2 gadgets on the EGI gadgets website;
>
> 2.2) create a new entry under 'Support Services' of the EGI website related to 'Scientific databases';
>
> 2.3) implement a communication plan including the preparation of a blog post, newsletter item and/or leaflet to attract new users (both end-users and grid-dabase service providers);
>
> 2.4) work with the NILs on identifying and linking scientific databases to EGI with GRelC;
>
> 2.5) send to the scientific user communities a revised version of the questionnaire prepared during Y1, to collect new requirements and needs regarding grid-database management and define/implement new use cases.

Since 2007, the GRelC software is a production-level grid-database service compliant with gLite. In this regard, a release plan towards the compatibility of this software with the EMI distribution will be defined by the end of PQ10. The deadline to implement the plan is the end of Y3.

2.2 HYDRA file encryption service

Sensitive data (related to individuals' health and possibly identifiable data) is often manipulated in the Life Sciences area. In an open environment such as a grid, only encryption of data through robust cryptography methods can ensure confidentiality of patient data. HYDRA is a secure file encryption service designed in the context of the EGEE series of project to fulfil that requirement.

2.2.1 Service overview

Hydra is a file encryption/decryption tool developed at CERN as part of the gLite middleware, enabling encryption of sensitive files stored on storage resources. It implements the Shamir's secret key sharing algorithm [R1], a key-based encryption solution where encryption keys are stored in a distributed key store, ensuring availability and robustness against attacks. An overview of the Hydra software is available here [R2].

The Hydra software is part of the regular gLite User Interfaces and Worker Nodes through the command line interface (glite-data-hydra-cli package).

2.2.2 Hydra Client configuration and usage

From the user's point of view, the service may be configured following two methods: a local-file based method, or through the service unique id published in the BDII. The latter is the preferred and most simple solution. To do so users should set the following variables in the user' profile:

```
export GLITE_SD_PLUGIN=bdii
export GLITE_SD_VO=biomed
```

The command-line interface is available under regular gLite User Interface hosts as binaries prefixed with "glite-eds-": glite-eds-chmod, glite-eds-getacl, glite-eds-rm, glite-eds-decrypt, glite-eds-key-register, glite-eds-setacl, glite-eds-encrypt, glite-eds-key-unregister, glite-eds-get, glite-eds-put.

A user with a valid proxy certificate usually interacts with the Hydra service through the glite-eds-put and glite-eds-get commands. The former encrypts and uploads a file to a Storage Element. Conversely the latter downloads and decrypts the file (if the user is authorized to access the file encryption keys).

The encryption keys access control is ACL-based. The glite-eds-get/setacl commands are used to control these rights. While the glite-eds-put/get commands coherently manipulate a file and its associated encryption key, an atomic operation on the hydra server can be performed through the glite-eds-key-register/unregister commands (to register/unregister new encryption keys) and glite-eds-encrypt/decrypt (to locally encrypt or decrypt a file). These interactions shall typically follow the examples below:

1. Create a key in Hydra, e.g. with id fmichel-id:

```
# glite-eds-key-register -v fmichel-id
A key has been generated and registered for ID 'fmichel-id'
```

2. Encrypt a local file:

```
# echo test > test.txt
# glite-eds-encrypt fmichel-id test.txt test.txt.encrypted
File 'test.txt' has been successfully encrypted
        with key 'fmichel-id'
        and written to 'test.txt.encrypted'.
```

3. Decrypt a local file:

```
# glite-eds-decrypt fmichel-id test.txt.encrypted text_decrypted.txt
File 'test.txt.encrypted' has been succesfully decrypted
        with key 'fmichel-id'
        and written to 'text_decrypted.txt'.
```

2.2.3 Service delivery status

The experimental Hydra service deployed within the first year of the EGI-InSPIRE project on a single gLite release 3.1 UI has been migrated to two gLite release 3.2 servers with the help of CERN experts from the EMI project. See [R3]. The service is usable for test purposes.

A new partner was identified to provision the third Hydra key store, in order to replace the server initially provisioned by the HealthGrid association, after it was dissolved. This partner, CNRS "Institut Pluridisciplinaire Hubert Curien" (IPHC), has taken this opportunity to start updating Quator scripts to deploy the current non-official release of the Hydra key store, that should be part of a later EMI2 delivery (not scheduled yet).

The Hydra service relies on the fact that the Hydra client software be (i) installed on all sites where Worker Nodes may be required to access the Hydra service (presumably all sites accessible to the LS

HUC VOs), or (ii) installed and published by means of runtime environment tags on those sites that wish to support the service. However, a survey has revealed that lots of production sites were misconfigured, not having deployed the Hydra client, having deployed an older version of the Hydra client, or publishing Hydra tags that are not consistent with the deployed client if any. Consequently, during this period, a negotiation was led with each site publishing Hydra tags, or providing Hydra client without tags, to clear off the situation.

Still, additional concerns hamper the deployment of the service in production: on one hand the Hydra client currently deployed on the production infrastructure is available only with gLite 3.2, which security support will terminate by September 2012. On the other hand, there is no planned date for the delivery of the Hydra client developed as part of the EMI middleware, and first tests show that its dependencies are incompatible with the gLite 3.2 release. Hence, along with resource providers choosing to migrate their Worker Nodes to EMI before the Hydra client is released, we can expect the number of appropriate WN supporting Hydra to decrease.

Therefore, in such an unclear situation, the service delivered today remains a test service mostly used for the validation of the functionality delivered and the testing of the deployment procedures.

2.2.4 Future work

A close follow-up of EMI release plans is also organized to ensure that the client be integrated in future EMI release and deployed as part of the standard distribution. This work is bound to the frequency of EMI release schedules though.

3 SERVICES FOR ADMINISTRATORS

3.1 Dashboard

Virtual Research Communities (VRCs) are a new concept in the EGI-InSPIRE user communities' organization. The middleware developed so far is predominantly relying on a separation of users through various Virtual Organizations (VOs). VRC management is a complex duty for which appropriate support tools still need to be developed. In particular, integrated tools are needed to provide a VRC-wise vision of the activity and facilitate the VRC administration, help VOs of the community to mutualise efforts and leverage common tools to avoid duplicating efforts.

3.1.1 Service overview

Among existing projects, the VO Operations Dashboard [16] (part of the EGI Operations Portal [14]) is a promising tool expected to become the daily tool for VO operation support teams. It covers several important operational needs from the LSGC point of view such as:
- The visualization and classification of alarms from monitoring sources (Nagios);
- The integration with the GGUS ticketing system to assist in the creation and follow-up of GGUS team tickets;
- Cross information with resource status from GOCDB;
- Support shifts team organization and takeover reports;
- As well as possible advanced features like the detection of flapping service or report builder.

Taking the biomed VO as a significant use case, the Life Sciences Grid Community (LSGC) directly interacts with the development team of the VO Operations Dashboard, suggests and discusses features, and provides feed-back as a beta-tester.

The Life Sciences dashboard intends to integrate various VO management services into a single portal. From the experience gained in operating Life Sciences VOs during the first two years of the project, some services have been deployed and are now used on a daily basis:
1. Community files management: monitor storage space consumed VRC-wise, anticipate problems of storage resources filling up, handle SEs decommissioning, file migration procedures.
2. Centralized view of VO resources that are currently not up and running (downtimes, not in production...)
3. Monitoring of computing resources availability (to complement current Nagios probes with alternative metrics)
4. Miscellaneous tools for facilitating daily follow-up of issues, manual checks etc.

3.1.2 Service delivery status

The tools described in 3.1.1 are documented and currently available from the biomed support team wiki. They were designed in a generic way, and may be reused by any VO inside and outside of the LSGC. So far, they are used daily by the biomed support team (see 3.2.1).

The tools are available for download from a github repository[1]. Some of them have been integrated into the biomed support team wiki, as a convenient way to assist the support team in its daily tasks. See examples at: http://lsgc.org/en/Biomed-Shifts:Index#STATUS_OF_BIOMED_RESOURCES

So far, wiki remains a convenient framework to integrate the tools of the LSGC dashboard as part of the overall documentation of the team. Somehow, the features described in the future work below shall need a more flexible development framework.

3.1.3 Future work

The following service are currently considered for future development and integration into the LSGC dashboard:

1. User management tools, covering users life cycle management and grid activity follow-up (see section 3.1 for details).
2. VRC-wide accounting, needed to deliver statistics at the VRC level. The current EGI accounting portal [R9] only provides per-VO accounting information.

In addition, the LSGC holds an ongoing discussion with storage resource providers and the HEP community to identify best practices in terms of community data management, and extend procedures and tools developed so far to more general procedures tailored for small communities like the LSGC VOs. A global effort to monitor the amount of computing resources used available is currently under investigation: its main objective is to detect possible bottle necks among computing resources, and eventually address this issue at different levels, for instance in job submission policies.

3.2 Life Science Virtual Research Community

Following the recommendations of the Life Sciences Grid Community (LSGC) board, effort was invested in the organization and the development of appropriate tools to better manage the users' community. The LSGC board organizes monthly phone meetings whose minutes are available at the community wiki [R4]. Most effort during the project first year has been invested in the set-up of community communication channels and technical assistance to the end users. The development of user management tools is planed to ease the task of VO administrators and VRC leaders.

3.2.1 Services delivered

Per-VO and VRC-wide mailing lists, initially maintained by the HealthGrid association, have been migrated to the GoogleGroups service. These communication channels are essential in maximizing the impact of the tools and services developed within the community, as well as negotiating and managing resources with the NGIs.

A redundant VOMS server has been provisioned and installed by the HealthGrid association for the biomed VO, which goal is to help shielding the users from downtimes of the main server. With HealthGrid being dissolved, the redundant VOMS operation will end in October 2012. A new partner needs to be identified to host the service after that date.

The LSGC receives support from the NGIs in term of grid resources and manpower. A technical support team of volunteer members from the most active users of the biomed VO [R11] assists the LS users. The support is organized in duty shifts, its functions are to monitor resources dedicated to the community, address problems reported either by the users or by the site administrators. The communication is ensured by the GGUS front-line support system. The technical support team

---

[1] https://github.com/frmichel/biomed-support-tools

actively probes most services for the proper VO operation, in particular using a dedicated SAM Nagios server [R12]. Additional tools, specifically developed by the team, help to proactively monitor well-known issues, thus anticipating service failures. Procedures have been defined to react to regular maintenance events such as resource decommissioning operations.

3.2.2 Future work

The design of a user management database has been delayed due to the end of the HealthGrid association. This database should facilitate liaising with hundreds of users registered in the affiliated Virtual Organizations. The database schema was specified, and state diagrams have been drafted. This service will interface to Virtual Organization Membership Service (VOMS) servers as well as the EGI applications database, to avoid replicating existing information. It will complement the VOMS and applications database with extra-information on the users and their affiliations.

The delivery of a redundant LFC server is postponed for now. This decision leans upon technical reasons discussed with LFC administrators, UCST, and HEP VOs. (i) The LFC proves to be very stable software. Therefore a redundant node would be more focused on dealing with hardware failures or catastrophic events. (ii) The node currently used is hosted in a secure environment, regularly backed up, thus building up a new one would not take a long time. (iii) A second LFC node could only be read-only (no active-active configuration has ever been attempted), which means that each users application would need to be specifically adapted. Therefore, so far, the conclusion is that the effort required is not worth the interest it would bring. However this could be considered again in the future as user applications may converge on common technical frameworks, which would make it easier to enforce common policies.

In parallel, the migration of the production VOMS node to the recent EMI2 version is currently tested. This migration will allow benefiting from new features that should facilitate usual operation tasks.

4 CONCLUSION

Within the context of the EGI-InSPIRE project, some services are provisioned for the exploitation by the Life Sciences Heavy Users Community: the GRelC database access interface and the Hydra file encryption services. This effort is complementary to the work endorsed by the Life Sciences Grid Community, which similarly aims at provisioning community-level services to ease grid applications development and grid exploitation [R4].

In addition, the emergence of Virtual Research Communities, which encompasses multiple VOs, triggers the need for new community monitoring and management tools, and proper interfaces to such tools.

The Life Sciences Grid Community covers a very wide area of health-related computational activities. An exhaustive list of services that would be useful for such a broad and heterogeneous community cannot be provided. However, the community has invested significant work in identifying the fundamental requirements for the underlying grid infrastructure and middleware. This effort resulted in a comprehensive list of core technical requirements that have been endorsed by the User Community Board. This list of requirements is indexed on the LSGC wiki [R13].

5 REFERENCES

R1	A. Shamir. "How to share a secret" in Communications of the ACM (CACM), vol. 22, pages 612-613, 1979.
R2	Hydra service overview. https://twiki.cern.ch/twiki/bin/view/EGEE/DMEDS
R3	MS607: Hydra service deployment on a multi-servers configuration, https://documents.egi.eu/document/327
R4	Life Sciences Grid Community, http://lsgc.org/en/LSGC:home
R5	GRelC documentation, http://grelc.unile.it/grelc_source/grelc-dais-all-guides.zip
R6	GILDA (Grid INFN Laboratory for Dissemination Activities), https://gilda.ct.infn.it/
R7	GILDA Wiki, https://grid.ct.infn.it/twiki/bin/view/GILDA/GRelCProject
R8	Nagios IT infrastructure monitoring tool, http://www.nagios.org/
R9	EGI accounting portal, http://accounting.egi.eu/egi.php
R10	EGI Application Database, http://appdb.egi.eu
R11	Biomed VO, http://lsgc.org/en/Biomed:home
R12	Biomed Nagios server, https://grid04.lal.in2p3.fr/nagios (accessible to EGI certificate owners only)
R13	LSGC list of requirements, http://lsgc.org/en/LSGC:home#List_of_requirements_submitted_to_EGI
R14	EGI Central Operation Portals, https://operations-portal.egi.eu
R15	The GRelC Project, http://grelc.unile.it/home.php
R16	VO Operations Dashboard, https://operations-portal.egi.eu/voDashboard
R17	S. Fiore, A. Negro, S. Vadacca, M. Cafaro, M. Mirto, G. Aloisio, "Advanced Grid DataBase Management with the GRelC Data Access Service", IEEE Proceedings of the 5th International Symposium on Parallel and Distributed Processing and Applications (ISPA07), August 29-31, 2007, Niagara Falls (Canada), LNCS 4742, pp. 683-694.
R18	DashboardDB, http://adm05.cmcc.it:8080/dashboardDB/Home.action
R19	GRelC website - "Registry", http://grelc.unile.it/registry.php
R20	GRelC website - "Monitoring", http://grelc.unile.it/monitoring.php