E-Infrastructures
H2020-EINFRA-2016-2017

EINFRA-11-2016: Support to the next implementation phase of Pan-European High Performance Computing Infrastructure and Services (PRACE)

PRACE-5IP

PRACE Fifth Implementation Phase Project

Grant Agreement Number: EINFRA-730913

D6.1

First Annual Report on the PRACE Operational Services
Final

Version: 1.0
Author(s): Filip Stanek, IT4Innovations
Date: 20.04.2018

Project and Deliverable Information Sheet

PRACE Project	**Project Ref. №: EINFRA-730913**	
	Project Title: PRACE Fifth Implementation Phase Project	
	Project Web Site: http://www.prace-project.eu	
	Deliverable ID: D6.1	
	Deliverable Nature: Report	
	Dissemination Level: PU*	**Contractual Date of Delivery:** 30 / 04 / 2018
		Actual Date of Delivery: 27 / 04 / 2018
	EC Project Officer: Leonardo Flores Añover	

* - The dissemination level are indicated as follows: **PU** – Public, **CO** – Confidential, only for members of the consortium (including the Commission Services) **CL** – Classified, as referred to in Commission Decision 2005/444/EC.

Document Control Sheet

Document	**Title: First Annual Report on the PRACE Operational Services**	
	ID: D6.1	
	Version: 1.0	**Status:** ***Final***
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2013	
	File(s): PRACE-5IP-D6.1.docx	
Authorship	**Written by:**	Filip Stanek, IT4Innovations
	Contributors:	Eoin McHugh, ICHEC Ralph Niederbeger, JUELICH Xavier Delaruelle, CEA Martin Bidner, HLRS Cristiano Padrin, CINECA Javier Bartolome, BSC Zoltan Kiss, KIFÜ Lukasz Flis, Cyfronet Miroslaw Kupczyk, PSNC
	Reviewed by:	Stephen Booth, EPCC Florian Berberich, JUELICH
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	07/03/2018	Draft	Structure and first elements of the document
0.2	08/03/2018	Draft	First draft of operational services
0.3	09/03/2018	Draft	Inputs added for AAA, Security, Network and User Services
0.4	27/03/2018	Draft	Inputs added for Data and Compute services
0.5	03/04/2018	Draft	Inputs added for Generic services
0.6	04/04/2018	Draft	Inputs added for Monitoring
0.7	04/04/2018	Draft	Draft final version
0.8	04/04/2018	Draft	Wording, polishing of the text itself
0.9	04/04/2018	Draft	Draft version ready for internal review process
1.0	20/04/2018	Final version	Final version ready for EC submission

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, , Operational Services, Tier-0 Systems, Tier-1 Systems, KPI

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° EINFRA-730913. 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.

Table of Contents

List of Figures

List of Tables

References and Applicable Documents

[1] http://www.prace-project.eu
[2] PRACE-3IP Deliverable D6.1.1, First Annual Operation Report, PRACE, June 2013

[3] PRACE-4IP Deliverable D6.2, Final Report on the PRACE Operational Services, PRACE, April 2017

[4] http://doc.aris.grnet.gr/system/hardware/

[5] https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/SystemIntegration

[6] https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/d622792/Information for new sites.ppt

[7] https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/SystemUpgradeProcedure

[8] https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/SystemUpgradeTemplate

[9] https://github.com/globus/globus-toolkit/blob/globus_6_branch/support-changes.md

[10] http://gridcf.org/

[11] https://github.com/gridcf/gct

[12] https://en.wikipedia.org/wiki/Continuous_integration

[13] AARC Project: https://aarc-project.eu/

[14] lRFC 5280, Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile: https://tools.ietf.org/html/rfc5280

[15] EUGridPMA: https://www.eugridpma.org/

[16] IGTF: http://www.gridpma.org/

[17] Globus GSI-OpenSSH: http://www.globus.org/toolkit/security/gsiopenssh

[18] GSISSH-Term: https://www.lrz.de/services/compute/grid_en/software_en/gsisshterm_en/

[19] DART: http://prace-ri.eu/Accounting-Report-Tool

[20] PRACE Helpdesk: https://tts.prace-ri.eu/

[21] Best Practical Request Tracker: https://www.bestpractical.com/rt-and-rtir

[22] PRACE SVN repository: https://prace.osd.surfsara.nl/svn/

[23] MarkDown: https://en.wikipedia.org/wiki/Markdown

[24] Environment Modules: http://modules.sourceforge.net/

List of Acronyms and Abbreviations

AAA	Authorization, Authentication, Accounting
AARC	Authentication and Authorisation for Research and Collaboration
AEGIS	AARC Engagement Group for Infrastructures
aisbl	Association International Sans But Lucratif (legal form of the PRACE-RI)
APGridPMA	The Asian Pacific Grid Policy Management Authority
BGP	Border Gateway Protocol
BSS	Batch Scheduling System
CA	Certificate Authority
CLI	Command Line Interfaces
CoE	Center of Excellence
CPU	Central Processing Unit
CP/CPS	Certificate Policy/Certification Practice Statement
CSIRT	Computer Security Incident Response Team
CUDA	Compute Unified Device Architecture (NVIDIA)
DARPA	Defense Advanced Research Projects Agency
DART	Distributed Accounting Reporting Tool
DEISA	Distributed European Infrastructure for Supercomputing Applications EU project by leading national HPC centres
DoA	Description of Action (formerly known as DoW)
EC	European Commission
EESI	European Exascale Software Initiative
EoI	Expression of Interest
ESFRI	European Strategy Forum on Research Infrastructures
EUDAT	European Data Infrastructure
EUGridPMA	European Grid Policy Management Authority
GB	Giga (= 2^{30} ~ 10^9) Bytes (= 8 bits), also GByte
Gb/s	Giga (= 10^9) bits per second, also Gbit/s
GB/s	Giga (= 10^9) Bytes (= 8 bits) per second, also GByte/s
GCT	Grid Community Toolkit
GÉANT	Collaboration between National Research and Education Networks to build a multi-gigabit pan-European network. The current EC-funded project as of 2015 is GN4.
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
GPU	Graphic Processing Unit
GridCF	Grid Community Forum
GSI	Grid Security Infrastructure
GT	Globus Toolkit
HET	High Performance Computing in Europe Taskforce. Taskforce by representatives from European HPC community to shape the European HPC Research Infrastructure. Produced the scientific case and valuable groundwork for the PRACE project.

HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
HPL	High Performance LINPACK
HTML	HyperText Markup Language
IGTF	Interoperable Global Trust Federation
ISC	International Supercomputing Conference; European equivalent to the US based SCxx conference. Held annually in Germany.
KB	Kilo (= 2^{10} ~10^3) Bytes (= 8 bits), also Kbyte
KPI	Key Performance Indicator
LDAP	Lightweight Directory Access Protocol
LINPACK	Software library for Linear Algebra
MB	Management Board (highest decision making body of the project)
MB	Mega (= 2^{20} ~ 10^6) Bytes (= 8 bits), also MByte
MB/s	Mega (= 10^6) Bytes (= 8 bits) per second, also MByte/s
MD-VPN	Multi Domain Virtual Private Network
MFlop/s	Mega (= 10^6) Floating point operations (usually in 64-bit, i.e. DP) per second, also MF/s
MOOC	Massively open online Course
MoU	Memorandum of Understanding.
MPI	Message Passing Interface
NDA	Non-Disclosure Agreement. Typically signed between vendors and customers working together on products prior to their general availability or announcement.
OS	Operating System
PA	Preparatory Access (to PRACE resources)
PATC	PRACE Advanced Training Centres
PCPE	PRACE Common Production Environment
PKI	Public Key Infrastructure
PMA	Policy Management Authority
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PRACE 2	The upcoming next phase of the PRACE Research Infrastructure following the initial five year period.
PTC	PRACE Training Centres
RHEL	Red Hat Enterprise Linux
RI	Research Infrastructure
RT	Request Tracker, same as TTS
SCI	Security for Collaborating Infrastructures
SSH	Secure Shell
SVN	SubVersioN: software versioning and revision system
TAGPMA	The Americas Grid PMA
PRACE BoD	PRACE Board of Directors
PRACE TB	PRACE 5IP Technical Board (group of Work Package leaders)
TB	Tera (= 2^{40} ~ 10^{12}) Bytes (= 8 bits), also TByte
TCO	Total Cost of Ownership. Includes recurring costs (e.g. personnel, power, cooling, maintenance) in addition to the purchase cost.
TDP	Thermal Design Power
TFlop/s	Tera (= 10^{12}) 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
TTS	Trouble Ticket System, same as RT
UCC	UNICORE Commandline Client
UNICORE	Uniform Interface to Computing Resources. Grid software for seamless access to distributed resources.
URC	UNICORE Rich Client
VPN	Virtual Private Network
WISE	Wise Information Security for collaborating E-infrastructures
WP	PRACE Work Package

List of Project Partner Acronyms

BADW-LRZ	Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, Germany (3rd Party to GCS), short name LRZ
BILKENT	Bilkent University, Turkey (3rd Party to UYBHM)
BSC	Barcelona Supercomputing Center - Centro Nacional de Supercomputacion, Spain
CaSToRC	Computation-based Science and Technology Research Center, Cyprus
CCSAS	Computing Centre of the Slovak Academy of Sciences, Slovakia
CEA	Commissariat à l'Energie Atomique et aux Energies Alternatives, France (3rd Party to GENCI)
CESGA	Fundacion Publica Gallega Centro Tecnológico de Supercomputación de Galicia, Spain, (3rd Party to BSC)
CINECA	CINECA Consorzio Interuniversitario, Italy
CINES	Centre Informatique National de l'Enseignement Supérieur, France (3 rd Party to GENCI)
CNRS	Centre National de la Recherche Scientifique, France (3 rd Party to GENCI)
CSC	CSC Scientific Computing Ltd., Finland
CSIC	Spanish Council for Scientific Research (3rd Party to BSC)
CYFRONET	Academic Computing Centre CYFRONET AGH, Poland (3rd party to PNSC)
EPCC	EPCC at The University of Edinburgh, UK
ETHZurich (CSCS)	Eidgenössische Technische Hochschule Zürich – CSCS, Switzerland
FIS	FACULTY OF INFORMATION STUDIES, Slovenia (3rd Party to ULFME)
GCS	Gauss Centre for Supercomputing e.V., Germany
GENCI	Grand Equipement National de Calcul Intensiv, France
GRNET	Greek Research and Technology Network, Greece
INRIA	Institut National de Recherche en Informatique et Automatique, France (3rd Party to GENCI)
IST	Instituto Superior Técnico, Portugal (3rd Party to UC-LCA)
IT4Innovations	IT4Innovations National supercomputing centre at VŠB-Technical University of Ostrava, Czech Republic, short name IT4I
IUCC	INTER UNIVERSITY COMPUTATION CENTRE, Israel

JUELICH	Forschungszentrum Juelich GmbH, Germany
KIFÜ (NIIFI)	Governmental Information Technology Development Agency, Hungary
KTH	Royal Institute of Technology, Sweden (3rd Party to SNIC), short name PDC
LiU	Linkoping University, Sweden (3rd Party to SNIC)
NCSA	NATIONAL CENTRE FOR SUPERCOMPUTING APPLICATIONS, Bulgaria
NTNU	The Norwegian University of Science and Technology, Norway (3rd Party to SIGMA)
NUI-Galway	National University of Ireland Galway, Ireland, short name ICHEC
PRACE	Partnership for Advanced Computing in Europe aisbl, Belgium
PSNC	Poznan Supercomputing and Networking Center, Poland
RISCSW	RISC Software GmbH
RZG	Max Planck Gesellschaft zur Förderung der Wissenschaften e.V., Germany (3 rd Party to GCS)
SIGMA2	UNINETT Sigma2 AS, Norway
SNIC	Swedish National Infrastructure for Computing (within the Swedish Science Council), Sweden
STFC	Science and Technology Facilities Council, UK (3rd Party to EPSRC)
SURFsara	Dutch national high-performance computing and e-Science support center, part of the SURF cooperative, Netherlands
UC-LCA	Universidade de Coimbra, Labotatório de Computação Avançada, Portugal
UCPH	Københavns Universitet, Denmark
UHEM	Istanbul Technical University, Ayazaga Campus, Turkey
UiO	University of Oslo, Norway (3rd Party to SIGMA)
ULFME	UNIVERZA V LJUBLJANI, Slovenia
UmU	Umea University, Sweden (3rd Party to SNIC)
UnivEvora	Universidade de Évora, Portugal (3rd Party to UC-LCA)
UPC	Universitat Politècnica de Catalunya, Spain (3rd Party to BSC)
UPM/CeSViMa	Madrid Supercomputing and Visualization Center, Spain (3rd Party to BSC)
USTUTT-HLRS	Universitaet Stuttgart – HLRS, Germany (3rd Party to GCS), short name HLRS
WCNS	Politechnika Wroclawska, Poland (3rd Party to PNSC), short name WCSS

Executive Summary

The objective of this deliverable is to present the activity done in the reporting period (January 2017 - April 2018) to operate and coordinate the common PRACE Operational services, foreseen by Task 6.1 of WP6 in PRACE-5IP project. The operation of the PRACE distributed HPC infrastructure involves the coordination of a set of services which integrate the Tier-0 systems and a number of national Tier-1 systems, providing services for Tier-0, in a "single" pan-european HPC infrastructure.

This work is the continuation of the work done by Task 6.1 in the previous PRACE-IP projects to give continuity to the PRACE Operational services for the HPC Eco-system.

Eight Tier-0 systems were operational in the first reporting period of the PRACE-5IP project period:

- JUQUEEN at GCS@JUELICH;
- CURIE at GENCI@CEA;
- HAZELHEN at GCS@HLRS;
- SuperMUC at GCS@LRZ;
- SuperMUC phase2 at GCS@LRZ;
- MARCONI (BDW & KNL) at CINECA;
- MareNostrum4 at BSC;
- PizDaint at ETHZurich@CSCS

Furthermore, operational support has been provided to 23 national Tier-1 systems that provide services for Tier-0 (i.e. used from SMEs for the SHAPE activity, or as stepping stone towards Tier-0 systems, or to prototypes and asses new operational services). These Tier-1 systems are distributed among 16 different countries, ensuring a wide distribution of the European HPC eco-system.

A new version of the PRACE Service Catalogue, which describes the PRACE common services, has been submitted for approval by the PRACE BoD to further guide the operational activity. The Service Catalogue represents a living document and, as a result of the successful evaluation in the first reporting period of PRACE-5IP, it will be revised during the second reporting period.

Based on the procedures for incident and change management the complete set of PRACE common services as defined in the Service Catalogue (Networking, Data, Compute, AAA and Security, User, Monitoring and Generic) have been operated and monitored on a day-by-day basis to assure continuity and integrity of the services.

The Security Forum, responsible for all security related activities, is also coordinated by Task 6.1, with periodic teleconferences to monitor the infrastructure and prevent possible incidents which could cause vulnerability on the PRACE RI.

1 Introduction

This deliverable describes the activities done in Task 6.1 "Operation and coordination of the comprehensive common PRACE operational services" of WP6 "Operational Services for the HPC eco-system" in PRACE-5IP. This task is responsible for the operations of the set of common services, which presents the PRACE Tier-0 systems as an integrated pan-European HPC eco-system [1]. The operational services are extended also to national Tier-1 systems, essential as a stepping-stone towards Tier-0 systems (Tier-1 for Tier-0). Examples of Tier-1 for Tier-0 activities are the SHAPE activity toward SMEs, the prototyping and assessment of new operational services investigated in Tasks 6.2 and 6.3 of PRACE-5IP WP6, the testing and utilization of specific architectures and technologies, which are only available in specific countries.

The operation and coordination of the common PRACE operational services provided in Task 6.1 go on with a well-established management procedures and organisation as set up already since PRACE-1IP, [1], [2]. The task further continues the implementation of the roadmap to a professional service level of sustainable services with a defined quality of service.

This report mainly focuses on the activities done in the first reporting period (April 2017-April 2018) and complements the activity undertaken during the second year of PRACE-4IP and documented in PRACE-4IP Deliverable 6.2 [3].

In this reporting period, the operation of the common PRACE operational services has been coordinated and monitored constantly, by means of biweekly teleconferences, attended by all the partners involved in the Task 6.1 activity. Furthermore, two successful WP6 Face-to-Face meetings have been organised, in Bologna (Italy) on 10th of October 2017 and in Poznan (Poland) on 16th – 17th of January 2018. These meetings have been very useful to discuss the status of the operational activity, to plan the activity for the subsequent periods and to agree on the teams involved in the different activities. Both meetings had around 30 attendees and most of the different sites involved in WP6 have been represented.

Section 2 describes the status of the Tier-0 systems and the Tier-1 systems involved in the Tier-1 for Tier-0 activity, composing the PRACE HPC eco-system. Section 3 gives a status overview of the common services for the different service areas:

- Network services (MD-VPN network provided by GEANT connecting Tier-0 and major Tier-1 centres);
- Data services;
- Compute services;
- Authorization, Authentication and Accounting;
- Operational security;
- User services;
- Monitoring services for operations;
- Generic services.

Finally, Section 4 represents conclusions.

2 PRACE HPC Ecosystem: Tier-0 and Tier-1 sites, system upgrades and new systems

The first part of this section presents the changes implemented during the current reporting period (April 2017 – April 2018) concerning the status of Tier-0 sites, and the Tier-1 national sites providing Tier-1 for Tier-0 services. In the second part, the chronology and the status of the performed system upgrades is being detailed.

2.1 Maintaining the service

In the current reporting period, the main tasks of the operational procedures used to offer the PRACE services have been kept similar to what we had in the previous PRACE-4IP Project. Accordingly, the work from the point of view of managing the procedures and operability resulted simplified and well tested as was the continuation and evolution of the activity already in progress.

Nevertheless, some minor changes to the operational procedures have been made and communicated to partners accordingly. Mainly these changes modified the communication between partners when a quick response becomes necessary for a reported issue, in order to improve the productivity and reduce the waiting time. However, the procedure has not been yet consolidated, and we aim to come to a full operability in the next period of the PRACE-5IP Project.

2.2 Status of Tier-0 & Tier-1 sites

The Tier-0 and Tier-1 systems constitute a HPC eco-system offering high level services to the European computational community. At present 8 Tier-0 systems and 23 Tier-1 systems are in production. Their status is monitored and the operational quality is assured by employing a specific regular activity provided daily by the members of PRACE-5IP WP6 Task 6.1. This “On-duty” activity, is described below.

2.2.1 On-duty Activity

The On-duty Activity is supplied following a weekly schedule, and it is guaranteed by all the partners who provide effort and/or systems in the WP6. The issues reported through this service are mainly related to operational issues and activities needed to maintain the distributed infrastructure in good shape. Specific requests from users are rare and normally are redirected to the local Help Desk of each individual site.

A report on the On-duty Activity is produced every week, reporting the status of the infrastructure, all occurred problems, and any other warning regarding the core services.

The related documentation on the Wiki site is constantly updated, using the report template agreed at the beginning of PRACE-4IP.

Starting from 1st April 2017, a schedule planning has been defined for the 22 PRACE partners involved in the On-duty Activity. Each of them is in charge for monitoring the infrastructure, and reporting the related issues using the Trouble Ticketing tool (TTS). The 22 partners involved in the schedule are reported in Table 1 below.

1 BSC	12 EPCC
2 JUELICH	13 ICHEC
3 UIO	14 UHEM
4 CASTORC	15 ETH
5 IDRIS	16 SURFSARA
6 CEA	17 NIIF
7 CINES	18 RZG
8 CINECA	19 CSC
9 NCSA	20 CYFRONET
10 HLRS	21 PSNC
11 VSB-TUO	22 PDC

Table 1: PRACE partners involved in the on duty activity

Any given partner is involved in the shift one week out of 22, i.e. every 6 months approximately. The weekly reports of the On-duty Activity are stored on the PRACE Operation Wiki.

The TTS tool used by the On-duty Activity staff is the 'Best Practical' Request Tracker (RT) version 4.2.8, updated during the PRACE-4IP project, and hosted by CINECA.

During the reporting period, the following KPIs have been applied to better supervise the On-Duty Activity (it is important to emphasize that, in principle, the activity is related to the traffic on the General Queue where the tickets are normally created):

Tickets resolving efficiency	
Description:	Number of tickets resolved by the Service Desk
Calculation:	O+R R/(O+R)*100
Inputs:	Number of open tickets reported (O) Number of resolved tickets reported (R)
Outputs:	Total number of worked tickets Percentage of resolved tickets respect worked tickets (%)
Time-interval:	Monthly every first working day (update during the first PRACE Operations meeting of the following month) Annual report
Threshold:	Percentage of resolved tickets respect worked tickets > 90% (Annual threshold)
Tools:	RT
ITIL Category:	Service Operation – Incident Management
'KPI Lead':	Cristiano Padrin (CINECA)
Implementation plan:	Request Tracker (RT) provides all data necessary for computing this KPI. The necessary input data can be extracted using the Query Builder of RT.

Table 2: Tickets resolving efficiency

Average Initial Response	
Description:	Average time taken between the time a user creates a ticket and the time that the Service Desk opens the ticket
Calculation:	for i=1,(O+R) { T(i) = ST(i)-CT(i) } maximum value of vector T [sum of T(i)] / (O+R)
Inputs:	Number of open tickets (O) Number of resolved tickets (R) Vector of Creation times (CT) Vector of Starting times (ST) Vector of Initial Response times (T)
Outputs:	Maximum Initial Response Average Initial Response
Time-interval:	Monthly every first working day (update during the first PRACE Operations meeting of the following month)
Threshold:	< 1 working day without a motivation - unlimited with motivation
Tools:	RT
ITIL Category:	Service Operation – Incident Management
'KPI Lead':	Cristiano Padrin (CINECA)
Implementation plan:	Request Tracker (RT) provides all data necessary for computing this KPI. The necessary output data can be extracted using the Query Builder of RT and managing the chart.

Table 3: Average Initial Response

In the last 12 months, 24 tickets have been created and resolved in the General queue, with a percentage of resolved tickets respect to the worked tickets equal to 100,00%, with an Average Initial Response of 27h 25m 58s. We can consider this KPI respected (Threshold < 1 working day without a motivation).

In the same period, 99 tickets have been created in the site queues and 98 have been monitored by the Operator and resolved by the sites watchers.

2.2.2 Production systems

At present (March 2018), the Tier-0 ecosystem is made up of 8 systems, distributed in 7 sites, operated by 7 different partners, provided by the 5 Hosting Members (France, Germany, Italy, Spain and Switzerland as reported in Table 4).

The theoretical peak performance of the Tier-0 systems ranges from more than 1.7 PFlop/s up to (around) 25 Pflop/s for the new Piz Daint system in Switzerland. One Tier-0 provide GPU accelerators: ETH/Piz Daint with Nvidia Tesla P100.

All the Tier-0 systems are ranked in the Top500 (November 2017) and four of them are in the first 20 positions (seven in the first 50 position). Vendor distribution is: 3 systems Lenovo, 2 systems IBM/Cray, 1 system Bull.

Partner	**Country**	**Tier-0**	**Architecture** **CPU**	**Rpeak (Tflop/s)** **Top500#(nov17)**
GCS-JUELICH	Germany	Juqueen	IBM BlueGene/Q IBM PowerPC A2; 16-core; 1.60GHz	5872.0 22
GCS-LRZ		SuperMUC	IBM iDataPlex Intel Sandy Bridge-EP; 8-core; 2.7GHz	3185.0 44
		SuperMUC phase 2	Lenovo NeXtScale Intel Haswell-EP 14 core; 2.6GHz	3580.0 45
GCS-HLRS		Hazel Hen	Cray XC40 Intel Haswell E5 2680v3; 12-core; 2.5GHz	7420.0 19
GENCI	France	Curie (Thin)	Bull Bullx B510 Intel Sandy Bridge-EP; 8-core; 2.70GHz	1741.8 93
BSC	Spain	Mare Nostrum	Lenovo SD530; Intel Xeon Platinum 8160; 24-core; 2.1 GHz	11150.0 16
ETH	Switzerland	Piz Daint	Cray XC50 Intel Xeon E5 2690v3; 12-core; 2.6GHz; NVIDIA P100 GPUs	25326.3 3
CINECA	Italy	Marconi	Lenovo NextScale (Intel OmniPath Cluster) Intel Xeon E5-2697 v4 (Broadwell); 2x18-core; 2.30 GHz +	1000.0

Partner	Country	Tier-0	Architecture CPU	Rpeak (Tflop/s) Top500#(nov17)
			Lenovo Adam Pass (Intel OmniPath Cluster) Intel Xeon Phi 7250 (Knights Landing) 1x68-core; 1.40 GHz + Lenovo Stark (Intel OmniPath Cluster) Intel Xeon Phi 8160 (Skylake) 2x24-core; 2.10 GHz	72 + 13000.0 14 + 7000.0 (n.a.)

Table 4: PRACE Tier-0 systems

As far as the Tier-1 ecosystem is concerned, 23 systems are operating as Tier-1 for Tier-0 services. These Tier-1 systems are distributed in 17 different PRACE sites, operated by 17 partners, in 17 different European Countries. Table 5 presents the list of the Tier-1 systems.

The theoretical peak performance ranges from very small system partitions (<10 TFlop/s) up to large systems in excess of 2.5 Pflop/s. Nine systems deliver more than 1 Pflop/s.

Eight of the Tier-1 systems are accelerated with Nvidia accelerators, three with Xeon Phi KNC coprocessors. Several different vendors and architectures are present and this is a real advantage for the PRACE HPC infrastructure: HP, SGI, Bull, Cray, IBM and Lenovo are nearly equally represented; BlueGene, iDataPlex/NeXtScale, Bullx are the most popular architectures.

Partner	Country	Tier-1	Architecture CPU	Rpeak (Tflop/s)
GCS – RZG	Germany	Hydra	IBM iDataPlex Intel Sandy Bridge-EP; 8core; 2.6GHz + Intel Ivy Bridge; 10-core; 2.8GHz, Nvidia GPU, Intel PHI	1463.2 (Ivy) + 1013.1 (GPU)
GENCI	France	Turing	IBM BlueGene/Q IBM PowerPC A2; 16-core; 1.60GHz	1258.0
EPCC	UK	Archer	Cray XC30 Intel Ivy Bridge; 12-core; 2.7GHz	2550.5
BSC	Spain	Minotauro	Bull Bullx B505 Intel Westmere-EP; 6-core; 2.53GHz, Nvidia/M2090 and K80	339.54

Partner	Country	Tier-1	Architecture CPU	Rpeak (Tflop/s)
CSC	Finland	Sisu	Cray XC40 Intel Haswell; 12-core; 2.60GHz	1688.0
SURFsara	Netherlands	Cartesius phase 1 + 2	Bull Bullx B720/B710 Intel Haswell; 12-core; 2.6GHz + Intel Ivy Bridge; 12-core; 2.4GHz+ Intel Sandy Bridge; 8-core; 2.7GHz	1349.0
		Cartesius GPU extension	Bull Bullx B515 Intel Ivy Bridge 8 core; 2.5GHz Nvidia K40	210.0
		Cartesius Sequana extension 1	Bull Sequana X1110 Intel Broadwell 16 core; 2.6 GHz	236.0
		Cartesius Sequana extension 2	Bull Sequana X1210 Intel Knights Landing 64 core; 1.3 GHz	48.0
SNIC	Sweden	Beskow	Cray XC40, Intel Haswell; 32-core; 2.3GHz	1973.0
PSNC Cyfronet	Poland	Zeus BigMem	HP BL685c G7 AMD Interlagos; 16-core; 2.3GHz	61.2
		Zeus GPGPU	HP SL390s Intel Westmere-EP; 6-core; 2.45GHz, Nvidia/M2090	136.8
		Eagle	Intel Cluster, Intel Haswell E5-2697 v3; 14-core; 2.60GHz	1380.0
		Supernova	HP Cluster Platform 3000 BL2x220 Intel Westmere-EP; 6-core; 2.67GHz	30.0
		Prometheus	HP Apollo 8000, Intel Xeon E5-2680v3; Nvidia Tesla K40 XL	2400.0
SIGMA	Norway	Abel	MEGWARE MiriQuid Intel Sandy Bridge-EP; 8-core; 2.6GHz	178.6
NUI Galway	Ireland	Fionn	SGI ICE X Intel Ivy Bridge; 12-core; 2.4GHz	147.5
CaSToRC	Cyprus	CyTera	IBM IdataPlex Intel Westmere; 12-core; 2.67GHz, Nvidia M2070	35.0
NCSA	Bulgaria	EA-Ecnis	IBM BlueGene/P IBM PowerPC 450; 4-core; 850MHz	27.8

Partner	Country	Tier-1	Architecture CPU	Rpeak (Tflop/s)
IT4I-VSB	Czech Rep	Anselm Solomon	Bull Bullx B510/B515, Intel Sandy Bridge-EP; 8-core; 2.4GHz SGI ICE-X Intel Xeon E5-2680v3; 12-core; 2.5GHz, Intel PHI	66.0 2000.0
NIIF	Hungary	NIIFI SC Seged Leo PHItagoras	HP Cluster Platform 4000SL AMD Magny-Cours; 12-core; 2.2GHz HP Cluster Platform 4000SL AMD Magny-Cours; 12-core; 2.2GHz, Nvidia M2070 HP SL250s Intel Xeon E5-2650 v2; 2.60GHz, Nvidia K20, K 40 HP SL250s Intel Xeon E5-2680 v2; 2.80GHz, Intel/PHI 7120	5.48 14.0 254.0 27.0
CCSAS	Slovakia Rep	Aurel	IBM Power 775, IBM Power7; 8-core; 3.84GHz	128
GRNET	Greece	ARIS	Heterogeneous cluster from multiple vendors with SandyBridge, IvyBridge and Haswell CPUs, NVIDIA K40 and Xeon Phi 7120 accelerators. See [4] for details.	445.55

Table 5: PRACE Tier-1 systems

2.3 System Upgrades

In this section, we describe the activity related to the system upgrades and integration in the PRACE ecosystem, being Tier-0 or Tier-1 systems involved in the Tier-1 for Tier-0 services.

In the current period 2017-2018, this activity has been included under WP6 subtask 6.1.2 and it is led by Barcelona Supercomputer Centre (BSC).

2.3.1 Operational procedures for new systems and system upgrades

Current operational procedures are well documented in the PRACE Wiki and BSCW and available at the following links:

- Integration of new Tier-0/Tier-1 sites
  - Procedure/Template:
    - WIKI template that new sites must fulfil in order to complete their integration into PRACE infrastructure ([5])
  - Information for new Tier-1 sites
    - BSCW presentation about basic concepts that new sites should know before starting their integration into PRACE infrastructure ([6])
- Upgrade of systems (Tier-0/Tier-1)
  - Procedure for upgrade of systems:
    - WIKI guide with information to upgrade systems in the PRACE infrastructure ([7])
  - Report template for upgrades
    - WIKI template for the completion of the upgrade procedure ([8])

This documentation ensures that all systems follow equal procedures and are in line with the Service Catalogue requirements.

2.3.2 Overview of System Upgrades in PRACE-5IP

From the last PRACE-4IP report only one Tier-0 system was upgraded and completely integrated into the PRACE Infrastructure, this is MareNostrum4 Tier-0 supercomputer of BSC.

In the next table, a brief overview of the upgraded systems and the completion dates is shown:

System	Site	Tier	Status	Last Check
MareNostrum4	BSC	Tier-0	Completed (01/07/2017)	01/07/2017

Table 6: Tier-0 systems upgrades

2.3.3 Details of upgraded sites

The full list of upgrades during this first period of PRACE-5IP time frame is reported below.

1. MareNostrum4 – BSC – Tier-0

The successful upgrade of the Tier-0 system MareNostrum at BSC has evolved into it's 4th generation (thus sometimes called MareNostrum4). The main compute part is now a Lenovo

general purpose cluster based on latest Intel Skylake-SP processors with total theoretical peak performance of 11.15 Pflop/s. The system was already fully integrated into PRACE in July 2017 including connection to the new MD-VPN based PRACE Network and installing all services according to the valid Service Catalogue.

2.4 New Tier-0/Tier-1 sites and systems

In this section, we present the new sites and systems that are going to be integrated in the PRACE infrastructure.

2.4.1 Overview of new Tier-0/Tier-1 sites & systems on PRACE-5IP

In the first period of PRACE5-IP, eight new systems have been planned to be included in the PRACE infrastructure and to start providing services to it, as shown in the table below. One system finished their integration, five systems are in the integration phase and two systems are scheduled in the next months (UHEM in Turkey and CEA in France).

Site	System Name	System Details URL	Tier	Integration Start Date	Status	Last Check
GRNET (Greece)	ARIS	http://doc.aris.grnet.gr	Tier-1	01/09/2015	Complete (12/2017)	08/02/2018
Cyfronet (Poland)	Prometheus	https://kdm.cyfronet.pl/portal/Prometheus	Tier-1	aprox. 09/2016	In progress	28/09/2016
CCSAS (Slovakia)	Aurel	http://vs.sav.sk/?lang=en&section=departments&sub=vvt&sub2=config	Tier-1	01/05/2017	In progress	03/05/2017
UL (Slovenia)	HPCFS	http://hpc.fs.uni-lj.si/hardware	Tier-1	aprox. 07/2017	In progress	01/06/2017
CESGA (Spain)	FiniTerrae	https://www.cesga.es/en/infraestructuras/computacion/FinisTerrae2	Tier-1	01/04/2016	In progress	09/02/2018
IUCC (Israel)	Moab	n/d (soon)	Tier-1	½/2017	In Progress	20/02/2018
UHEM (Turkey)	Sariyer	n/d	Tier-1	n/d	Planned	
CEA (France)	Irene	http://www-hpc.cea.fr/en/complexe/tgcc-Irene.htm	Tier-0	21/03/2018	Planned	

Table 7: New Tier-0 and Tier-1 systems

2.4.2 Details of new Tier-0/Tier-1 sites

The full list of the new systems integrated in the first period of PRACE-5IP time frame is reported below.

1. **ARIS – GRNET – Tier-1**

ARIS is a very heterogeneous system combining different generations of Intel CPUs. It has 426 generic nodes with Intel IvyBridge processors and 96GB memory. Intel Haswell processors are in 44 Nvidia K40 GPU accelerated nodes and 18 nodes with Xeon Phi 7120P coprocessors. Intel SandyBridge processors feature 44 nodes with 256GB memory. All of these compute elements are interconnected using Infiniband FDR and accessing 750TB of shared GPFS storage.

The integration into PRACE has been completed in February 2017.

2. **Prometheus – Cyfronet – Tier-1**

Prometheus consists of more than 2,200 servers based on the HP Apollo 8000 platform, interconnected with InfiniBand FDR network. Nodes are equipped with Intel Haswell generation processors offering more than 53,000 cores in total. These are accompanied by 279 TB RAM in total, and by two storage file systems of 10 PB total capacity, and 180 GB/s access speed. Prometheus has been also equipped with 144 NVidia K40 GPUs.

Full integration of the system is still ongoing, currently coping with missing connection to the new MD-VPN PRACE Network.

3. **Aurel - CCSAS – Tier-1**

Aurel is an IBM Power 755 system composed by 2 frames, 16 drawers, 128 physical nodes (130 logical nodes), with a total of 4096 cores, which provides a total of 128 TFlops. Each compute node holds 256 GB of main memory. The cluster is interconnected using Infinband QDR and is using AIX 7.1 as operating system. Shared storage of 600TB is provided to the system.

Most of the tasks are pending to be done for the integration

4. **HPCFS – UL – Tier-1**

HPCFS is a system with 768 cores of Intel Westmere processors, total 3TB of memory and 20TB of disk space. An extension of 480 cores on Intel Haswell processors was recently added, plus one node with 3 Nvidia K80 accelerators and one visualization node with Nvidia Quadro K40 card. The nodes are interconnected using Infiniband QDR.

Most of the tasks are pending to be done for the integration

5. **Finis Terrae – CESGA – Tier-1**

Finis Terrae is heterogeneous cluster, with an Infiniband FDR network interconnecting 317 computing nodes based on Intel Xeon Haswell processors. Together, these nodes are able to provide a computing power of 328 TFLOPS, 44.8 TB of memory and 1.5 PB of disk capacity. The system includes a high performance parallel storage system able to achieve a speed of more than 20 GB/s.

CESGA is in the process of the integration and is pending to be integrated into the PRACE network.

6. **IUCC – Tier-1**

Partial integration has been done of the IUCC new partner, no network connection has been yet established through PRACE private network, but other services have been installed using the public Internet connection, namely GridFTP and Unicore.

7. **Sariyer – UHEM – Tier-1**

Sariyer is system of Intel Haswell and Broadwell processors with a total of 108 compute nodes and 2964 cores, using an interconnect of FDR network. Four nodes are accelerated using Nvidai K20 accelerators. This system is planned to be integrated to PRACE infrastructure during 2018.

8. **Irene – CEA – Tier-0**

Irene will be the next Tier-0 cluster for CEA, it will be a 9 PFlop system with two partitions, one with 1656 compute nodes based on Intel SkyLake processors and 666 compute nodes based on Intel KNL processors.

SkyLake partition will be connected using Infiniband EDR network, meanwhile KNL partition will be using new Bull eXascale Interconnect (BXI).

Integration is planned to start at the end of March.

3 Operational Services

Common services are divided into thematic categories: Network, Data, Compute, AAA, User, Monitoring and Generic. Each service category has a responsible person who is in charge of managing all the information and decisions related to a specific service area.

The common services are listed in the PRACE service catalogue and responsibility for a given service is taken by the respective service area.

The following sections provide an update of the status of each service category and the main achievements within the current reporting period.

3.1 Network services

The main task within network services handled in the first reporting period of PRACE-5IP has been the general operation of the PRACE-VPN network including optimization of network throughput by adjustment of the relevant network parameters. Furthermore, user support concerning optimal network usage has been a main task.

During the first reporting period, also additional partners (RZG, GRNET, PSNC, CYFRONET, WCSS, LECAD and SURFsara) were connected to the PRACE-VPN network. Therefore, the old dedicated network was not used anymore and was shut down and uninstalled.

Following is the status from the network monitoring reflecting the state of the connected sites using the new MD-VPN service.

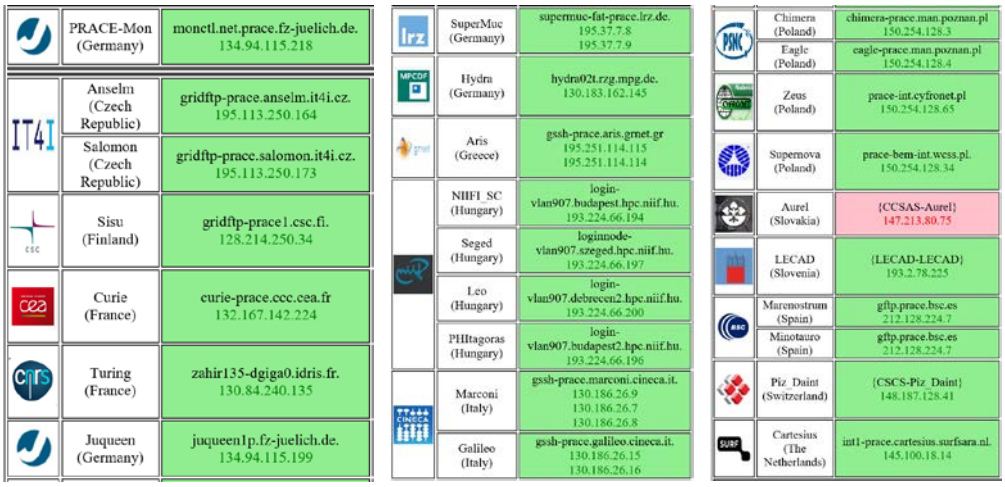

Figure 1: Overview of the PRACE Network by sites connected via MD-VNP

Discussions on new monitoring solutions via PERFsonar with close collaboration of GÈANT have been started. First partners as test candidates for this monitoring solution have been contacted. Since the network is used by the European project HBP also, the partners BSC, CINECA, and CSCS would be ideal candidates, since they are involved in both projects. Until now no final decision has been made about who will, beneath Jülich, be a test partner.

3.2 Data Services

Data services within PRACE-5IP are delivered by GridFTP as a core service, gtransfer as an additional service and Unicore as an optional service.

3.2.1 GridFTP, gtransfer and Gsatellite

GridFTP is a core service and should be installed on every PRACE site attached to the PRACE network. Deployment status currently is:

The following Tier-0 sites - BSC, CEA, CINECA, JUELICH, HLRS- have GridFTP installed in a split configuration. CSCS has not yet installed GridFTP.

The following Tier-1 sites have installed GridFTP:

CASTORC, CSC, CYFRONET, EPCC, GRNET, ICHEC, IDRIS, IUCC (Testing Phase), NCSA (Testing Phase), NIIF, PSNC, RZG, SURFSARA, UHEM, UIO-SIGMA, VSB-TUO, WCSS

Eight of them run a split configuration, all of them have upgraded to versions >= GT6.0

CESGA, KTH-SNIC, UC-LCA, UL, UPM do not run GridFTP at the moment.

Gtransfer as an additional service is not actively tracked, but is used on both tiers of PRACE systems, where Tier-0 is represented by HLRS and Tier-1 by both clusters of IT4I. At IT4I gtransfer was successfully used in the last two PRACE-EUDAT data pilots, namely DECI-13 and DECI-14.

Gsatellite was successfully used in a WP6.2 prototype project in previous PRACE-IP and was suggested as a new additional data service into the PRACE service catalogue. Currently only the final PRACE BoD approval is missing for this status.

3.2.2 Globus Toolkit – end of life

The announcement of the end of support for the Globus Toolkit made near the end of the second quarter of 2017 ([5]) marked an important change for GridFTP - a major part of the Globus Toolkit and the core service for data transfers in PRACE. The Globus team of the University of Chicago - the main developers of the Globus Toolkit - will discontinue all support of the Globus Toolkit at the end of 2018. Specifically they already no longer provide any support apart from security updates and critical bug fixes since February 2018.

These support changes for the Globus Toolkit imposed a serious issue because there was and still is no other data transfer service available that offers the level of integration GridFTP currently offers in PRACE: We have tools to quickly set up GridFTP servers, we have tools to measure data transfer performance, we have tools to easily set up reliable and performing data transfers, we have the expertise and the documentation ready. Moreover, GridFTP can perform well and can exploit the advantages of parallel file systems. To throw all this away would be simply a waste of resources and money. Because GridFTP is in use everywhere around the world, PRACE is not alone with this issue, as it affects other e-Infrastructures and user communities in similar ways.

In order to decide a mitigation strategy, joint discussions between different user communities took place in the second half of 2017. Possible impacts of the support changes and dependencies on the Globus Toolkit in different communities were reviewed there and also ways out of this issue. E.g. it was proposed by us and other stakeholders to continue the maintenance and support of the Globus Toolkit after 2018 as a community. The outcome of these discussions between representatives of CERN, EGI, EUDAT, OSG, PRACE, XSEDE and other communities was the formation of the Grid Community Forum (GridCF) ([6]) in November 2017 whose objectives are to:

- Maintain the quality and security of critical functionality such as the Grid Security Infrastructure (GSI) and an open-source GridFTP implementation.
- Partner across several grid organizations to share the burden of supporting this core functionality.
- Provide an open and welcoming forum for external contributors to improve the software.

The first work result of the GridCF was the creation of a fork of the soon unsupported Globus Toolkit named the Grid Community Toolkit (GCT) ([7]). It is available from a Git repository hosted on GitHub and already made outstanding progress: For example a continuous integration ([8]) service was set up that checks if the GCT can still be built with the proposed changes before they are reviewed and actually included into the code base. In addition, source packages and source RPM packages are automatically built and provided on each release as basis for pre-compiled native operating system packages (for e.g. RHEL and compatible GNU/Linux distributions). It is expected that native OS packages currently based on the Globus Toolkit will soon switch to the GCT as basis. The GCT also already moved forward and ahead of the Globus Toolkit about correctness as it also includes community provided fixes for functional errors experienced during operation of the Globus Toolkit in various fields of application that up until now were ignored by the Globus team. This example alone already shows the value and the advantages of a community

driven toolkit. Updates and fixes still provided by the Globus team from time to time are of course also promptly included into the GCT. Current and future tasks include:

- The revising and updating of the old Globus Toolkit documentation and the hosting of the resulting GCT documentation under the GridCF website.
- The setup of user support facilities like issue tracking and mailing lists and a contact point for reporting security problems under the GridCF umbrella.
- The early involvement in the solution of community reported problems with the Globus Toolkit.
- The preparation of the first official GCT release which for example requires updates to meta data like URLs and email addresses that still relate to Globus.

All the effort and interest by various user communities that is currently going into this endeavor feed the expectation of a smooth transition from the Globus Toolkit to the GCT and of making the support changes announced by the Globus team a non-issue for PRACE.

3.3 Compute Services

Compute services provide interfaces between users and computing capabilities. The target of this activity is to find out what services can be commonly provided in PRACE and try to maintain the uniformity among all sites providing computational hours.

In the PRACE-5IP timeframe, the work consisted on maintaining the different components that are provided in the Service Catalogue, mainly the batch scheduling systems and Unicore. An update of the activity done in the reporting period for these services is provided in the following.

One of the major changes has been the movement of the Unicore service as an optional service inside the PRACE Service Catalogue.

3.3.1 Local batch scheduling

Local batch scheduling systems identify the software tools responsible for managing user jobs. The service consists on maintaining a list of different software components that current supercomputers deploy now, thus providing a statistic of which are the most used ones. In the next table, it is possible to see the relationship between supercomputer and batch system. One of the most interesting systems is SLURM, that is step by step overtaking others. SLURM is an Open Source software maintained by SchedMD and one of its strong points is its modularity and flexibility, making relatively easy to adapt source code to site needs. These characteristics ensures a rapid growth and adaption to new systems, platforms and architectures and we think this is the reason that many sites are using it. The code of this software is open so it is possible to see that it has a very high quality. Sites are getting support from the community but also directly from the developers and moreover there is the possibility of having a contract with SchedMD for enterprise support. All these features and many other, make SLURM the preferred solution among different sites.

Another responsibility of Compute Services in PRACE-5IP has been the maintenance of a set of scripts to wrap the different inputs that different systems could take as parameters, motivated by the differences in interaction with batch systems for different products. A scripted file must be

edited with a full description of a job, e.g. resources like core/cpu-hours, number of cores/cpu, I/O staging, service level, and all needed actions for preparing the execution environment.

Some IBM systems have migrated their batch scheduling system from Loadleveler to IBM LSF, as development was stopped some years ago, and now only one system keeps LoadLeveler as batch scheduling system.

These differences made it interesting to develop a set of wrappers with a common syntax for describing a job around the different batch systems and platforms. Such wrappers should make it easy for users to migrate their work from one system to another. Nowadays, given the complexity of systems and architectures, with different features, the work should be enhanced and reviewed during actual project to better enable the wrappers and adapt to different types of systems.

Partner	Site	Machine Name	Tier	Batch System
1. JUELICH (de)	JUELICH	Juqueen	Tier-0	IBM LSF
2. GCS (de)	LRZ	SuperMUC	Tier-0	IBM LSF
2. GCS (de)	LRZ	SuperMUC phase 2	Tier-0	IBM LSF
2. GCS (de)	HLRS	Hazelhen	Tier-0	Torque/Moab
2. GCS (de)	RZG	Hydra	Tier-1	IBM LSF
3. GENCI (fr)	CEA	Curie (Fat)	Tier-0	SLURM
3. GENCI (fr)	CEA	Curie (Hybrid)	Tier-0	SLURM
3. GENCI (fr)	CEA	Curie (Thin)	Tier-0	SLURM
3. GENCI (fr)	IDRIS	Turing	Tier-1	LoadLeveler
4. EPCC (uk)	EPCC	ARCHER	Tier-1	PBSPro
5. BSC (es)	BSC	Marenostrum4	Tier-0	SLURM
5. BSC (es)	BSC	Minotauro	Tier-1	SLURM
6. CSC (fi)	CSC	Sisu	Tier-1	SLURM
7. ETH (ch)	CSCS	Piz-Daint	Tier-0	SLURM
8. SURFSARA (nl)	SURFsara	Cartesius	Tier-1	SLURM
10. SNIC (se)	PDC	Beskow	Tier-1	SLURM
11. CINECA (it)	CINECA	Marconi	Tier-0	PBS Pro
11. CINECA (it)	CINECA	Galileo	Tier-1	PBS Pro
12. PSNC (pl)	CYFRONET	Zeus (BigMem)	Tier-1	PBS Pro
12. PSNC (pl)	CYFRONET	Zeus (GPGPU)	Tier-1	PBS Pro
12. PSNC (pl)	PSNC	Eagle	Tier-1	SLURM
12. PSNC (pl)	WCSS	Bem	Tier-1	PBS Pro
13. Sigma (no)	UIO	Abel	Tier-1	SLURM
16. NUI (ie)	ICHEC	Fionn	Tier-1	Torque/Moab
18. CaSToRC (cy)	CaSToRC	CY-TERA	Tier-1	SLURM
19. NCSA (bg)	NCSA	EA-ECNIS	Tier-1	IBM LSF
20. IT4I (cz)	IT4I/VSB-TUO	Anselm	Tier-1	PBS Pro
20. IT4I (cz)	IT4I/VSB-TUO	Salomon	Tier-1	PBS Pro
22. NIIF (hu)	NIIF	NIIFI SC	Tier-1	SLURM
22. NIIF (hu)	NIIF	Seged	Tier-1	SLURM
22. NIIF (hu)	NIIF	Leo	Tier-1	SLURM
22. NIIF (hu)	NIIF	Phitagoras	Tier-1	SLURM
24. GRNET(gr)	GRNET	Aris	Tier-1	SLURM
25. CCSAS (sk)	CCSAS	Aurel	Tier-1	IBM LSF

Table 8: Batch schedulers on partner systems

3.3.2 Unicore

Another way to create and operate on a common layer is through an even higher abstraction level. The UNICORE software, which is part of the PRACE software portfolio since the beginning, and which was originally adopted in the DEISA projects, accomplishes this. It allowed a user to manage single jobs as well as a workflow of jobs remotely through a Java-based graphical interface (URC) and/or a command line client (UCC). Access to end-systems relies on account information stored in the PRACE central LDAP, while resources are published on a central directory service by sites or resource providers.

Despite of the interesting features that UNICORE provided in the past to the infrastructure, in WP6, asking to the different Tier-0 and Tier-1 sites, we have noticed that there is very small usage of this service from the PRACE community. As commented before, different architectures and systems make it complex to group all the possible system features under a unique set of wrappers or under a single piece of software. This led us to evaluate the convenience of keeping this software as a core service in the catalogue. After a wide discussion and inquiry on the different PRACE sites, WP6 decided to move this software as an optional service for the Tier-0 and Tier-1 systems.

This is the actual installation of Unicore components that is needed in the PRACE infrastructure to provide this service:

Component	Description	Deployment
REGISTRY	Directory service publishing Tier-0 resources.	JUELICH (Primary), CINECA (Backup)
UNICORE/X	Translate abstract jobs into concrete jobs for a specific target system. Job submission and monitoring	Tier-0 System
GATEWAY	Main entrance to each Tier-0 system. Client connections go through the gateway, which forwards them to internal components, and vice versa.	Tier-0 System
XUUDB	User Database for authentication and authorization. It is synchronized with the PRACE LDAP.	Tier-0 System
TSI	Interface with the local batch system and storage resources	Tier-0 System
URC	Graphical user client (based on Eclipse Java)	Client-side
UCC	Command-line user client	Client-side

Table 9: Unicore components deployment description

3.4 AAA Services

The AAA activity is responsible for services, which provide Authentication, Authorization and Accounting facilities on the infrastructure. These services play a fundamental role in the operation of other PRACE services and the provision of access to PRACE resources. As a result, changes made to these services are normally of a routine nature and follow pre-documented policy to ensure continued stability and reliability of AAA services.

While the operational AAA services continue to be managed in this stable fashion, PRACE also needs to look forward to the future, particularly where collaborations with other external projects and e-Infrastructures may be involved. PRACE and GÉANT have previously discussed the feasibility of a more general federated authentication and authorization infrastructure. This idea is being further pursued with the AARC Project [9] to investigate how the AARC Blueprint Architecture might be applied to provide future PRACE AAA services. PRACE AAA services have a representative on the AARC Engagement Group for Infrastructures (AEGIS) to provide direct consultation into their project activities.

3.4.1 Public Key Infrastructure - PKI

Several PRACE services rely on X.509 certificates [10] for authentication and authorization. These certificates must be issued by entities trusted by the service providers. PRACE relies on the Certificate Authorities (CA) accredited as a member by the EUGridPMA, the European Policy Management Authority (PMA) [11], or by one of the two other organizations, the TAGPMA and the APGridPMA, all three federated in the IGTF [12]. These PMAs require a minimum set of requirements for the CP/CPS (Certificate Policy/Certification Practice Statement) of the member CAs, as published and maintained by the PMAs in profile documents. PRACE itself is a member of the EUGridPMA as Relying Partner and participates in its activities.

For PRACE partners CA information is maintained at a central repository. The information is based on the IGTF distribution and updates are made available to the PRACE partners for download in several formats. To guarantee a high level of availability, the distributions are stored on a virtualized webserver that can migrate between two physical hosts at SURFsara. New IGTF distributions are, and have been, provided on an almost monthly basis.

3.4.2 User Administration

Information about users and their accounts is maintained in a Lightweight Directory Access Protocol (LDAP) based repository. This facility is used to share among PRACE partners the authorization information needed by PRACE services and is used to retrieve information about users and their projects. Authorization information is provided among others for data transfers with GridFTP, interactive access through GSI-SSH, job submission with UNICORE, accounting services and access to the helpdesk facilities.

A single LDAP server is used for PRACE Tier-0 accounts. For almost all Tier-1 accounts the same server is used, however some partners host a local LDAP server for the registration of Tier-1 users. At present only IDRIS chose to make use of this external hosting option. Separate databases are used for Tier-0 and Tier-1 accounts. An overview of the LDAP namespace is shown in Figure 1. For Tier-0 accounts the “ou=ua,dc=prace-ri,dc=eu” part of the name space is used and for Tier-1 accounts “ou=ua,dc=deci,dc=org” is used as top part. The Tier-1 accounts registered by other LDAP servers can be accessed through referrals.

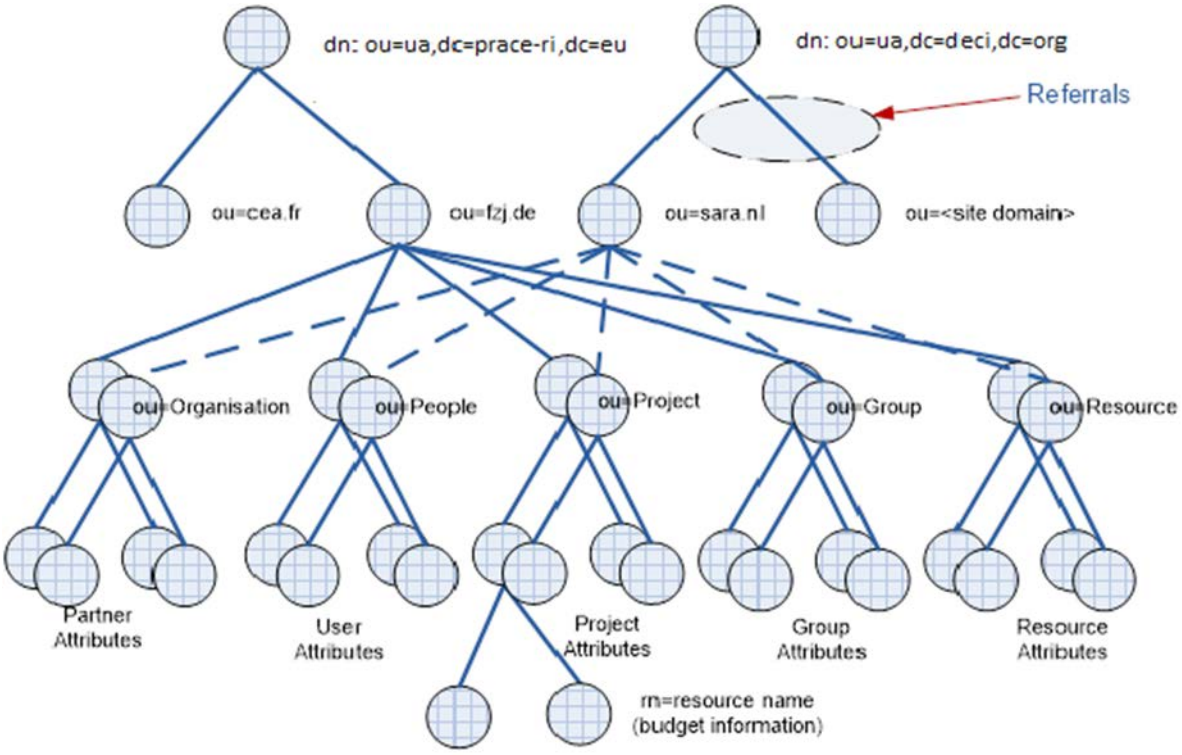

Figure 2: PRACE LDAP directory tree

All Tier-0 sites manage their own branch in the LDAP repository. They register the users who have access to their site in their LDAP branch, which starts at ou=<site domain>.

Each Tier-1 site also 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 of the user. The Home site in general is the site from the country where the user is working and not the site where Tier-1 resources have been granted to the user (the Exec site). The Home site mostly already knows the user and this makes the registration procedure easier. The Exec site retrieves the information from the LDAP server for the creation of a local account.

The main PRACE LDAP server and a replica server are operated by SURFsara. Both servers are virtualized and migratable. To avoid dependency on a single site for this core service, an additional remote replica server is operated by HLRS. The replica servers can be used if the primary server is unavailable for some reason. It is the distributed responsibility of the collective of Home sites to enter the correct information pertaining to their respective users in the central LDAP. It is the responsibility of each Exec site to retrieve the relevant information from the central LDAP and disperse it into their local user administration.

There is no specific prescription how to do this, merely that it must be done and that the delay should not be more than 24 hours. The Exec site may choose to fully automate this process in a way that fits its local user administration, or it may care of it entirely by a manual operation. This abstention from any specific definition of a required interoperability interface between the PRACE LDAP and a site-specific user administration in practice had made it easy to integrate new sites.

Like Exec sites, door nodes for interactive access, explained below, must process PRACE LDAP mutations in their local user administration. Unlike Exec sites, due to their function, they cannot be selective and must always keep track of the complete active user community.

Occasionally, when a new partner site joins PRACE, a new branch to the LDAP is added. In the reporting period, two new Tier-1 LDAP branches have been added for IUCC (IL) and UL (SI).

No changes have been made to the LDAP schema.

A minor policy change has been made to allow re-allocation of a reserved but unused uidNumber / gidNumber range for PRACE Preparatory Access Type-D projects.

3.4.3 Interactive Access

Interactive access to the Tier-0 systems is provided either by the SSH (Secure Shell) facilities, provided by most distributions of operating systems, or X.509 based SSH facilities. The latter are mainly used for access from other PRACE sites using the dedicated PRACE network. On PRACE Tier-0 and Tier-1 sites the recommended X.509 based SSH facility is GSI-OpenSSH [13] (GSI-SSH for short), distributed by the Globus community. All sites that support GSI-SSH are required to use Globus Toolkit 6.0 with OpenSSH 7.1 or newer.

For access from their workplace users can use GSISSH-Term, a GSI-SSH Java based client, which is supported by the PRACE partner LRZ [14].

Not all sites support GSI-SSH based access directly from the internet. Therefore, two partners, IT4I (CZ) and SURFsara (NL), provide GSI-SSH based access for all PRACE users who can use these sites as door nodes to other PRACE sites using the PRACE network. Door node services were previously provided by CINECA but IT4I recently took over this responsibility from them.

3.4.4 Accounting Services

Users can get accounting information locally at the sites at which they consume resources. Each site runs a standard PRACE accounting web service that can be accessed easily using GSI authentication. In addition, PRACE users and staff can get accounting information from several sites in a uniform way with the DART tool [15]. With this tool users can display their accounting information for Tier-0 systems and Tier-1 sites.

3.5 Operational Security and Security Forum

The main task of operational security within PRACE-5IP includes the user support on IT security issues within the PRACE dedicated network as well as the operation of the PRACE CSIRT itself. No major intrusions have been monitored. No services have been on risk.

One further task is the discussions on operational security within PRACE for potential future PRACE services. Here the PRACE Security Forum had several video conferences, discussing on these issues.

One major point is the PRACE Accreditation to GÈANT's Trusted Introducer program: Here, all required information has been collected, so that in the near future a "listing" of the PRACE CSIRT team can be requested. A listing of the team is a prerequisite for the final accreditation. It is planned to get listed within the next weeks. The accreditation is planned within the next three months.

Furthermore, contributions have been made to the WISE group (Wise Information Security in collaborating Environments), a community of security experts working on security issues in multi organizational infrastructures. Here PRACE signed at the TNC17 conference in June 2017 the SCI version 2 framework endorsement. This framework helps to generate trust relationships between collaborating infrastructures, which PRACE naturally is one.

At a joint PRACE-GÈANT meeting in Cambridge also further collaboration in the AAI environment, Authentication and Authorization Infrastructure, has been discussed. Here, a close collaboration with the AARC2 project has been agreed on.

3.6 User Services

User services subtask is composed of the following four activities: the PRACE Common Production Environment (PCPE), the Trouble Ticket System (TTS), Helper scripts and PRACE User Documentation.

3.6.1 PRACE Common Production Environment

The PRACE Common Production Environment (PCPE) presents a common application development environment interface across all PRACE execution sites regardless of underlying architecture. It guarantees availability of a minimum set of software components to support the running of highly-scalable parallel simulation software. Thus, it allows automated monitoring of key software components across all PRACE execution sites.

PCPE allows users to gain many benefits, such as improved efficiency in porting and running applications due to commonality of interface across PRACE; confidence that the required software for their research is in place at the start of their project; automated access to optimization best practice for local compilers, irrespective of vendor and hardware architecture.

An updated specification of PCPE (v1.1) has been released during the period. Previous specification (v1.0) was found complex and hard to implement at sites and to monitor properly. Therefore, the main goal of this new specification is to clarify things and help PRACE partner to understand what has to be done to setup PCPE and what has to be done to monitor its correct installation over time.

A focus has also been put on the “module” tool command which is at the root of the PCPE to enable or disable parts of this software collection. The module command has different implementations and at the time of the DEISA project the implementation called Modules-Tcl was the one chosen as the default to handle the DEISA Common Production Environment (DCPE). This Modules-Tcl implementation comes from the Environment Modules project [24], which also delivers a Modules implementation written in C language and widely available in Linux distributions. CEA has taken over the leadership of the overall Environment Modules open source project this year. Work has been achieved to publish a combined release of “Modules” and “Modules-Tcl”, making the later the default implementation provided by the project. Thus clarifying choices available when deploying a module command at sites.

3.6.2 Trouble Ticket System

The centralised Helpdesk [20] was deployed as part of the PRACE-1IP project. It is an important tool for the PRACE project staff to communicate among sites about problems. The Helpdesk is

also available for users. However, Tier-0 users are advised to contact the helpdesk of the Tier-0 sites directly.

TTS is based on the Request Tracker (RT) issue tracking system [21]. Since its deployment in PRACE-1IP, this service is hosted and operated at CINECA.

No relevant change was made on the PRACE TTS as no new need arise during this first period of PRACE-5IP.

3.6.3 Helper scripts

User services subtask provides helper scripts for PRACE users and staff to easily access the PRACE services in operations. Among these scripts, the *prace_service* utility gives access to the directory of services helping users and staff to get information on server addresses and ports for the various services.

This year the *prace_service* configuration has been updated many times to include the new PRACE Tier-0 and Tier-1 systems or to remove the systems that have been decommissioned.

No change was made on the *prace_service* script itself, as no new need arise during this first reporting period.

3.6.4 PRACE User Documentation

The User documentation for PRACE is available online on the PRACE-RI website [1]. Source document are stored in the PRACE SVN repository [22] in the MarkDown [23] format. During the PRACE-4IP project, a publication process was established to define how documentation changes made in the PRACE SVN repository are validated and uploaded on the PRACE-RI website.

This first period of activity, some updates of documents happened mainly to add references to the new PRACE Tier-0 and Tier-1 systems and to remove descriptions of systems that have been decommissioned.

3.7 Monitoring Services

Monitoring services sub-task is responsible for maintenance and development of PRACE-RI resource monitoring system and for supporting all parties involved in the monitoring process including sites (site representatives and generic service maintainers) and users including OoD/HdD on duty.

Monitoring functionality is provided by Icinga2 software with all changes required by PRACE project developed internally as well as metrics and custom methodology of check execution based on GSI-SSH protocol and proxy distribution.

Currently, the monitoring service is being hosted in PSNC and consists of two production machines:

- **mon.prace-ri.eu** - hosting web frontend and customised icinga2 instance
- **mon02.internal** - machine responsible for authentication and PRACE ldap interaction on behalf of the frontend node

After heavy development and testing phase in PRACE-4IP, Monitoring Services activity entered stabilization phase at the beginning of PRACE-5IP. Entering production state of a service changed the way in which service is being developed and operated. Monitoring service availability and stability became a priority over the new functionalities. On the other hand, to enable future development, a new testbed consisting of two machines has been created in PSNC to allow introduction of the new features and deep code changes without disturbing production instance of the service. The changes on production instance are now less frequent but more significant and better tested.

One of two main important activities mentioned before is maintenance. Current effort of keeping production state of a service and maintaining availability is a constant process consisting in:

- maintaining of underlying hardware and network infrastructure
- keeping systems secure by monitoring security aspects of operating systems, system software, middleware and service software and applying security updates and fixes as soon as possible to reduce the risk of system compromise
- updating icinga2 service software to address issues and problems found in the code
- solving issues reported by monitoring system clients in the frontend, backend, monitoring scripts, overall system behaviour or other components behaviour
- assuring that changes in PRACE infrastructure are reflected in the monitoring system including new system/service additions as well as removals of decommissioned systems and services;
- helping sites and supporters with solving problems indicated by failed checks in the monitoring system to increase infrastructure availability.

Development activity focuses on introducing necessary changes resulting from project requirements. During the reporting period, monitoring system code has been altered to enable integration with PRACE LDAP. The change resulted in an addition of new functionality of X.509 certificate based authentication and authorization of users together with SSL secured access.

New features which development started in the reporting period include:

- Configuration generator: fix of service dependencies to reduce number of alarms issued for failed services
- Generation of site availability reports based on data collected by monitoring service
- Inclusion of new security and LDAP infrastructure checks

Planned completion and production deployment of above functionalities is due by the end of Q2/2018

During the current reporting period, there were no additions of new systems. At the time of writing, one site has requested assistance (IUCC) and the integration is in-progress state.

Older systems have been decommissioned on request of their owners due to end-of-life and removed from the monitoring system. The requests were in form of RFC documents.

All availability and historical data related to those systems have been retained for availability calculation in the future. The list of removed systems include:

- UHEM-Karadeniz

- CINECA-Galileo
- JUELICH-Juqueen
- PSNC-Chimera
- NIIF

Current monitoring concerns the following systems:

No	Site/System	Long name of site	gsissh address
1	bsc-marenostrum	Barcelona Supercomputing Center	gftp.prace.bsc.es -p 2222
2	bsc-minotauro	Barcelona Supercomputing Center	gftp.prace.bsc.es -p 2222
3	castorc-cytera	CaSToRC CY-TERA High Performance Computing Center	login2-p.cytera.cyi.ac.cy -p 2222
4	cea-curie	Commissariat a l'energie atomique et aux energies alternatives	curie-prace.ccc.cea.fr -p 2222
5	cineca-marconi	Consortium of Italian universities	gssh-prace.marconi.cineca.it -p 2222
6	csc	The Finnish IT center for science	sisu-prace.csc.fi -p 2222
7	epcc	Edinburgh Parallel Computing Centre, ARCHER/UK-RDF	dtn01-prace.rdf.ac.uk -p 2222
8	ichec	Irish Centre for High-End Computing	not available
9	idris	Institut du Developement et des Ressources en Informatique Scientifique	turing2-d.idris.fr -p 2222
10	lrz-supermuc-fat	Leibniz Supercomputing Centre, SuperMUC Fat Node Island	supermuc-fat-prace.lrz.de -p 2222
11	lrz-supermuc	Leibniz Supercomputing Centre, SuperMUC Thin Node Islands	supermuc-prace.lrz.de -p 2222
12	ncsa	National Center for Supercomputing Applications	not available
13	psnc-eagle	Poznan Supercomputing and Networking Center	eagle-prace.man.poznan.pl -p 2222

No	Site/System	Long name of site	gsissh address
14	rzg-hydra	Rechenzentrum Garching of the Max Planck Society and the IPP	hydra01t.rzg.mpg.de -p 2223
15	surfsara	SURFsara Computing and Networking Service	int1-prace.cartesius.surfsara.nl -p 2222
16	uio	University of Oslo	not available
17	wcss	Wroclaw Centre for Networking and Supercomputing	prace-bem-int.wcss.pl -p 2222
18	cyfronet	Academic Computer Centre CYFRONET	prace-int.cyfronet.pl -p 2222
19	seged	SEGED Supercomputing Center of NIIF Institute	seged-login.sc.niif.hu -p 2222
20	leo	DEBRECEN Supercomputing Center of NIIF Institute	leo-login.sc.niif.hu -p 2222
21	phitagoras	Supercomputing Center of NIIF Institute	phitagoras.sc.niif.hu -p 2222
22	it4i-anselm	IT4Innovations VSB-TUO	anselm-prace.it4i.cz -p 2222
23	it4i-salomon	IT4Innovations VSB-TUO	salomon-prace.it4i.cz -p 2222
24	grnet	Greek Research and Technology Network	gssh-prace.aris.grnet.gr -p 2222

Table 10: Monitored Tier-0 and Tier-1 systems

The following table presents number of objects reflecting currently monitored infrastructure size:

Monitored object	Amount
monitored hosts	84
service instances	1036
generic services	10
user accounts	270

Table 11: Number of entities in monitoring system

3.8 Generic Services

In general, all services that need an operational basis and a centralized distribution for the PRACE project (or a part of it) could be assumed as Generic Services.

The goal of this task is the provisioning of these services and the supervision of their operation, as they are crucial for the day-by-day work of the project.

3.8.1 Roles and actions

The leader of generic services is acting as a liason / consultant to consult with

- PRACE aisbl,
- other Work Packages (WP),
- external providers,

upon request when IT or operational issues of these entities arise.

Examples from the reported period are:

- Questions of another WP related to the operation of new or existing services:
  - Suggested operational aspects for new PRACE Events and Training portal with WP4.
  - Consulted further integration plans between Training, Events portals and PRACE Web site.
  - Negotiated *repository.prace-ri.eu* hosting and features with WP4/WP7.
  - Negotiated proposal of storage of web content in markdown format with WP3.
- Questions of PRACE aisbl when negotiating with service providers offering service for PRACE:
  - Helped with consulting mailing lists provider mitigating spam issues.
  - Consulted operational aspects with service hosts.
  - Helped managing domain portfolio.
  - Helped with certificate installation.
  - Helped detecting and mitigating security incident on website.
- Helping internal workflows of using a shared PRACE-owned resource:

- Handled **.prace-ri.eu* wildcard certificate, general help provided, helped with new certificate requests.
- Handled *prace-ri.eu* domain: subdomain requests, helped with resolving domain issues related to a change.

Participation to discuss hosting centralisation plans of PRACE aisbl and definition of potential alternatives of hosting e.g. websites.

In connection with the upcoming GDPR requirements, an action plan on preparation of implementing new data protection measures was designed and coordinated by PRACE aisbl, BoD and MB with the help of Generic Services leader. The process was started with service overview and by creating an assessment survey. Based on results of the survey review of aspects of the plan was done.

Migration from PRACE SVN to Git has started by planning the migration, creating a private test project and migrating WP6 related branch to PRACE Gitlab with all commits tracked and maintained.

4 Conclusions

In this reporting period, Task 6.1 has continued the successful operation of the PRACE common services for the Tier-0 sites and the Tier-1 sites providing services for Tier-0. The operational procedures have continued to be successful in maintaining a reliable and available set of integrated services. The on-duty activity has a weekly schedule and continuously supervises the HPC infrastructure to assure a more concrete and punctual support on the day-by-day operation monitoring actions and support. This activity involves 22 PRACE partners with weekly shifts. Every week a report on the duty activity is produced, reporting the status of the infrastructure, of the core service and the problems occurred.

To ensure a seamless use of the PRACE distributed research infrastructure the PRACE common services are available on all PRACE Tier-0 systems and on the identified Tier-1 systems providing services for Tier-0. Through this common service provision, the Tier-0 and Tier-1 systems are presented as a single infrastructure, which enables a smooth interoperation of Tier-0 and Tier-1 systems.

A new version of the Service catalogue was given PRACE BoD for approval to update the status of few services (Unicore, gsatellite and the new generic services).

Thanks to the procedures for incident and change management, Task 6.1 operates and monitors on a day-by-day basis the complete set of PRACE common services, as defined in the Service Catalogue. This activity will continue and improve in PRACE-5IP in the context of the PRACE 2 programme, where new powerful Tier-0 and Tier-1 systems will further improve the infrastructure and enrich the whole European HPC eco-system at the service of Science.