



**SEVENTH FRAMEWORK PROGRAMME  
Research Infrastructures**

**INFRA-2011-2.3.5 – Second Implementation Phase of the European High  
Performance Computing (HPC) service PRACE**



**PRACE-2IP**

**PRACE Second Implementation Phase Project**

**Grant Agreement Number: RI-283493**

**D6.3  
Second Annual Operations Report of the Tier-1 Service**

***Final***

Version: 1.0  
Author(s): Jules Wolfrat, SURFsara  
Date: 20.08.2013

## Project and Deliverable Information Sheet

PRACE Project	Project Ref. №: RI-283493	
	Project Title: PRACE Second Implementation Phase <b>Project</b>	
	Project Web Site: <a href="http://www.prace-ri.eu">http://www.prace-ri.eu</a>	
	Deliverable ID: < D6.3 >	
	Deliverable Nature: <DOC_TYPE: Report >	
	Deliverable Level: PU *	Contractual Date of Delivery: 31/08/2013
		Actual Date of Delivery: 31/08/2013
EC Project Officer: Leonardo Flores Añover		

\* - The dissemination level are indicated as follows: **PU** – Public, **PP** – Restricted to other 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).

## Document Control Sheet

Document	Title: Second Annual Operations Report of the Tier-1 Service	
	ID: D6.3	
	Version: 1.0	Status: Final
	Available at: <a href="http://www.prace-project.eu">http://www.prace-project.eu</a>	
	Software Tool: Microsoft Word 2007	
	File(s): D6.3.docx	
Authorship	Written by:	Jules Wolfrat, SURFsara
	Contributors:	Ralph Niederberger (FZJ), Tilo Eissler (LRZ), Gabriele Carteni (BSC), Thomas Bönisch (HLRS), Andreas Schott (RZG), Liz Sim (EPCC), David Cable (STFC), Thomas Röblitz (UiO), Lukasz Flis (CYFRONET), Antun Balaz (IPB), Filip Stanek (VSB-TUO), Maciej Cytowski (ICM)
	Reviewed by:	Hervé Lozach, CEA; Dietmar Erwin, FZJ
	Approved by:	MB/TB

## Document Status Sheet

Version	Date	Status	Comments
0.1	05/July/2013	Draft	TOC
0.2	08/August/2013	Draft	Contributions for Network, Data, Compute, AAA and Monitoring services and sites UiO, VSB-TUO,

			STFC (HARTREE), IPB, CYFRONET, ICM
0.3	09/August/2013	Draft for internal review	User services still missing
0.4	10/August/2013	Draft for internal review	Includes User services
1.0	20/August/2013	Final version	

## Document Keywords

<b>Keywords:</b>	PRACE, HPC, Research Infrastructure
------------------	-------------------------------------

### Disclaimer

This deliverable has been prepared by the responsible Work Package of the Project in accordance with the Consortium Agreement and the Grant Agreement n° RI-283493. It solely reflects the opinion of the parties to such agreements on a collective basis in the context of the Project and to the extent foreseen in such agreements. Please note that even though all participants to the Project are members of PRACE AISBL, this deliverable has not been approved by the Council of PRACE AISBL and therefore does not emanate from it nor should it be considered to reflect PRACE AISBL's individual opinion.

### Copyright notices

© 2013 PRACE Consortium Partners. All rights reserved. This document is a project document of the PRACE project. All contents are reserved by default and may not be disclosed to third parties without the written consent of the PRACE partners, except as mandated by the European Commission contract RI-283493 for reviewing and dissemination purposes.

All trademarks and other rights on third party products mentioned in this document are acknowledged as own by the respective holders.

## Table of Contents

Project and Deliverable Information Sheet .....	i
Document Control Sheet.....	i
Document Status Sheet .....	i
Document Keywords .....	iii
Table of Contents .....	iv
List of Figures.....	v
List of Tables.....	v
References and Applicable Documents .....	vi
List of Acronyms and Abbreviations.....	vii
Executive Summary .....	1
<b>1 Introduction .....</b>	<b>2</b>
<b>2 Operation and coordination of comprehensive common Tier-1 services .....</b>	<b>3</b>
<b>2.1 Overview of Tier-1 infrastructure.....</b>	<b>3</b>
<b>2.2 Network Services .....</b>	<b>6</b>
<b>2.3 Data Services.....</b>	<b>8</b>
2.3.1 <i>Status of Deployment</i> .....	9
2.3.2 <i>Advanced User Tools</i> .....	10
2.3.3 <i>Multi-cluster Shared File System – MC-GPFS</i> .....	11
<b>2.4 Compute services .....</b>	<b>11</b>
<b>2.5 AAA services .....</b>	<b>14</b>
2.5.1 <i>Public Key Infrastructure - PKI</i> .....	14
2.5.2 <i>User Administration</i> .....	14
2.5.3 <i>Interactive access</i> .....	15
2.5.4 <i>Accounting services</i> .....	16
<b>2.6 User Services .....</b>	<b>17</b>
2.6.1 <i>PRACE Common Production Environment</i> .....	17
2.6.2 <i>User Documentation</i> .....	18
2.6.3 <i>PRACE Trouble Ticket System and Helpdesk</i> .....	18
Helpdesk Statistics .....	19
<b>2.7 Monitoring Services.....</b>	<b>20</b>
<b>2.8 Internal services.....</b>	<b>21</b>
<b>3 Integration of new Tier-1 systems in the PRACE distributed RI .....</b>	<b>22</b>
<b>3.1 Introduction .....</b>	<b>22</b>
<b>3.2 Overview and status .....</b>	<b>23</b>
3.2.1 <i>UiO</i> .....	23
3.2.2 <i>IPB</i> .....	24
3.2.3 <i>VSU-TUO</i> .....	26
3.2.4 <i>STFC (Hartree)</i> .....	27
3.2.5 <i>ICM</i> .....	28
3.2.6 <i>CYFRONET</i> .....	29
<b>4 Annex: PRACE Service Catalogue .....</b>	<b>31</b>
<b>Introduction .....</b>	<b>31</b>
<b>Classification of services.....</b>	<b>32</b>

Core services .....	32
Additional services .....	32
Optional services .....	32
Update procedure.....	33
PRACE Services .....	34
Uniform access to HPC.....	34
PRACE internal interactive command-line access to HPC .....	35
PRACE external (user) interactive command-line access to HPC .....	35
Project submission.....	35
Data transfer, storage and sharing .....	36
HPC Training.....	37
Documentation and Knowledge Base.....	37
Data Visualization .....	37
Authentication .....	38
Authorization .....	38
Accounting .....	39
Information Management.....	39
Network Management .....	40
Monitoring.....	40
Reporting .....	41
Software Management and Common Production Environment .....	41
First Level User Support .....	42
Advanced User Support.....	42

## List of Figures

Figure 1 Overview PRACE network August 2013 .....	7
Figure 2 PRACE Path Discovery web page.....	8
Figure 3 Batch Systems distribution map on PRACE Tier-1 for DECI 10.....	12
Figure 4 UNICORE deployment on PRACE Tier-1 infrastructure.....	13
Figure 5 PRACE LDAP directory tree.....	15
Figure 6 Accounting architecture .....	16
Figure 7 Number of helpdesk tickets for Tier-1 sites.....	19
Figure 8 Distribution of helpdesk tickets over service areas.....	19
Figure 9 Average resolution time of tickets .....	20
Figure 10: PRACE Service provision scheme and contracts to its users.....	31
Figure 11: Update procedure Service Catalogue.....	34

## List of Tables

Table 1 Available Tier-1 systems August 2013 .....	5
Table 2 Provision of DECI resources, an “x” indicates participation .....	6
Table 3 Overview of installed GridFTP and Globus Toolkit versions at Tier-1 sites.....	9
Table 4 gtransfer deployment status August 2013 .....	10
Table 5 PRACE <i>modulefiles</i> .....	17
Table 6 Provision of DECI resources by new partners, x = participation. With start dates of resource allocations given below DECI-x .....	22
Table 7 Status of the integration for UiO .....	24

Table 8 Status of the integration for IPB.....	25
Table 9 Status of the integration for VSB-TUO.....	26
Table 10 Status of the integration for the Hartree Centre.....	28
Table 11 Status of the integration of ICM's Boreasz system.....	29
Table 12 Status of the integration for CYFRONET.....	30
Table 13 Classification of PRACE Services as part of the PRACE Service Catalogue.....	32
Table 14 Overview of PRACE services, categories and product classes.....	44

## References and Applicable Documents

- [1] <http://www.prace-project.eu>
- [2] PRACE-2IP deliverable D6.1 [First Annual Operations Report of the Tier-1 Service](#)
- [3] Globus GridFTP: <http://www.globus.org/toolkit/docs/latest-stable/gridftp/>
- [4] gtransfer: (<https://github.com/fscheiner/gtransfer>)
- [5] uberftp: [http://www.globus.org/grid\\_software/data/uberftp.php](http://www.globus.org/grid_software/data/uberftp.php)
- [6] UNICORE: <http://www.unicore.eu/>
- [7] Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile RFC 5280: <https://datatracker.ietf.org/doc/rfc5280>
- [8] EUGridPMA: <http://www.eugridpma.org>
- [9] IGTF: <http://www.igtf.net/>
- [10] PRACE distribution of CA information: <http://winnetou.sara.nl/deisa/certs>
- [11] Globus GSI-OpenSSH:  
<http://www.globus.org/toolkit/docs/5.2/5.2.4/gsiopenssh/#gsiopenssh>
- [12] Usage Record – Format recommendation: <http://www.ogf.org/documents/GFD.98.pdf>
- [13] Accounting Facilities in the European Supercomputing Grid DEISA, J. Reetz, T. Soddemann, B. Heupers, J. Wolfrat eScience Conference 2007 (GES 2007)  
<http://edoc.mpg.de/get.epl?fid=36146&did=316646&ver=0>
- [14] DART: <http://prace-ri.eu/Accounting-Report-Tool>
- [15] PRACE User Documentation: <http://www.prace-ri.eu/Documentation-and-User-Support>
- [16] Inca home page: Inca: <http://inca.sdsc.edu/>
- [17] BSCW - Basic Support for Cooperative Work: <http://public.bscw.de/en/index.html>
- [18] TWiki® - the Open Source Enterprise Wiki and Web 2.0 Application Platform  
<http://twiki.org/>
- [19] TRAC <http://trac.edgewall.org/>
- [20] Subversion <http://subversion.tigris.org>

## List of Acronyms and Abbreviations

AAA	Authorization, Authentication, Accounting.
AMD	Advanced Micro Devices
API	Application Programming Interface
BSC	Barcelona Supercomputing Center (Spain)
BSCW	Basic Support for Cooperative Work
CA	Certificate Authority
CEA	Commissariat à l'Energie Atomique (represented in PRACE by GENCI, France)
CINECA	Consorzio Interuniversitario, the largest Italian computing centre (Italy)
CINES	Centre Informatique National de l'Enseignement Supérieur (represented in PRACE by GENCI, France)
CLE	Cray Linux Environment
CP/CPS	Certification Policy and Practice Statement
CPU	Central Processing Unit
CSC	Finnish IT Centre for Science (Finland)
CSCS	The Swiss National Supercomputing Centre (represented in PRACE by ETHZ, Switzerland)
DDR	Double Data Rate
DN	Distinguished Name (as used in X.500 style directories)
DECI	Distributed Extreme Computing Initiative
DEISA	Distributed European Infrastructure for Supercomputing Applications. Former EU project by leading national HPC centres.
DIT	Directory Information Tree (LDAP)
DMA	Direct Memory Access
DPMDB	DECI Project Management Database
DRAM	Dynamic Random Access memory
EC	European Community
EP	Efficient Performance, e.g., Nehalem-EP (Intel)
EPCC	Edinburg Parallel Computing Centre (represented in PRACE by EPSRC, United Kingdom)
EPSRC	The Engineering and Physical Sciences Research Council (United Kingdom)
ETHZ	Eidgenössische Technische Hochschule Zuerich, ETH Zurich (Switzerland)
ESFRI	European Strategy Forum on Research Infrastructures; created roadmap for pan-European Research Infrastructure.
FC	Fiber Channel
FDR	Fourteen Data Rate
FP	Floating-Point
FZJ	Forschungszentrum Jülich (Germany)
GB	Giga ( $= 2^{30} \sim 10^9$ ) Bytes ( $= 8$ bits), also GByte
Gbps	Giga ( $= 10^9$ ) bits per second, also Gb/s and Gbit/s
GB/s	Giga ( $= 10^9$ ) Bytes ( $= 8$ bits) per second, also GByte/s
GCS	Gauss Centre for Supercomputing (Germany)
GÉANT	Collaboration between National Research and Education Networks to build a multi-gigabit pan-European network, managed by DANTE. GÉANT2 is the follow-up as of 2004.
GENCI	Grand Equipement National de Calcul Intensif (France)
GFlop/s	Giga ( $= 10^9$ ) Floating point operations (usually in 64-bit, i.e. DP) per second, also GF/s



GHz	Giga (= $10^9$ ) Hertz, frequency = $10^9$ periods or clock cycles per second
GigE	Gigabit Ethernet, also GbE
HP	Hewlett-Packard
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
HT	HyperTransport channel (AMD)
IB	InfiniBand
IBA	IB Architecture
IBM	Formerly known as International Business Machines
ICE	(SGI)
IDRIS	Institut du Développement et des Ressources en Informatique Scientifique (represented in PRACE by GENCI, France)
IGTF	International Grid Trust Federation
I/O	Input/Output
ISTP	Internal Specific Targeted Project
IPP	Max Planck Institute for Plasmaphysics
JSC	Jülich Supercomputing Centre (FZJ, Germany)
KB	Kilo (= $2^{10} \sim 10^3$ ) Bytes (= 8 bits), also KByte
KTH	Kungliga Tekniska Högskolan (represented in PRACE by SNIC, Sweden)
LDAP	Lightweight Directory Access Protocol
LRZ	Leibniz Supercomputing Centre (Garching, Germany)
MB	Mega (= $2^{20} \sim 10^6$ ) Bytes (= 8 bits), also MByte
MB/s	Mega (= $10^6$ ) Bytes (= 8 bits) per second, also MByte/s
MFlop/s	Mega (= $10^6$ ) Floating point operations (usually in 64-bit, i.e. DP) per second, also MF/s
MHz	Mega (= $10^6$ ) Hertz, frequency = $10^6$ periods or clock cycles per second
Mop/s	Mega (= $10^6$ ) operations per second (usually integer or logic operations)
MoU	Memorandum of Understanding.
MPI	Message Passing Interface
MPP	Massively Parallel Processing (or Processor)
NFS	Network File System
NIC	Network Interface Controller
NUMA	Non-Uniform Memory Access or Architecture
OS	Operating System
PCPE	PRACE Common Production Environment
pNFS	Parallel Network File System
PMA	Policy Management Authority
PoP	Point of Presence
POSIX	Portable OS Interface for Unix
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PRACE-PP	PRACE Preparatory Phase Project
PRACE-RI	PRACE Research Infrastructure
PRACE-1IP	PRACE First Implementation Phase
PRACE-2IP	PRACE Second Implementation Phase
PSNC	Poznan Supercomputing and Networking Centre (Poland)
QDR	Quad Data Rate
RI	Research Infrastructure
RT	Request Tracker – software for trouble ticket handling
RZG	Rechenzentrum Garching (RZG) of the Max Planck Society and the IPP (Germany)

SAN	Storage Area Network
SURFsara	Dutch national High Performance Computing & e-Science Support Center
SGI	Silicon Graphics, Inc.
SNIC	Swedish National Infrastructure for Computing (Sweden)
SSH	Secure Shell
STFC	Science and Technology Facilities Council (represented in PRACE by EPSRC, United Kingdom)
TB	Tera (= 240 ~ 10 <sup>12</sup> ) Bytes (= 8 bits), also TByte
TFlop/s	Tera (= 10 <sup>12</sup> ) Floating-point operations (usually in 64-bit, i.e. DP) per second, also TF/s
Tier-0	Denotes the apex of a conceptual pyramid of HPC systems. In this context the Supercomputing Research Infrastructure would host the Tier-0 systems; national or topical HPC centres would constitute Tier-1
TSI	Target System Interface, an UNICORE facility
UHeM	National Center for HPC of Turkey (Formerly UYBHM)
UNICORE	Uniform Interface to Computing Resources. Grid software for seamless access to distributed resources.
UCC	UNICORE Command line Client
URC	UNICORE Rich Client
VSb-TUO	Vysoká škola báňská – Technical University of Ostrava (Czech republic)
WCSS	Wrocławskie Centrum Sieciowo-Superkomputerowe (WCNS, Wrocław Centre for Networking and Supercomputing, Poland)
X.509	IETF standard used for certificates in Public Key encryption, RFC 5280

## Executive Summary

The primary objective of WP6 'European HPC Infrastructure Operation and Evolution' in PRACE-2IP is the operation of an infrastructure of national HPC Tier-1 systems in Europe which provides resources to users through PRACE calls for proposals for Tier-1 access (the DECI calls). The Tier-1 infrastructure is tightly integrated with the PRACE Tier-0 infrastructure and for this reason this task is closely collaborating with the corresponding task of PRACE-3IP responsible for the operations of the Tier-0 infrastructure. The services operated by this work package are based on the services listed in the Service Catalogue, which has been defined by WP6 in PRACE-1IP, and which has been updated this year by WP6 of PRACE-3IP. The Service Catalogue is waiting for approval by the AISBL. The Project Management Board has agreed that WP6 should continue to work on the implementation of the services as if the Catalogue had been approved. The management structure of operations is the same as for the Tier-0 infrastructure and the same persons are responsible for particular services on both the Tier-0 and Tier-1 infrastructure. The well established operational procedures and facilities have been successful in the provision of a reliable and highly available infrastructure.

In this project period the PRACE Tier-1 infrastructure has increased from 18 to 21 sites. Five new sites were integrated and two Tier-0 sites were removed as Tier-1 resource. Six sites upgraded their systems this project year. The total peak performance of the Tier-1 infrastructure has increased by more than 50% from 3080 to 4865 TFlop/s.

The dedicated 10Gbps network is operated by PRACE in close collaboration with GÉANT. Two more sites from Poland have been connected with 10Gbps this year using the existing 10Gbps link to Poznan. Four sites have been connected with 1Gbps to the PRACE network through the newly developed IPSEC service. The network monitoring facilities have been enhanced providing link status information for sites connected through the PRACE network.

GridFTP has been deployed and supported at all sites with a link to the PRACE network as the high throughput data management facility for transfers between PRACE sites as well as between PRACE sites and external sites. In addition, some sites support the *gtransfer* tool to ease the management of GridFTP data transfers.

This task supports job submission to the local batch systems and successfully deploys the current version of UNICORE to submit jobs remotely.

A common user administration facility for the distribution of information needed for the authentication and authorisation for PRACE services has been maintained and extended with new partners.

The helpdesk handled more than 380 Tier-1 related support tickets. The User documentation for PRACE services has been updated with the changes in the infrastructure.

The monitoring facilities provided with the Inca tools have been enhanced with several new features and additional facilities, one of which is the information on the version of the deployed services at sites.

The second objective is the integration of new Tier-1 sites. Five new sites which started to provide resources for the DECI calls in this project period have been integrated, while for one more site the preparations started but only can be finished later because of a delay in the delivery of their system. This brings the total number of successful integrations of new Tier-1 sites for the whole PRACE-2IP project period to eleven, so almost doubling the total number of Tier-1 sites.

## 1 Introduction

The objectives of WP6 are:

- Provision and operation of comprehensive common services at the Tier-1 level, and integration of those Tier-1 resources and services in the PRACE distributed Tier-0 infrastructure;
- Integration of new Tier-1 services and systems

For both objectives a dedicated task is defined. Task 6.1 is responsible for the operation of the Tier-1 services and the achievements and status are described in section 2.

Section 3 describes the achievements and status of task 6.2, Integration of new Tier-1 systems in the PRACE distributed RI.

In the annex 4 the current version of the Service Catalogue is included.

## 2 Operation and coordination of comprehensive common Tier-1 services

This section describes the work done by task WP6.1. The objective of this task is to support the coordinated and integrated operations of the comprehensive and common Tier-1 services. These services include:

- Network services – network connectivity through the GÉANT network and a dedicated network provided by GÉANT, monitoring of the network, and collaboration with the GÉANT NOC;
- Data services – GridFTP, GPFS and data management tools;
- Compute services – Local batch schedulers, UNICORE, Globus GRAM;
- AAA – PKI, user administration, accounting and interactive access;
- User services – Common production environment based on modules, first and second level support and a help desk, user documentation and advanced application production assistance;
- Monitoring – Service availability with Inca, maintenance information;
- Internal or generic services – Collaborative workspaces, software versioning facility, DECI project management database.

The set of services and software tools supported are described in the Service Catalogue, which lists and describes the complete set of operational services that the PRACE RI is providing, both for Tier-0 and Tier-1 services. The purpose of the PRACE Service Catalogue is:

- To describe all PRACE operational services
- To define PRACE service categories, and classify all PRACE services accordingly

Three service classes have been defined: Core services, Additional services and Optional services. The Service Catalogue has been updated in the past year in discussions with all Tier-0 and Tier-1 sites in the operations meetings. The new version has been submitted by WP6 of PRACE-3IP to the PRACE hosting members for approval. The Project Management Board has agreed that WP6 should continue to work on the implementation of the services as if the Service Catalogue had been approved. This version is appended in the Annex (section 4). The Service Catalogue also provides for every service the actual product or products used. These products have been classified too and separate classifications are given for the Tier-1 and Tier-0 infrastructure, as the classification can differ between the two infrastructures.

Bi-weekly video conferences are attended by the service leaders to discuss all operational issues: like the status of the infrastructure, and the status of proposed and planned changes. Minutes of all meetings are published on the BSCW pages.

### 2.1 Overview of Tier-1 infrastructure

Table 1 gives an overview of the available resources of the Tier-1 infrastructure by the end of the second project year, August 2013. Compared to the available resources at the end of the first project year the total compute capacity increased from 3080 to 4865 TF/s, an increase of almost 60 percent. A large part of this increase is due to the addition of one system, the BlueGene/Q of STFC's Hartree centre in the UK. The number of Tier-1 sites increased from 18 to 21 in this project period. Five new sites were integrated in the infrastructure: CYFRONET, ICM, STFC, VSB-TUO and UiO. Two sites, HLRS and LRZ, only offer Tier-0 resources at the moment. IDRIS is not in the table because they are in the middle of a migration. Six sites upgraded their systems in this period: BSC, CINECA, CSC, EPCC, RZG

and SURFsara. Some sites also provide access to more than one system. The total amount of memory only increased with just more than 10 percent while the number of cores increased with more than 40 percent.

It's interesting to see that the distribution of systems over vendors is quite evenly spread, with IBM and SGI both counting for five systems and Bull, CRAY and HP each for four systems. The only other vendor is MEGWARE with one system. So, the Tier-1 infrastructure is offering a wide portfolio of system architectures, not dominated by a single vendor. The CPU offering is clearly dominated by Intel.

Site	Architecture	TF/s	#Cores	Total Memory (TB)
BSC (Spain)	Bull Bullx B505 CPU: Intel Westmere-EP 6-core 2.53GHz + Nvidia Tesla M2090 Interconnect: InfiniBand QDR / 2-plane Full Fat Tree	185.8	1536	3.1
CINECA (Italy)	IBM iDataPlex DX360M3 CPU: Intel Westmere-EP 6-core 2.40GHz + Nvidia Tesla M2070 Interconnect: InfiniBand QDR / Fat Tree	293.1	3288	12.8
CINES (France)	SGI Altix ICE 8200 CPU: Intel/Xeon Harpertown Quad Core/3.0 GHz Interconnect: Infiniband DDR / Dual plane Hypercube	147.3	122800	48.0
	SGI Altix ICE 8200 CPU: Intel/Xeon Nehalem Quad Core/2.8 GHz Interconnect: Infiniband DDR / Dual plane Hypercube	120.4	10752	47.2
CSC (Finland)	Cray XC30 CPU: Intel Sandy Bridge-EP 8-core 2.6GHz Interconnect: Cray Aries / Dragonfly	244.9	11776	23.0
CSCS (Switzerland)	Cray XE6 CPU: AMD Interlagos 16-core 2.1GHz Interconnect: Cray Gemini / 3D Torus	400.0	47872	46.75
CYFRONET (Poland)	HP BL685c G7 CPU: AMD Interlagos 16-core 2.3GHz Interconnect: Infiniband QDR	61.2	6656	26.0
	HP SL390s CPU: Intel Westmere-EP 6-core 2.45GHz + Nvidia Tesla M2090 Interconnect: Infiniband QDR	136.8	528	3.6
EPCC (UK)	Cray XE6 CPU: AMD Interlagos 16-core 2.3GHz Interconnect: Cray Gemini / 3D Torus	360.0	44544	59.4
FZJ (Germany)	Bull NovaScale R422-E2 CPU: Intel Nehalem-EP quad-core processors 2.93 GHz Infiniband QDR Full Fat Tree topology	207.0	17664	53.0

ICHEC (Ireland)	SGI Altix ICE 8200 CPU: Intel Xeon Woodcrest X5650 2.67GHz hex-core Interconnect: Infiniband DDR / dual plane Hypercube	41.0	3840	7.7
ICM (Poland)	IBM Power 775 CPU: IBM Power7 8-core 3.836GHz Interconnect: HFI / 2-level direct graph	78.4	2560	10.2
KTH (Sweden)	Cray XE6 AMD Magny-Cours 12-core 2.1GHz Interconnect: Cray Gemini / 3D Torus	305.6	36384	47.4
NCSA (Bulgaria)	IBM BlueGene/ P CPU: IBM PowerPC 450, 850MHz Interconnect: IBM BG/P 3D Torus	27.85	8192	12.0
PSNC (Poland)	SGI UV 1000 CPU: Intel Westmere-EX 8-core 2.67GHz Interconnect: NUMalink® 5, 2D torus	21.8	2048	16.0
	SGI/Rackable C1103-G15 CPU: AMD Interlagos 12-core 2.40GHz + Nvidia Tesla M2050 Interconnect InfiniBand QDR, Fat Tree	224.3	5448	10.6
RZG (Germany)	IBM iDataPlex cluster CPU Intel Sandy Bridge-EP 2x8-core 2.6GHz Interconnect: InfiniBand 4xFDR14 / Fat Tree	203.0	9760	40.3
STFC (UK)	IBM BlueGene/Q CPU: IBM PowerPC A2 16-core 1.60GHz Interconnect:	1258.3	98304	96.0
SURFsara (Netherlands)	Bull Bullx B510 CPU: Intel Ivy Bridge en Intel Sandy Bridge Interconnect: InfiniBand 4 × FDR	271.0	13984	41.75
UHeM (Turkey)	HP Proliant BL460 Cluster CPU: Nehalem-EP 4-core 2.67GHz. Interconnect: InfiniBand DDR / Fat Tree	2.5	256	0.75
UiO (Norway)	MEGWARE MiriQuid CPU: Intel Sandy Bridge-EP 8-core 2.6GHz Interconnect: Infiniband FDR/ fat tree	178.6	9984	39.4
VSB-TUO (Czech Republic)	Bull Bullx B510 and B515 CPU: Intel Sandy Bridge-EP 8-core 2.4GHz (180 nodes) and 2.3GHz (27 nodes) Interconnect: Infiniband QDR / Fat Tree	66.0	3312	13.2
WCSS (Poland)	HP Cluster Platform 3000 BL2x220 CPU: Intel Westmere-EP 6-core 2.67GHz Interconnect: Infiniband DDR Full fat-tree	30.0	4848	9.6
Totals		5012	466336	668

Table 1 Available Tier-1 systems August 2013

Table 2 gives an overview of the DECI commitments of all Tier-1 sites from DECI-7 to DECI-10. For each call the start date of access to the resources is given. Five or more percent of the Tier-1 resources are available for DECI calls on a yearly basis. Not all sites participate in each call because of system upgrades and migrations.

Site (Country)	DECI-7	DECI-8	DECI-9	DECI-10
start date	November 2011	May 2012	November 2012	May 2013
BSC (Spain)	x	-	x	-
CINECA (Italy)	x	x	x	x
CINES (France)	x	x	x	-
CSC (Finland)	x	x	x	x
CSCS (Switzerland)	-	x	x	x
CYFRONET (Poland)	-	-	-	x
EPCC (UK)	x	x	x	x
FZJ (Germany)	x	x	x	x
HLRS (Germany)	x	x	-	-
ICHEC (Ireland)	x	x	x	x
ICM (Poland)	-	-	-	x
IDRIS (France)	x	x	x	-
KTH (Sweden)	x	x	x	x
NCSA (Bulgaria)	x	-	x	
LRZ (Germany)	x	x	-	
PSNC (Poland)	x	x	-	x
RZG (Germany)	x	x	x	x
STFC (UK)	-	-	x	x
SURFSARA (Netherlands)	x	x	x	x
UHEM (Turkey)	-	x	x	x
UIO (Norway)	-	-	x	x
VSB-TUO (Czech Republic)	-	-	x	x
WCSS (Poland)	-	x	-	x

Table 2 Provision of DECI resources, an “x” indicates participation

## 2.2 Network Services

The PRACE network services, which are based on the development work in previous projects, are evolving over time. The work of the second year was primarily focused on the decommissioning of old systems, the integration of new systems at already connected sites and the establishment of connections to new sites. Furthermore, the deployment of new technologies (100 Gb/s) at NRENs and GÉANT led to a lot of scheduled maintenances which, dependent on used technology, resulted in connectivity breaks.

Within the last twelve months the PRACE backbone has been extended by further links. Besides the already connected sites PSNC at Poznan and WCSS in Wroclaw, also the Polish



sites of ICM and CYFRONET were connected in the last quarter of 2012 via the already existing 10 Gb/s link to Poznan and the PIONIER national optical network.

In parallel the IPSEC infrastructure has been set up, so that until now four sites could be connected: UHeM, Istanbul, Turkey, VSB-TUO, Ostrava, Czech Republic, NCSA, Sofia, Bulgaria and CaStoRC, Nicosia, Cyprus.

An overview of the network status in August 2013 is shown in Figure 1.

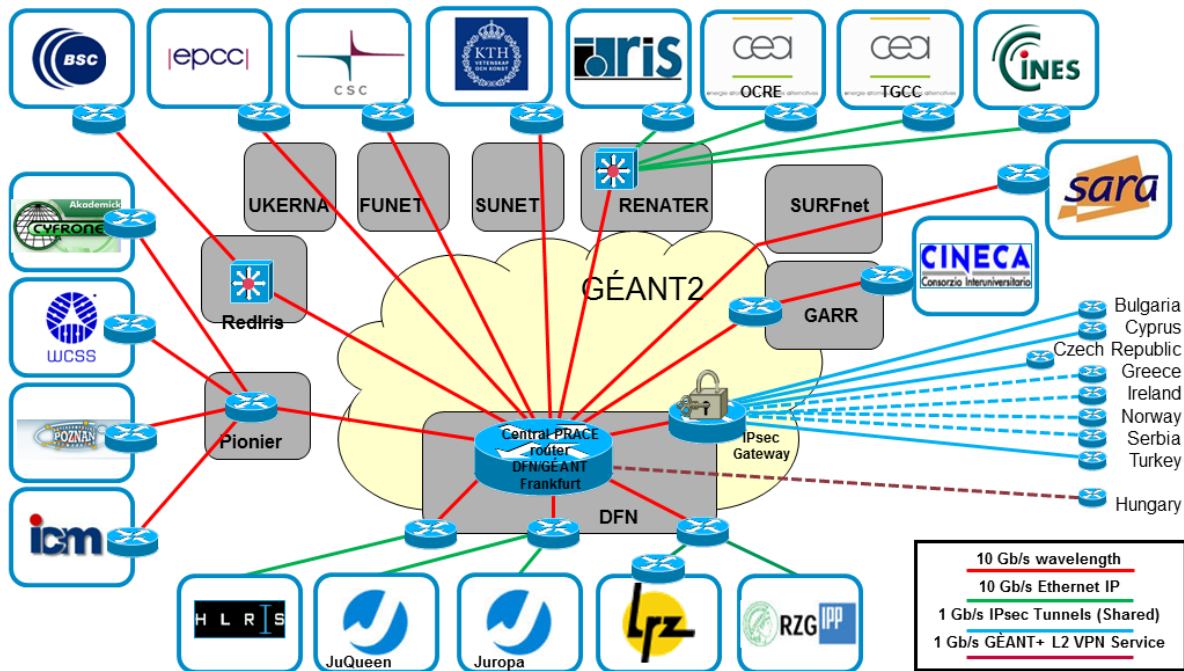


Figure 1 Overview PRACE network August 2013

Network staff of all partners has been engaged in solving network problems, i.e. interrupts because of network hardware, fibre cuts, interface errors, low throughput, routing problems, errors in network configurations, firewall settings, i.e. ports which had not been opened, as well as global network status monitoring and logging.

Within the reporting period more than 50 network maintenance warnings and connectivity outages have been reported and resolved. Most of them did not have any impact on the network itself since the links were optically protected. Where real network cuts had to be announced, those have been repaired within hours, so only small outages occurred. All maintenance has been announced in advance and mostly been scheduled during night times (22:00 – 6:00 CET). Only a small amount of them have been emergency maintenances during daytime hours. During those outages, thanks to the star topology of the PRACE network, only single sites did not have connectivity to the rest of the PRACE partners. The PRACE router hardware used at the central PoP in Frankfurt is working since eight years now, but has run without any problems during the whole reporting period. It is assumed that in the future the hardware will have to be replaced by new components. Alternatively different concepts providing network connectivity between PRACE partners have to be developed. A first analysis of current status and possibilities for future network designs and capacity planning has been provided in PRACE-2IP deliverable D6.2 “Network Capacity Planning”.

Network Monitoring has been enhanced by providing a web page displaying current status of network links. This function has become very valuable, since the End to End Monitoring (E2EMon) service from GÉANT has been cancelled. The E2EMon service collected data about multi-domain circuits which had been composed out of segments. The status data were

provided to service requestors, so that the provider of a service in error could be easily determined.

Since this service is not available anymore a status page displaying ping and traceroute info for all HPC systems has been developed (see Figure 2).

PRACE-Router-Frankfurt				wavelength and/or path				remote PRACE site				
Interface to	Port 10GE	VLAN or Tunnel #	IP Tunnel Source	Link via GEANT2 and/or local NREN				IP Tunnel Destination	link to HPC		HPC address	
CEA-OCRE	1/2	893	10.31.1.129	Frankfurt - Paris via GEANT2 wavelength	Layer2 Switch Renater	Ethernet Paris - Orsay - Bruyeres via Renater		10.31.1.130	Bruyeres-le-chatel - CEA		132.167.248.110	
CEA-TGCC		896	10.31.1.161			Ethernet Paris - Orsay - Bruyeres via Renater		10.31.1.164	Bruyeres-le-chatel - CEAT0		132.167.143.254	
CINES		895	10.31.1.169			Ethernet Paris - Montpellier via Renater		10.31.1.170	Montpellier - CINES		195.83.184.4	
IDRIS		894	10.31.1.73			Ethernet Paris - Orsay via Renater		10.31.1.74	Orsay - IDRIS		130.84.240.135	
HLRS	1/4	406	10.31.1.49	Frankfurt - Stuttgart via DFN wavelength				10.31.1.50	Stuttgart - HLRS		193.196.155.183	
NCSA	1/3	432	10.31.2.1	Frankfurt (Cat2ASR) via local cable	10.31.2.2	10.31.2.9	tunnel # 20 to NCSA	10.31.2.10	Sofia - NCSA		194.141.80.193	
CastoRC		432	10.31.2.1		10.31.2.2	ASR1004	10.31.2.17	tunnel # 30 to CastoRC	10.31.2.18	Nicosia - CastoRC		127.0.0.2
VSU-TuO		432	10.31.2.1		10.31.2.2	Frankfurt	10.31.2.28	tunnel # 40 to VSU-TuO	10.31.2.26	Ostrava - VSU-TuO		195.113.250.161
UHeM		432	10.31.2.1		10.31.2.2		10.31.2.73	tunnel # 100 to UHeM	10.31.2.74	Istanbul - UHeM		160.75.120.190
EPCC	2/1	414	172.31.1.137	Frankfurt - Edinburgh via GEANT2 and Janet wavelength				172.31.1.138	Edinburgh - EPCC		193.62.216.69	
PDC	2/2	417	10.31.1.145	Frankfurt - Stockholm via GEANT2 and NORDUnet wavelength				10.31.1.146	Stockholm - PDC		130.237.230.196	
FZJ-T0	2/3	407	10.31.1.57	Frankfurt - Juelich via DFN wavelength				10.31.1.58	Juelich - FZJ-T0		134.94.115.199	
FZJ-T1									Juelich - FZJ-T1		134.94.115.201	
FZJ-Mon									Juelich - FZJ-Mon		134.94.115.218	
RZG	3/1	404	10.31.1.25	Frankfurt - Garching via DFN wavelength				10.31.1.26	Garching - RZG		130.183.162.38	
LRZ									routed IP RZG - LRZ	130.183.13.2	Garching - LRZ	195.37.7.53
SARA	3/3	405	145.100.9.79	Frankfurt - Amsterdam via GEANT2 and Surfinet wavelength				145.100.9.78	Amsterdam - SARA		145.100.18.14	
PSNC	4/1	418	10.31.1.177	Frankfurt - Poznan via GEANT2 and Pioneer wavelength				10.31.1.178	routed IP Poznan - Poznan	150.254.128.1	Poznan - PSNC	150.254.128.1
ICM									routed IP Poznan - Warsaw	150.254.128.126	Warsaw - ICM	150.254.128.97
CYFRONET									routed IP Poznan - Krakow	150.254.128.65	Krakow - CYFRONET	150.254.128.65
WCSS									routed IP Poznan - Wroclaw	150.254.128.60	Wroclaw - WCSS	150.254.128.33
CINECA	4/2	413	10.31.1.113	Frankfurt - Milano via GEANT2 wavelength				10.31.1.114 GARR	routed IP Milano - Bologna via GARR	10.31.1.122	Bologna - CINECA	130.186.26.1
BSC	4/3	409	10.31.1.41	Frankfurt - Madrid via GEANT2 wavelength	Layer2 Switch Rediris	Ethernet Madrid - Barcelona via Rediris		10.31.1.42	Barcelona - BSC		212.128.224.2	
CSC	4/4	412	10.31.1.89	Frankfurt - Espo via GEANT2, Nordunet and Funet wavelength				10.31.1.90	Espo - CSC		128.214.250.51	

Figure 2 PRACE Path Discovery web page

The information is gathered through "Traceroute" and "ping -R" to HPC addresses from FZJ-Mon, A Prace Monitoring Server located at FZJ, so checking connectivity is going from FZJ-Mon (134.94.115.218) via Frankfurt link to PRACE Frankfurt on the left and then via the different NRENS and GEANT to the specific test IP on the right.

If the font is green then the IP-address is reachable.

If the cell is green then the IP-address has been found on the routing path ("traceroute" and/or "ping -R"). Because of access-list issues some hosts will never become a green cell.

## 2.3 Data Services

GridFTP [3] is a data transfer protocol that can fully utilize the high bandwidths between the PRACE computing centres, so it has been chosen as standard for data transfers between the sites.

GridFTP supports parallel TCP streams and multi-node transfers to achieve a high data rate via high bandwidth connections. Furthermore, third-party transfers can be established, which is very useful for the PRACE users.

In the PRACE-2IP project it has been agreed that GridFTP is a core service at all Tier-1 sites which have a direct connection to the dedicated PRACE network.

### 2.3.1 Status of Deployment

Table 3 gives the deployment status in August 2013 for the sites which deploy GridFTP Services for their Tier-1 system.

Site	Version (GridFTP/GT)	splitting Frontend/Backend	Multiple Backends
BSC (Minotaur)	6.19/GT5.2.4	x	
CEA (Inti)	6.5/GT5.2.0		
CINECA (PLX)	6.14/GT5.2.4	x	
CINES	6.19/GT5.2.4		x
CSC	6.19/GT5.2.3	x	x
Cyfronet	6.19/GT5.2.3		
EPCC	3.28/GT5.0.3		
FZJ (JuRoPa)	3.15/GT4.2.1		
ICHEC	6.10/GT5.2.2		
IDRIS	6.19/GT5.2.4	x	
NCSA	6.10/GT5.2.2		
RZG	6.19/GT 5.2.4	x	
SURFSARA	6.19/GT5.2.4		
UHeM	6.10/GT5.2.1		
UiO	6.10/GT5.2.2		
WCSS	6.05/GT5.2.0		
VSU-TUO	6.32/GT 5.2.4	x	x

**Table 3 Overview of installed GridFTP and Globus Toolkit versions at Tier-1 sites.**

The up-to-date status of the deployment is maintained on the inventory page on the PRACE Wiki. The following new Tier-1 sites have deployed GridFTP successfully during the second year of PRACE-2IP: CYFRONET, NCSA, UIO and VSU-TUO.

As shown in Table 3, 16 sites successfully deploy GridFTP. Five have chosen a split configuration, where the frontend node resides on another machine than the backend node. This is beneficial for security reasons: the backend nodes must have the supercomputer filesystems mounted, but are not directly visible.

Some sites run multiple backend processes, which is an advantage for performance. Only with multiple backend processes, the "-stripe" option can be used for file transfers. This distributes the load of the file streams to different processor cores, if available. Experiments at HLRS have shown that at least two backend processes should be used. Increasing the number of backend processes further may however increase the data transfer performance as more CPU

cores can be utilised then Some sites even distribute those backend processes to more than one backend system.

User documentation has been renewed in August 2013 to reflect the updated deployment status. The internal installation documentation has been kept up-to-date as well.

### 2.3.2 Advanced User Tools

*Gtransfer* [4] is a wrapper script for *tgftp*, a tool for benchmarking, testing and data transfers which itself uses *globus-url-copy* (*guc*), and provides an advanced CLI for performing GridFTP transfers. It also uses *uberftp* [5] to work around some issues in *guc*.

Its functionality and user interface overcome some disadvantages of less advanced clients. For example the user interface includes command line completion of host addresses and remote directory browsing. In addition *gtransfer* also includes the following functionalities:

- data transfers with optimized performance by using pre-optimized data transfer parameters (a user only needs to provide the source and destination of a transfer, additional knowledge is not required)
- multi-step transfers which can bridge different network domains (e.g. PRACE network and Internet)

*Gtransfer* is developed and supported by Frank Scheiner from HLRS.

Late 2012 *gtransfer* has passed the acceptance procedure for a new service and is now an additional service.

In the second year of PRACE-2IP several Tier-1 sites have deployed *gtransfer* on their systems. For August 2013 the *gtransfer* deployment status is given in Table 4.

Site / Tier-1 System	Version
CEA / Inti	v0.0.10a
CINECA / PLX	v0.1.2
CINES / Jade	v0.1.2
CYFRONET / Zeus	v0.1.2
ICHEC / Stokes	v0.0.10a
SURFSARA / Cartesius	v0.1.2
WCSS / Supernova	v0.0.10

**Table 4** *gtransfer* deployment status August 2013

It has to be noted that an additional service is not mandatory to be deployed whenever an unreasonable effort is needed to provide the service.

During the reporting period *gtransfer* was enhanced with additional functionalities that allow for:

- Automatic optimizations depending on the size of files in a transfer
- Interruption and later continuation of transfers
- Automatic retrying of failed transfers

Development of *gtransfer* continues and a new version (0.2.0) is distributed in August 2013 with support for host aliases and persistent identifiers as used and provided by EUDAT.

### 2.3.3 Multi-cluster Shared File System – MC-GPFS

Due to several machine upgrades, partly with architectural changes, the number of sites providing I/O-servers for MC-GPFS decreased down to one site, so that the feature of a common file system spanning across many sites is in the meanwhile depending on the provisioning of I/O server capacity at a single site, namely RZG. RZG will continue to provide the I/O server with backend storage for the currently running DECI-projects. The future of the usage of MC-GPFS depends on strategic decisions concerning future data handling in PRACE. The technical infrastructure is available and the MC-GPFS software is further developed by IBM. Because new systems have a much higher number of nodes, a common shared file system would practically only be provided on the login nodes or some special data-nodes. Furthermore, in order to cope with the growth of the data-volume, investments into larger storage capacities at one or more sites for supporting such a shared file system would be necessary.

## 2.4 Compute services

This activity focuses on three principal compute services that provide job management to the PRACE DECI users. Two of them are considered as core, while the third one is optional:

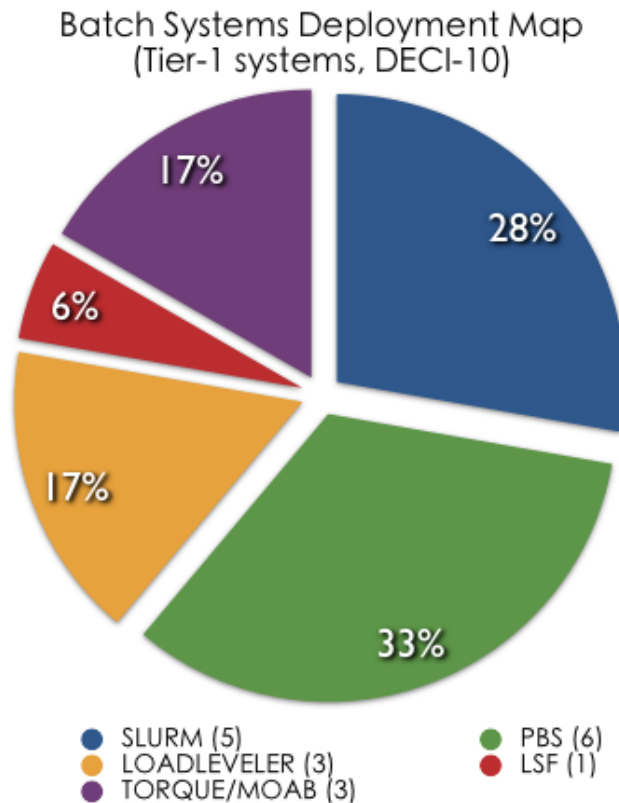
- Local Batch Systems (core)
- UNICORE (core)
- Globus GRAM (optional)

Batch Systems provide users with job management capabilities and interactive command line interfaces. Every supercomputer, and in general every distributed system has such a service for managing its resources (computing units, memory, network and storage) and scheduling user requests asking for an allocation.

Supercomputer centres autonomously manage their specific local batch systems by configuring and tuning them for an optimised delivery of computing services. Since PRACE-1IP, an inventory of all local batch systems deployed in PRACE is maintained on the PRACE wiki. This allows to:

- Keep track of what specific batch system is offered to DECI users;
- Inform other partners using the same technology and enable knowledge sharing and mutual support;
- Collect statistics about what software solutions are deployed in PRACE (such information can be useful for market watch as well as future procurements).

Figure 3 shows the distribution of resource manager and scheduling systems among the 21 PRACE Tier-1 systems participating in the DECI-10 call for which the allocation period started May 2013.



**Figure 3 Batch Systems distribution map on PRACE Tier-1 for DECI 10**

UNICORE [6] offers a uniform interface to all systems which have it installed. It is based on a client/server architecture, based on Java, and it is an alternative solution to the interaction with batch systems.

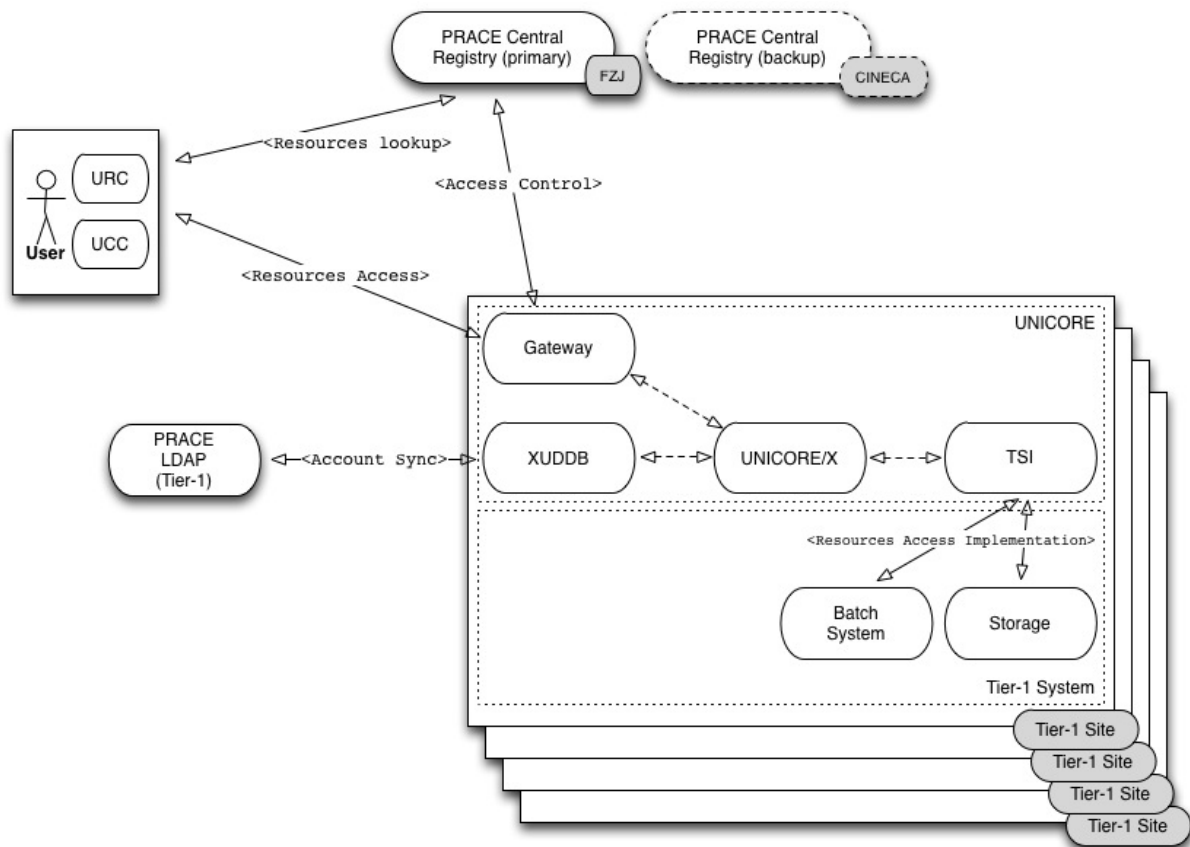
The default version is 6.4.2 although some sites have already deployed version 6.5 or 6.6.

Deployment of UNICORE in PRACE relies on a central registry hosted by FZJ and mirrored by CINECA. On this central registry, which is a directory service for users, sites publish their resources (compute, storage, applications, file transfer protocols, etc...) and how to access them. Users get access to the central registry with their X.509 certificate and manage jobs on different systems via a graphical (URC client) and/or a command line interface (UCC).

A basic design of the infrastructure with the UNICORE components is shown in Figure 4.

It has been discussed to set up an automatic logging mechanism able to distinguish between jobs submitted by UNICORE and those directly submitted by users through a local batch system. This would allow producing a report about usage of UNICORE. The implementation however is not straightforward and it should be considered by PRACE-3IP.

UNICORE shows its maximum potential in a distributed computing infrastructure where users can submit jobs on a wide range of systems. This is not always the case for the Tier-1 infrastructure where multiple allocations (one project running on multiple and different resources) are not as frequent as in a Grid infrastructure. This is also the reason that some new sites only consider the installation of UNICORE if there is a request from users.



**Figure 4 UNICORE deployment on PRACE Tier-1 infrastructure**

A uniform interface to different batch systems is a valuable service since it is the only real alternative to the traditional way for submitting jobs, where you first have to login to each system where you want to submit jobs.

For the installation of UNICORE, specific support is needed for the TSI (Target System Interface) component, which is the interface to the specific local batch system. Implementations are not available for all batch systems, so effort is needed to adapt the TSI to the local environment.

Globus GRAM5 is considered as an optional service, i.e. each PRACE Tier-1 site optionally decides to provide it. The main issue, especially from the performance point of view, is the efficient retrieval of the status of the running jobs. Interactions with different local Resource Managers, which are responsible to provide information about the status of jobs, have been tested and deployed except for LoadLeveler for which the parsing of the log file is still under investigation. Only two Tier-1 systems offer this service to DECI-10 users.



## 2.5 AAA services

The AAA subtask is responsible for services which provide Authentication, Authorization and Accounting facilities on the infrastructure. This includes the support of a Public Key Infrastructure for authentication and a user administration service for the authorisation for services.

### 2.5.1 *Public Key Infrastructure - PKI*

Several PRACE services rely on X.509 certificates [7] for authentication and authorization. These certificates must be issued by entities which are trusted by the service providers. PRACE relies on the Certificate Authorities (CA) accredited as a member by the EUGridPMA, the European Policy Management Authority [8], or by one of the two sister organizations TAGPMA and APGridPMA, all three federated in the IGTF [9]. These PMAs all require a minimum set of requirements for the CP/CPS of the member CAs, as published in a profile document. PRACE is a Relying Party member of EUGridPMA, and as such participates in the discussion of problems and requirements that we have with the deployment of the PKI.

For PRACE a distribution of CA information is maintained at a central repository [10]. The distribution is provided in several formats because services have different requirements for the presentation of the information. This distribution is based on the monthly release of a new distribution by the IGTF. Sites are informed about new updates and must update the local repositories used by services.

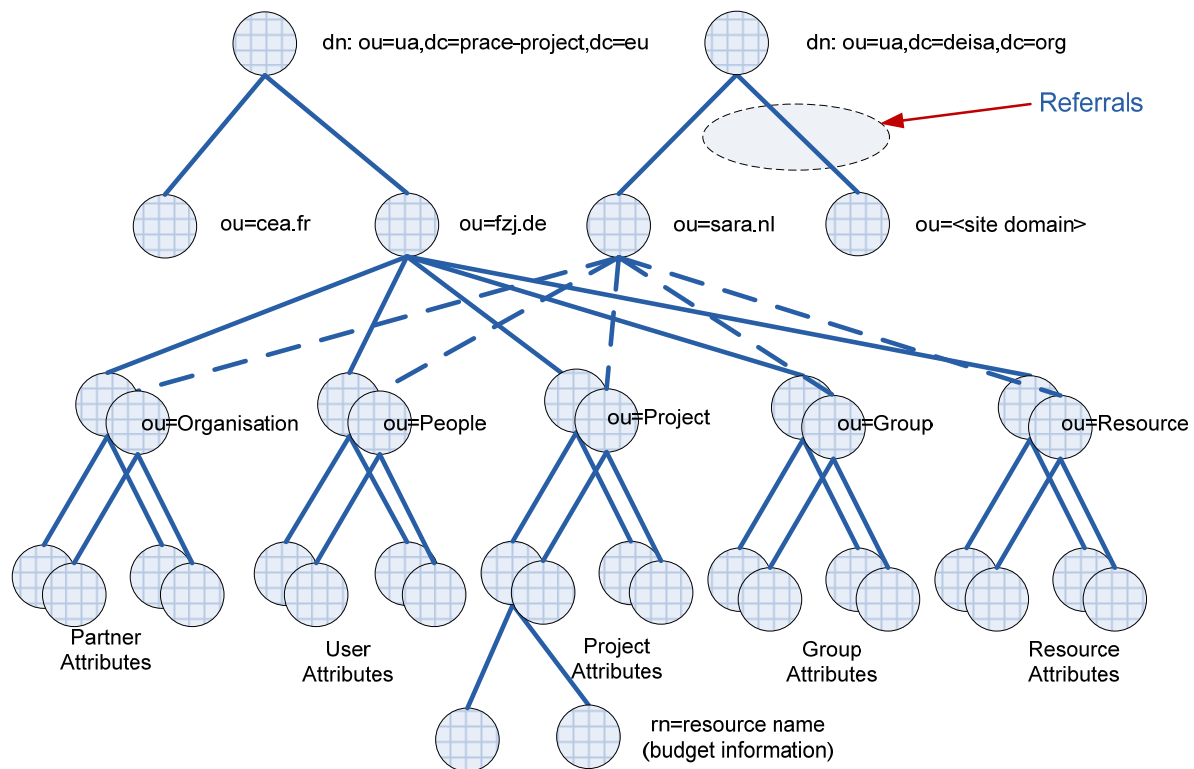
### 2.5.2 *User Administration*

Information about users and their accounts is maintained in an LDAP based repository. This facility is used to update the authorization information needed by services and can be used to retrieve information about users and the projects that they are affiliated to. Authorization information is provided among others for interactive access through GSI-SSH, job submission with UNICORE, accounting services and access to the helpdesk facilities.

The current LDAP infrastructure for Tier-0 and Tier-1 accounts is shown in Figure 5. The top part of the name space for PRACE Tier-0 accounts is “ou=ua,dc=prace-project,dc=eu” and for Tier-1 accounts “ou=ua,dc=deisa,dc=org”. This means that two databases are used, however because the same LDAP schemas are used the same tools can be used to update and retrieve information. The difference in the top level, the suffix, only has an historical reason and no functional reason.

Each Tier-1 site manages its own branch in the LDAP Directory Information Tree (DIT), which is everything below the ou=<site domain> attribute. Each user of Tier-1 resources is assigned to a Home site, which is one of the Tier-1 partners and it's the Home site which creates and updates the account information for the user.





**Figure 5 PRACE LDAP directory tree**

Most partners use the same main LDAP server for managing their branch. Only two partners use their own LDAP server and these can be accessed through referrals from the main server. Three partners have migrated their LDAP branch to the main server in this period. In total 27 branches for Tier-1 accounts are operational, which includes seven sites that were added to the Tier-1 infrastructure this year. The integration of two of these new sites is managed by WP6 of PRACE-3IP.

The main server is operated by SURFsara and for high availability a local replica server is operational too. A remote replica server is operational at HLRS. This server can be used in case the production servers at SURFsara cannot be reached.

In this period the implementation of a new procedure to improve the privacy of users has been started. A new attribute, `praceAccountStatus`, has been introduced in the PRACE ldap schema for this purpose. The new attribute can have the string values “active” for an active account, “closed” for an inactive account and “deleted” for an account which is inactive and for which personal data is removed. The latter is used if a partner has removed all personal data because of privacy requirements. The basic account information will be kept to prevent reuse of the account. The Home site must change status to “deleted” as required by legislation and internal rules. All sites of course must follow the local rules for the deletion of local personal information after an account has been deactivated, also if the status is “closed” and not only “deleted” in LDAP. The new attribute was needed because the existing attribute for the account status was of type Boolean, so only could have two values. The old attribute can be removed once all partners have adapted their procedures to using the new attribute.

The PRACE AAA Administration guide has been updated for the above change.

### 2.5.3 Interactive access

Interactive access to the Tier-1 systems is a basic requirement. For the home site this is in general provided using the SSH (Secure Shell) facilities or the GSI-OpenSSH facility of the

Globus Toolkit [11], or GSI-SSH or *gsissh* for short. This is a core service provided for all systems.

GSI-SSH is also used for interactive access between PRACE sites, using the PRACE network. GSI-SSH uses X.509 certificates for the authentication. The authorisation for interactive access is based on information distributed through the LDAP facilities. Eight Tier-1 sites provide access in this way from other PRACE sites. Three sites – CINECA, LRZ, and SARA – also provide interactive access from outside the PRACE infrastructure with GSI-SSH for all PRACE users. Using these sites users can then access other sites using the internal PRACE network. In addition five other partners provide *gsissh* access from outside the PRACE network for users which have an account on their system.

GSI-SSH\_Term, a GSI-SSH JAVA based client for use on the end user systems, is supported by the PRACE partner LRZ.

#### 2.5.4 Accounting services

Information about the usage of resources is important for users, Principal Investigators (PIs), PRACE sites, and the resource managers, e.g. the DECI call team. PRACE provides facilities to publish and display usage with the following characteristics: 1) the usage of resources is published in a common format, which follows the recommendations of OGF's UR-WG (Usage Record Working Group) [12]; 2) access is based on the authorizations of the requestor, e.g. a normal user can only see his/her personal usage while the principal investigator of a project can see the usage of the whole project. Detailed information about the design considerations can be found in [13].

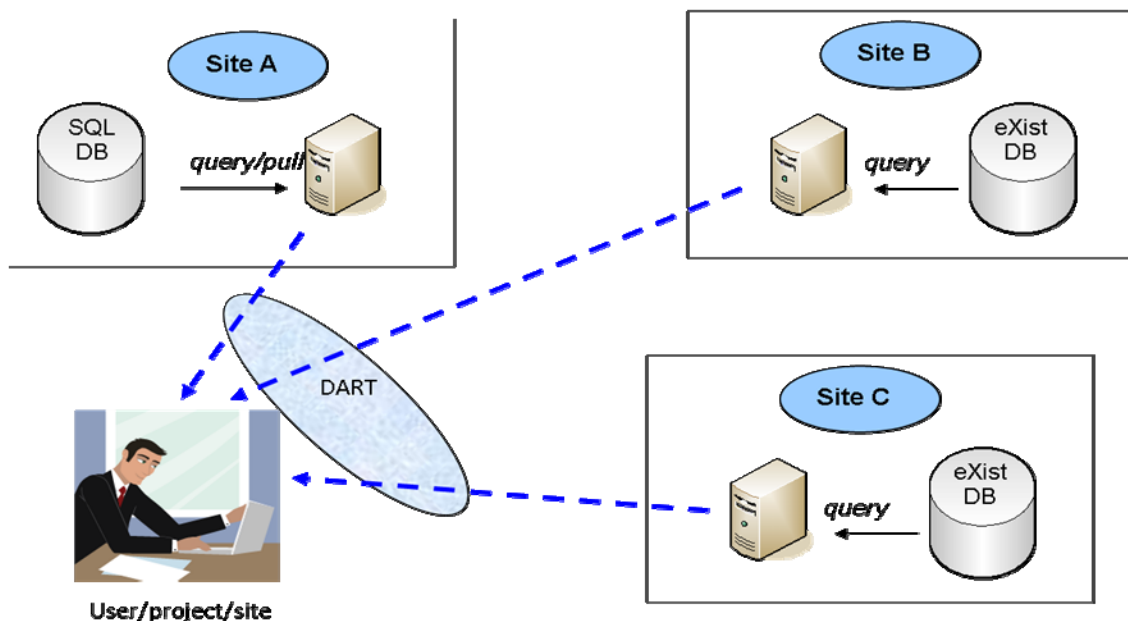


Figure 6 Accounting architecture

Figure 6 shows the basic setup of the facilities. Each site stores usage records for PRACE users in a local database, this can be an eXist database (sites B and C in the figure) or an SQL based database (site A). An Apache/CGI web interface is available which will publish data to authorized clients. The authorization is based on X.509 certificates and the access rights are given by the attribute *deisaAccountRole* of the user administration service. DART [14] is a Java Webstart tool, which can be used by a client to retrieve and to display the information from different sites.

13 sites are fully operational. Most of the new Tier-1 sites are still in the implementation phase of the service.

## 2.6 User Services

### 2.6.1 PRACE Common Production Environment

The PRACE Common Production Environment (PCPE) distribution has been updated to allow for installation on Tier-1 as well as Tier-0 sites. Due to the larger range of architectures present at the Tier-1 level compared to Tier-0 this has entailed introducing more flexibility for individual sites on deciding how the components of the PCPE are implemented locally.

One module called 'prace' is defined that enables access to the PCPE modules and sets the PRACE environment variables. The current set of *modulefiles* is given in Table 5.

Tool	Description
bash	BASH Shell
tcsh	TC Shell
OpenSSH	Secure Shell
Emacs	Emacs Text Editor
nedit	Text Editor
C	C Compiler
C++	C++ Compiler
Fortran	Fortran Compiler
Java	Java Compiler
Perl	PERL Programming Language.
Python	Python Programming Language
TCL	Tool Command Language
TK	CL GUI Toolkit
gmake	GNU make
MPI	Message Passing Interface
BLACS	Basic Linear Algebra Communications
BLAS	Basic Linear Algebra Subroutines
LAPACK	Linear Algebra Package
ScaLAPACK	Scalar Linear Algebra Package
FFTW 2	Fast Fourier Transform in the West (v2)
FFTW 3	Fast Fourier Transform in the West (v3)
HDF5	Hierarchical Data Format
NetCDF	Network Common Data Format

Table 5 PRACE *modulefiles*

The main work in the period was the configuration of the PCPE for the new Tier-1 systems. The installation status of the PCPE on the new Tier-1 sites is included in the tables in Section 3.2. The PCPE is now installed on the majority of sites but some are still in the integration phase.

Work was also completed in the period to integrate the monitoring of the PCPE status into the PRACE Inca monitoring tool.

In addition to deploying the PCPE on the new Tier-1 sites, additional user functionality was added to the PCPE. During the period a new version of the *'prace\_service'* tool was implemented. The *'prace-service'* command is of benefit to users when running any Globus based commands. The user may implement the Globus command without the need to know the details of the destination server or port numbers. For example *gsissh `prace\_service -i -s surfsara`* would invoke a *gsissh* session to the Tier-1 system at SURFsara using the correct server and port details for that site.

Templates for modulefiles are available for download from the PRACE SVN repository which can be utilised by each of the sites. Examples for both linux and Cray XE platforms are currently available. Step-by-step guides are available on our internal PRACE Wiki to aid sites in the configuration process. The aim is to make the PCPE as flexible, but also as straightforward as possible to implement.

### 2.6.2 *User Documentation*

The User documentation for PRACE is available online on the PRACE website [14]. The PRACE documentation is subject to change as new Tier-1 sites come online, or when existing sites deploy a new or modified architecture. All such changes to the documentation are managed following the agreed PRACE Documentation Review Panel procedures.

The user documentation has been updated to include the key information users need in order to access the new Tier-1 sites. This includes details on the interactive access methods, batch systems, and data transfer capabilities. For site specific components which are not part of the PRACE RI, users are directed to the detailed support documentation provided by the Tier-1 sites.

### 2.6.3 *PRACE Trouble Ticket System and Helpdesk*

The centralised PRACE Helpdesk was deployed as part of PRACE-1IP. Those Tier-1 sites providing compute resources for the DECI-9 and DECI-10 calls have now also been integrated into the Helpdesk.

Support staff for the Helpdesk is provided by the PRACE partners on a rotational basis, with each site manning the Helpdesk for one week at a time. A number of the new Tier-1 sites now form an integral part of the rota, providing user support for PRACE.

The PRACE Helpdesk software has been very reliable throughout the reporting period. There have been no major technological changes applied; however the queue configuration has been updated to include the new Tier-1 sites. More than 380 Tier-1 related incidents have been reported and handled in the reporting period.

## Helpdesk Statistics

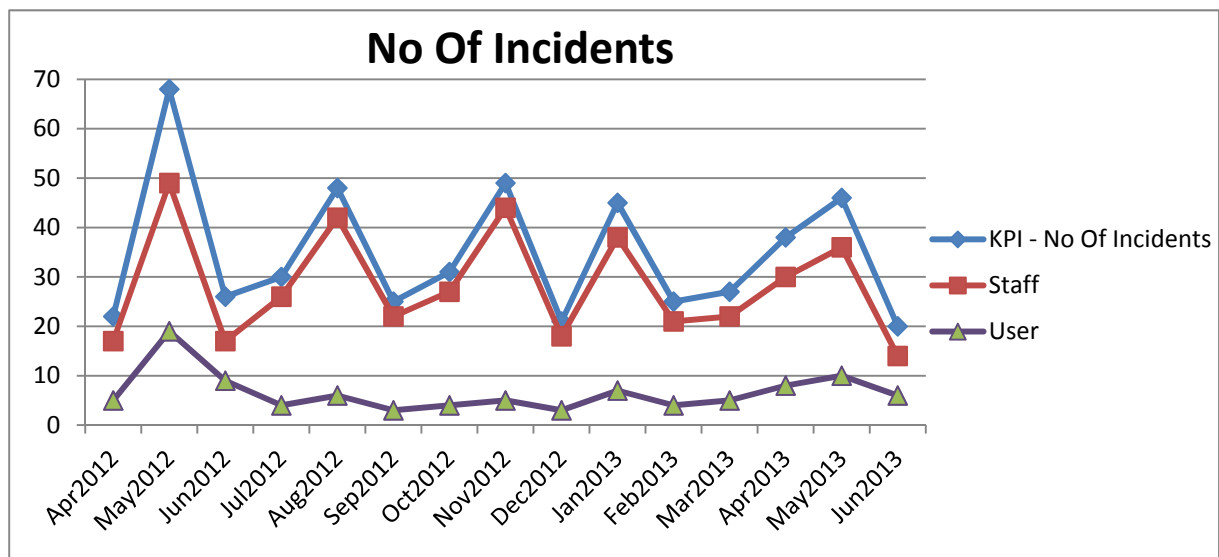


Figure 7 Number of helpdesk tickets for Tier-1 sites

Figure 7 shows the number of newly opened helpdesk tickets each month for Tier-1 sites. Despite the increase in the number of Tier-1 sites, the overall volume of incidents has decreased as the PRACE RI has matured. Peaks of incidents sometimes occur at the start of DECI calls, but this is to be expected as new users are created and there is a temporary increase in associated low level requests. This is further illustrated in the data in Figure 8, which shows the breakdown of Helpdesk tickets by area. The peaks in tickets which occur around the start of DECI calls are generally in the area of AAA.

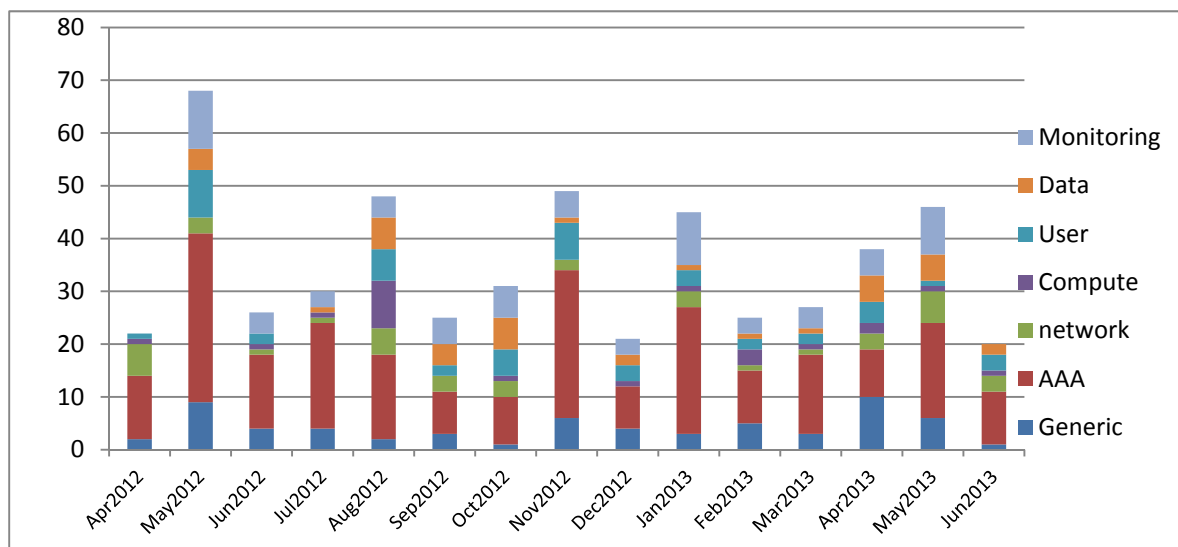
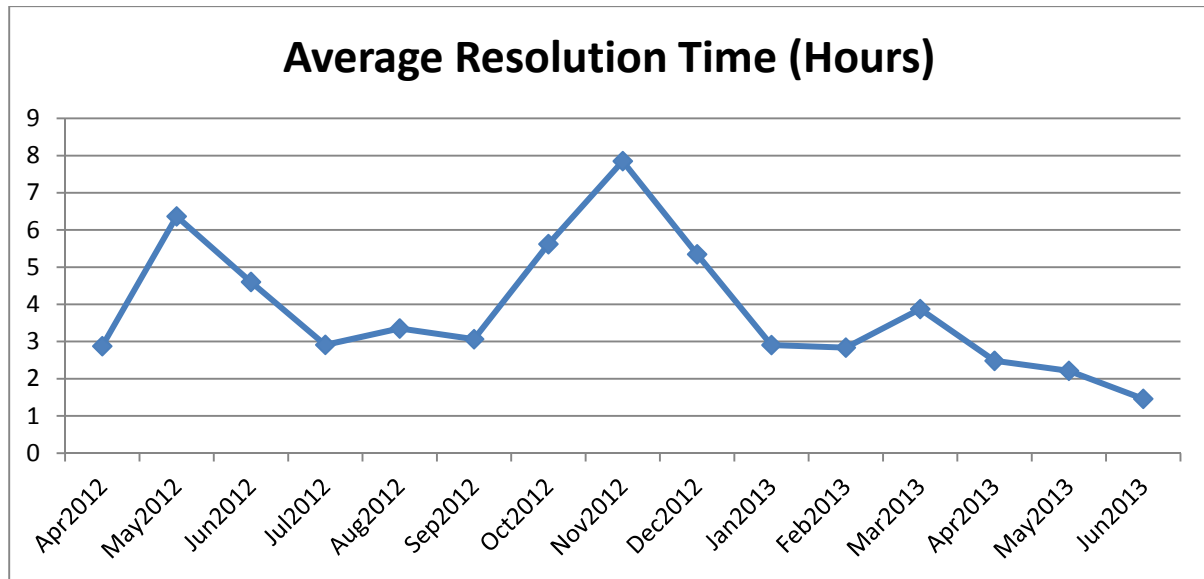


Figure 8 Distribution of helpdesk tickets over service areas



**Figure 9 Average resolution time of tickets**

The average resolution time as shown in Figure 9 also shows a downward trend since the start of 2013. Support staff at the PRACE sites now have the experience to deal with issues quickly. Close working relationships have been built between site support staff and task leaders which also improves the service which can be delivered to the end user. The deployment of common services across all sites ensures a breadth of knowledge, enabling sites to share their experiences at both the regular operational meetings and via dedicated email threads.

As detailed in Section 3.1, all new Tier-1 sites are assigned a support site within PRACE, and this also helps to ensure that any issues users of the new resource face can be addressed as quickly as possible. Support sites also assist the new DECI sites to integrate with the Helpdesk itself.

## 2.7 Monitoring Services

The monitoring subtask implements, deploys and operates tools for the monitoring of PRACE infrastructure services. The subtask focuses on availability and functionality monitoring of services from infrastructure and user perspectives. For this PRACE utilizes the user level monitoring application Inca [16].

The main activity was to keep the Inca monitoring configuration in sync with the existing Tier-1 infrastructure as well as adding new and improved tests.

The configuration was extended by new Tier-1 sites and resources as well as removing Tier-1 resources which were decommissioned.

Further adjustments to the existing configuration for different resources and test suites have been made to improve the Inca monitoring and to make it conform to the new naming convention used within operations, which was implemented this period.

Details about all changes, including the affected sites and resources, are documented in Change Management documents in the internal PRACE wiki. The list of major changes is:

- Added new reporters for middleware tools (PRACE service script version, PRACE service script config version, Globus Toolkit version, GridFTP client version, GSISSH version, gtransfer version, myproxy version);

- The update of the Inca configuration to be consistent with PRACE naming convention;
- Renaming SARA to SURFSARA;
- Added more tests for the tools/libraries defined in the PRACE common production environment (PCPE) (library tests for fftw, hdf5 and netcdf);
- LDAP availability and audit series configured for CYFRONET, GRNET, ICM, IPB, NIIF, VSB-TUO;
- Split of Door Node tests into Door Node, which means all users have access and Open Access, which means only access for those users that have resource allocations for that site;
- The removal of monitoring for resources which went out of business (CSC Louhi, HLRS Laki, RZG Genius and VIP, SURFSARA Huygens);
- Added new systems (CSC Sisu, NCSA EA-Ecnis, PDC Lindgren, PSNC Cane and Chimera, RZG Hydra, SURFSARA Cartesius, UIO Abel);
- Improving Inca maintenance information integration and displaying RTT Ticket number;
- Various Inca frontend corrections;
- Various modifications of existing configuration.

## 2.8 Internal services

For sharing documents among the WP6 partners the PRACE internal collaborative workspace based on BSCW [17] is used.

A wiki server, based on the TWiki® software [18], is used to share information among partners for the installation and configuration of the production services and for other operational activities which need regular updates:

- Change management;
- Maintenance information;
- Operational security;
- Help-desk duty reports;
- Operator on duty reports;
- User documentation reviews.

The service is hosted by FZJ and is shared between the Tier-0 and Tier-1 infrastructure. The service is only used for internal communication.

Based on the *Subversion* software [20] a software version control service is used for the distribution and management of software tools and configuration files that are maintained by PRACE partners. The service is based on the TRAC software [19], which integrates the Subversion software. The service is hosted by SURFsara and is shared between the Tier-0 and Tier-1 infrastructure.

WP6 is responsible for the support of the DPMDB facility, the database which is used to maintain information about the DECI projects. The DPMDB server is hosted by SURFsara.

### 3 Integration of new Tier-1 systems in the PRACE distributed RI

#### 3.1 Introduction

This task is responsible for the integration of new Tier-1 sites into the PRACE Tier-1 infrastructure. Eleven new partner sites have committed resources for DECI project calls in the two years of the PRACE-2IP project. A twelfth new partner, IPB from Serbia, had to postpone their commitments till after the end of PRACE-2IP because of a delay in the delivery of their system. Table 6 gives an overview of the new sites and their commitment of resources for the active DECI calls in the second project year. For each new site a support site is assigned. These act as a contact point for the integration process and are shown in the second column of the table.

Site	Support Site	DECI-7 November 2011	DECI-8 May 2012	DECI-9 November 2012	DECI-10 May 2013
CINES (France)	IDRIS (France)	x	x	x	-
CYFRONET (Poland)	CSC (Finland)	-	-	-	x
ICHEC (Ireland)	EPCC (UK)	x	x	x	x
ICM (Poland)	CSC (Finland)	-	-	-	x
NCSA (Bulgaria)	CINECA (Italy)	x	-	x	-
PSNC (Poland)	CSC (Finland)	x	x	-	x
STFC (UK)	EPCC (UK)	-	-	x	x
UHeM (Turkey) (Formerly UYBHM)	LRZ (Germany)	-	x	x	x
UiO (Norway)	FZJ (Germany)	-	-	x	x
VSB-TUO (Czech Republic)	HLRS (Germany)	-	-	x	x
WCSS (Poland)	CSC (Finland)	-	x	-	x

**Table 6 Provision of DECI resources by new partners, x = participation. With start dates of resource allocations given below DECI-x**



### 3.2 Overview and status

For six sites the integration already had been completed or almost completed at the start of the second project year. These are the sites that participated to the DECI-7 call and or DECI-8 call, which both started allocations before the second project year. Details about these sites and their systems can be found in the first annual report of WP6 of PRACE-2IP [2].

In this section we give an overview of the remaining six new sites that for the first time joined a DECI call in this project year or will join later, as is the case for IPB. A short site description is given and the status of the PRACE services is given based on a table from the Service Catalogue.

#### 3.2.1 UiO

The University of Oslo (UiO) provides Tier-1 services to the PRACE infrastructure with a Linux cluster called Abel and supporting infrastructure. The Abel cluster is based on over 600+ Intel Sandy Bridge based nodes, with additional nodes providing GPU and large memory resources. All nodes are interconnected with FDR InfiniBand.

Service	Product	Service class	Status
<b>Network Management, Monitoring</b>			
Dedicated Network	PRACE Network	additional	Deployment of IPsec solution in progress (finishing in Q8)
Network management, Monitoring	Network monitoring	core	Deployment after IPsec solution in place (foreseen for Q8)
<b>Data</b>	<b>Data</b>		
Data transfer, storage & sharing	GridFTP	core	Available (since DECI-9)
	UNICORE	additional	Deployment in progress (finishing in Q8)
	Gtransfer	Additional	Available (since DECI-10)
	MC-GPFS	optional	Not planned
<b>Compute</b>	<b>Compute</b>		
Uniform access to HPC	Local batch systems	core	Available (SLURM)
	UNICORE	core	Deployment in progress (finishing in Q8)
	Globus GRAM	optional	Not planned
<b>AAA</b>	<b>AAA</b>		
Authentication	PKI	core	Available (since DECI-9)
Authorization, Accounting	User Administration (LDAP)	core	Using the PRACE central server (since DECI-9)
Accounting,	Accounting/Apache/CGI DART	core	Manually transferring data;

Service	Product	Service class	Status
Reporting			Automatic reporting to be deployed (foreseen in Q8)
PRACE internal interactive access	GSISSH	core	Deployment after IPsec solution in place (foreseen for Q8)
	X.509-based SSH	optional	Not planed
PRACE external interactive access	at least one of SSH, GSISSH, X.509-based SSH	core	Available (since DECI-9)
<b>User</b>	<b>User</b>		
Software management & common production environment	Modules	core	Available (since DECI-9)
	PCPE	core	Available (since DECI-9)
Data visualization	Data visualization services & tools	optional	Not planned
<b>Monitoring</b>	<b>Monitoring</b>		
Monitoring	Inca	core	Available (since DECI-9)

Table 7 Status of the integration for UiO

## 3.2.2 IPB

Institute of Physics Belgrade (IPB) will provide Tier-1 services to the PRACE infrastructure with PARADOX Linux cluster which is going to be based on 106 nodes, each with 32 GB of RAM, two eight-core Sandy Bridge processors and one NVIDIA Tesla M2090 GPU. All nodes in PARADOX cluster will be interconnected with QDR Infiniband. Table 8 shows the status of the integration. The preparations have been done by this task and it is expected that this system will become operational in October 2013.

Service	Product	Service class	Status
<b>Network Management, Monitoring</b>			
Dedicated Network	PRACE Network	additional	Planned via shared 1Gbit link, subject to availability of the PRACE IPSEC hardware.
Network management, Monitoring	Network monitoring	core	TBI
<b>Data</b>	<b>Data</b>		
Data transfer, storage & sharing	GridFTP	core	Planned for November 2013
	UNICORE	additional	TBD

Service	Product	Service class	Status
	Gtransfer	Additional	TBD
	MC-GPFS	optional	Not planned
<b>Compute</b>	<b>Compute</b>		
Uniform access to HPC	Local batch systems	core	TBI, Planned for October 2013 – Torque/Maui
	UNICORE	core	TBD
	Globus GRAM	optional	Not planned
<b>AAA</b>	<b>AAA</b>		
Authentication	PKI	core	TBI, Planned for December 2013
Authorization, Accounting	User Administration (LDAP)	core	Available on the PRACE Central server.
Accounting, Reporting	Accounting/Apache/CGI DART	core	TBI, Planned for January 2014
PRACE internal interactive access	GSISsh	core	TBI, Planned for January 2014
	X.509-based SSH	optional	TBD
PRACE external interactive access	at least one of SSH, GSISsh, X.509-based SSH	core	TBI, Planned for February 2014
<b>User</b>	<b>User</b>		
Software management & common production environment	Modules	core	TBI, Planned for December 2013
	PCPE	core	TBI, Planned for January 2014
Data visualization	Data visualization services & tools	optional	Not planned
<b>Monitoring</b>	<b>Monitoring</b>		
Monitoring	Inca	core	TBI, Planned for February 2014

Table 8 Status of the integration for IPB

## 3.2.3 VSB-TUO

Service	Product	Service class	Status
<b>Network Management, Monitoring</b>			
Dedicated Network	PRACE Network	additional	Connected July 15 <sup>th</sup> 2013.
Network management, Monitoring	Network monitoring	core	Set up July 26 <sup>th</sup> 2013.
<b>Data</b>			
Data transfer, storage & sharing	GridFTP	core	Planned (August 2013).
	UNICORE	additional	Planned (August 2013).
	Gtransfer	additional	Planned (September 2013).
	MC-GPFS	optional	Not planned.
<b>Compute</b>			
Uniform access to HPC	Local batch systems	core	Set up July 1 <sup>st</sup> 2013.
	UNICORE	core	Planned (September 2013).
	Globus GRAM	optional	Planned (fall 2013).
<b>AAA</b>			
Authentication	PKI	core	Planned (August 2013).
Authorization, Accounting	User Administration (LDAP)	core	Available. Provided via the central PRACE LDAP server.
Accounting, Reporting	Accounting/Apache/CGI DART	core	Planned (fall 2013).
PRACE internal interactive access	GSISSH	core	Set up July 15 <sup>th</sup> 2013.
	X.509-based SSH	optional	Not planned.
PRACE external interactive access	at least one of SSH, GSISSH, X.509-based SSH	core	Set up July 1 <sup>st</sup> 2013 (SSH).
<b>User</b>			
Software management & common production environment	Modules	additional	Set up June 24 <sup>th</sup> 2013.
	PCPE	core	Set up June 24 <sup>th</sup> 2013.
Data visualization	Various services & tools	optional	Planned (beginning 2014).
<b>Monitoring</b>			
Monitoring	Inca	core	Planned (fall 2013).

Table 9 Status of the integration for VSB-TUO

## 3.2.4 STFC (Hartree)

The Hartree Centre provides Tier-1 services to the PRACE infrastructure with an IBM BlueGene/Q Linux cluster called BlueJoule and supporting infrastructure. BlueJoule has six racks, and contains 98.304 processors. All nodes are interconnected with the proprietary BlueGene 5-dimensional torus. File services are provided by an IBM GPFS parallel filesystem, connected to BlueJoule via an Infiniband network. Table 10 shows the status of the integration.

Service	Product	Service class	Status
<b>Network Management, Monitoring</b>			
Dedicated Network	PRACE Network	additional	Not planned
Network management, Monitoring	Network monitoring	core	TBI (to be implemented)
<b>Data</b>	<b>Data</b>		
Data transfer, storage & sharing	GridFTP	core	Planned
	UNICORE	additional	Not planned
	Gtransfer	Additional	Not planned
	MC-GPFS	optional	Not planned
<b>Compute</b>	<b>Compute</b>		
Uniform access to HPC	Local batch systems	core	Available (IBM LoadLeveler)
	UNICORE	core	Not planned
	Globus GRAM	optional	Not planned
<b>AAA</b>	<b>AAA</b>		
Authentication	PKI	core	Available
Authorization, Accounting	User Administration (LDAP)	core	Planned to be on the PRACE central server (TBI/TBD)
Accounting, Reporting	Accounting/Apache/CGI DART	core	Planned
PRACE internal interactive access	GSISsh	core	Planned
	X.509-based SSH	optional	Not planned
PRACE external interactive access	at least one of SSH, GSISsh, X.509-based SSH	core	Available
<b>User</b>	<b>User</b>		
Software management & common production	Modules	core	Available

Service	Product	Service class	Status
environment	PCPE	core	Available
Data visualization	Data visualization services & tools	optional	Not planned
<b>Monitoring</b>	<b>Monitoring</b>		
Monitoring	Inca	core	Planned

Table 10 Status of the integration for the Hartree Centre

## 3.2.5 ICM

The Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw (ICM) provides Tier-1 services to the PRACE infrastructure with an IBM Power 775 (IH) system called Boreasz. The Boreasz system is based on 80 nodes, each with 32 Power7 compute cores (128 hardware threads) and 128 GB of memory. All nodes are interconnected with IBM dedicated interconnect based on HFI technology. Boreasz runs under the AIX 7.1 operating system. Table 11 shows the status of the integration.

Service	Product	Service class	Status
<b>Network Management, Monitoring</b>			
Dedicated Network	PRACE Network	additional	Available
Network management, Monitoring	Network monitoring	core	Available
<b>Data</b>			
Data transfer, storage & sharing	GridFTP	core	Planned (September 2013)
	UNICORE	additional	Planned (September 2013)
	Gtransfer	additional	Not planned.
	MC-GPFS	optional	Not planned.
<b>Compute</b>			
Uniform access to HPC	Local batch systems	core	Available.
	UNICORE	core	Planned (September 2013)
	Globus GRAM	optional	Not planned.
<b>AAA</b>			
Authentication	PKI	core	Available.
Authorization, Accounting	User Administration (LDAP)	core	Available. Provided via the central PRACE LDAP server.
Accounting, Reporting	Accounting/Apache/CGI DART	core	Planned (October 2013)
PRACE internal interactive access	GSISSH	core	Planned (September 2013)
	X.509-based SSH	optional	Not planned.

Service	Product	Service class	Status
PRACE external interactive access	at least one of SSH, GSISsh, X.509-based SSH	core	Available.
<b>User</b>			
Software management & common production environment	Modules	additional	Available.
	PCPE	core	Planned.
Data visualization	Various services & tools	optional	Not planned.
<b>Monitoring</b>			
Monitoring	Inca	core	Planned (October 2013)

Table 11 Status of the integration of ICM's Boreasz system

### 3.2.6 CYFRONET

Academic Computing Center CYFRONET provides Tier-1 services to the PRACE infrastructure.

All PRACE related services delivered by the CYFRONET Zeus computer, an x86\_64 based cluster machine with a QDR Infiniband interconnect and supporting storage infrastructure. Zeus is a heterogeneous computing platform consisting of 20000+ Intel Westmere and AMD Interlagos cores, 208 GPGPU accelerators. Resources have been integrated with PRACE infrastructure before the beginning of DECI-10 call. The following table represents the current integration status.

Service	Product	Service class	Status
<b>Network Management, Monitoring</b>			
Dedicated Network	PRACE Network	additional	Completed. 10GBe based, dedicated link for PRACE network.
Network management, Monitoring	Network monitoring	core	Completed
<b>Data</b>			
Data transfer, storage & sharing	GridFTP	core	Completed
	UNICORE	additional	Not required for current allocations. TBD in the future
	Gtransfer	Additional	Completed
	MC-GPFS	optional	Not planned. May change in the future if required by allocated projects

Service	Product	Service class	Status
<b>Compute</b>	<b>Compute</b>		
Uniform access to HPC	Local batch systems	core	Completed (Torque+MOAB)
	UNICORE	core	Not required for current allocations. TBD in the future
	Globus GRAM	optional	Not required for current allocations. TBD in the future
<b>AAA</b>	<b>AAA</b>		
Authentication	PKI	core	Completed
Authorization, Accounting	User Administration (LDAP)	core	Completed. Access to Central LDAP granted. Integration with local LDAP infrastructure has been finished.
Accounting, Reporting	Accounting/Apache/CGI DART	core	In progress.
PRACE internal interactive access	GSISSH	core	Completed
	X.509-based SSH	optional	Not planned until requested by DECI users.
PRACE external interactive access	at least one of SSH, GSISSH, X.509-based SSH	core	Completed
<b>User</b>	<b>User</b>		
Software management & common production environment	Modules	core	Available
	PCPE	core	Available
Data visualization	Data visualization services & tools	optional	Not planned
<b>Monitoring</b>	<b>Monitoring</b>		
Monitoring	Inca	core	Completed

Table 12 Status of the integration for CYFRONET



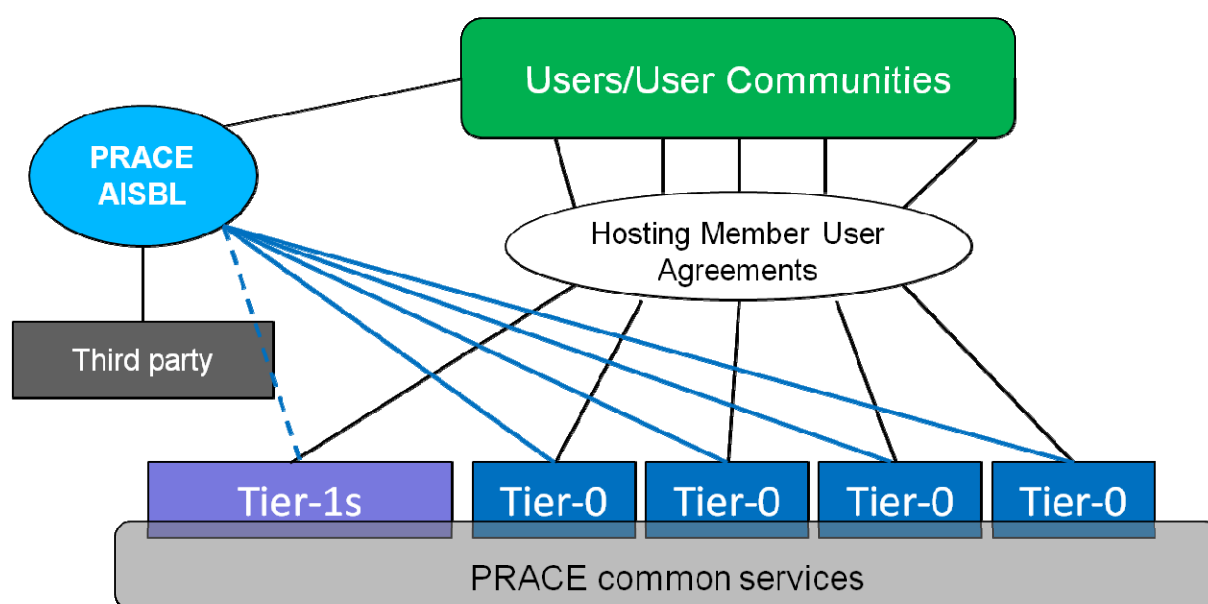
## 4 Annex: PRACE Service Catalogue

This version (2.4) is still under discussion by the PRACE MB.

### Introduction

The PRACE distributed research infrastructure provides a complete set of common services to its users. Service provision to users is done by the Tier-0 hosting partners, governed by the PRACE RI<sup>1</sup> statutes and the Agreement for the Initial Period. Relations between Tier-0 sites and their users are typically managed through specific User Agreements between them. PRACE RI gives advice to the hosting sites on the allocation of compute resources based on the pan-European PRACE Peer Review. For the execution of the peer review and other services such as Web, the PRACE RI also uses services provided by third parties. Other important services such as user support and operation of the distributed infrastructure are provided by the PRACE-IP projects.

Tier-1 partners provide services to users as part of the DECI program for access, currently within the Implementation Phase projects.



**Figure 10:** PRACE Service provision scheme and contracts to its users

To support a good and complete overview of all PRACE Operational Services, the PRACE Implementation projects have developed the PRACE Service Catalogue, which lists and describes the complete set of operational services that the PRACE RI is providing, from the point of view of PRACE as a service provider. In addition, Tier-1 services are added to this Service Catalogue to complete the picture of PRACE service provision.

The purpose of the PRACE Service Catalogue is:

- To describe all PRACE operational services
- To define PRACE service categories, and classify all PRACE services accordingly

In this way it describes the full PRACE service picture from hosting partners, other partners, the project and the PRACE RI.

<sup>1</sup> The PRACE RI is established as an international non-profit association (aisbl) with seat in Brussels

## Classification of services

An important aspect of the PRACE Service Catalogue is the classification of services. Three service classes have been defined: Core services, Additional services and Optional services. The availability and support for each of these service classes is defined and described in Table 1.

Core services	
<b>Availability:</b>	Robust, reliable and persistent technologies that must be implemented and accessible at all PRACE Tier-0/1 sites, or provided centrally.
<b>Support:</b>	Support for these services is provided during support hours, i.e. the normal working hours according to the usual working arrangements of the particular Tier-0/1 site.
Additional services	
<b>Availability:</b>	Robust, reliable and persistent technologies that must be implemented and accessible at all PRACE Tier-0/1 sites where possible. Reasons for the service not being implemented at a Tier-0/1 site include technical, legal, financial and policy limitations, whenever an unreasonable effort is needed to provide the service.
<b>Support:</b>	If applicable, support for these services is provided during support hours.
Optional services	
<b>Availability:</b>	Implemented optionally by PRACE Tier-0/1 sites. Availability and long-term support are not guaranteed by PRACE.
<b>Support:</b>	PRACE RI and/or Tier-1 partners provide support for these services on a case by case basis, in addition to any support provided directly by the specific site.

**Table 13 Classification of PRACE Services as part of the PRACE Service Catalogue**

Every PRACE service will be classified according to this classification. It should be noted that the service classes define the availability of the services at the hosting sites, and are not related to service levels.

The definition of the services in the PRACE Service Catalogue is achieved through six criteria:

- **Description:** A brief summary of the service, indicating its value and a general overview of its implementation.
- **Class:** Services are arranged according to their expected availability and support across PRACE hosting partners. This classification is composed of three levels that indicate how essential a service is for the PRACE RI: Core, Additional, and Optional.
- **Provider:** The person(s), group(s), site(s), or team(s) involved in and responsible for the correct implementation and operation of the services.
- **Reference:** Documents and agreements that contain more specific details and

information concerning the service provision.

- **Category:** Services are grouped into seven different categories, according to their specific domain: Compute, User, Data, Generic, AAA, Network, and Monitoring.
- **Service:** Concrete services and/or software products that have been chosen to implement the service. For each service/product its Service Class (core, additional, optional) is indicated for Tier-0, Tier-1 and/or PRACE RI or a single partner.

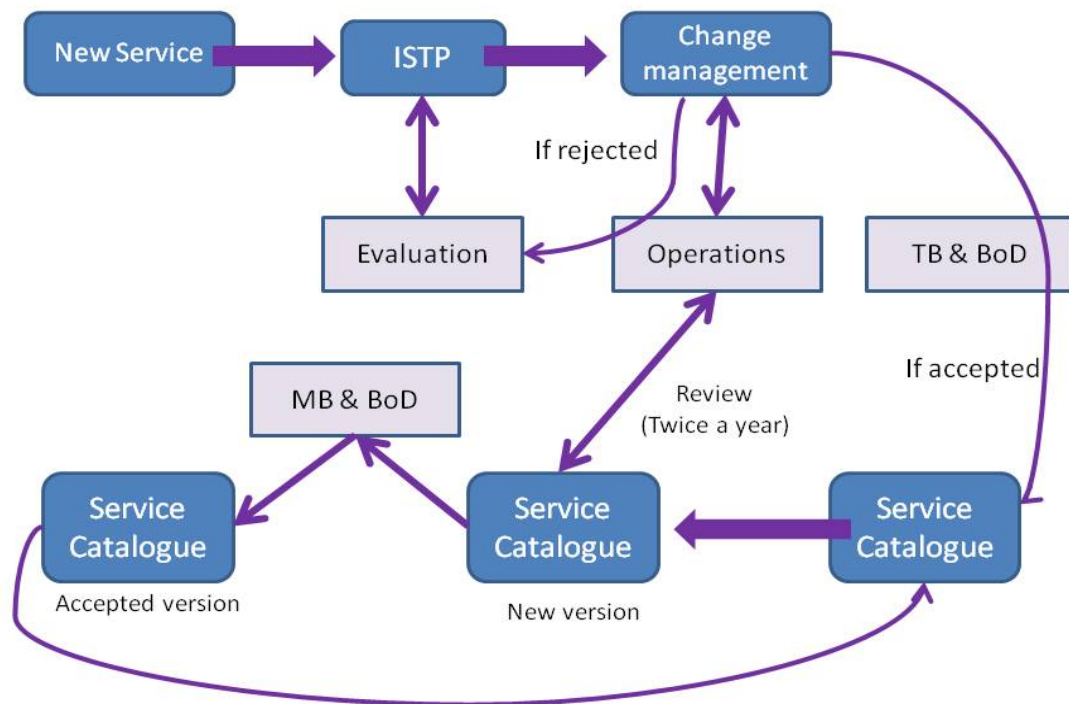
## Update procedure

The PRACE Service Catalogue will be regularly updated to document the actual status of all services and will be maintained as a living document. Status of services can change when new services are deployed, when levels of services are changed, when new service providers (i.e. new hosting partners) are integrated or when new software products are released. The document will at all times reflect the current situation of PRACE services, so that it can be used as the main reference document for service provision within PRACE.

New production services are proposed after an elaborate testing by evaluation teams, e.g. from the Implementation Projects. The evaluation is documented following the ISTP (Internal Specific Targeted Project) procedure. The acceptance as a production service is described by the Change Management procedure for PRACE operational services. The proposal for the new service also proposes the Service classification. In general a new service should be labelled additional or optional at the start. Later a service can be upgraded as core. After acceptance as a production service by the site representatives in PRACE operations this will be communicated to the Technical Boards of the IP projects and PRACE BoD with details of the new service. If within two weeks time after the announcement of the new service no objections have been received from the TB members or the BoD then the service will be considered as accepted by the projects and operations can proceed with the implementation. The new service will be added to a new version of the Service Catalogue. See Figure 11 for a schematic overview of the update procedure

Twice a year the Service Catalogue will be reviewed in a dedicated video/telcon with all site representatives invited. After the meeting the updated Service Catalogue will be distributed for final comments to all site representatives and is accepted if no objections are received within 14 days. A new round of discussion is needed if changes are proposed, either by e-mail or video/telcon. For the updated version again the 14 day period for objections applies. This can be repeated until there is agreement.

After acceptance by the site representatives the new Service Catalogue will be distributed to the PRACE MB and the PRACE BoD for final acceptance.



**Figure 11: Update procedure Service Catalogue**

## PRACE Services

Table 14 at the end of this section gives a concise overview of all services and products.

Uniform access to HPC				
<b>Description:</b>	Allows a user to execute code on PRACE Tier-0/1 systems, monitor its evolution and retrieve the results across Tier-0/1 systems.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 (compute services representative of the PRACE Operational Team)			
<b>Reference:</b>	Draft User Agreement			
<b>Category:</b>	Compute			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	UNICORE	core	core	-
	Globus GRAM	optional	optional	-
	Local batch system	core	core	-
<b>Remarks:</b>	The local batch system is included because it's a prerequisite for the provision of the other tools, however in general it will be different at each site.			

PRACE internal interactive command-line access to HPC				
<b>Description:</b>	Allows an employee or user of a PRACE partner to connect remotely to a Tier-0/1 system using the PRACE dedicated network and execute command-line instructions.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 (compute services representative of the PRACE Operational Team)			
<b>Reference:</b>	NA			
<b>Category:</b>	AAA			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	GSISsh	additional	core	-
	X.509-based SSH	optional	optional	-
<b>Remarks:</b>				

PRACE external (user) interactive command-line access to HPC				
<b>Description:</b>	Allows a user to connect remotely to a Tier-0/1 system and execute command-line instructions.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 (compute services representative of the PRACE Operational Team)			
<b>Reference:</b>	Draft User Agreement			
<b>Category:</b>	AAA			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	At least one of SSH, GSISsh, X.509-based SSH	core	core	-
<b>Remarks:</b>				

Project submission	
<b>Description:</b>	Provides Tier-0 users with a centralized point for submitting projects for Peer Review. In case of Tier-1 access, provision of a DECI database for project registration.
<b>Class:</b>	Core

<b>Provider:</b>	PRACE Peer Review Team			
<b>Reference:</b>	PRACE PP D2.4.2			
<b>Category:</b>	User			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	PRACE peer review tool (for Tier-0 access)	-	-	core
	DECI database (for Tier-1 access)	-	-	core
<b>Remarks:</b>	-			

### Data transfer, storage and sharing

<b>Description:</b>	Each PRACE User is provided a “home” directory and access to a project space shared with his User Group, at each of the assigned Tier-0/1 sites. The amount of space in each of these directories is indicated in Annex A of the User Agreement for Tier-0 sites. Data can be transferred to and from these directories.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 (data services representative of the PRACE Operational Team)			
<b>Reference:</b>	Draft User Agreement			
<b>Category:</b>	Data			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	MC-GPFS	optional	optional	-
	GridFTP	core	core	-
	UNICORE (UFTP)	additional	additional	-
	Gtransfer	additional	additional	
<b>Remarks:</b>	GridFTP is a core service for Tier-1 only if a dedicated network is available Gtransfer details are in change #26: <a href="https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/GtransferDeployment">https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/GtransferDeployment</a>			

HPC Training				
<b>Description:</b>	Provides training sessions and workshops for topics and technologies in high-performance computing, as well as online and offline education material.			
<b>Class:</b>	Core			
<b>Provider:</b>	PRACE 3IP WP4, PRACE 2IP WP4, Tier-0/1 site, PRACE Advanced Training Centres			
<b>Reference:</b>				
<b>Category:</b>	User			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	Training portal	-	-	core
<b>Remarks:</b>	-			

Documentation and Knowledge Base				
<b>Description:</b>	User documentation in the form of an online knowledge base, including manuals and other information and tools that are indispensable for the users.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE RI + PRACE 3IP WP6, WP7, WP3 + PRACE 2IP WP6, WP7, WP3			
<b>Reference:</b>				
<b>Category:</b>	User			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	CMS	-	-	core
	Plone	-	-	core
	DocBook	optional	optional	-
<b>Remarks:</b>				

Data Visualization				
<b>Description:</b>	Converts data into images as a tool to help users with analysis.			
<b>Class:</b>	Optional			
<b>Provider:</b>	Specific PRACE sites			
<b>Reference:</b>				
<b>Category:</b>	User			

Service:	Product/service	Tier-0	Tier-1	PRACE RI or single partner
	Various services and tools	optional	optional	-
Remarks:				

### Authentication

<b>Description:</b>	Confirm the identity of a user and bind that user to a new account. This involves the provision of credentials and identifying a user's X.509 certificate, creating a global PRACE RI account for the user on the central LDAP and making it available for distribution on all PRACE RI Resources.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 (AAA services representative of the PRACE Operational Team)			
<b>Reference:</b>	PRACE Security Policy and IGTF			
<b>Category:</b>	AAA			
Service:	Product/service	Tier-0	Tier-1	PRACE RI or single partner
	PKI	core	core	-
	MyProxy	additional	additional	core
Remarks:	My proxy is provided by multiple partners (e.g. as backup/disaster recovery). Each site must provide the Myproxy client functionality			

### Authorization

<b>Description:</b>	Specifies access rights for each user account created based on the content of the specific User Agreement and the PRACE Security Policy. Ensures that security rules and access rights are obeyed, and manages changes to these (based on new security policies or redefined User Agreements).			
<b>Class:</b>	Core			
<b>Provider:</b>	Peer Review Team + Security Forum + Tier-0/1 site + DECI team + PRACE 3IP/2IP WP6 (AAA services representative of the PRACE Operational Team)			
<b>Reference:</b>	Draft User Agreement, PRACE Security Policy, PRACE Acceptable Use Policy			
<b>Category:</b>	AAA			
Service:	Product/service	Tier-0	Tier-1	PRACE RI or single partner
	LDAP (user administration)	core	core	-



<b>Remarks:</b>	
-----------------	--

### Accounting

<b>Description:</b>	Keeps track of resource usage linked to an account for analysis by users and management. Guarantees that users are not exceeding their limits, as specified by their User Agreement.			
<b>Class:</b>	Core			
<b>Provider:</b>	Peer Review Team + DECI Team + Tier-0/1 site + PRACE 3IP/2IP WP6 (AAA services representative of the PRACE Operational Team)			
<b>Reference:</b>				
<b>Category:</b>	AAA			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	Apache/CGI DART	core	core	-
	LDAP (user administration)	core	core	-
	GridSAFE common accounting repository			additional
<b>Remarks:</b>	GridSAFE details are in change #69: <a href="https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/GridSAFEinProduction">https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/GridSAFEinProduction</a>			

### Information Management

<b>Description:</b>	Provides a common PRACE collaborative environment for sharing relevant information between PRACE sites (BSCW, wiki, subversion, ...).			
<b>Class:</b>	Core			
<b>Provider:</b>	WP6			
<b>Reference:</b>				
<b>Category:</b>	Generic			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	TWiki	-	-	core
	SVN	-	-	core
	BSCW	-	-	core
	Prace-ri website	-	-	core

	e-mail lists			core
<b>Remarks:</b>	-			

## Network Management

<b>Description:</b>	<p>Establishes and maintains network connections between all PRACE nodes (Tier-0 and Tier-1 systems). The PRACE Network Operations Centre (NOC) operates the PRACE backbone network and the corresponding network monitoring system. The PRACE NOC coordinates networking activities of PRACE partners, who are responsible for creation and management of network connection between the local resources and GÉANT (PRACE backbone).</p> <p>PRACE partner's local network specialists and the PRACE NOC should support PRACE users in using the PRACE network infrastructure.</p> <p>The PRACE backbone will be dedicated, whereas local site connectivity of HPC systems and PRACE servers to the global Internet are public.</p>			
<b>Class:</b>	Core			
<b>Provider:</b>	PRACE NOC and local NOCs of PRACE partners (at least one person per site should be also a network services representative of the PRACE Operational Team)			
<b>Reference:</b>	NA			
<b>Category:</b>	Network			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	PRACE dedicated network	core	additional	-
	DNS (PRACE RI domain management)	-	-	core
	PerfSonar framework (Iperf)	core	core	-
<b>Remarks:</b>	<p>Dedicated network is an additional service for Tier-1 partners. This means that a dedicated network is required unless unreasonable effort or funding is required.</p> <p>PerfSonar framework is only a service if a dedicated network is available.</p> <p>Currently only Iperf is used for monitoring the network.</p>			

## Monitoring

<b>Description:</b>	Periodically presents and analyzes up-to-date essential PRACE parameters and service availability to keep track of the situation of the distributed RI, for example: system uptime/downtime and usage levels, network connections,
---------------------	--

	available software and service availability.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 (Monitoring services representative of the PRACE Operational Team)			
<b>Reference:</b>				
<b>Category:</b>	Monitoring			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	Inca	core	core	core
<b>Remarks:</b>	Inca for Tier-0/1 is for client support and PRACE-RI/single server for the server part			

## Reporting

<b>Description:</b>	Periodic reports of system utilization from the Tier-0/1 hosting partner to the PRACE RI.			
<b>Class:</b>	Core			
<b>Provider:</b>	PRACE RI + Tier-0/1 Hosting Partner			
<b>Reference:</b>				
<b>Category:</b>	Monitoring			
<b>Service:</b>	<b>Product/service</b>	<b>Tier-0</b>	<b>Tier-1</b>	<b>PRACE RI or single partner</b>
	Apache/CGI DART	core	core	-
	GridSAFE common accounting repository			additional
<b>Remarks:</b>	-			

## Software Management and Common Production Environment

<b>Description:</b>	Provides software, tools, libraries, compilers, and uniform mechanisms for software and environment configuration. Presents users with a uniform environment across PRACE Tier-0/1 systems, hiding inessential details such as software installation paths.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 + WP7			
<b>Reference:</b>	NA			
<b>Category:</b>	Generic			

Service:	Product/service	Tier-0	Tier-1	PRACE RI or single partner
	Modules	core	core	-
	PCPE	core	core	-
Remarks:	The module command can be implemented in different ways, e.g. by wrapper scripts			

### First Level User Support

<b>Description:</b>	Each PRACE User has access to a centrally managed Helpdesk. Issues raised to the Helpdesk are routed to the appropriate First Level Support team. First Level support is responsible for gathering the user's information and determining their issue by identifying what the user is trying to accomplish, analyzing the symptoms and figuring out the underlying problem.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 (User services representative of the PRACE Operational Team)			
<b>Reference:</b>	Draft User Agreement, PRACE 1IP D6.1			
<b>Category:</b>	User			
Service:	Product/service	Tier-0	Tier-1	PRACE RI or single partner
	Helpdesk (RT-TTS)	-	-	core
Remarks:	For Tier-0 users the helpdesk of the Tier-0 site can be the first level support address			

### Advanced User Support

<b>Description:</b>	Provision of support above and beyond basic problem analysis including but not limited to applications porting, performance tuning, pre-post processing, data access. Higher Level support receives issues that are escalated and routed from First Level User Support.			
<b>Class:</b>	Core			
<b>Provider:</b>	Tier-0/1 site + PRACE 3IP/2IP WP6 + WP7			
<b>Reference:</b>	Draft User Agreement, PRACE 1IP D6.1			
<b>Category:</b>	User			
Service:	Product/service	Tier-0	Tier-1	PRACE RI or single partner
	Helpdesk (RT-TTS)	-	-	core
Remarks:	The helpdesk is needed to support this service			

Service	Service class	Product	Tier-0	Tier-1	PRACE RI or single partner
<b>Network management, Monitoring</b>					
Dedicated network	core	PRACE Network	core	additional	
		DNS			core
Network management, Monitoring	core	PerfSonar framework (lperf)	core	core	
<b>Data</b>	-				
Data transfer, storage & sharing	core	MC-GPFS	optional	optional	
		GridFTP	core	core	
		UNICORE	additional	additional	
		gtransfer	additional	additional	
<b>Compute</b>	-				
Uniform access to HPC	core	Local batch systems	core	core	
		UNICORE	core	core	
		Globus GRAM	optional	optional	
<b>AAA</b>	-			-	
Authentication	core	PKI	core	core	
Authentication	core	MyProxy	additional	additional	core
Authorization, Accounting	core	User Administration (LDAP)	core	core	
Accounting, Reporting	core	Apache/CGI DART	core	core	
		GridSAFE accounting repository			additional
PRACE internal interactive access	core	GSISsh	additional	core	
		X.509-based SSH	optional	optional	
PRACE external interactive access	core	at least one of SSH, GSISsh, X.509-based SSH	core	core	
<b>User</b>	-			-	
Software management & common production environment	core	Modules	core	core	
		PCPE	core	core	

First level user support, advanced user support	core	RT-TTS			core (tool)
Documentation and knowledge base	core	CMS, Plone, DocBook			core
Project submission, Accounting	core	PRACE peer review tool (for Tier-0 access)			core
Project submission, Accounting	core	DECI database (for Tier-1 access)			core
HPC Training	core	Training portal			core
Data visualization	optional	Various services & tools	optional	optional	
<b>Monitoring</b>	-				
Monitoring	core	Inca	core	core	core
<b>Generic</b>					
Information management	core	TWiki			core
		SVN			core
		BSCW			core
		prace-ri website			core
		e-mail list service			core

Table 14 Overview of PRACE services, categories and product classes