EGI-InSPIRE

HUC CONTACT POINTS AND THE SUPPORT MODEL

EU DELIVERABLE: MS601

Document identifier:	EGI-SA3-WP6-MS601-v1 1.odt
Date:	**21-July-2010**
Activity:	**SA3**
Lead Partner:	**EGI.eu**
Document Status:	**In progress**
Dissemination Level:	**PUBLIC**
Document Link:	https://documents.egi.eu/document/

Abstract

This milestone delivers the contact points and the Support Model for the Heavy User Community (HUC).

Delivery Slip

	Name	Partner/Activity	Date
From			
Reviewed by	**Moderator:** **Reviewers:**		
Approved by	**AMB & PMB**		

Document Log

Issue	Date	Comment	Author/Partner
0.1	08-July-2010	First draft	Vera Hansper/CSC
0.2	13-July-2010	Second draft	Vera Hansper/CSC, Ivan Degtyarenko/CSC
0.5	14-July-2010	Updated with HEP input	Jamie Shiers/CERN
0.6	15-July-2010	GrelC, GridWay, DRMAA, Kelpler	Marcin Plociennik, Giovanni Aloisio, Antonio Gómez
0.7	07/20/10	Updated GrelC, added DIANE, Hydra	Viet Tran, Johan Montagnat, Sandro Fiore
0.8	07/21/10	General editing	Vera Hansper/CSC
1.0	07/23/10	Restructured document, above contributions changed to appendices	Vera Hansper/CSC
1.1	08/24/10	Reviewer comments incorporated	Pekka Lehtovuori/CSC

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 the 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 the ESFRI projects. 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 − structured international user communities − that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

TABLE OF CONTENTS

1. INTRODUCTION....4
2. SUPPORT MODEL....6
3. CONCLUSIONS....7
4. APPENDIX....7
DOCUMENT AMENDMENT PROCEDURE....12
TERMINOLOGY....12
5. REFERENCES....12

1 INTRODUCTION

1.1. PURPOSE

The purpose of this document is to provide contact points for the Heavy User Community (HUC) and to create a support model for this community.

The communities identified as Heavy Users Communities (HUCs) within this proposal are:

- High Energy Physics (HEP)
- Life Sciences (LS)
- Astronomy and Astrophysics (A&A)
- Computational Chemistry and Materials Sciences and Technologies (CCMST)
- Earth Sciences (ES)
- Fusion (F)

1.2. INVOLVED PARTNERS (CONTACTS / MANAGERS)

Germany	KIT-G	Horst Schwichtenberg, (horst.schwichtenberg@scai.fraunhofer.de)
Spain	CSIC	Isabel Campos Plasencia (isabel@campos-it.es)
Spain	CIEMAT	francisco.castejon@ciemat.es (Paco Castejón) antonio.gomez@ciemat.es (Antonio Gómez)
Finland	CSC	Pekka Lehtovuori (Pekka.Lehtovuori@csc.fi)
France	CNRS	Johan Montagnat (johan@i3s.unice.fr)
Italy	INFN	Laura Perini (laura.perini@mi.infn.it)
Ireland	TCD	Brian Coghlan (coghlan@cs.tcd.ie]), John Walsh (John.Walsh@cs.tcd.ie)
Poland	ICBP (PSNC)	Pawel Wolniewicz (pawelw@man.poznan.pl), Marcin Plociennik (marcinp@man.poznan.pl)
Slovenia	IJS	Andrej Filipcic (andrej.filipcic@ijs.si)
Slovakia	UI SAV	Ladislav Hluchy (ladislav.hluchy@savba.sk)
UK	EMBL	Peter Rice (pmr@ebi.ac.uk)
CERN		Jamie Shiers (Jamie.Shiers@cern.ch), Maria Girone (Maria.Girone@cern.ch)

1.3. USER SUPPORT CONTACTS

TSA3.2.1:	Dashboards	Jamie Shiers (Jamie.Shiers@cern.ch)
TSA3.2.2:	Applications	Andrej Filipcic (andrej.filipcic@ijs.si)
TSA3.2.3:	Services	Giovanni Aloisio, deputy:Sandro Fiore (sandro.fiore@unile.it) for GrelC (SPACI)
TSA3.2.4:	Workflows and Schedulers	Francisco Castejon (francisco.castejon@ciemat.es), Antonio Gomez (antonio.gomez@ciemat.es) (CSIC), Tapani Kinnunen (tapani.kinnunen@csc.fi) (CSC), Marcin Plociennik (marcinp@man.poznan.pl) (PSNC)
TSA3.2.5:	MPI	Enol Fernandez (enolfc@ifca.unican.es), John Walsh, (John.Walsh@cs.tcd.ie) (TCD), David O'Callaghan (david.ocallaghan@cs.tcd.ie), Antonio Lagana' (CCMST) (INFN)

1.4. DESCRIPTION OF TOOLS

• **TSA3.2.1 Dashboards**: Dashboards provide a generic framework to monitor sites and their services within a VO using tests specific to that community. Dashboards have emerged from within the HEP community, and are now being adopted by the LS community, to monitor their resources.

• **TSA3.2.2 Applications**: The **GANGA** and **DIANE** tools are both part of the EGEE RESPECT programme which recognises software that builds on top of the gLite platform. Although initially developed for the HEP community, these tools have now gained traction in other communities, as they provide simple environments to manage large collections of tasks. The requirements of these tools will be integrated into the work plan.

• **TSA3.2.3 Services**: **HYDRA** and **GReIC** are services that have emerged from a single community and which show potential for adoption in other communities. HYRDA is an encrypted file storage service. In HYDRA encryption key for the files is securely stored on distributed servers in order that storage elements can be used to store confidential data. Secure datamanagement is critical service for the medical community. The GReIC service provides uniform relational and non-relational access to heterogeneous data sources and is currently being used to support bioinformatics and Earth Observation Systems.

• **TSA3.2.4 Workflow and Schedulers**: These tools are critical in integrating complex processes, generally involving multiple data sources and different computational resources, as needed within many disciplines. **SOMA2** is a web-based workflow tool used for computational drug design and general molecular modelling. **TAVERNA** is used extensively by the bioinformatics community. The combination of the Kepler workflow engine and the Migrating Desktop platform are used by the Fusion community to run workflows requiring visualisation and interactive access on gLite and UNICORE-enabled resources. For simpler workflows and metascheduling scenarios the **GridWay** system is used by the Fusion community. Effort is provided to maintain the integration of these tools with the different systems.

• **TSA3.2.5 MPI**: Support for parallel computing (MPI) applications are critical for many user communities but the integration of this capability into the general infrastructure has been difficult. This task will focus on the improvement of the core services and software needed to support MPI, while engaging with two representative user communities (CCMST & Fusion) to ensure that the offered support meets their requirements.

.

2. SUPPORT MODEL

2.1 USING EXISTING TOOLS

A variety of tools are available to the HUC and support for these is generally through a combination of submitting a GGUS[i] ticket and having the VO manager act on the ticket.

The advantage of using GGUS as a communication method for alerting the VO to a problem is that the problem can be allocated and tracked appropriately.

The end user communicates with the VO manager via the GGUS ticket, who then has the responsibility of forwarding the issue to the correct site or administrator. If necessary, the VO manager could also forward it to the appropriate tools support unit.

2.2 PREPARATION FOR SUPPORT MODEL

A number of details will need to be set up to prepare the HUC for the envisaged support model. These will include

- collecting the VO names and the task descriptors
- collecting the VO manager contacts and their duties
- ensuring that the VO managers have the correct rights in the GGUS system, and that the correct support units (SU) exist in GGUS.

End users may need to apply for an account to access GGUS.

2.3 TRAINING FOR SUPPORT MODEL

As many end users are already familiar with GGUS as a system, it is anticipated that training in the use of GGUS would be minimal. There will be a need to train new users, and training can often be done by an experienced user, via online help and tutorials. There should be a possibility to train at EGI technical forums.

Training of the VO managers for their duties and responsibilities is the responsibility of the task manager for that VO.

2.4 DEVELOPER SUPPORT

Developers should use the savannah bug tracking tool[ii] to track issues. It would be best that the VO manager submits feature requests for any tool so as to minimise duplication of the effort. Users wishing new features, or encountering issues with a tool should create a GGUS ticket which the VO manager can then use to submit a bug report or similar.

3. CONCLUSIONS

This is a living document to describe the heavy user community (HUC) services and contact points. Document will be updated constantly and new versions will be available for each project year (PM1, PM 13, PM 25). This document lists the HUC services and contact information in the beginning of the EGI Inspire), more up to date information will be always available in the wiki at https://wiki.egi.eu/wiki/WP6:_Services_for_the_Heavy_User_Community.

4. APPENDIX

The following provides some extra information about the Tools used by the HUC.

DASHBOARDS:

The Dashboard system provides multiple applications for monitoring sites and applications.. Depending on the information source, some of these are shared by several virtual organizations (VOs) whereas others are VO-specific. There are 16 different categories, 1 for the Dashboard framework and 15 for various Dashboard applications.

Bugs and feature requests for the system are tracked through Savannah. In addition to the Savannah system there is a dashboard support list where users can ask their questions or request help.

The dashboard team organizes regular tutorials for users of the dashboard applications and takes part in the VO tutorials for Grid users.
The link to the material of the user tutorial sessions can be found here.

GANGA:

New feature and development requests for Ganga are tracked via Savannah (via items called “Feature Requests”). This is also used to do a basic prioritization of new features and follow up the evolution of the tool. The priorities are discussed during the weekly Ganga meetings. This link points to the open feature requests in Savannah.

Presently there are ~60 open feature requests and ~260 closed ones logged in the system (for reference: there are currently ~70 open bug reports and ~960 closed ones). Some special views (display forms) have been developed by the Ganga project in order to track the “originator” of a request (user community) and map them against a given (or future) release. In this view one can see bug reports (with corresponding fixes for closed reports) and feature requests alike. N.B. the “originator” of a feature request is only an approximate indication of the request's area of impact. For example, certain functionalities prototyped in ATLAS or LHCb may be re-factored into the Core. The same applies to bug fixes: for example, a bug found in ATLAS may become critical because of a potential big impact on LHCb.

DIANE:

DIANE provides a reliable execution tool for a distributed simulation in a master-worker style. The scheme has proved to be very efficient for parametric study applications. It uses Ganga as an execution backend. The lastest version is 2.0-beta20.

Bugs and feature requests are tracked through Savannah https://savannah.cern.ch/bugs/?group=diane. There are 12 open items of 15 total. Supports are also provided via email with developers.

There is a quick tutorial at https://twiki.cern.ch/twiki/bin/view/ArdaGrid/DIANETutorial. Other documents, including an installation guide and technical details are also available at the DIANE main page http://cern.ch/diane.
Demonstrations of DIANE have been made at various past events, including EGEE conferences and User forums, where were active discussions occur between the developers and users.

HYDRA:

Hydra is a file encryption/decryption tool developed as part of the gLite middleware. Hydra is a special secure metadata catalog designed to hold encryption keys. The Hydra functionality is accessible in the regular gLite UI through the command line interface. Hydra may be deployed as a single key store or as a distributed key store, implementing the Shamir's secret key sharing algorithm, for improved availability and higher robustness against attacks.

The installation and deployment procedures for the Hydra keystore are described in this gLite documentation:

`http://glite.web.cern.ch/glite/packages/R3.0/R20060502/doc/installation_guide_3.0-2.html#_Toc135537608`

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 Hydra catalog will be deployed within the first year of the EGI-InSPIRE project as a service for the life sciences community. Support for the Hydra software is available through the GGUS global gLite support.

GREIC:

The GRelC service is a grid database management service aiming at providing access and management functionalities related to relational and non-relational databases in a grid environment. The GRelC project provides a website (www.grelc.unile.it) where users can find several information about the status of the project, new releases, installation guides, software

development kit, rpms, etc. ~~Since 2007,~~ the project provides a mailing list (grelc-user@sara.unile.it) where people can ask support in terms of new features, use cases, gridification of existing databases, tutorial material, etc.

In terms of documentation, support is and will be provided through updated material related to:

- the GRelC service deployment (for system administrators)
- the GRelC service administration (for service administrators)
- the GRelC service security insights (for service administrators
- the GRelC service access (CLI for end users)
- the GRelC service access (SDK for developers).

All of these documents can be found at the following link: http://grelc.unile.it/documentation.php

Regarding C and Java SDK two online documentation guides are available:

1. http://grelc.unile.it/docs/grelcC/ (for C developers)
2. http://grelc.unile.it/docs/grelcjproxy/ (for Java developers)

Additional information and support will be provided in the Italian Grid Initiative (IGI) web site where new contents about the GRelC service are now being added.

Support is also provided through some tutorials available on the GILDA website where people can learn more about the GRelC service, the GRelC Portal, the Command Line Interface, and so on.

Tutorials can be found at the following link:

https://grid.ct.infn.it/twiki/bin/view/GILDA/GRelCProject

and are organized in order to provide different skills as could be needed according to different users' requests. These links will be updated when needed and extended/improved taking into account comments coming from the user community.

Requests about new features will be prioritized taken into account:

7. the impact of the new feature on the specific application that is requiring it (low, medium, high);
8. the impact of the new feature at the VO level (are there any other potential users in the same VO/domain-specific community? How many?);
9. the impact of the new feature at the community level (are there any other VOs which could be potentially attracted by this new feature? Which one?).
10. the impact of the new feature at the GRelC service level (how well does the new feature fit in the overall service picture and the planned roadmap?);

In the last years the GRelC Team has regularly organized several tutorials to disseminate information about the GRelC service and attract people and communities.

Last tutorials on the GRelC service has been held at the following three conferences in 2009:

1 International Conference for Internet Technology and Secured Transactions , IEEE, (ICITST 2009) - November 9-12, 2009, London, UK

2 Grid and Pervasive Computing 2009 Conference & Swiss Grid School (GPC09) – May 8, 2009, Geneve, Switzerland.

3 Parallel and Distributed Computing and Networks, PDCN 2009 – February 17, 2009, Innsbruck, Austria.

Moreover we plan to organize new tutorials in the context of the EGI-Inspire project, as we did in 2008 in Istanbul at the EGEE08 Conference:

- "Tutorial on GRelC" - Introduction to gLite & RESPECT tools at EGEE'08 conference" - September 20-21, 2008 - Istanbul, Turkey.

User requirements will be gathered starting from specific use cases defined with the user communities. Some application-domain use cases needing database access and management functionalities in a grid environment will be defined with the user communities to help them in implementing new scenarios or extending the existing ones.

What it is expected is to define telco and/or F2F meetings with the user communities to start such a crucial process. In the Earth Science community, new use cases will extend the work that has been done in the EGEE project, in the context of the Climate-G VO/testbed. Some updates regarding the setup of this environment have been already sent to the people involved in this activity in July. In the Climate-G use case, several GRelC instances are right now running at several sites: SPACI (Italy), Fraunhofer-SCAI (Germany), CNRS-IPSL (France) and University of Cantabria (Spain). A metadata harvesting scenario has been already identified to integrate metadata information distributed among the four sites. A F2F meeting will be scheduled in September with Earth Science people, to better define and collect the requirements coming from this scenario to start the implementation phase in the near months.

SOMA2:

SOMA2 is a web browser-based workflow environment for computational drug design and general molecular modelling. The purpose of the SOMA2 environment is to provide users with easy access to computational tools. SOMA2 hides all technicalities related to execution of scientific applications in complex computing facilities allowing users to focus on their actual scientific tasks.

Currently CSC provides a website (http://www.csc.fi/soma) where users can access SOMA2 related material. The provided material consists of general SOMA2 information, a detailed user manual with several screenshots and tutorials. For developers, the SOMA2 website provides source code and technical documentation. The website also provides access to SOMA2 mailing lists (soma2discussion@postit.csc.fi, soma2updates@postit.csc.fi) which are aimed for general SOMA2 related discussions, including support requests for end users and developers, bug reports and feature requests.

TAVERNA:

Currently, information for TAVERNA is available from:

http://www.taverna.org.uk/about/contact-us/

SUPPORT FOR PARALLEL COMPUTING (MPI)

As part of the MPI related tasks of user support, the recommendation document from the EGEE MPI Working Group has been updated following the middleware developers suggestions. The new recommendation maintains the existing user requirements for executing parallel applications without creating clashes in the job description attributes that could lead to confusion or misinterpretation by the developers. This recommendation is now final and is available at http://grid.ie/mpi/wiki/WorkingGroup/

The MPI Support Activity has participated at a training event at the Grids & e-Science 2010 course (Valencia, Spain). Training material from the course is available for download.

SUPPORT MODEL – SERVICES FOR HIGH ENERGY PHYSICS

The support model for High Energy Physics will be described in the document MS603 – Services for High Energy Physics.

GRIDWAY

User suuport for GridWay has three main parts: the support for the GridWay platform, the support for the developments using GridWay and the support to the developments using the DRMAA API.

The support for the GridWay platform and developments includes:

1. GridWay Installation Guide. This guide contains general information about the installation of GridWay on a grid infrastructure.
2. GridWay Configuration Guide. This document details how to customize GridWay.
3. GridWay EGEE Configuration. It describes how to configure GridWay in a EGEE-based infrastructure.
4. GridWay User Mailing List. This list is used to get information about problems with the installation as well as problems with the GridWay commandline interface. There is also a mailing list archive to search for past issues.
5. The bugs are tracked through the GridWay tracking system.

DRMAA

1. DRMAA bindings for C and JAVA. They describe all the options of the API.
2. Some examples of DRMAA are provided for C and JAVA. These examples show the typical use cases of DRMAA.

3. DRMAA mailing list and archived messages from the list.

KEPLER

User support for Kepler can be seen as two separate parts: support for the Kepler main platform and engine, and support for the Grid extensions and related workflows.

The Kepler core platform support includes:

1. Kepler User FAQ , which contains information about the project, getting started, and common usage questions and problems;
2. Kepler User Documentation, which includes a getting started guide, an actor reference, and a complete user manual;
3. Kepler User Mailing List, which is used to share questions and disseminate information among Kepler users and Kepler team members. One can also peruse archived messages from the list;
4. Kepler bug base. For a bug report in the core software, or an idea for a new feature or enhancement, there is Kepler's Bugzilla bug base.

In terms of the bugs and issues related to support the Grid/HPC access related actors and workflows:

1. Bugs and feature requests are tracked through JIRA
2. Additionally, there is training material available from the different courses which have been run.

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

TERMINOLOGY

A complete project glossary is provided in the EGI-InSPIRE glossary:
http://www.egi.eu/results/glossary/.

5. REFERENCES

[i] GGUS: https://gus.fzk.de/pages/home.php

[ii] Savannah: https://savannah.cern.ch/bugs/?group=dashboard