GES 2007 German e-Science

Available online at http://www.ges2007.de



# DEISA and D-Grid: using UNICORE in production Grid infrastructures

M. Rambadt[1], R. Breu[1], L. Clementi[2], T. Fieseler[1], A. Giesler[1], W.Gürich[1], P. Malfetti[2], R. Menday[1], J. Reetz[3], A. Streit[1]

[1]Central Institute for Applied Mathematics,
Forschungszentrum Jülich GmbH
D-52425 Jülich, Germany

[2] CINECA
Via Magnanelli 6/3,
40033 Casalecchio di Reno, Italy

[3] Rechenzentrum Garching
Max-Planck-Institute for Plasmaphysics
D-85748 Garching, Germany

*Email:*
{m.rambadt,r.breu,t.fieseler,a.giesler,w.guerich,
r.menday,a.streit}@fz-juelich.de,
l.clementi@cineca.it
johannes.reetz@rzg.mpg.de

*phone:* (+49 2461) 61 4340, *fax:* (+49 2461) 61 6656

**Abstract**

DEISA is a consortium of leading national supercomputing centres in Europe that is operating and enhancing a persistent, production quality, distributed supercomputing environment with continental scope in Europe. D-Grid is the central German Grid initiative that builds up and operates a sustainable Grid infrastructure in Germany and establishes methods of e-science in the German scientific community. To integrate their resources, both the DEISA and D-Grid communities have adopted the most advanced Grid middleware systems and applications currently available. D-Grid uses UNICORE, Globus and LCG/gLite for job submission. The DEISA consortium decided to embrace UNICORE as a job submission interface for the DEISA Grid infrastructure. Both projects focus on the production use of Grid middleware. This paper describes the experiences with UNICORE as Grid middleware in e-infrastructures comprising systems in production use.

## 1 Introduction

Since “The Grid: Blueprint for a New Computing Infrastructure" [2] many national and continental Grid projects around the world exist and the number of funded projects even seem to increase every year. While in the beginning the projects concentrated on building a working Grid middleware, today we observe more and more projects targeting production ready Grid environment for scientists. Among others this production scope is defined in DEISA and D-Grid, too.

The DEISA (Distributed European Infrastructure for Supercomputing Applications) project [1] started in May 2004 to provide a persistent and production quality, distributed supercomputing environment. The members of the consortium wish to improve the level of exploitation of their systems and, at the same time, to provide a higher Quality of Service to the users, by being able to offer them a larger joint resource pool [2]. When building such an infrastructure the DEISA partners considered several applications and middleware technologies that are providing the functionalities necessary to integrate their high-performance computing systems.

The DEISA consortium decided to use UNICORE (UNiform Interface to Computing REsources) [3] to establish a Grid infrastructure. UNICORE is one of the leading Grid middleware systems used in for production usage at a number of supercomputing centres and Grid infrastructures [4]. It hides the complexity of the underlying systems and architectures providing users with a seamless environment to work in and it uses a single sign-on mechanism based on X.509 certificates from a Public Key Infrastructure (PKI).

D-Grid [5] started as the national German Grid initiative in September 2005 combining seven community projects dealing with astronomy, climate research, high energy physics, engineering, medical research, humanities, and energy meteorology. In D-Grid the available middleware is Globus [12], LCG/gLite [6] and UNICORE. Thus the scientists are free to choose the best middleware for their application or just the middleware they are familiar with. This paper describes the experiences with UNICORE in production within DEISA and D-Grid and the features of a Grid middleware to meet the user requirements.

## 2 UNICORE

UNICORE provides a seamless interface for preparing and submitting jobs to a wide variety of heterogeneous distributed computing resources and data storages. It supports users for running scientific and engineering applications in a heterogeneous Grid environment.

The UNICORE software was initially developed in the UNICORE and UNICORE Plus projects [4,9] funded by the German Ministry of Education and Science (BMBF) until the end of 2002. After that, its functionalities and its robustness were enhanced within several EU-funded projects, for example, EUROGRID [6] and OpenMolGRID [7]. Since 2004, several supercomputing centres are employing UNICORE in production.

In UNICORE every job is represented by a Java based abstract job formulation, the so-called Abstract Job Object (AJO). This gives the user the possibility to prepare jobs on an abstract level without having to know specific details of a particular target system. With the abstract formulation, the job can be submitted to different target architectures running different batch schedulers without changes.

### 2.1 UNICORE components

UNICORE possesses a vertically integrated architecture. It provides both client and server components. In the current, production ready version, UNICORE 5, the server-side consists of the Gateway, Network Job Supervisor (NJS) including an Incarnation Database (IDB), the UNICORE User Database (UUDB), and the Target System Interface (TSI). All components are written in Java and Perl allowing UNICORE to be installed on a large variety of operating systems.

The user only has to install the UNICORE client, and apart from some simple location configuration, has to know little regarding the functionality of the server components.

The UNICORE client GUI is used for the preparation, submission, monitoring, and administration of complex multi-site and multi-step jobs. It provides the user with an extensible application support, resource management of the target system, and integrated security mechanism.

Every submitted abstract job request (AJO) is signed using the personal X.509 certificate of the user. Thus, other UNICORE server components can perform authentication and authorisation relying on the public key infrastructure (PKI) in use. The NJS translates the abstract job information to a concrete job which is executed on the target system then. The client also provides data management and transfer functionality through an intuitive GUI.

## 2.2 Monitoring of UNICORE servers with SIMON

For production Grids it is crucial to monitor the functionality of all involved components. Administrators should always be aware of problems before the user is. Therefore, DEISA developers implemented SIMON (Site Monitor for UNICORE Grids). SIMON is based on implementations coming together with DESHL (see section 2.3) from the DEISA JRA7 group. It monitors the state of a UNICORE based Grid deployment by executing user-defined test suites periodically and automatically against selected sites. SIMON provides both availability and functionality tests of the UNICORE server components.

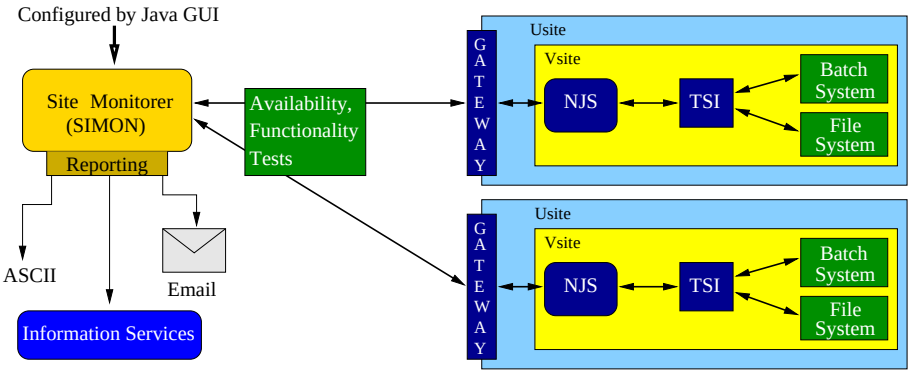


**Fig. 1.** Schema of SIMON functionality

Functionality tests include job execution monitoring as well as file imports to remote file systems. SIMON is easy to install and to configure via a Java based graphical user interface and thus is very flexible to add UNICORE sites or extend the parameters of an already existing site. SIMON records both expected and unexpected outcomes. That information can be used to build detailed reports regarding the health status of the Grid. The test results might be extracted and prepared in various final presentation formats or for instance they can be emailed to the UNICORE administrators. SIMON is used in both DEISA and D-Grid to monitor the UNICORE components. Currently this is in a pilot phase, but will move to full production status very soon.

### 2.3 DESHL (DEISA Services for the Heterogeneous management Layer)

DESHL is a set of tools which allow users and their applications to manage batch jobs and data across a computing Grid. DESHL offers a command line environment to users. The option of having command line access to Grid infrastructures is extremely attractive for many users, but often not at the expense of the graphical option, and DESHL nicely compliments the full featured UNICORE graphical client. The development of DESHL takes place within the JRA7 activity of the DEISA project and whilst DEISA users are currently the main DESHL users, there is no impediment preventing the software being used on other infrastructures, such as D-Grid.

DESHL is currently used across the UNICORE 5 based infrastructure of DEISA, and does not require any additional software installation or configuration on the server-side, on top of the normal UNICORE installation and configuration. Thus, the seamless environment enjoyed by a GUI UNICORE user is also available to the DESHL user.

In addition to the command line API support DESHL can be used from within applications through an implementation of the SAGA API. Finally, DESHL is distributed with excellent documentation and a graphical installer.

Two key software abstraction layers further enhance the portability of DESHL. From the implementation perspective, the Roctopus [11] API is used to isolate the specific characteristics of a particular Grid middleware, and this will be exploited in future to provide DESHL support for Web-Service based UNICORE 6 Grid deployments. From an application developer and command line user perspective, file and job management use the emerging SAGA (Simple API for Grid Applications) [16] Grid standard. This includes using SAGA directives as the basis of the job description document. The programmer API which users can use within their own applications is based on the SAGA API. It is noteworthy that DESHL was the first available Java implementation of the emerging SAGA standard.

## 3 Production environment in D-Grid and DEISA

A Grid middleware can only be as powerful as the underlying software, hardware and network is. So it is essential to achieve a smoothly functioning combination of all aspects. How this fits into the production character of D-Grid and DEISA will be shown in the next sections.

### 3.1 D-Grid user management infrastructure

In order to make Grid resources available to users via UNICORE or other middleware, a Grid infrastructure is required. For this, a point of information comprising a user portal and a provider portal has been installed for D-Grid users and resource providers. Users obtain the information to get a user Grid certificate, to register for a Virtual Organization (VO), to install and to use Grid middleware clients (i.e. the UNICORE client). Certificates can be obtained from Registration Authorities (RAs) run by all partners of the core D-Grid, by the DFN and by GridKA. The registration of users is done using VOMRS [13]. For each community and for the core integration project an own VO with an own instance of a VOMRS server has been configured. Users register to the VOMRS server of the corresponding VO. The UNICORE client is provided in a version which is specially pre-configured for the use in D-Grid.

For resource providers information is supplied to install the Grid middleware and to integrate the resources into the D-Grid infrastructure. The integration comprises the registration of a resource to the D-Grid resource management system, where a database with information about all resources is stored. To help the local administrator in managing D-Grid users at his site, the mapping of certificate DNs (Distinguished Names) to Unix accounts for D-Grid users of all VOs which are allowed to access these resources, can (regularly) be generated by a script which contacts the D-Grid resource management system. The information concerning which VOs have access to a resource is stored in the database of the resource management system. This information is used as a pointer to find the corresponding VOMRS database which obtains the information about users who are allowed to access the resources of the VO. Additionally the resource management system provides the DNs of the Grid server certificates of the UNICORE Network Job Supervisor (see section 2.1) of other installations within D-Grid. The information about user mapping together with the DNs of NJS server certificates are used to update the UNICORE User Data Base (UUDB) of the local installations.

## 3.2 DEISA user management infrastructure

To achieve a higher level of interoperability between the different resources, the DEISA partners decided to harmonise their user management systems and to establish a DEISA user administration system by deploying a distributed network of LDAP [15] servers. They are used to propagate information about DEISA users from the user's home site to all the partner sites. A standardisation of the naming schema for DEISA users and the assignment of site-specific ranges of UIDs and GIDs ensure that DEISA user accounts are replicable on every system belonging to the DEISA infrastructure.

A user who wants to use DEISA resources needs to apply for an account only at his home site. The user record information (user name, UID, GID, the subject of his certificate, etc.) propagates via LDAP from his home site to all the other DEISA sites.

Every night the local UUDB is updated automatically with the new user information coming from LDAP. As an improvement of the UNICORE authorization system the DEISA consortium requested to implement a modification in the UUDB internal management. The standard UUDB implementation maps the complete public part of the user's certificate to a user's account, while the modified DEISA UUDB checks only whether the Distinguished Name (DN) of the certificate used to sign the UNICORE Job (see section 2) is present in the UUDB. With support of the UNICORE developers, the implementation of the UUDB authorisation mechanism has been adapted accordingly. A new release of the UUDB also allows the mapping of one DN to more than one user ID and/or project. This is interesting for users working in several projects and who do not want to apply for a new certificate each time.

Nearly all the resources of the DEISA infrastructure are now integrated also by means of a shared file system (fig. 2). GPFS-MC (General Parallel File System-Multi Cluster) [10] allows achieving a transparent high-performance data access over the Wide Area Network.

When submitting a job, UNICORE creates a temporary working directory (called USPACE) at the target system where, among others, batch scripts, input, output and error files are placed. At first, DEISA partners did not require a common path for the USPACE. It was simply located on a site local file system. However, when submitted UNICORE jobs are to be migrated to

other clusters by the Multi-Cluster LoadLeveler, the USPACE at the originating cluster needs to be transparent.

As a solution, the different partners have decided to configure UNICORE in order to use a common USPACE path on the GPFS-MC. In this way, UNICORE jobs submitted to the homogeneous super-cluster (see the dotted lines in fig. 2) have always a consistent reference to the USPACE and thus to the files needed by the NJS at the originating site for monitoring the job status and fetching the output. The implementation of this solution required the modification of some TSI scripts.

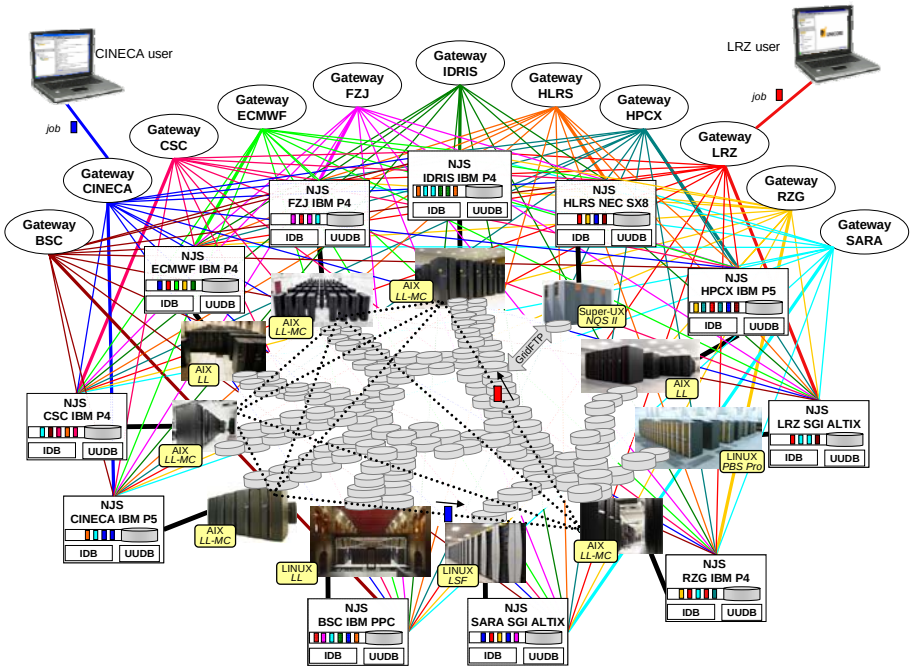


**Fig. 2.** Schema of the DEISA UNICORE deploymen. The dotted lines connect the sites of the homogeneous super-cluster. These sites employ the multi-cluster LoadLeveler that allows migrating jobs directly from site to site.

## 4 Experiences from using UNICORE in production Grid infrastructures

For more than three years now, UNICORE has been used in production in several European universities, companies and research centres. Mid 2005 the DEISA Extreme Computing Initiative (DECI) was launched aiming at enabling new challenging supercomputer applications. This initiative was set up to enable a number of those “grand challenge” applications in all areas of science and technology. These leading, pioneering applications depend on the DEISA infrastructure because they are dealing with complex, demanding and innovative simulations, and benefit from the exceptional resources provided by the Consortium.

### 4.1 User requirements to a Grid middleware

From the DEISA and D-Grid experiences first of all users need a working Grid middleware and infrastructure. The software has to be stable and available preferably all the time. In DEISA and D-Grid the UNICORE developers established fall back options e.g. for the UNICORE Gateway. As the central entry point for every user request, the Gateway plays a crucial role. In DEISA every partner site has installed a Gateway and all target systems are

available via all those Gateways. So, if one Gateway should crash, the user is still able to connect to another one to submit his jobs successfully. In D-Grid there are two identically configured Gateways running at different institutions to avoid a single point of failure, too.

In case of any trouble with the server components a monitoring tool for Grid components detecting problems like SIMON is essential, as problems should be fixed as soon as possible.

Furthermore scientists have to be interested in protecting their data from unauthorized access. It would have dramatic consequences, if, for example, patent details would be stolen while the job is submitted to the Grid. Hence, security is essential for a production Grid middleware. UNICORE uses the standard X.509 PKI for authentication, authorisation, and data integration. Authentication is done in the UNICORE gateway which does the following checks:

- Is the user's certificate still valid?
- Is the certificate on the current Certification Revocation List (CRL) of the signing CA?
- Has it been signed by a trusted CA?

Both DEISA and D-Grid only accept certificates signed by an EUGridPMA [14] trusted national Grid CA. The UNICORE security model for authentication is implemented in the UNICORE NJS which operates as a UNICORE scheduler. All public user certificates are stored in the UNICORE UUDB and they are mapped to an existing account on the target system. Every time the NJS receives an AJO, it checks if the signer's certificate distinguished name (DN) is stored in the UUDB, and if so the job is forwarded to the target system and assigned to the corresponding user account.

In addition scientists submitting their jobs to a Grid often want their jobs not only to be computed on one computing platform but on different ones. They want to use an as much abstract job description as possible in their Grid client. UNICORE provides this with the AJO model which allows the user to generate his jobs and workflows very easily and platform independent.

Often only initial hurdles prevent users from using Grid middleware. So difficulties occur even before the Grid middleware is used the first time. E.g. a criterion for users is how easy it is to adapt applications or to deal with and manage certificates. In order to deal with the first aspect, UNICORE provides the plug-in concept to adapt user's applications into UNICORE in a comfortable way. Many Scientist are not familiar with certificates as user credentials. With the integrated UNICORE Keystore Editor the user can easily manage all his certificates or request for new ones.

The acceptance of a Grid Middleware looms large with the convenience to access distributed resources regarding getting user accounts etc. The user administration of DEISA and D-Grid reduces the user's administrational efforts to a minimum.

Users' customs differ e.g. concerning graphical user interfaces and command line tools for job submission. So it is precious to provide both with the same functionality like with DESHL and UNICORE.

Our experiences with Grid infrastructures like DEISA and D-Grid show that fulfilment of functional requirements is not sufficient. Users want

- 24 hours a day and 7 days a week availability of the Grid infrastructure
- 24 hours a day 7 days a week availability of the Grid experts

Even if this is a ideal scenario which can't be established at every Grid providing site for cost reasons DEISA and D-Grid established multiple ways to

provide multiple ways of user support. Both projects introduced a user and administrator help desk, central mailing lists and support hotlines at each participating site. In case of any problems the user contacts the list at their home site to solve the problem locally or to forward the request to another site. The quality of problem management is up to the time needed to solve the request.

User support is absolutely essential but not only after a problem has occurred but also before. So detailed documentation of all used Grid middleware is necessary. Job examples have to be available and an additional FAQ section. DEISA and D-Grid provides this information on their web pages http://www.deisa.org and http://www.d-grid.de.

## 5 Conclusions

Grid has to mature more and more into a production status as it gets more and more desirable for scientists. There are various numbers of established Grid middleware systems which already satisfy the production interests of scientists and there is still much effort and funding to be spend on further development and integration of additional features, etc. However, in all those activities the developers should always focus on the user requirement; Grid middleware can only be successful if it is accepted and used by the end users, i.e. the scientists.

Scientists have to be stimulated and encouraged to use Grid middleware for applications, computations, data transfer and access to resources. They also have to be encouraged to adapt and integrate their applications to/into Grids. But once convinced, they likely use it further on.

In the meantime developers focus on Grid interoperability as a user should be able to submit his job transparently and independently of the underlying Grid middleware.

## Acknowledgements

A special thank goes to all the DEISA and D-Grid partners that contributed their effort, solutions, and ideas to the current deployments and configurations.

## References

1 http://www.deisa.org – Distributed European Infrastructure for Supercomputing Appications
2 I. Foster and C. Kesselman (Eds.). The Grid 2: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers Inc. San Fransisco (2004)
3 http://www.unicore.eu
4 D. Erwin (Ed.): UNICORE Plus Final Report - Uniform Interface to Computing Resources. Forschungszentrum Jülich (2003)
5 A. Streit, D. Erwin, Th. Lippert, D. Mallmann, R. Menday, M. Rambadt, M. Riedel, M. Romberg, B. Schuller, and Ph. Wieder. UNICORE - From Project Results to Production Grids. L. Grandinetti (Edt.), Grid Computing: The New Frontiers of High Performance Processing, Advances in Parallel Computing 14, Elsevier, 2005, pages 357-376
6 https://www.d-grid.de – German Grid Initiative
7 K. Nowinski, B. Lesyng, M. Niezgódka, and P. Bala: Project EUROGRID. Proceeding of the PIONIER 2001, Poznan (2001)
8 S. Sild, U. Maran, M. Romberg, B. Schuller, and E. Benfenati: OpenMolGRID: Using Automated Workflows in GRID Computing Environ-

ment. Proceedings of the European Grid Conference 2005, Amsterdam, (02/2005)
9 D. Erwin and D. Snelling: UNICORE: A grid computing environment. Proceedings of Euro-Par 2001, Springer, Machester (08/2001)
10 F. Schmuck and R. Haskin: GPFS: A Shared-Disk File System for Large Computing Clusters. Proceedings of the FAST, Monterey, (01/2002)
11 B. Hagemeier, R. Menday, B. Schuller, and A. Streit A Universal API for Grids. Proceedings of Cracow Grid Workshop (CGW'06), Cracow, Poland, 2006,
12 I. Foster: Globus Toolkit Version 4: Software for Service-Oriented Systems. International Conference on Network and Parallel Computing, Springer LNCS, (2005)
13 http://computing.fnal.gov/docs/products/vomrs/vomrs1_2/pdf/booktitle.pdf
14 Profile for Traditional X.509 Public Key Certification Authorities with secured infrastructure. EUGridPMA, Version 4.0 (10/2005)
15 K. Zeilenga and OpenLDAP foundation: Lightweight Directory Access Protocol (LDAP): Technical Specification Road Map. RFC 4510, (06/2006)
16 http://forge.ogf.org/sf/projects/saga-core-wg