CONTRACT NUMBER RI-222919

DEISA 2
DISTRIBUTED EUROPEAN INFRASTRUCTURE FOR SUPERCOMPUTING APPLICATIONS

European Community Seventh Framework Programme
RESEARCH INFRASTRUCTURES
Integrated Infrastructure Initiative

Initial Report on Operations

Deliverable ID: DEISA2-D3-1.1
Due date: October 31st, 2008
Author: Jules Wolfrat, SARA

Project start date: May 1st, 2008
Duration: 3 years

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2011)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Executive Summary

The main objective of the WP3 activity is the operation of the DEISA infrastructure including all services. This activity builds upon four years of experience, best practices and collaboration among all DEISA partners. The start of the DEISA2 project has been a natural point in time for reflection, refocus and implementing some of the lessons learned over the last four years in operating the DEISA infrastructure. This concerns the start of the implementation of re-grouping of the various services with reference to the Description of Work (DoW) [1], the alignment of those with WP4 (Exploring New Technologies) and the formation of new teams and communication in accordance with that. This also concerns the start of the implementation of more distinct procedures for incident- and change management, monitoring services, and the integration of new associated partners joining the DEISA infrastructure.

All connected centres are very experienced in the management of large HPC facilities locally and basically this management also extends to the local services which are operated for the DEISA infrastructure. These DEISA services have obviously dependencies between services running at different sites, making it important to have a well established organisation between the operational staff of the different centres.

For a professional operations team it is important that good internal communication facilities and procedures are in place. With the start of the current operations activity additional facilities have been created to improve the information exchange about the status of the production service. The already existing trouble ticket system is now in use by operational staff for reporting problems and an easy to use collaborative work environment for the internal exchange of information is taken in production, e.g. for publishing maintenance schedules and other service interruptions. The monitoring information provided through INCA, a user level grid monitoring facility, has been improved and extended through the addition of new monitoring information providers.

The operational services are now divided in five domains: network, data, computing, AAA (Authentication, Authorization and Accounting), and user environment services. For each of these domains a subtask is defined and for each the status and developments in this project period are discussed. All operational services, as retained at the start of DEISA2, have proven to be stable. Several partners have upgraded or installed new systems in this period and all these changes have been successfully incorporated into the DEISA infrastructure.

Equally important for a reliable service is that change management is a well defined process. This is needed both for changes within production services as well as for new services coming into production. For the latter a separate task was defined. This task implies interaction with the technology activity (WP4), which investigates and prepares new services for production. This directly refers to the second objective of this activity, the adoption of new technologies from the technology activity. In this reporting period an initial implementation of the procedures has been completed. These procedures are thus in place for the first technologies from WP4 that will be accepted for production.

The third and final objective of this activity is the integration of new sites from new partners within the infrastructure. Preparations for the integration of the new partners are well underway, i.e. network connections are negotiated, information on software installation is

provided, and contacts between new partners and supporting partner are established. CEA has been added to the list of associated partners to be integrated in the DEISA infrastructure.

This document describes the WP3 organisation and the status of the production services in the first six months of DEISA2. In section 2 the facilities are described that were set up to facilitate the first objective, the operation of the infrastructure. In section 3 the status and changes of the production services are described. Section 4 discusses the second objective of this activity, the adoption of new technologies from WP4. Section 5 discusses the integration of new partners. In section 6 a quality plan is discussed.

1.2 References and Applicable Documents

[1] DEISA2 Description of Work (Annex I of the Grant Agreement)
[2] DEISA web pages http://www.deisa.eu
[3] Deliverable DEISA2-D1.1: Initial Report on Management
[4] Request Tracker (RT) http://bestpractical.com/rt
[5] DEISA Trouble Ticket System (DTTS) https://tts.deisa.eu/
[6] DEISA trusted CA certificate repository http://winnetou.sara.nl/deisa/certs/
[7] EUGridPMA The European Policy Management Authority for Grid Authentication
[8] BSCW - Basic Support for Cooperative Work http://public.bscw.de/en/index.html
[9] TWiki® - the Open Source Enterprise Wiki and Web 2.0 Application Platform http://twiki.org/
[10] DART http://www.deisa.eu/usersupport/tools/dart
[11] Deliverable DEISA2-D6.1: Initial Report on Environment and User Related Application Support
[12] Modules http://modules.sourceforge.net/
[13] NICE http://www.nice-italy.com
[14] JSR-168 standard http://jcp.org/en/jsr/detail?id=168
[15] Globus http://www.globus.org/
[16] TeraGrid project http://www.teragrid.org/

1.3 Document Amendment Procedure

This document is prepared according to the guidelines defined by the management of DEISA2. These rules can be found in section 2.7 of the deliverable DEISA2-D1.1 [3].

1.4 List of Acronyms and Abbreviations

AAA	Authentication, Authorization & Accounting
BSCW	Basic Support for Cooperative Work
CA	Certificate Authority
CRL	Certificate Revocation List
DART	DEISA Accounting Report Tool
DEISA	Distributed European Infrastructure for Supercomputing Applications
DUAS	DEISA User Administration System
DCPE	DEISA Common Production Environment
DEC	DEISA Executive Committee
DoW	Description of Work

DTTS	DEISA Trouble Ticket System
Ganglia	A scalable distributed monitoring system
GÉANT2	European backbone Network connecting national research networks
GPFS	General Parallel File System
GridFTP	FTP version enabled for use with X.509 certificates
GSI-SSH	Secure Shell (OpenSSH) using Grid Security Infrastructure technology
INCA	User Level Grid Monitoring facility
ITIL	IT Infrastructure Library
LDAP	Lightweight Directory Access Protocol
Nagios	A host and service monitor tool
NREN	National Research Network
OpenSSH	Open Source Implementation of Secure Shell
TSI	Target System Interface – A UNICORE service
X.509	format used in public key encryption

2 Management of the Operations activity

2.1 Introduction

The main objective of the activity as defined in the DoW is operating the infrastructure including all services. Most of the work of this work package is devoted to this objective.
The services to support are those that were made available by DEISA and eDEISA and which were accepted as required and ready for production at the start of DEISA2. In addition the activity has to support those services which will be proposed by the WP4 activity and accepted as a production service by the project management.

Based on the experience from the previous projects DEISA/eDEISA combined with the long standing know-how in the local organization of an HPC centre, five logically separable tasks have been identified for this WP, which cover the whole range of services to be provided for the project:

(1) Network-Related Services
(2) Data-Related Services
(3) Compute-Related Services
(4) AAA-Related Services
(5) User-Related Services
(6) Integration of new partners

Common to all these tasks is that for a high quality support of the production services well defined procedures and facilities must be in place. Much effort has been dedicated to setting up communication facilities, a collaborative work space, procedures for management of the trouble ticket system, and change management. An “operator on duty” task is defined for overseeing the operational status of the infrastructure. The monitoring facilities are continuously improved and expanded. Details of these facilities are described in the next sections.

2.2 Management organisation

2.2.1 Organisation

The basic management structure is given by Figure 1. Each site has a site representative, the deputy, who is responsible for the status of the DEISA services at the site. Any issues with the site can be discussed between the WP leader and the site deputy as a first escalation step. Table 1 gives the list of site deputies.

On a horizontal level there are subtask leaders for the coordination of the activities needed for the operation of the services belonging to that task For some of the subtasks more than one leader is appointed, each responsible for a subset of the services. The list is given in Table 2. The integration subtask is organised differently. For each new site which will be integrated a supporting site is appointed from the existing sites. And the person responsible for the integration for a particular site will come from the supporting site.

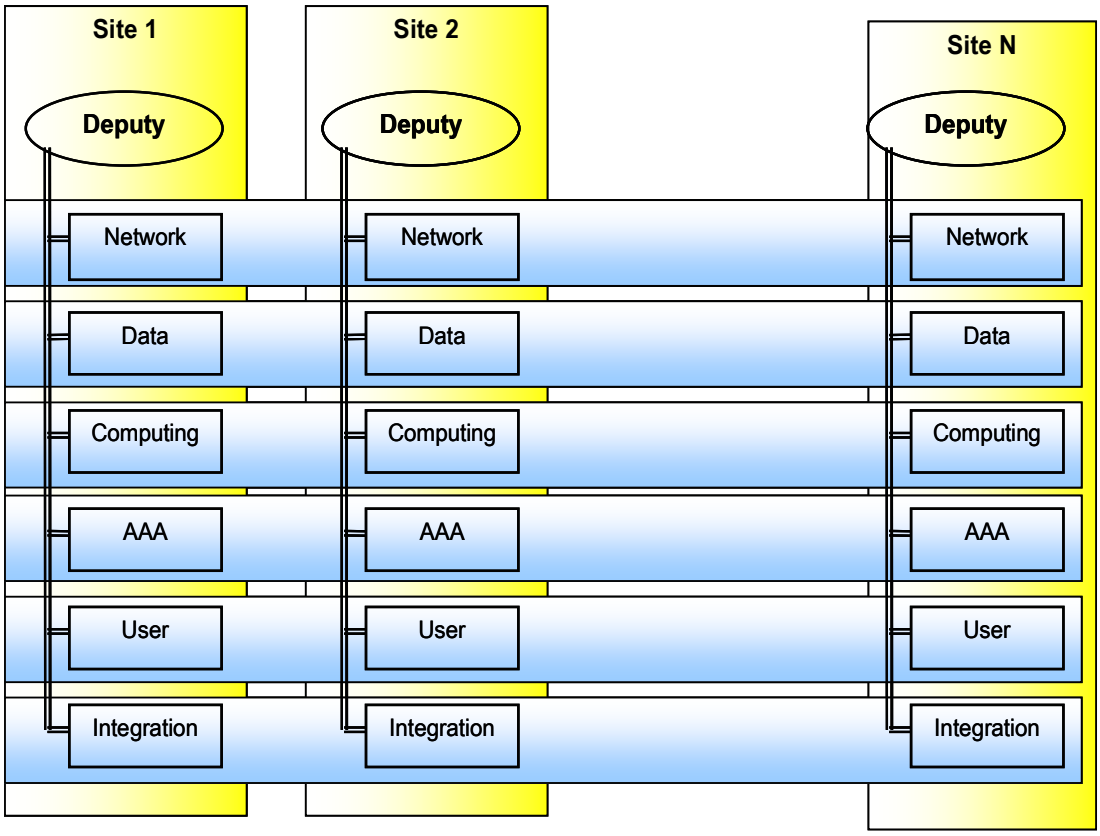


Figure 1 – WP3 structure and internal relations

Site	Name
BSC	Javier Bartolomé
CINECA	Stefano Martinelli
CSC	Pekka Lehtovuori
ECMWF	Cristian Mezzanotte
EPCC	Jeremy Nowell
FZJ	Klaus Wolkersdorfer
HLRS	Thomas Bönisch
IDRIS	Philippe Collinet
LRZ	Anton Frank
RZG	Andreas Schott
SARA	Vincent van den Elzen

Table 1 – List of Site Deputies

Subtask	Name
WP3-1 Network	Ralph Niederberger
WP3-2 Data	Andreas Schott, Frank Zeller
WP3-3 Compute	Philippe Collinet, Michael Rambadt, Helmut Heller
WP3-4 AAA	Jules Wolfrat
WP3-5 User	Denis Girou

Table 2 – List of Subtask Leaders

2.2.2 Internal Communication

2.2.2.1 E-mail lists

Efficient e-mail communication is important in a distributed organisation like DEISA. Several lists for e-mail have been defined for internal use. All lists are defined and maintained on a central majordomo server hosted by RZG. All lists have deisa2.eu as domain. The lists in use are given in Table 3. The list for the user support task, WP3.5-User, is an alias for the WP6 task. It is important that WP6 and subtask 3.5 share information because these two activities have many dependencies between each other. For the Integration subtask no separate list is defined as the communication mainly will be between the supporting site and the site to be integrated. General issues with regard to this subtask can be discussed on the WP3 list.

Name of list	Description
WP3	For general WP3 discussions and announcements (meetings, documents, procedures etc.).
Operations	For all operational announcements (maintenance, broken services, and upgrades) with all operational staff from partners subscribed
WP3.1-network	discussion of network issues
WP3.2-data	discussion of data issues
WP3.3-computing	discussion of computing issues
WP3.4-AAA	discussion of AAA issues
WP3.5-User/WP6	discussion of user support issues

Table 3 – List of WP3 e-mail list for internal use

2.2.2.2 Video Conferences and face-to-face meetings

Bi-weekly video conferences (VCs) are organised to discuss the status of the infrastructure. These VCs are open for everyone involved in the activity, but all subtask leaders and site representatives are expected to attend. Other persons can be invited if needed for the discussion of specific issues. The agenda of these meetings has since several months now the following items:

- Agenda
- Minutes of preceding meeting
- Discussion of operational problems based on reports of operator on duty (by responsible sites)
- Discussion of action list, see minutes preceding minutes
- Reports by subtask leaders
- Any Other Business (AOB)

The role of the "operator on duty" task is discussed in section 2.3.4. A list is maintained with short term actions, e.g. improvements of tools. Actions to solve operational problems are tracked using the DEISA Trouble Ticket System (DTTS, see section 2.3.2). Finally in the reports by subtask leaders all issues, problems or changes in the services can be discussed. Minutes of these meetings are published on the DEISA internal BSCW [8].

Also face to face meetings are organised, in principle two are planned for each year. The first meeting was held in Amsterdam, on July 16 and 17. It was a combined meeting with WP4 because of the many interactions between these two work packages. For the next meeting also WP6 will be involved because of the common interests between WP6 and task 3.5 (User environment) of WP3.

2.3 Internal services

In this section we discuss several facilities which are used by WP3 to operate the infrastructure.

2.3.1 INCA

INCA is an application used for monitoring of the computing infrastructure operated by DEISA. The main goal of the application is to determine availability and functionality of a system from a user point of view, however at the moment it is used only internally. INCA, developed by the San Diego Supercomputer Center initially for the TeraGrid project [16], was chosen in 2005 by DEISA to monitor the DEISA Common Production Environment (DCPE), see deliverable DEISA2-D6.1 [11].

INCA implements a client-server model in which clients that run on monitored DEISA resources periodically execute small scripts called reporters in order to test a specific functionality of the local environment. Results of these tests are sent to the INCA server, operated by LRZ, where they are stored and automatically analysed. The results are available to all DEISA staff through a graphical web interface, using data consumers which allow the definition of customised presentation layers, that implements certificate based access control. An additional important feature from the operational point of view is that an automatic e-mail notification is sent to the affected site if a failure is just diagnosed.

INCA is an on-going development, with in particular a complete rewriting in 2006 for the 2.0 version that DEISA deployed in 2007. Since the early stage of our usage of INCA, we maintain a very close and efficient relationship with the development team in SDSC. On their side, they give us very good and fast support, both for the problems that we diagnose in the software and for the customizations that we need due to some specifics of our DEISA infrastructure or to some special wishes that we sometimes have for the presentation layer. In the beginning of September 2008, the INCA team in SDSC organized an international workshop in San Diego and the two main DEISA staff members in charge of INCA participated to this workshop (one physically in San Diego and the other one by videoconference for most of the sessions). This provided us first-hand experience with the latest INCA release and the opportunity to give the development team feedback and a list of some desired features for future versions of the software.

The current implementation of INCA displays both the version numbers of the software available in the so-called *Software Stacks* and the real status of their usage, based on small test codes included in the *DEISA Applications Test Suite* (DATS), which check the behaviour of the compilers, the communication libraries, as well as some scientific libraries and third party applications. But already some years ago, the usage of INCA inside DEISA was progressively extended to also monitor the availability and status of various middleware, or at least, in some other cases, to display their status according to the information collected by their dedicated monitoring facilities:

- LoadLeveler batch filters for the IBM platforms using the *DEISA LoadLeveler Test Suite* (DLLTS),
- UNICORE using its dedicated Simon monitoring tool,
- Globus services deployed on the infrastructure (gsi-SSH, GridFTP, and RFT)
- LDAP user administration service.

At this moment INCA is running more than 140 reporters monitoring all these services and the availability of the components of the DCPE.

The details of the evolutions and improvements in the INCA framework during the reporting period are described in the section 3.6.3 and some ongoing or planned enhancements, like the integration of few other services in the monitoring environment, are explained.

2.3.2 DEISA Trouble Ticket System

2.3.2.1 Overview

The DEISA Trouble ticket system (DTTS) is an application which allows reporting about problems and their follow-up discussion in a central repository. It is used both for internal problem tracking by DEISA staff and for tracking of problems reported by users.
Users can interact with the DTTS by e-mail to support@deisa.eu and through a web interface (https://tts.deisa.eu).

The DTTS is set up with the Request Tracker (RT) open source product. The basic concepts of this application and user-documentation can be found at [4]

2.3.2.2 Installation and configuration

The application has been slightly modified in order to increase its usability within DEISA as well as making its web presentation similar to that of the official DEISA website [1]. Care has been taken that all the modifications easily can be applied with upgrades in the future.
The DTTS runs on a dedicated virtual host (1 CPU, 512MB RAM) with the DNS name tts.deisa.eu using a host-certificate issued by the NorduGrid Certification Authority for host identification.

Client authentication is based on X.509 certificates and root certificates of trusted CAs are updated once a day from the repository at SARA [6]. The corresponding certificate revocation lists (CRLs) are updated shortly afterwards using the "fetch-crl" script from EUGridPMA [7]. The database of authorized users is derived from the DEISA LDAP repository, from which it is updated once a day. Only persons registered in the DEISA LDAP system are authorised to use the service, with DEISA staff having more permissions than the users.

For web access a valid certificate which is registered in the DEISA-LDAP is mandatory. E-mails are accepted if they are signed with a valid and registered certificate. The use of certificates for e-mail signing can be cumbersome for users. So in addition unsigned e-mails are also accepted if they fulfil certain requirements. These requirements enable that a user should be able to e-mail to the DTTS, but that Spam is blocked.

2.3.2.3 Basic usage

The DTTS has a queue for each DEISA site, where all tickets belonging to a site are attached to. Furthermore it has one "General" queue where all global DEISA related tickets are attached to. Besides this geographically based division, tickets can also be tagged with technical keywords, so that tickets can also be grouped by their technical nature. Currently keywords are defined for the five service subtasks defined for WP3 (network, data management, middleware, AAA, and user environment). Members of the DEISA staff can search in the DTTS for tickets by queue or by keyword. More complex searches are possible by a SQL-like language within the web-interface, see Figure 2.

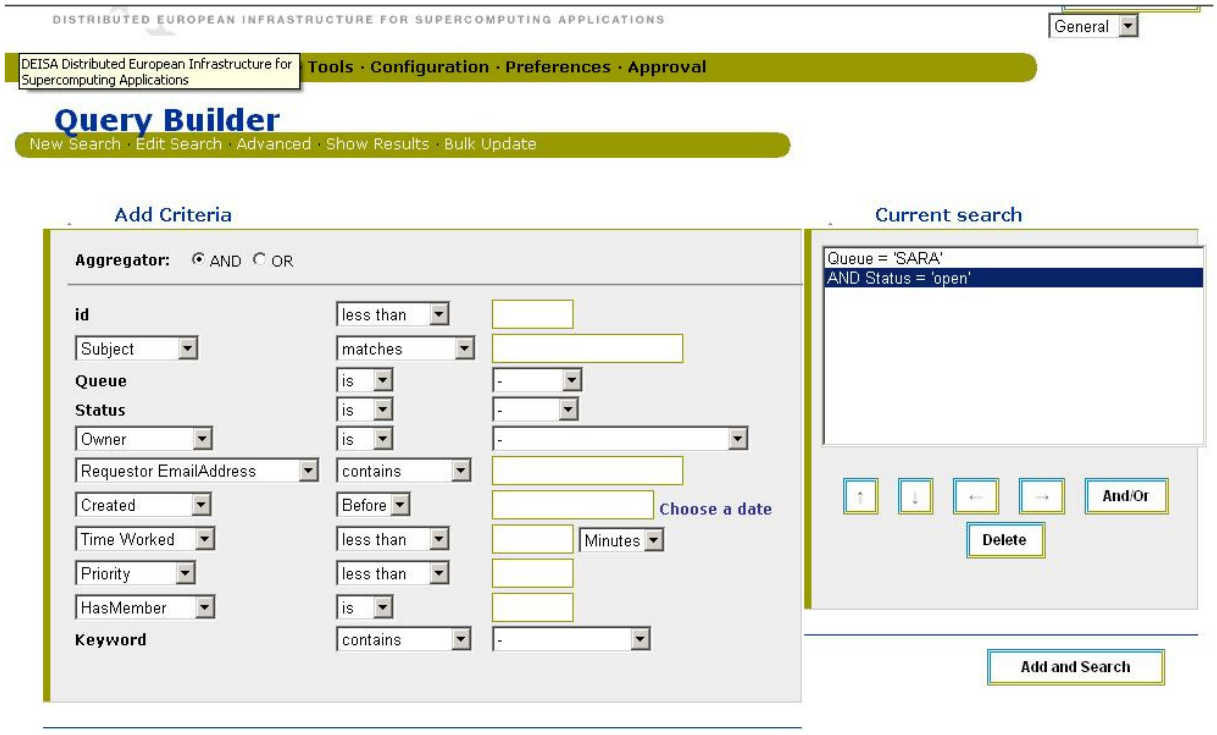


Figure 2 – Example of Query Building

When a ticket is created by web or e-mail within a certain queue DEISA staff is notified by e-mail about the event if they subscribe to that queue as a so called watcher. These persons are then responsible to assign keywords to the ticket and to either work on the ticket or to delegate it to another staff member within the DTTS. So each site must subscribe to their queue as watcher.

The general queue is monitored by the user support subtask (T3.5), i.e. by the Helpdesk person on duty for a specific week. This person checks this queue for the incoming tickets, moving them to the appropriate site-specific queue and labelling them with the relevant keyword from the predefined list. After the assignment of the tickets, WP6 (task T6.3) takes charge of them. For details, see the deliverable D6.1 of WP6 [11].

2.3.2.4 Future work

The current set-up is a starting point and will be adapted based on the experience that will be gained in the coming months. An extension which already is considered is the introduction of an escalation mechanism. This will bring stalled tickets to the attention of a defined group like e.g. the operations team or the operator on duty (see section 2.3.4).

Changes are also to be expected when the DTTS is used for general user support within DEISA.

2.3.3 Collaborative work space (BSCW and Wiki)

DEISA2 continues to use BSCW [8] as the tool for document sharing between partners and the scheduling of meetings (calendar). CSC is hosting the service and provides support for the service (user registration, problem solving). However BSCW is not well suited to quickly update information, you cannot edit directly information on the server. So documents always have to be replaced by updated documents. For some purposes this is not very convenient, e.g. if just a small change is needed. An example is the publishing of maintenance schedules and other service interruptions. A site must be able to easily update this information, either by manual editing of the information or by automatic update remotely.

For the purpose of a collaborative workspace with easy update facilities DEISA2 has set up a wiki service based on the TWiki® software [9]. This service is also hosted and supported by CSC. Authentication and authorisation is based on a PKI using X.509 style certificates. DEISA staff registered in the DEISA LDAP repository is automatically registered as user of the TWiki® facility. It is used by WP3 for publishing maintenance information, internal reporting, configuration information etc.

2.3.4 Operator on Duty

The "Operator on Duty" task is that each week a site is responsible for overlooking the operational status of the infrastructure. This task was already defined by the eDEISA operations group and is fully operational now in DEISA2. The schedule of which site is responsible which week is maintained on BSCW.
The tasks defined are:

- Daily look at monitoring information.
- Create a ticket in the DTTS if problems seen in the monitoring are persistent. Also look at the service interruption information if a problem already is signaled. Assign the ticket to a site queue. If more than one site is involved the general queue can be used and the ticket can be assigned to the subtask leader responsible for that service. But in the latter case it must be verified that this person is available.
- At the end of the week a report is produced mentioning major problems and issues that require attention. Also the status of open tickets is included in this report. The report is appended on Twiki. So we have a log on Twiki.
- The publication of a status report of trouble tickets. Currently tickets without owner are always mentioned, e.g. the ticket is assigned to a queue but nobody has taken ownership of the ticket. Trouble tickets without any activity for longer than two weeks must be reported too. The status report is discussed in the bi-weekly VC.

It must be noted that each site is responsible in the first place for taking action for problems with the DEISA services at their site. These services must be an integral part of their operational service. The operator on duty task is part of the quality assurance of the production service.

2.3.5 Change management

Procedures for changes in the operational status of the production services must be well defined in order to maintain the reliability of these services. In the first place proposed changes must be documented and tested. Also user documentation must be available and accepted if applicable. The document describing the change and test results must be placed on BSCW in a dedicated folder. Then the operations team can be informed and if no objections are received the change can be planned. The decision to implement the change can be agreed in the bi-weekly VC or by e-mail.

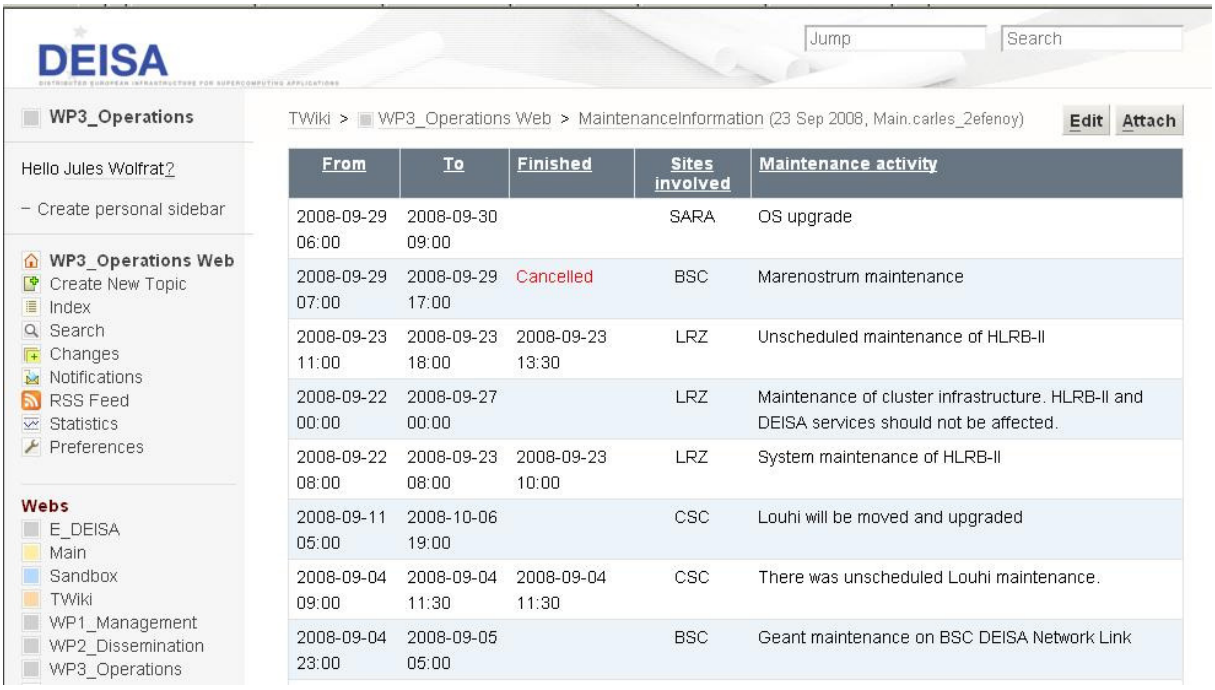
A very specific change in the infrastructure is the replacement of a system or the introduction of a new system. In the first six months of DEISA2 three systems have been replaced. FZJ, RZG, and SARA updated their IBM system from Power4 or Power5 to Power6 based architectures. In addition FZJ introduced a new BlueGene system. These system changes mean that for most services the configuration has to be adapted, e.g. change of network addresses, target host names for services etc. All these changes have taken place without many problems, which demonstrates that the different services are well documented with regard to configuration etc. For the coordination of such changes a template is produced of all actions needed for the different services.

Changes in the production infrastructure as a result of the WP4 (new technologies) activity are discussed in section 4. The change management procedure from the WP3 point of view can be the same for these changes as for the internal changes.

For the future we plan to look at ITIL based procedures for the management of changes in the production service.

2.3.6 Maintenance and incident reporting

For project internal use scheduled maintenance and unscheduled service interruptions on the systems in the infrastructure are published with the TWiki® based service (see example in Figure 3). In parallel it can be announced on the operations e-mail list. It is planned that at least part of this information also will be publicly published for access by the users. Also enabling electronic exchange of this information between different locations is on the list of requirements. So the DEISA web site can be automatically updated from a site specific location and vice versa. A format based on XML for the exchange of the information will be defined.



From	To	Finished	Sites involved	Maintenance activity
2008-09-29 06:00	2008-09-30 09:00		SARA	OS upgrade
2008-09-29 07:00	2008-09-29 17:00	Cancelled	BSC	Marenostrum maintenance
2008-09-23 11:00	2008-09-23 18:00	2008-09-23 13:30	LRZ	Unscheduled maintenance of HLRB-II
2008-09-22 00:00	2008-09-27 00:00		LRZ	Maintenance of cluster infrastructure. HLRB-II and DEISA services should not be affected.
2008-09-22 08:00	2008-09-23 08:00	2008-09-23 10:00	LRZ	System maintenance of HLRB-II
2008-09-11 05:00	2008-10-06 19:00		CSC	Louhi will be moved and upgraded
2008-09-04 09:00	2008-09-04 11:30	2008-09-04 11:30	CSC	There was unscheduled Louhi maintenance.
2008-09-04 23:00	2008-09-05 05:00		BSC	Geant maintenance on BSC DEISA Network Link

Figure 3 – Maintenance schedules published on TWiki® web

Incidents at a site which require the involvement of other partners in the infrastructure must be entered in the DTTS and assigned to those partners. Of course e-mail notification or contacts by telephone always can be used in parallel. Further, sites must create tickets in the DTTS for problems at their site which can not be solved immediately.

3 Discussion of production services

3.1 Introduction

In this section the status of the operational services is discussed in detail. In the DoW for DEISA2 a division is made in subtasks for the operational management of the different services. Each subtask is responsible for part of the operational services. In the DoW for each task a description was given of the services that would be supported by these tasks. However not all of these services had full operational status at the start of DEISA2. Only the services that are accepted as fully operational are described here. The list of operational services has been discussed at the start of DEISA2 at the kick-off meeting in Barcelona in May.

Table 4 gives an overview of the production services and some additional services in brackets which at this moment are not in production. These latter are not discussed here as the evaluation and test work is the responsibility of WP4 of DEISA or eSA3 of eDEISA. In sections 3.2 through 3.6 the status of the core services for the Network, Data, Compute, AAA and User subtasks is discussed. The integration subtask is discussed in section 5 and the internal services already have been discussed in section 2.

DEISA services	Production service
Network (section 3.2)	Dedicated high speed network
Data (section 3.3)	MC-GPFS-(HSM), GridFTP data staging
Compute (section 3.4)	Local batch systems, UNICORE
AAA (section 3.5)	LDAP, PKI, interactive login, Accounting
User (section 3.6)	DCPE, modules, INCA, user support & help desk, trouble ticket system, DESHL, Portals
Integration (section 5)	Additional partners
Internal (section 2)	BSCW, security alert, resource monitoring and information

Table 4 – Service matrix

3.2 Network Related Services

As described in the DoW for DEISA2 the basic goal of this task is connecting the different HPC systems of the partners with a dedicated network, mostly with 10Gbit/s, in order to form a tightly coupled European-wide HPC infrastructure.

3.2.1 Network contracting with NRENS central hardware

The DEISA backbone consists of network links provided by GÉANT2, the corresponding NRENs and for some sites local providers. Since most of the links have been operational for several years, only some contract extensions had to be made. Within the reporting period contract negotiations have been started with the Swiss service provider SWITCH for connecting the new partner CSCS, UKERNA connecting the Hector system of EPCC at Edinburgh and NORDUnet providing connectivity to the new partner KTH. Very recently

also CEA joined the DEISA2 community as new Associate Partner. This led to additional negotiations with the French NREN RENATER.

3.2.2 Network configuration and monitoring

The DEISA network has been operational for several years already. The main work within this time period has been the maintenance of a high quality, error free production network infrastructure. Additionally to this the connections to the new partners and associate partners had to be planned and partly configured. The network connectivity for KTH is in a very early stage. It has not yet been decided which bandwidth (1 Gbit/s or 10 Gbit/s) will be chosen. KTH has started some negotiations with local providers, but has not yet signed any contracts. CSCS is also in negotiations with their NREN SWITCH and with GÈANT2 currently. EPCC and CEA have made a lot more progress. EPCC has ordered a wavelength from London to Frankfurt, which is working quite well already. Only some local provider configurations have to be completed to get the link into production. The same applies to CEA. Here it has been decided to connect CEA via the already existing trunk used by IDRIS. The 10 Gbit/s link from Paris to Frankfurt will be shared between these two installations. It is assumed, that connectivity can be established very soon.

From the operational point of view there have been a lot of short interrupts to the network or parts of the network within the first half year of the project lifetime. Fortunately most of those interruptions could be solved very quickly by the GÉANT2 and NREN staff. To react on those interrupts several monitoring procedures have been developed and installed.

3.3 Data Related Services

3.3.1 MC-GPFS services

The DEISA2 project started seamlessly with the configuration of the Global Files System GPFS achieved in the FP6 funded project DEISA.

During the first half year of the DEISA2 project many HPC systems of the participating sites have been upgraded. Thus, a smooth transition of the compute nodes as well as the disk server nodes providing the GPFS disk space had to be assured. In addition a lot of sites not only upgraded their systems but added new systems, namely the BlueGene/P systems at FZJ, IDRIS and RZG. In the following the upgrade activities, some still ongoing, are summarized per site:

- CINECA is unchanged in the hardware configuration, but all changes in the other sites had to be reflected by modifications in the configuration.
- ECMWF provides still the same GPFS-servers and also had to adapt the configurations.
- IDRIS has upgraded its Power4 system to a combination of a Power6 system and and a BlueGene/P system.
- FZJ has installed a small Power6 system substituting the old Power4 system. The file-servers have been moved to a Power5 system, with the disk-space increasing from 4 TB to 18TB. The major HPC system is a BlueGene/P, which is not directly mounting the DEISA-GPFS but the data can be moved on a gateway machine from the DEISA-GPFS to the BlueGene/P internal file-system.
- RZG has switched the file-server and the HPC computing functionality for DEISA in mid-July from Power4 to Power6. The new disk systems provide with 78TB more then tenfold of the old DEISA-GPFS. Furthermore, RZG is still hosting the data for all sites, which do not provide their own disk-servers. The DEISA-GPFS file-system is mounted on the Power6 cluster and also mounted directly on the BlueGene/P system.

- SARA has undergone an upgrade from Power5 to Power6 processors, including the DEISA-GPFS file-servers. The provided disk-space remained unchanged at 12 TB.

Figure 4 shows the current configuration of the DEISA-GPFS.

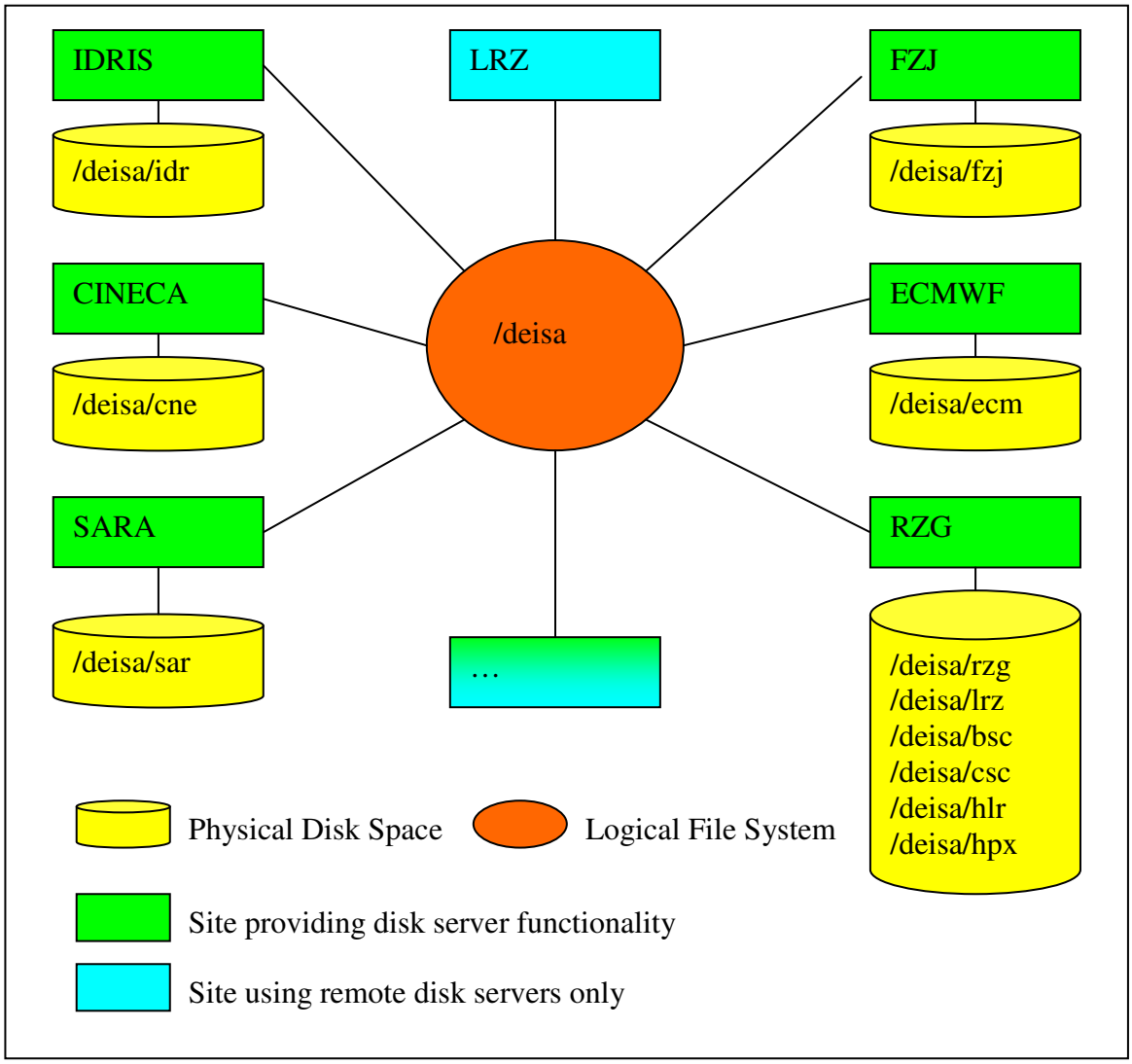


Figure 4 – GPFS Provision throughout DEISA

The following Table 5 summarizes the systems currently connected to the DEISA-GPFS:

Site	Fileservers	Storage in TB	Systems Connected
CINECA	2	5	Power5
ECMWF	2	1	Power6
FZJ	2	18	Power6, BlueGene/P (via GW)
IDRIS	2	2	Power6
LRZ	0	0	Altix
RZG	2	78	Power6, BlueGene/P
SARA	2	12	Power6

Table 5 – Current hardware of the DEISA sites integrated with MC-GPFS

The currently integrated sites are still the same which were integrated at the end of the DEISA project, but now with much more disk space and much more powerful systems connected.

There are still some sites not integrated. In the following the situation and plans are described for each of these sites:

- BSC is still running an incompatible version of GPFS on the MareNostrum system. It is discussed to configure a small gateway system, which provides DEISA-GPFS at BSC. From this gateway the data-transfer to the local GPFS could be achieved. Another option, still under discussion is an upgrade of the GPFS version. Furthermore other technical alternatives based on NFS or pNFS are investigated in WP4.
- CSC has upgraded its Cray system to a XT4 + XT5 combination. For this system currently the GPFS-client is not yet available, but it is expected to reintegrate CSC during the next 6 month.
- EPCC for a long time had only private IP addresses available, which made it very difficult to integrate the HPCx system into the DEISA-GPFS. By now it would possible to integrate the old IBM system. But in the near future EPCC is switching its HPC system to a Cray XT system. Therefore it is planned to concentrate on the integration of this system as a client system analogous to CSC.
- HLRS with its vector system provides a very special hardware, for which no GPFS client is available. Considerations of attaching this system with NFS to clients running on different hardware are investigated in WP4.

3.3.2 GridFTP services

The major changes of the GridFTP infrastructure within the reporting period (which are the past six months) are related to the recent installation of new computing systems at some of the DEISA sites.

In May, within the scope of the eDEISA activity, EPCC has provided a dedicated GridFTP service for DEISA on HPCx. To make GridFTP a full DEISA production service that is provided by all sites, ECMWF (who was not a member of the eDEISA Global Data Management working group) has installed GridFTP at their site and is currently in the process of testing the installation in order to make it work with all other DEISA sites.

In the June / July 2008 timeframe new computing systems were installed at FZJ, RZG and SARA. As a consequence, GridFTP had to be re-installed and configured at those sites. In all cases GridFTP has soon afterwards been brought back into full operation.

Another update was carried out at LRZ. In order to increase transfer performance with their file systems, LRZ has deployed an additional GridFTP server that uses striping based on four data nodes.

As mentioned above, EPCC completed the deployment of GridFTP on HPCx. They now provide GridFTP access to the HPCx system via the DEISA 10 Gb/s network, which enables users with a DEISA account to transfer data between the HPCx GPFS service and other file systems that are available on the DEISA network. The service is provided on a node of the HPCx system that is connected to the DEISA network, using a 10 Gb/s interface. It is the same server that supports GPFS, allowing fast access to the file system. In addition to the internal GridFTP server, HPCx also provides a public GridFTP service that is visible on the internet. This service is hosted on the login node and is available to all users of the system. Thus, EPCC belongs to the sites that allow all DEISA users to transfer data between their local sites and the DEISA infrastructure using the efficient GridFTP file transfer mechanism. The other sites are currently CINECA, IDRIS, LRZ and RZG.

Besides this deployment work, effort is and will be undertaken to improve serviceability as well as usability. To ease serviceability, the GridFTP service is monitored within the INCA framework used by DEISA. Currently, it is monitored whether a GridFTP server is up or down and the results are published to DEISA administrators. A future point of discussion is

whether we will provide more sophisticated information such as performance values and who will be allowed to access which kind of information. Overall, GridFTP has turned out to give reasonable performance figures and proved to be a reliable tool where it was used in production.

To improve on usability, a GridFTP command-line tool developed by LRZ with scp-like syntax has been deployed across the sites.

This work has been done partly by the WP4 activity. Although WP3 already takes responsibility for the GridFTP service as available today the official acceptance as full production service must still take place. This is planned for the end of 2008.

3.4 Compute Related Services

System monitoring tools (Ganglia, Nagios) are mentioned in the DoW as services belonging to this subtask. These tools are however still under evaluation and so not described here as they don't have production status.

3.4.1 Local batch system adaptations for DEISA

If adaptations have taken place they are described under the services for which they were required.

3.4.2 MC-LL configuration for DEISA

LoadLeveler Multi-Cluster has a full production status since 2005, with thousands of DEISA jobs already executed. Since then LL-MC is intensively used locally by AIX sites and is operated as a low-level resource management facility within a stressed environment without any failure. It is a core channel for efficient execution of DECI projects by offering the maximum robustness possible. It is also directly or indirectly used by middleware (on higher job management level) like UNICORE, DESHL, and Portals, if jobs are submitted by the middleware to LoadLeveler sites.

LL-MC is customized to operate an AIX super-cluster, allowing remote site job management (submission, monitoring, control) and also job migration across sites. The customisation is realized by maintaining configuration files and utilizing DEISA filters, which have been developed to keep the local resource management policies compatible with the DEISA LL job description policy (a set of keywords agreed by all sites documented in the DEISA Primer Guide).

Besides the functionality of the DEISA filters to interface with local filters in order to enforce the DEISA policy and to adapt local site policies to the DEISA one (allowing remote job submission with a flexible policy), further exit routines have been implemented in order to enhance security and to control the job routing processes.

Table 6 shows the DEISA sites currently interconnected by LoadLeveler Multi-Cluster with the product and filter releases.

Site	RZG	FZJ	CINECA	IDRIS
OS	AIX	AIX	AIX	AIX
Processor Type	POWER6	POWER6	POWER5	POWER 4
Batch System	LL-MC	LL-MC	LL-MC	LL-MC
Batch Version	3.4.3.1	3.4.3.3	3.4.2.2	3.4.1.2
DEISA Filters	V2.02 / V2.03	V2.06 / V2.05	V2.06 / V2.05	V2.06 / V2.05

Table 6 – LoadLeveler versions and filters deployed

New AIX POWER6 platforms have been recently installed in a number of sites, requiring filter modifications in order to handle new batch keywords compatibility.

A survey is performed to look at the possibility to integrate in the DEISA LoadLeveler super-cluster sites running LoadLeveler on non AIX or non IBM POWER platforms.

3.4.3 UNICORE services

UNICORE is not anymore the only middleware providing access to the systems in the infrastructure. In DEISA2 users can access all systems also interactively. The workflow functionality is an additional feature interesting for some applications. UNICORE proved to be stable in the reporting period and no general changes for this service were required. FZJ, RZG, and SARA updated their IBM system from Power4 or Power5 to Power6 based architectures. The adaptation of the involved UNICORE TSIs has been done at all three sites. FZJ introduced their new BlueGene system as the 12th target system into the DEISA environment. It also has been integrated in the DEISA UNICORE infrastructure in the meantime.

Some users may have access to more than one userid on the systems. This requires an update of the UNICORE user database (UUDB) which currently is only able to map an X.509 certificate to exactly one userid only. A new version of the UUDB allows handling of more than one userid for one certificate/identity. This update is prepared and can be implemented after approval by WP3 management.

3.5 AAA Related Services

3.5.1 LDAP based User Administration System

The DEISA User Administration System (DUAS) is not only used for the administration of user accounts, but it also provides information which is used by other DEISA services. Examples are the registration of authorized users of the DTTS, the DEISA BSCW service, the TWiki® information and authorization for access to accounting information. So in general it is the source of authorization information for DEISA services. In order to enhance the functionality it is under discussion to define role attributes in the LDAP schemas used. An example is the introduction of an attribute which will specify the accounting role (which permissions) apply for a person. Currently authorization for access to accounting information depends on different attributes. The introduction of such a new attribute would make it easier to introduce new roles or authorization levels.

3.5.2 Accounting Services

The accounting service is in production at all sites and access for users is provided through the DART tool [10]. A minor bug in DART was repaired. For some months the number of records taken into account was wrong.
The implementation of a single repository in which summary records of DEISA users can be stored for long term archival is planned. This probably will be a WP4 activity as this requires a major effort.

3.6 User Related Services

3.6.1 Modules Framework

The Modules tool [12] was chosen four years ago as the tool for providing the common interface to access the software of the DEISA Common Production Environment (DCPE). This simplifies both the access and use of the software, offering a uniform and simple method to access the software in the same manner on all systems. And it also prevents that a user has to find out the exact location of the various executables, libraries, utilities, etc. made available and physically installed in different file systems and directories on each system, according to the local preferences of the various sites.

Among the two flavours of the Modules tool available in the public domain, we chose the Tcl reimplementation of the original C version. This implementation has the benefit of a higher level language and it was at that time the only version maintained. However, it appeared that the maintenance was problematic, with the backward compatibility not preserved by the current maintainer, which forced us to spend some time in DEISA on a frozen old version. This is why we decided at the end of 2007 that in DEISA2 we will maintain our own flavour of this tool, staying inline with the official development branch but both re-implementing an important feature for our usage which was suddenly withdrawn and adding some new features which appeared to be useful for our usage.
This is why we use now (beginning of October 2008) the current version 1.113 of the Tcl implementation, with the behaviour of the switch command like it was three years ago, and with some new features:

- The "silent" option to run the commands but not displaying any informational messages,
- The "default" option of the avail command, to display only the default versions of each software, ignoring the other versions which possibly may be available too,
- The "category" option of the avail command, to display only the available versions (or only the default one if the "default" option previously described is used concurrently) of the available software of one category (like compilers, libraries, tools, etc.)
- The "lookup" command to display, among the loaded module files (or, using the "all" option, among all the available ones) which ones change the specified environment variable, to allow to understand the origin of some changes in the environment variables of a user session.

3.6.2 DEISA Common Production Environment

The maintenance of the DEISA Common Production Environment (DCPE) is part of the T3.5b subtask, but this is deeply linked with the task T6.1 of the work package 6. This is why the general activities around the maintenance of the DCPE will be described in the D6.1 deliverable [11], to avoid duplicate information.

3.6.3 Monitoring facilities (INCA)

One of the objectives of the user subtask is to provide information to the users about the status of the DEISA facilities, available software, versions installed etc. INCA is the application used for monitoring of the computing infrastructure operated by DEISA. The main goal of the application is to determine availability and functionality of a system from a user point of view, however at the moment it is used only internally. A detailed description of the tool and the set-up implemented for DEISA is given in section 2.3.1. Here we describe the developments for this reporting period.

INCA monitoring infrastructure is constantly being improved. During the last six months new reporters were developed and installed. A number of existing reporters was updated. Globus GSI-SSH and GridFTP reporters were optimized in order to decrease system load and increase reliability. The INCA web interface was enhanced by the introduction of new views allowing clearer presentation of data. Stalled reporters were originally marked with a star symbol; a different colour code for such reporters was introduced.

During the next months development of the monitoring infrastructure will continue. Several new reporters, for example DEISA GPFS, are currently being implemented. Inca will be upgraded to the latest release, which includes new features and important bug fixes. New DEISA high-performance computing systems installed during the part months will be integrated in the monitoring infrastructure. Inca will be configured to display resource maintenance information.

New functionality will be added to the Inca user interface. As mentioned above Inca test results are available only to the DEISA staff. DEISA users could also benefit from having access to this information. For achieving that an additional web interface has to be deployed. Implementation details are currently being discussed. The required functionality should be available to the users in the near future.

3.6.4 Portals

Since April 2008, the DEISA Life Science Portal is operated as a production service. A test has been implemented to monitor the different portal submission nodes and agents spread on the DEISA infrastructure and integrated into the INCA monitoring framework. The currently available applications are BLAST, NAMD, RAxML, DOCK and PHYML.

As part of the eDEISA project the portal working group is currently working on the development of a new open source based portal. The portal deployed currently is based on a proprietary technology, EnginFrame, developed by the NICE Company [13]. In order to be more compliant to standards, and in particular to the JSR-168 standard [14], defining the Portlets specifications, it was decided to develop a new open source based portal. The future portal will interface to the DESHL/Unicore layer and thus become batch agnostic. The working group has just finished a comparison analysis of the different existing open-source portlets frameworks and it is now designing the portal software architecture.

3.6.5 DESHL

The DESHL (DEISA Services for the Heterogeneous management Layer) was developed by the JRA7 activity during the initial 4 year DEISA contract. The DESHL provides command line access to the DEISA UNICORE infrastructure. It is based on a number of Grid standards including the Open Grid Forum's Simple API for Grid Applications also known as SAGA. Currently the DESHL is only capable of interacting with UNICORE 5 based grids like the existing DEISA infrastructure. However as part of the third year of the extended eDEISA

contract, an adaptation of the DESHL for the Web Services-based UNICORE 6 is being evaluated.

The DESHL is in full operation across all DEISA sites and a number of DECI projects (including Helium and Einstein) have used the DESHL for job submission.

During the reporting period there has been discussion on whether or not to integrate some dedicated sensors (reporters in the INCA vocabulary) to monitor the DESHL service in the general DEISA middleware monitoring framework. Since the technology underlying the DESHL is UNICORE, which is already integrated in the monitoring framework, and as the independent DESHL Java client is in principle used on the personal computer of the users and not on the supercomputers, it was decided that additional monitoring for DESHL would not be useful.

3.6.6 First Level Help Desk

The general description of the DEISA Trouble Ticket System (DTTS), both used to keep track of the operational problems diagnosed internally and of the ones reported by the users when using the DEISA infrastructure in production mode, was given in the section 2.3.2, with a technical description of its configuration and usage.

The DTTS is the central tool used by the Help Desk to register and manage all the requests for information and all the problems reported by the users if their project is in production mode. There is a strong link between the task T3.5f, in which the problem tickets are monitored when they arrive - and are distributed to the relevant sites with the specific keyword which categorizes the request - and the task T6.3, which is the Help Desk, which takes care, after that they were assigned, of the registered tickets until their resolution. To avoid duplicate information the activities of the Help desk are described in the D6.1 deliverable [11], in the section for the task T6.3.

4 Adopting new technologies from WP4

The second objective for WP3 is defined in the DoW as "Adopting new technologies from WP4". As new or enhanced developments provided by WP4 need to become part of the production infrastructure, the Operations Team of WP3 integrates these in the production service with assistance from the Technology Team (WP4) and operates and supports them thereafter.

After a successful evaluation by WP4 of a new technology for use by DEISA the service based on this new technology must be prepared by WP4 to become production ready. In this preparatory phase WP4 will address the following issues:

- Identification of the necessary components (both hardware and software) for the service, their configurations (e.g. machine names, software configurations and customization), and policy agreements (e.g. agreements on permissible fields, their semantics, permissible defaults)
- operational procedures for start up and shut down of the service components
- development of monitoring sensors for the service components
- operational procedures for handling alarms generated by monitoring sensors
- the test suites necessary for operational use (e.g. stress test suite for workload aspects or periodic test suite)
- duration of pre-production sub-infrastructure testing and main results/observations; conclusions (e.g. to go or not into operation).

The procedure for taking new services into production by WP3 will be the same procedure as defined for the internal change management (see section 2.3.5) in many aspects. But in addition an assessment of all deliveries by WP4 will have to take place. This procedure has to be defined yet.

As WP4 just started the evaluation of some new services only one new service is now proposed by WP4 as production service:

GSI-SSH - The new service proposed is to provide interactive access to all sites using an openSSH version with GSI functionality of the Globus toolkit version 4 [15]. The reason that this already could be proposed is that most of the preparation already has been done by the eDEISA project. It is expected that WP3 soon can start planning the production status of this service.

5 Integration of Associate Partners

The third objective for this work package is defined as **"**Advancing Operations as a turnkey solution for the future European HPC ecosystem". In the description of work this objective is described as follows: "By improving the operational model it will become straightforward to integrate new sites. Thus the integration of Associate Partners, who have not previously been part of the DEISA infrastructure, is a test of the consortium's ability to perfect a process for the addition of new HPC centers to the operational infrastructure. This would later allow the European HPC ecosystem (supporting everything from Tier-0 down to Tier-2 centers) to adopt DEISA2's operational model. The role of the Associate Partners is to serve as a prototype for the integration of new partners in the HPC ecosystem. The level of integration (tight or loose) is depending on the available network access and the financial engagement of the respective associate partners. Therefore the new partners ETH-CSCS and KTH-PDC will be candidates for a tight integration, while the Russian Academy JSCC will be coupled only loosely via the standard network and by using parts of the middleware layer."

5.1 Progress in first half year

First discussions with both PDC and CSCS have taken place. A first priority is the connection of their facilities with a dedicated network link. Both centers will be able to connect either with 1 or 10 Gbit/s dedicated links. The links are expected to be implemented before the end of 2008.

Also the integration of these sites with other DEISA services has been discussed and documentation has been provided. Most of the implementation work still has to start. The integration of PDC must be completed by project month 9. For CSCS this is project month 15. For JSCC it currently is only planned that access through UNICORE will be implemented for members of this partner. Integration of this partner is planned for month 21.
After the start of the project it has been agreed to integrate another partner, CEA from France, into the infrastructure. Discussions already started and IDRIS is assigned as supporting DEISA partner.

6 Quality Plan

For the operations of the DEISA production infrastructure the DEISA project established a European-wide operations team. This team integrates the local systems and DEISA services into a European wide supercomputing service. This requires intensive synchronisation and tuning of service levels, operational procedures as well as common and best practices among all eleven sites. This is a tremendous challenge that requires much time and effort for evolution. The establishment of the DEISA operations team has started in 2006, during the eDEISA project, and much progress has been made since. In DEISA2 this evolution will be continued with a strong focus of DEISA wide service levels and common operational procedures, like distinct incident-, change and release management. To be able to assess the planned evolution as well as the quality of service of the provision of the DEISA operational services, in DEISA2 quality control and quality assurance measures will be introduced and established through a quality plan.

Quality control and quality assurance requires that:

(1) a service level for every single DEISA service is determined and described
(2) metrics are defined to measure the quality of operational service for every service
(3) instruments are developed to measure the defined metrics for every service at every site; this will include development and implementation of various automatic monitoring, logging and reporting tools
(4) reference measurement of these metrics are done to determine the starting point level of service
(5) procedures are defined to monitor, control and assure the service level for every single DEISA service (these will include procedures for incident-, change- and release management, operator on duty etc.)
(6) quality control and assurance procedures are implemented across all sites

The milestones for the implementation of this quality plan for operations are:

(1) determined and described until M12
(2) determined and described until M12
(3) developed until M15
(4) done during M15-M18
(5) defined until M18
(6) implemented until M24

The planning for the complete implementation of this quality plan is obviously crude and will overall not be sequential with respect of the total set of DEISA services. Obviously this will depend on many factors for a particular service like the current differences in service level, common practices and procedures across sites, the current monitoring and reporting tools already implemented. For the implementation of this quality plan priority will be given to the DEISA core services. The plan will be elaborated and refined by internal discussions among all partners.

Despite the planning of the implementation of this complete quality plan, many quality control and quality assurance measures have already been taken for a number of services within DEISA. These include monitoring of most of the DEISA services; most of the information is provided through INCA. For most of the services the availability of the service is reported. One quality assurance measure already implemented at this moment to ensure high availability of the services is the “operator on duty” task (section 2.3.4).

7 Conclusions

The operations activity in DEISA2 is built upon the experience and best practices of the operations of the DEISA infrastructure in the last four years. Much effort has been spent in the first six months of the project on the organisation of the new team, the set-up and improvement of internal services, and the definition of procedures. Much progress has been made as shown by the results described in section 2; this is an ongoing effort. In the next 6 months emphasis will be put on, e.g. the change management procedures.

The production services of the DEISA infrastructure in general have been stable in this reporting period. Most service interruptions were due to scheduled maintenance or complete system upgrades or replacements. The new systems introduced in this period were integrated in the infrastructure without large problems, which also demonstrates that maintenance procedures for the different services are well documented. For the future we plan to have more concrete statistics on the availability of the services. This is part of the quality plan. Details on the status of the different production services are given in section 3.

The adoption of new technologies provided by WP4, the task described in section 4, has only just been started as the first results of WP4 have to be delivered yet.

Preparations for the integration of new partner sites have been started, among others most importantly the provision of dedicated network links for some of these sites to the DEISA dedicated network. CEA has been added to the list of associated partners. Details are given in section 5.

The outline of a quality plan is given in section 6.