



**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

D6.5

First Report on Collaborations with e-Infrastructures and CoEs
Final

Version: 2.0
Author(s): Olivier Rouchon, CINES
Date: 28.02.2018

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: < D6.5 >	
	Deliverable Nature: <Report>	
	Dissemination Level: PU*	Contractual Date of Delivery: 28 / February / 2018
		Actual Date of Delivery: 28 / February / 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 Report on Collaborations with e-Infrastructures and CoEs	
	ID: D6.5	
	Version: <2.0>	Status: Final
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2013	
	File(s): PRACE-5IP-D6.5.DOCX	
Authorship	Written by:	Olivier Rouchon, CINES
	Contributors:	Lilit Axner, KTH Stéphane Coutin, CINES Giuseppe Fiameni, CINECA Ralph Niederberger, FZJ
	Reviewed by:	Alexandre Strube, FZJ Veronica Teodor, FZJ
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	31/January/2018	Draft	Initial draft
0.2	05/February/2018	Draft	Added FZJ contribution on security
1.0	12/February/2018	Final version	Minor changes following internal review
2.0	28/February/2018	Final version	Update after MB/TB review

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Collaboration, Cooperation, Data management, Security, Training
------------------	--

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.

Copyright notices

© 2018 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 EINFRA-730913 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

Document Control Sheet.....	i
Document Status Sheet	i
Document Keywords	ii
List of Figures	iv
List of Tables.....	iv
References and Applicable Documents	iv
List of Acronyms and Abbreviations.....	vi
List of Project Partner Acronyms.....	vii
Executive Summary	1
1 Introduction.....	2
2 Collaboration with other Infrastructures.....	3
2.1 EUDAT	3
2.2 GÉANT	3
2.3 EGI.....	4
2.4 Human Brain Project	4
2.5 XSEDE / RIST	4
2.6 Other	4
3 Collaboration with Centres of Excellence.....	5
3.1 Energy oriented Centre of Excellence	5
3.2 Performance Optimisation and Productivity Centre of Excellence.....	6
3.3 BioExcel Centre of Excellence	6
4 Collaboration on the Data Management Domain	7
4.1 Leveraging EUDAT Services.....	7
4.2 DECI-13 and DECI-14 Pilots.....	8
4.2.1 <i>CHARTERED2 Data Pilot</i>	<i>9</i>
4.2.2 <i>subGridEoR Data Pilot.....</i>	<i>10</i>
4.3 Collaboration with EUDAT on the training domain.....	10
4.4 Leveraging security collaboration.....	11
5 Collaboration on the security domain.....	12
5.1 Collaboration with WISE	12

5.2	Incident Handling – Trusted Introducer (TI) programme.....	12
5.3	Federated Authentication and Authorisation	13
6	Conclusions.....	14
7	Annex	15
7.1	CHARTERED2 data pilot – data management plan	15

List of Figures

Figure 1	: EUDAT Data Domain modelled on the ANDS[20] Data Curation Continuum	7
Figure 2	: Data Flow diagram for CHARTERED data pilot.....	9

List of Tables

Table 1	List of Centres of Excellence	5
Table 2:	DECI-13 call data pilots.....	8
Table 3:	DECI-14 call data pilots.....	9

References and Applicable Documents

- [1] <http://www.prace-project.eu/>
- [2] EGI <https://www.egi.eu/>
- [3] EUDAT project <http://www.eudat.eu/>
- [4] GEANT Project <http://www.geant.org/>
- [5] Human Brain Project <https://www.humanbrainproject.eu/>
- [6] Centres of Excellence <http://ec.europa.eu/programmes/horizon2020/en/news/eight-new-centres-excellence-computing-applications>
- [7] Energy Oriented CoE for computing applications <http://www.eocoe.eu/>
- [8] POP CoE in computing applications <https://pop-coe.eu/>
- [9] BioExcel CoE for Computational Biomolecular Research <https://bioexcel.eu/>
- [10] XSEDE <https://www.xsede.org/>
- [11] RIST <http://www.tokyo.rist.or.jp/rist/en/>
- [12] EUDAT service descriptions <http://www.eudat.eu/services>
- [13] Data Management Plan <https://www.eudat.eu/d22-data-management-plan>
- [14] eduGAIN <http://services.geant.net/edugain/Pages/Home.aspx>
- [15] AARC project <https://aarc-project.eu/>
- [16] WISE <https://wise-community.org/>
- [17] PRACE-4IP D6.6: http://www.prace-ri.eu/IMG/pdf/D6.6_4ip.pdf
- [18] PRACE-4IP D6.5: http://www.prace-ri.eu/IMG/pdf/D6.5_4ip.pdf
- [19] perSONAR <https://www.perfsonar.net/>

- [20] Australian National Data Service organization <http://www.ands.org.au/>
- [21] WISE <https://wise-community.org/>
- [22] WISE SCIV2 workgroup <https://wiki.geant.org/display/WISE/SCIV2-WG>
- [23] EUDAT data management training : <https://www.eudat.eu/training/research-data-management>
- [24] TF-CSIRT Trusted Introducer: <https://www.trusted-introducer.org/index.html>
- [25] Accreditation process: <https://www.trusted-introducer.org/processes/accreditation.html>
- [26] RFC 2350: <https://www.ietf.org/rfc/rfc2350.txt>

List of Acronyms and Abbreviations

aisbl	Association International Sans But Lucratif (legal form of the PRACE-RI)
ANDS	Australian National Data Service
BCO	Benchmark Code Owner
CoE	Center of Excellence
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture (NVIDIA)
DARPA	Defense Advanced Research Projects Agency
DECI	DEISA Extreme Computing Initiative
DEISA	Distributed European Infrastructure for Supercomputing Applications EU project by leading national HPC centres
DMP	Data Management Plan
DoA	Description of Action (formerly known as DoW)
EC	European Commission
EESI	European Exascale Software Initiative
EGI	European Grid Infrastructure
EoI	Expression of Interest
ESFRI	European Strategy Forum on Research Infrastructures
EUDAT	European Collaborative Data Infrastructure
GB	Giga ($= 2^{30} \sim 10^9$) 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
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
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.
HMM	Hidden Markov Model
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
HPL	High Performance LINPACK
ISC	International Supercomputing Conference; European equivalent to the US based SCxx conference. Held annually in Germany.
KB	Kilo ($= 2^{10} \sim 10^3$) Bytes ($= 8$ bits), also Kbyte
KNL	Knights Landing, an Intel Xeon Phi microarchitecture
LINPACK	Software library for Linear Algebra
MB	Management Board (highest decision making body of the project)
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
MOOC	Massively open online Course
MoU	Memorandum of Understanding.
MPI	Message Passing Interface
NAMD	Nanoscale Molecular Dynamics software
NDA	Non-Disclosure Agreement. Typically signed between vendors and customers working together on products prior to their general availability or announcement.
NREN	National Research & Education Network
PA	Preparatory Access (to PRACE resources)
PATC	PRACE Advanced Training Centres
POP	3Performance Optimisation and Productivity
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.
PRIDE	Project Information and Dissemination Event
RI	Research Infrastructure
TB	Technical Board (group of Work Package leaders)
TB	Tera (= $2^{40} \sim 10^{12}$) 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
UNICORE	Uniform Interface to Computing Resources. Grid software for seamless access to distributed resources.
VMD	Visual Molecular Dynamics, molecular graphics software

List of Project Partner Acronyms

BADW-LRZ	Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, Germany (3 rd Party to GCS)
BILKENT	Bilkent University, Turkey (3 rd 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'Énergie Atomique et aux Énergies Alternatives, France (3 rd Party to GENCI)
CESGA	Fundacion Publica Gallega Centro Tecnológico de Supercomputación de Galicia, Spain, (3 rd 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 (3 rd Party to BSC)
CYFRONET	Academic Computing Centre CYFRONET AGH, Poland (3 rd 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 (3 rd Party to ULFME)
GCS	Gauss Centre for Supercomputing e.V., Germany
GENCI	Grand Equipement National de Calcul Intensif, France
GRNET	Greek Research and Technology Network, Greece
INRIA	Institut National de Recherche en Informatique et Automatique, France (3 rd Party to GENCI)
IST	Instituto Superior Técnico, Portugal (3 rd Party to UC-LCA)
IT4Innovations	IT4Innovations National supercomputing centre at VŠB-Technical University of Ostrava, Czech Republic
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 (3 rd Party to SNIC)
LiU	Linköping University, Sweden (3 rd Party to SNIC)
NCSA	NATIONAL CENTRE FOR SUPERCOMPUTING APPLICATIONS, Bulgaria
NTNU	The Norwegian University of Science and Technology, Norway (3 rd Party to SIGMA)
NUI-Galway	National University of Ireland Galway, Ireland
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 (3 rd 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, Laboratório de Computação Avançada, Portugal
UCPH	Københavns Universitet, Denmark
UHEM	Istanbul Technical University, Ayazaga Campus, Turkey
UiO	University of Oslo, Norway (3 rd Party to SIGMA)
ULFME	UNIVERZA V LJUBLJANI, Slovenia
UmU	Umea University, Sweden (3 rd Party to SNIC)
UnivEvora	Universidade de Évora, Portugal (3 rd Party to UC-LCA)

D6.5

First Report on Collaborations with e-Infrastructures and CoEs

UPC	Universitat Politècnica de Catalunya, Spain (3 rd Party to BSC)
UPM/CeSViMa	Madrid Supercomputing and Visualization Center, Spain (3 rd Party to BSC)
USTUTT-HLRS	Universitaet Stuttgart – HLRS, Germany (3 rd Party to GCS)
WCNS	Politechnika Wroclawska, Poland (3 rd Party to PNSC)

Executive Summary

This deliverable presents the existing and new activities between PRACE and complementary e-Infrastructures with the aim to build an integrated Europe-wide e-Infrastructure for our users. The focus has been on the continuation and further development of the collaboration on security and data management topics.

PRACE has been collaborating on security with other e-Infrastructures from its beginning. This includes cooperation with the major European organizations (EGI [2], EUDAT [3], GÉANT [4], Human Brain Project [5]), as well as with Centres of Excellence (CoEs [6]) such as EoCoE [7], POP [8] or BioExcel [9]. There are also some ongoing preliminary discussions with worldwide consortiums (XSEDE [10] and RIST [11]).

Workshops attended by representatives from EGI, EUDAT, GÉANT, NRENs, PRACE as well as US participants, are held regularly to organize these partnerships. A Steering Committee, in which PRACE is actively engaged, coordinates the follow-up activities of this forceful collaboration, organised in working groups, which cover specific topics such as networks, authentication and security, data storage, training, etc. PRACE has a long history of collaboration with these e-Infrastructures, and has been using the services these e-Infrastructures provide in various domains in which it has interest.

For data services, PRACE collaborates with EUDAT, which offers a service catalogue [12] complementing PRACE in order to fulfil user needs during the full research data life cycle. During the past year, some integration in the areas of data transfer, storage and sharing has been implemented as four data pilots were run as part of the DECI-13 call, and another two have been launched after the DECI-14 call. A Data Management Plan template [13] has been used to ease requirements gathering during the early phase of each pilot and its systematic use is now highly recommended. The feedback and results of the first four pilots have been presented during the EUDAT conference in Porto in January 2018. The need for cross PRACE and EUDAT trainings has been identified, some have been delivered to support data pilots. This collaboration is increasing with EUDAT delivering some sessions in PRACE training events and vice versa.

Regarding authentication and authorisation services, PRACE is considering to use services provided by the Identity Provider (IdP) federations which exist for educational organisations in several countries and which are also federated in eduGAIN [14]. Working with external IdPs can enhance the security of the PRACE infrastructure because users do not have to use separate credentials for PRACE access. Also, credential misuse may be noticed earlier as more services will rely on the same credentials. This task follows other activities in this area, especially those of the AARC [15] project which is coordinated by GÉANT. Feedback has been provided on the requirements PRACE has for the acceptance of authentication services provided by external IdPs.

For security alert and incident management, PRACE has been collaborating with EGI and EUDAT for a long time. This cooperation will continue through the WISE [16] community initiative.

1 Introduction

One crucial direction given as part the H2020 programme (EOSC/EDI) is to allow for seamless access among the major research e-Infrastructures. In this integrated vision, a scientific workflow should be able to use services from PRACE as well as from other European providers such as EUDAT, EGI or GÉANT, or even from other continents (XSEDE for example).

The objective of this report is to provide a status about ongoing collaboration activities at the beginning of the PRACE-5IP project. The overall goal is to implement the collaboration as much as possible on the basis of user driven pilots.

An overview of the PRACE collaboration on the operational domain is provided in Section 2, on a per partner basis including active and potential cooperation, and is followed by the description of the activities with Centres of Excellence in Section 3. The two ensuing sections will give a deeper dive into the key collaboration domains.

Section 4 focuses on a vertical collaboration, the interoperability with EUDAT data services. We begin with a presentation of the benefit brought to PRACE users by using EUDAT data services and different outcomes of the security collaboration. We then describe the way we have identified relevant pilots, and for the running ones, provide a status including requirements, technical solutions and future evolution. Finally, we summarise the collaboration we have established around HPC and data management training.

In Section 5, we describe activities on security, without which there is no chance of building a sustainable collaboration. Activities with the major European e-Infrastructures have been ongoing for several years. We describe these ongoing activities as well as their evolution. We also present our involvement in the AARC project.

2 Collaboration with other Infrastructures

In this chapter the organisations PRACE is having or is planning to have collaboration for operations with are identified. For each of them, formal agreements such as a Memorandum of Understanding (MoU) and collaboration domains are described. Further details are also provided unless the area is more detailed in the next chapters.

2.1 EUDAT

Two MoU have been signed between PRACE and EUDAT. The initial one goes back to March 2014, and was intended to allow EUDAT to use the PRACE network to transfer data. The most recent one was signed off in June 2016 and covers, in addition to the network usage, the definition of governance and support processes, the identification of training and methods as well as the implementation of pilot projects.

The data service related collaboration has been an important activity and is detailed in the section 4. The data pilots highlighted the need for training, and a fruitful collaboration started (see subsection 4.3).

EUDAT is also involved in the cross organisation security workgroups (see section 4.4).

Following the need for a PRACE service explicitly requested by the European Commission for PRACE-4IP to offer easy and highly customisable search function of outcomes of PRACE funded activity (e.g. reports on results of PRACE resource allocation, white papers, best practice guides or results of PRACE IP Projects), it has been agreed to use the EUDAT B2SHARE service. This will be implemented as part of PRACE-5IP project.

2.2 GÉANT

The collaboration with GÉANT is probably the longest in PRACE history, and goes back to 2006. At that time, PRACE overtook the network put in place as part of the DEISA project, which was based on the GÉANT footprint and is still in operation. A close cooperation has been in place since then, which has been detailed in the previous PRACE-4IP reports [17][18]. A MoU has been in place since March 2016 in order to redesign the network layout.

Recently, a meeting involving PRACE-GÉANT representatives took place in Cambridge (September 2017), where further collaboration opportunities were reviewed: network (remote VPN service, perfSONAR monitoring), authentication and authorisation infrastructure (Trusted introducer service), and training (MOOC) were among the topics discussed.

perfSONAR [19] is a tool developed by Internet2, ESnet, Indiana University and GÉANT for end-to-end monitoring and troubleshooting of multi-domain network performance. It provides network engineers with the ability to test and measure network performance, as well as to archive data in order to pinpoint and solve service problems that may span multiple networks and international boundaries. The PRACE VPN is an ideal candidate to stress the perfSONAR idea and demonstrate the proper operation over several network providers. First discussions and ideas have been exchanged, but due to personal bottlenecks the collaboration has not yet started.

2.3 EGI

EGI has been involved for years with PRACE in different workgroups related to operational security and trust models. In order to develop further this collaboration, a MoU has been signed in March 2016. In addition to security related topics, it targets the ability to build an integrated service catalogue and more interoperability to support data driven science. At this stage, security related activities are active (see section 3), the others have not yet started.

2.4 Human Brain Project

The PRACE-HBP MoU was signed in November 2014 and focuses on the PRACE network usage by the HBP community. The access to the PRACE infrastructure seems obvious as the main HBP HPC partners (BSC, CINECA, CSCS, and JUELICH) are already part of PRACE.

2.5 XSEDE / RIST

XSEDE experts already contribute to the WISE (see section 5) community activities, so there is already some ongoing security collaboration.

A MoU was initiated by PRACE, RIST and XSEDE senior management following a meeting during SC'16 in Austin, with the objective to define a scientific steering committee, prepare for joint allocations, and align peer-review systems. It has been signed-off in April 2017.

2.6 Other

The PRACE-5IP Management Board approved the principal to use Tier-1 resources for the cooperation pilots with large scale scientific infrastructures during a meeting held in September 2017. Two collaborating institutions have been identified: the European Synchrotron Radiation Facility (ESRF) and the European Organization for Nuclear Research (CERN).

In November 2017, a call for volunteer has been published to identify Tier-1 centre willing to provide computing hours on an in-kind basis and implement a first pilot based on the pre-data acquisition simulations made at ESRF. At this stage, three institutions have offered to participate: CINES, CYFRONET and PSNC. The initiative is in the requirements gathering phase, the first meetings have been held to discuss the various constraints and assign the simulation codes to be executed to the centres where would fit. The next steps will be to define a roadmap for the deployment phase. In the meantime, the scope for the collaboration with CERN is being discussed.

3 Collaboration with Centres of Excellence

Nine Centres of Excellence (CoEs) for computing applications have been funded as part of the EC Horizon 2020 programme. The WP6 team had several meetings with their representatives (as reported in previous reports, e.g. PRACE-4IP D6.5 [16]) in order to understand their needs, give them access to the relevant PRACE services, or organise joint events. The list of CoEs is shown below.

Coordinator / Organisation	Country	Acronym	Title
COMMISSARIAT A L'ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES	FR	EoCoE	Energy oriented Centre of Excellence for computer applications
KUNGLIGA TEKNISKA HOEGSKOLAN	SE	BioExcel	Centre of Excellence for Biomolecular Research
MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	DE	NoMaD	The Novel Materials Discovery Laboratory
CONSIGLIO NAZIONALE DELLE RICERCHE	IT	MaX	Materials design at the eXascale
DEUTSCHES KLIMARECHENZENTRUM GMBH	DE	ESiWACE	Excellence in Simulation of Weather and Climate in Europe
UNIVERSITY COLLEGE DUBLIN, NATIONAL UNIVERSITY OF IRELAND, DUBLIN	IE	E-CAM	An e-infrastructure for software, training and consultancy in simulation and modelling
BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	ES	POP	Performance Optimisation and Productivity
UNIVERSITAET POTSDAM	DE	COEGSS	Centre of Excellence for Global Systems Science
University College London (UCL)	UK	CompBioMed	Centre of Excellence for Computational Biomedicine

Table 1 List of Centres of Excellence

The collaborative activities of other PRACE-5IP Work Packages with CoEs have been checked regularly, and led to activities for this task. This is particularly the case with BioExcel, which comes from a WP4 initiative. For this reason, the commitment to support the CoEs in T6.3 will continue for the duration of the project.

3.1 Energy oriented Centre of Excellence

The EoCoE is a hub coordinating a pan-European network of partners which are strongly engaged in both scientific simulation and energy fields. It aims to foster and accelerate the European transition to a reliable and low carbon energy supply using HPC.

In October 2017, a “Scientific Applications towards Exascale” workshop was organised in CINES. The general topic of the first one-and-a-half days were HPC scientific applications on new

computing architectures with a specific focus on the Intel Xeon Phi Knight's Landing processor (KNL), in order to prepare potential users for the KNL-based cluster that had been installed at CINES as part of the PRACE-PCP initiative. The last day was dedicated to a hands-on session on this machine, where participants will be able to try porting their code and make first performance and energy measurements on the ATOS-BULL PCP pilot system. During this session, BEO, HDEEViz and SLURM energy plugin were presented and evaluated by a group of selected user. More than forty people participated at the event, and the feedback was rather positive, as the proposed tools were considered as useful to understand the power consumption behaviour of their own applications.

3.2 Performance Optimisation and Productivity Centre of Excellence

The POP CoE in Computing Applications provides performance optimisation and productivity services for academic and industrial code(s) in all domains.

Following a meeting in May 2017 in Barcelona during which a potential collaboration around performance assessment and optimisation was discussed, it has been agreed to put together a MoU to describe a pilot between POP and PRACE. The objective of this pilot is to set up the necessary tools and documentation to evaluate the performance of the codes (projects) running on PRACE machines. The work plan will include the selection of the most appropriate analysis tools, the documentation of the installation and usage of the selected tools, the installation of the tools in PRACE systems, the use cases of the pilot and the conclusions and recommendations.

The MoU is still in progress, activities are due to start when the MoU is signed.

3.3 BioExcel Centre of Excellence

The aim of the BioExcel CoE is to enable better science by improving the most popular biomolecular software and spreading best practices and expertise among the communities through consultancy and training. It should act as a facilitator, a central hub for biomolecular modelling and simulations.

The 2017 PRACE Spring School took place in April 2017 at KTH, with the objective to allow participants to get a comprehensive introduction to the different codes for molecular modelling and simulations (Gromacs, Amber, NAMD, VMD), understand their scalability and performance while avoiding potential issues, and learn the best practices about using them on HPC systems. Extensive hands-on sessions also covered more than half of time during the school. Fifty-five people attended the event, and more than 75% of the students gave the training a good or excellent rating in the feedback form. The balance of theoretical and practical content of the classes was also given a very good rating, and the overall workshop organisation was rated favourably.

Another joint training event is under preparation for May 2018, and will cover the performance optimisation tools by ARM as well as EUDAT data management techniques and tools.

4 Collaboration on the Data Management Domain

4.1 Leveraging EUDAT Services

One of the objectives of PRACE-5IP WP6 is to increase our portfolio in addition to the existing internal PRACE services and offer PRACE users some further data management services. This is done by developing collaborations with other e-Infrastructures, and particularly with EUDAT, by implementing pilots selected within the PRACE/DECI user projects.

EUDAT is seen as a perfect fit as it provides a collection of services oriented to Data Management [12], while PRACE is focused on HPC resources. In this context, pilot projects can take advantage of the computing power offered by PRACE in the phase of data production and of the storage and long-term data preservation by EUDAT in post-production. The most widely used EUDAT service is B2STAGE especially because it is easy to use and has reduced setup times. B2SHARE is also used to explore the possibilities for sharing sets of experimental data and post-production results.

The tight integration of PRACE/EUDAT services can be depicted as follows:

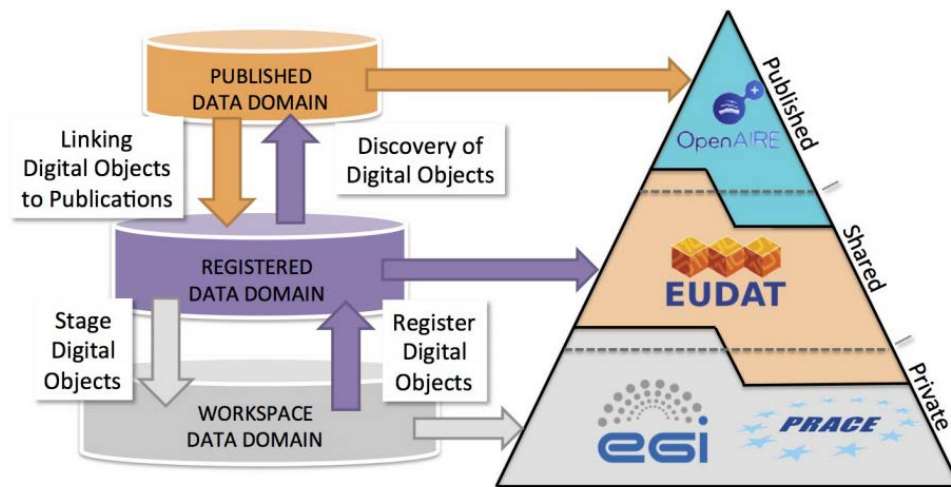


Figure 1 : EUDAT Data Domain modelled on the ANDS [20] Data Curation Continuum

The first step of such pilot is to gather the user requirements to understand the expected use cases and the nature of the data. The Data Management Plan (DMP) template elaborated by EUDAT has been used to assist this and ensure the full data life cycle is taken into account.

To fulfil the requirements we have identified some existing EUDAT services such as data transfer, data and metadata publication and search, or data preservation, but also described new ones like workspace storage or data processing at storage location.

It is crucial to work towards a standard service approach for such outsourcing. This will ensure an easy uptake and clear agreements for users and communities as well as inclusion into scientific workflow. Seamless services integration between e-Infrastructures must rely on robust technical solutions.

4.2 DECI-13 and DECI-14 Pilots

The DECI-13 and DECI-14 calls are issued as a pilot call jointly with EUDAT. They offer the opportunity to allocate compute and data resources to projects that also have needs for storage capacity and associated data services. Applicants may apply for both the DECI and EUDAT calls at the same time or may apply to just one of the calls.

Five projects had been selected as pilots during the DECI-13 call, they are shown below:

Project Acronym	Project Title	Field	Data req. in TB	PI country	Exec site	EUDAT site
HybTurb3D	Hybrid 3D simulations of turbulence and kinetic instabilities at ion scales in the expanding solar wind	Astro Sciences	140 TB	IT	SURFsara	CINECA
MULTINANO	Multiscale simulations of nanoparticle suspensions	Engineering	30 TB	IT	RZG	CINECA
HiResClimate	High Resolution EC-Earth Simulations	Earth Sciences	150TB	IE	KTH	EPCC
AFiD	Effect of rotation and surface roughness on heat transport in turbulent flow	Engineering	11TB	NL	EPCC	SURFsara
CHARTERED	Charge transfer dynamics by time dependent density functional theory	Materials Science	30TB	SE	IT4I	KTH/PDC

Table 2: DECI-13 call data pilots

The use case of these pilots – which were run from April 2016 to March 2017 – is described in the PRACE-4IP D6.6 deliverable [17], the assessment of the results by the EUDAT panel is ongoing. Initial feedback has been presented during the final EUDAT conference in Porto in January 2018, and the final evaluation will be reported in the EUDAT deliverable to be published in March 2018.

Three out of the applications requested for access to EUDAT resources during the DECI-14 call within the PRACE Optional Programme. At the end of the DECI appraisal process, only two candidates have been selected. One of them is the continuation of data pilot that ran as part of DECI-13. Both started in April 2017.

Project Acronym	Project Title	Field	Data req. in TB	PI country	Exec site	EUDAT site
subGridEoR	The effects of small-scale structure and halo stochasticity on Cosmic Reionization	Astrophysics	100TB	UK	surfSARA	STFC

Project Acronym	Project Title	Field	Data req. in TB	PI country	Exec site	EUDAT site
CHARTERED2	Charge transfer dynamics by time dependent density functional theory	Materials Science	30TB	SE	IT4I	KTH/PDC

Table 3: DECI-14 call data pilots

4.2.1 CHARTERED2 Data Pilot

The CHARTERED2 project is in the field of Materials Science, focusing on “Charge transfer dynamics by time dependent density functional theory”, and had initially been selected as part of the DECI-13 call. The project team is based in Uppsala University (Sweden). The PRACE execution site is IT4Innovations (Czech Republic) and the EUDAT site is PDC/KTH (Sweden).

The data management plan has been initiated during the first phase of the pilot, and completed iteratively along some regular calls. The current version is available in annex 6.1. The data sharing section of the document is not yet filled, as the results of the simulations need to be evaluated before publication. This will be done at a later stage.

The following data flow diagram describes the objective of the pilots based on the requirements we gathered.

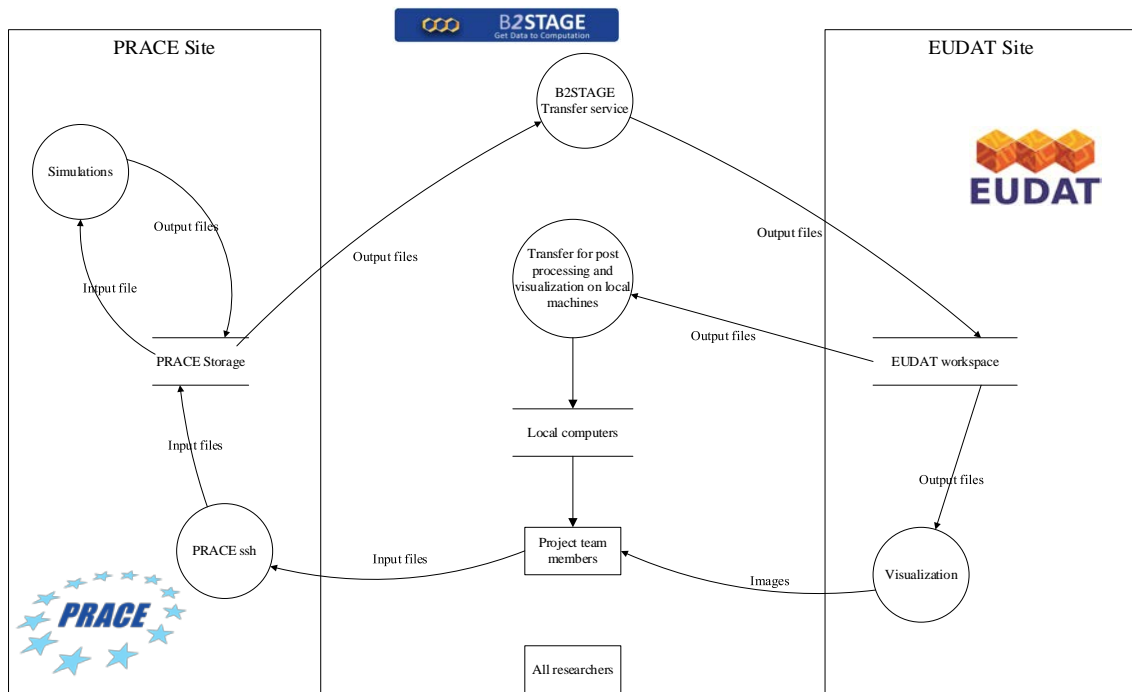


Figure 2 : Data Flow diagram for CHARTERED data pilot

There is now a good understanding about the data used in this project and the data flows. It is agreed with the project team that they will need to define what data could be shared and when. This will drive the definition of associated metadata.

4.2.2 *subGridEoR Data Pilot*

The pilot is currently on hold as STFC, the EUDAT partner that had to provide data resources, withdrew their commitment due to a change in their priorities. This has been escalated to the EUDAT Work Package leader and is being actively worked for resolution.

4.3 Collaboration with EUDAT on the training domain

While working on the data pilots described above, it appeared that PRACE users are usually small groups (a project and not a community) who produce large volume of data, as maturity levels within the users can vary: while they are subject matter experts, they lack data management skills and would need to be trained to good practices. A working group involving PRACE and EUDAT training managers, meeting via video conference, was established to facilitate this.

As a first step, the PRACE project staff and to data pilots users had some informal research data management training using some of the EUDAT material [23] and attended the EUDAT/OpenAire webinar on the same topic. The objective was to present research data challenges, the required services along the data life cycle, as well as the European Commission recommendations on the subject. The feedback was very positive, and it was agreed that this type of training should be the starting point.

To strengthen the collaboration, the workgroup has agreed on the cross reference of training websites from both e-Infrastructures. Training events from the partner are also adverted on each other mailing list.

Finally, it was agreed to include some PRACE content into EUDAT trainings and vice versa. The current calendar for this is as follows:

- During PRACE 2017 Spring School joint event with VI-SEEM, Cyprus, EUDAT led a session about ‘Data Management Plans – EUDAT best practices and case study’ as part of the Data/Computational Services for Scientific Communities stream (25 to 27 April 2017).
- During the HPC Summit Week 2017, Barcelona, PRACE contributed a workshop about “Coupling HPC and Data Resources and services together”, as part of a joint event during the PRACEdays17 (18 May 2017).
- During ISC 2017 conference, Frankfurt, PRACE organised the 3rd session in the series of the education and training related workshops (22 June 2017). A EUDAT training team member participated to the event.
- During the EUDAT & Research Data Management Summer School, Heraklion, partners involved in PRACE delivered several sessions about HPC data management (July 2017).
- During the PRACE joint event with BioExcel, Stockholm, focusing on ARM & EUDAT performance optimisation and data management tools, EUDAT will lead a session on DMP good practices (May 2018).

Now that this collaboration started, we will use feedback from the data pilots and the trainings to refine the training content. We have already gathered some good suggestions, particularly from the CHARTERED2 data pilot.

4.4 Leveraging security collaboration

It quickly appeared, when we reviewed the requirements, that solutions for security related elements like AAI and operational security are a cornerstone of a reliable integration. Furthermore, the solutions must be accepted by both partners. Thus, we want to leverage the work done in the collaboration for the security activities (WISE, AARC project, EuGridPMA, etc.).

The main area of focus is the AAI. The objective for a PRACE user is to use any of the authenticated EUDAT services with its PRACE credentials. Today, PRACE uses a dedicated X.509 certificate based identification for some services, while EUDAT implements a common framework for all its services named B2ACCESS, based on FIdM (federation of identity).

To cope with this, a `deisaUserProfile` attribute was added in the PRACE LDAP schema. It is set with value "EUDAT" for users who indicated consent that their profile information may be shared between PRACE and EUDAT. In collaboration with AARC and EUDAT projects, a prototype has been developed and tested on non-production environments. It replicates into EUDAT B2ACCESS the PRACE users with `deisaUserProfile = EUDAT`, enabling them to use their X.509 certificate to transfer data between both e-Infrastructures using GridFTP.

This has demonstrated the feasibility of an integrated user authentication. Based on this, the next steps will start with an end-to-end specification of AAI integration process.

5 Collaboration on the security domain

5.1 Collaboration with WISE

The WISE community [21] was born as the result of a workshop in October 2015, which was jointly organised by the GÉANT group SIG-ISM (Special Interest Group on Information Security Management) and SCI, the ‘Security for Collaboration among Infrastructures’ group of staff from several large-scale distributed computing infrastructures.

WISE provides a trusted global framework where security experts can share information on topics such as risk management, experiences about certification processes and threat intelligence.

WISE is governed by a steering committee and project-managed by staff from GÉANT.

A workshop was held in Amsterdam in March 2017, where PRACE was directly involved in the specific topics in scope in the four working groups:

- The Security for Collaborating Infrastructures (SCI), producing the SCIV2 framework [22];
- Training, e.g. security awareness training module for users (can be used by PATCs);
- Risk assessment and Audits by sites based on best practice;
- Confidentiality, integrity and availability of Big Data, working on a document “Definition of Big and Open Data” and gathering use cases.

PRACE endorsed the SCIV2 during the TNC conference in Linz (June 2017).

5.2 Incident Handling – Trusted Introducer (TI) programme

PRACE has been running an internal security team (PRACE CSIRT) for managing security incidents within the infrastructure for several years already.

As PRACE is a European wide infrastructure it is important to be able to collaborate with CSIRTs from all over Europe and beyond because of global collaborations. Therefore in 2016 the idea has been born to let the PRACE CSIRT team become accredited for the Trusted Introducer (TI) Service provided by GÉANT [24]. The main benefit of the accreditation is that it enables the collaboration with other CSIRTs and provides access to information related to incidents and vulnerabilities not available otherwise.

The TI service offers three status levels for member teams: listed, accredited and certified. For the PRACE CSIRT it is proposed to apply for accreditation, since this status provides the most benefit to PRACE concerning cost/benefit ratio.

The accreditation process [25] requires several prerequisites which are currently worked on together with GÉANT. Much of the required material is administrative information, e.g. several documents have to be prepared and made available to the public. PRACE PMO and BoD already supported the accreditation of the PRACE CSIRT team and PRACE aisbl will cover the costs.

Currently the PRACE CSIRT applies for the listed status for the TI services as prerequisite to become accredited. The RFC2350 document [26] to be filled out is in preparation. Also, the responsibilities within the PRACE CSIRT team are under discussion.

5.3 Federated Authentication and Authorisation

The Authentication and Authorisation for Research and Collaboration (AARC) [15] initiative was first launched in May 2015 to address the increased need for federated access and for authentication and authorisation mechanisms by research and e-Infrastructures.

Running for two years, AARC was funded by the European Union's Horizon 2020 research and innovation programme. A second phase of the project (AARC2) started in May 2017 to continue to develop and pilot an integrated cross-discipline authentication and authorisation framework, building on existing authentication and authorisation infrastructures (AAIs).

Already in the first AARC project, a close collaboration between the two projects, PRACE and AARC, has been initiated. Within AARC2 PRACE is actively participating to the AEGIS (AARC Engagement Group for Infrastructures), which brings together representatives from research- and e-Infrastructures, operators of AAI services and the AARC team to bridge communication gaps and make synergies happen.

6 Conclusions

In the general framework of the H2020 programme (EOSC/EDI), the aim is to provide to the users a seamless approach integrating all the major research e-Infrastructures. To this aim, the most important factor is to understand the interplay between the different service providers (e.g. EUDAT, EGI, GÉANT, etc.) and to contribute to the fruitful cooperation between them.

In this Deliverable we reported our work spent in this direction, by establishing relations with other projects and/or e-Infrastructures and CoEs. We reported about the ongoing collaborations at the beginning of the PRACE-5IP project based on user driven pilots.

We described the PRACE collaboration on the operational domain, in relation with overall HPC European Eco-System, with a particular mention to the CoEs. We investigated the potential interaction between PRACE operational services and CoEs and reported about the ongoing collaborations.

We presented in detail the interplay between PRACE and EUDAT. This enhanced the offer towards the users, permitting them to operate with more efficient and complete workflows. To demonstrate this, we report about the pilots that we ran highlighting the benefits coming from a common HPC and data management.

We discussed the activity on security, which is a key-factor and that requires to be treated as a common cross-activity between all the European e-Infrastructures in a persistent way. The collaboration with GÉANT, AARC and WISE are examples of work made in this direction.

The HPC European Eco-System is a complex landscape in which PRACE is working to actively collaborate with the other actors. This is crucial to provide the users with a modern and comprehensive service that will permit to better satisfy their computational needs. In this Deliverable we reported the activity from the beginning of the PRACE-5IP project so far. We are looking forward to make these collaborations more persistent and to establish new ones.

7 Annex

7.1 CHARTERED2 data pilot – data management plan

Data Collection

What data will you collect or create?

The project data set types are the following:

Data set type	Size	Format	Comments
Input files	10th of MB.	ASCII files	. One input file per simulation
Output files	Average is 1GB, biggest are 10 GB 30000 files	Specific Fortran binary ASCII files	These are raw data
Post processing results			They will be stored on local machine.

How will the data be collected or created?

Simulations:

All output files are created as results of the simulations ran on HPC as part of the DECI project.

Once simulations are run, a manual check is done to ensure the simulation was well executed / didn't crash.

The volume will grow on a linear way across time.

Post-processing:

The post-processing is some file adaptation and visualization.

Raw data are downloaded to local computers in their labs by researchers. For each job, the data volume downloaded is about 5 to 10 Gb Visualization is done with tools like vmd or vesta.

An alternative solution is to perform visualization at PDC/KTH to avoid file transfer (tools availability, network bandwidth ...). It appears that the required tools are available (see <https://www.pdc.kth.se/software/>)

Note also that there is no plan to perform visualization at IT4I. But it would be possible (https://docs.it4i.cz/salomon/list_of_modules#vis).

Documentation and Metadata

What documentation and metadata will accompany the data?

No metadata managed.

Ethics and Legal Compliance

How will you manage any ethical issues?

So far no ethical issue identified.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

This management has to be defined. Doesn't sound to be crucial at this stage.

Storage and Backup

How will the data be stored and backed up during the research?

In IT4I (PRACE site):

The required 30 TB storage space are available. They will be on the scratch space which is not backed up.

Available 3 months after end of DECI project.

The transfer should be triggered manually using a simple command.

In KTH/PDC (EUDAT site):

Storage is on disk, using B2SAFE service. No backup on tape so far

How will you manage access and security?

In IT4I: Permissions are associated to the PRACE login.

In KTH/PDC: permissions are assigned to PDC delivered login.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

What is the long-term preservation plan for the dataset?

Nothing planned so far. To be completed.

Data Sharing

How will you share the data? Are any restrictions on data sharing required?

Data sharing requirements are summarized in the following table

Data set type	Before resulting papers publication	After publication
Input files	Project team only	To be defined
Output files	Project team only	To be defined
Post-processing results	Project team only	To be defined