



**E-Infrastructures
H2020-EINFRA-2014-2015**

**EINFRA-4-2014: Pan-European High Performance Computing
Infrastructure and Services**

PRACE-4IP

PRACE Fourth Implementation Phase Project

Grant Agreement Number: EINFRA-653838

D6.5

Inventory of links with e-Infrastructures and CoEs

Final

Version: 1.0
Author(s): Stéphane Coutin, CINES
Jules Wolfrat, SURFsara
Date: 30.11.2015

Project and Deliverable Information Sheet

PRACE Project	Project Ref. №: EINFRA-653838	
	Project Title: PRACE Fourth 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: 30/11/2015
		Actual Date of Delivery: 30/11/2015
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 2991/844/EC.

Document Control Sheet

Document	Title: Inventory of links with e-Infrastructures and CoEs	
	ID: D6.5	
	Version: <1.0>	Status: <i>Final</i>
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2010	
	File(s): D6.5.docx	
Authorship	Written by:	Stéphane Coutin, CINES Jules Wolfrat, SURFsara
	Contributors:	Lilit Axner, KTH Andy Turner, EPCC Giuseppe Fiameni, CINECA
	Reviewed by:	F. Berberich, JUELICH S. Erotokritou, CaSToRC
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	01/October/2015	Draft	Initial document structure
0.5	03/November/2015	Draft	Draft version consolidating individual contributions
0.6	10/November/2015	Draft	Updated security contributions
0.7	10/November/2015	Draft	Submitted for internal review
0.8	16/November/2015	Draft	Ongoing update after

D6.5**Inventory of links with e-Infrastructures and CoEs**

			internal review
0.9	20/November/2015	For Approval	Addressed comments from internal review
0.92	30/November/2015	For Approval	Addressed comments from approvers
1.0	30/November/2015	Final version	

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Data management, Security, links with e-infrastructure, CoE
------------------	--

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-653838. 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

© 2015 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-653838 for reviewing and dissemination purposes.

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

Table of Contents

Project and Deliverable Information Sheet	i
Document Control Sheet.....	i
Document Status Sheet	i
Document Keywords	iii
Table of Contents	iv
List of Figures.....	iv
List of Tables.....	v
References and Applicable Documents	v
List of Acronyms and Abbreviations.....	v
List of Project Partner Acronyms.....	viii
Executive Summary	1
1 Introduction	2
2 Collaboration in the Security Domain.....	3
2.1 Collaboration with e-Infrastructures and NRENS	3
2.2 Incident Handling.....	4
2.3 Federated Authentication and Authorisation	4
3 Collaboration with EUDAT in the Data Domain	5
3.1 Data service interoperability guiding principles.....	5
3.2 EUDAT data services	5
3.3 Data pilots identification.....	7
3.3.1 PRACE calls for project access.....	7
3.3.2 DECI calls	8
3.3.3 Alternative sources	8
3.4 Identified Initial Pilots	9
3.4.1 Data drop generic pilot	9
3.4.2 Data publication for UK Turbulence community	9
3.4.3 Technical pre-requisites for pilots.....	9
3.5 Proposed Planning for Pilots	10
3.6 Strengthening Collaboration for Sustainability.....	10
3.6.1 Governance and contractual aspects.....	10
3.6.2 User training	11
3.6.3 User support.....	11
3.6.4 Synchronise services evolution.....	11
3.6.5 Scientific workflows.....	11
4 Other Potential Collaboration.....	12
4.1 Collaboration with CoEs.....	12
4.2 Collaboration with XSEDE and RIKEN	13

List of Figures

Figure 1: Overview of EUDAT services.....	6
---	---

List of Tables

Table 1 List of Centres of Excellence	12
---	----

References and Applicable Documents

- [1] <http://www.prace-project.eu>
- [2] WISE workshop: <https://www.terena.org/activities/ism/wise-ws/>
- [3] WISE workshop agenda <https://www.terena.org/activities/ism/wise-ws/agenda.html>
- [4] Trusted Introducer <https://www.trusted-introducer.org/>
- [5] SCI document: http://pos.sissa.it/archive/conferences/179/011/ISGC%202013_011.pdf
- [6] EUGridPMA <https://www.eugridpma.org/>
- [7] IGTF <http://www.gridpma.org/>
- [8] eduGAIN <http://services.geant.net/edugain/Pages/Home.aspx>
- [9] AARC <https://aarc-project.eu/>
- [10] Moonshot project <https://www.jisc.ac.uk/rd/projects/moonshot>
- [11] bwIDM <https://www.bwidm.de/>
- [12] EUDAT 2020 project - Deliverable 2.2 Data Management Plan (not published yet)
- [13] DMP Online <http://www.dcc.ac.uk/dmponline>
- [14] EUDAT project www.eudat.eu
- [15] EUDAT service descriptions www.eudat.eu/services
- [16] Memorandum of Understanding (MoU) between the PRACE-3IP and EUDAT projects for the usage of the PRACE dedicated network
- [17] PRACE 3IP deliverable : [D6.1.3 Final Operations Report including XSEDE Collaborative Services](#)
- [18] Centres of Excellence <http://ec.europa.eu/programmes/horizon2020/en/news/eight-new-centres-excellence-computing-applications>
- [19] <http://www.prace-ri.eu/exdci-coe-rome>

List of Acronyms and Abbreviations

AAI	Authentication and Authorisation Infrastructure
aisbl	Association Internationale Sans But Lucratif (legal form of the PRACE-RI)
API	Application Programming Interface
B2xxx	EUDAT service names start with B2, they are presented in the document
bwIDM	Föderiertes Identitätsmanagement der baden-württembergischen Hochschulen
CA	Certificate Authority
CoE	Center of Excellence
CPU	Central Processing Unit
CSIRT	Computer Security Incident Response Team
DoA	Description of Action (formerly known as DoW)
DMP	Data Management Plan

EC	European Commission
eduGAIN	Interconnects identity federations around the world, a service developed within the GÉANT Project
EGI	European Grid Initiative
ESFRI	European Strategy Forum on Research Infrastructures
EUDAT	EUropean DATa project
EUGridPMA	EUropean Policy Management Authority for Grid Authentication in e-Science
EXDCI	European eXtreme Data and Computing Initiative
FET	Future & Emerging Technologies
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
HBP	Human Brain Project
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
IdP	Identity Provider
IGTF	Interoperable Global Trust Federation
iRODS	integrated Rule-Oriented Data-management System
ISMS	Information Security Management System
IT	Information Technology
ITIL	Information Technology Infrastructure Library
KB	Kilo (= $2^{10} \sim 10^3$) Bytes (= 8 bits), also KByte
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
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
NREN	National Research and Education Network
PA	Preparatory Access (to PRACE resources)
PID	Persistent IDentifier
PMA	Policy Management Authorities

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.
RI	Research Infrastructure
RIKEN	Rikagaku Kenkyusho (Institute of Physical and Chemical Research, Japan)
SCI	Security for Collaborating Infrastructures
SP	Service Provider
TB	Technical Board (group of Work Package leaders)
TB	Tera (= 240 ~ 1012) 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.
TFlop/s	Tera (= 1012) 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.
WISE	WISE Information Security for collaborating E-infrastructures
WLCG	Worldwide Large Hadron Collider Computing Grid
XSEDE	Extreme Science and Engineering Discovery Environment project

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.
GENCI	Grand Equipement National de Calcul Intensif, France
GRNET	Greek Research and Technology Network, Greece
HBP	Human Brain Project
ICM	Warsaw University, Poland (3 rd party to PNSC)
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)
IUCC	INTER UNIVERSITY COMPUTATION CENTRE, Israel
JKU	Institut fuer Graphische und Parallele Datenverarbeitung der Johannes Kepler Universitaet Linz, Austria
JUELICH	Forschungszentrum Juelich GmbH, Germany
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
NIIF	National Information Infrastructure Development Institute, Hungary
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
UC-LCA	Faculdade Ciencias e Tecnologia da Universidade de Coimbra, 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)
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)
VSU-TUO	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA, Czech Republic
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. The collaboration has been extended now to GÉANT and the NREN communities. As a result, a first common workshop was organised in October 2015 under the name of WISE (WISE Information Security for collaborating E-infrastructures). This workshop was attended by representatives from EGI, EUDAT, GÉANT, NRENS, PRACE and also included US participation. As a result, several follow up activities have been agreed on. A Steering Committee will coordinate the follow-up activities of the new collaboration, among which is the setup of some working groups for specific topics. PRACE actively participated in the Programme Committee of the workshop and will continue to participate also in the Steering Committee and working groups.

PRACE has a long history of collaboration with EGI and EUDAT for security incident handling and this collaboration will continue and may be extended through the WISE initiative.

For authentication and authorisation services it can be of interest for PRACE 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. This task follows from activities in this area, especially those of the AARC project [9] which recently started. 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 sooner as more services will rely on the same credentials. Feedback has been provided on the requirements PRACE has for the acceptance of authentication services provided by external IdPs.

Regarding data services, the proposal is to collaborate with EUDAT. The requirements gathering and future pilot identification is done essentially through a call for data pilots included in PRACE and DECI calls. The planning work we have undertaken so far with resources from both projects has primarily focused on defining the multiple aspects of the services and the way to ensure we include them all in the integration design. We have already ongoing activities with two initial pilots and expect further ones to come out of DECI Call 13.

Links with CoEs and with the XSEDE project have been initiated and we expect specific activities to start early 2016.

As a result of the pilots we expect to provide PRACE management with some lessons learned and recommendations to build a sustainable collaboration and by this way fulfil user needs.

1 Introduction

One crucial direction given for the H2020 program 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 GEANT, or even from other continents (XSEDE for example).

In this document, we aim to describe ongoing collaboration activities and present planned work for the forthcoming months. A complete plan cannot be provided in this deliverable as some of the potential PRACE partners have just been started (CoEs for example), so we cannot understand their requirements yet. The goal is to implement the collaboration as much as possible on the basis of user driven pilots.

In Section 2 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 new AARC project.

Section 3 focuses on a vertical collaboration, the interoperability with EUDAT data services. We begin with what must be defined and setup for a successful integration and later identify which EUDAT services are already available. On this basis, we present the way we will identify relevant pilots and the plans we have for them. We then present some domains we need to tackle before we can progress with the integration.

We conclude with Section 4 which introduces some other organisations we potentially could collaborate with.

2 Collaboration in the Security Domain

Collaboration in the security domain with other large e-infrastructures is important because the users and the communities which use PRACE resources may also have access to resources of other e-infrastructures. Because of this, security incidents can easily spread from one infrastructure to another. In a similar manner, incidents at a user's organisation can also spread to infrastructures like PRACE. Because of this, collaboration with the security activities of the NRENs, to which these user organisations are in general connected, can help in mitigating security threats and containing security incidents.

These collaborations are also very useful for sharing best practices, e.g. on policies, procedures and risk assessments. In previous years the larger e-infrastructures have already collaborated under the Security for Collaborating Infrastructures (SCI) initiative, which among others resulted in a document which on a high level describes a recommended set of security policies and procedures [5]. The sharing of information on the status of recommended policies and procedures can help in the set-up of interoperable services among e-infrastructures by providing a trust basis for the collaboration. Furthermore the document can be used to define a minimum required set of security policies and procedures for partners of the PRACE infrastructure.

In the following sections the collaboration activities since the beginning of PRACE-4IP are described.

2.1 Collaboration with e-Infrastructures and NRENs

In the spring of 2015 discussions started for the organisation of a next security event among the larger e-infrastructures EGI, EUDAT, PRACE and WLCG. Later this was expanded to include GÉANT, the NREN community, and some US representatives. It was decided to organise a common workshop with e-infrastructure and NREN representatives. A Program Committee (PC) of around ten persons was formed, including two representatives from PRACE. It was also decided to organise a meeting under a new name, WISE (Wise Information Security for collaborating E-infrastructures). The first meeting was held between October 20-22 in Barcelona, hosted by the PRACE partner BSC, and sponsored by EGI, EUDAT, GÉANT, HBP and PRACE [2]. Topics on the agenda included an introduction on the security organisation of the different e-infrastructures, how to set up Information Security Management Systems (ISMS), experiences with Risk Assessments and how to share threat information in a secure way. Most of the presentations are publicly available on the agenda page [3]. Twelve participants, of the in total 49, were from PRACE partners. There were also several US participants, among others from XSEDE.

As this was the first meeting with representatives of both the e-infrastructure and NREN communities a special objective of the meeting was to discuss how a future collaboration can be organised and which topics should get attention. It was decided to create working groups for investigating the collaboration in different areas, such as training, risk assessments and sharing threat information. The WISE PC will be responsible to set up these working groups.

For training it was agreed at the meeting to develop a security awareness training module for users of the e-infrastructures. This will be based on already existing material. Within PRACE this material can be used in the trainings given to the users by the PATCs, but also by partners at other events with users. Awareness training of users is important as many security incidents are caused by careless behaviour of users.

It is the intention to organise a WISE event at least once a year.

2.2 Incident Handling

PRACE has already been collaborating for some years with the EGI CSIRT team on sharing information on security incidents and specific vulnerability risks for our respective infrastructures. For this purpose PRACE security members are subscribed to the EGI CSIRT list and EGI members are subscribed to the PRACE security information list with EUDAT also represented on both lists. This sharing of information has proven to be very valuable.

The PRACE Security Forum will investigate if the PRACE CSIRT team should register as a CSIRT team at the TERENA Trusted Introducer service [4]. Through this registration the PRACE CSIRT can more easily exchange information with other CSIRT teams. A first step will be to get listed as a CSIRT team.

2.3 Federated Authentication and Authorisation

Some PRACE services rely for authentication and authorisation on X.509 certificates issued by Certificate Authorities (CAs) which are members of the EUGridPMA [6] which is the organisation for the accreditation of CAs for use in distributed e-infrastructures.

PRACE is a member of EUGridPMA as a Relying Party. This involvement is important for monitoring the accreditation process of new CAs and the auditing of already accredited CAs. It is also used to provide feedback on the policies and procedures of the CAs as defined by the EUGridPMA and two other Policy Management Authorities (PMAs), one for the American continent and one for the Asia-Pacific region. Together, these three PMAs govern the trust base globally through the IGTF [7].

By using the IGTF CAs PRACE relies on identity information from external organisations, the CAs. This is an example of a set-up where Service Providers (SPs) make use of external Identity Providers (IdPs). For PRACE it can be interesting to use existing IdP federations for authentication and authorisation services, also referred to as Authentication and Authorisation Infrastructures (AAIs). Today there are many such national AAI federations specific for educational organisations and which are globally federated through eduGAIN [8]. Relying on these federations means that a user can use the same credentials to access a PRACE service as those used to access services from the user's home organisation. Additionally, the PRACE SP does not have to check some personal information on which access to resources is based and does not have to manage specific authentication credentials for the service. It also facilitates collaboration between SP and IdP in the investigation of security incidents.

There are several initiatives to implement facilities through which users can access services using federated IdPs. Web based services have more mature solutions at the moment. For command line services – such as Moonshot [10] and bwIDM [11] –, implementations are in general still in an early phase. In addition to the technological challenges to enable such a workflow there is also the need to build trust among the IdP federations and the SPs for allowing such scenarios.

In May 2015 the EC funded AARC project [9] started. AARC is a two year project which has the goal to develop and pilot an integrated cross-discipline authentication and authorisation framework, built on existing AAIs and on production federated e-infrastructures. To achieve these goals the project will set up (among others) pilot environments based on proposed architectures. The development of procedures and policies to enable the acceptance of such facilities by both IdPs and SPs is also an important activity. The results of this project may help PRACE in setting up access to services using existing AAIs. Two PRACE partners are involved in the AARC project, which may facilitate an early adoption of the project results. PRACE also provided feedback on an AARC survey on the policy requirements for the

adoption an AAI by PRACE. The answers provided have been distributed and discussed within the PRACE security Forum.

3 Collaboration with EUDAT in the Data Domain

3.1 Data service interoperability guiding principles

Our objective, defined by the DoA, is based on the following:

- Outsource data management functions for the benefit of the users,
- Provide an integrated vision and offer seamless access.

To fulfil this, it is crucial to define what a data service is and what services interoperability means.

In the IT world, a commonly agreed definition of a service is the ITIL one: “*A means of delivering value to customers by facilitating outcomes customers want to achieve without the ownership of specific costs and risks*”.

From this definition, we can extract some key directions. Firstly, we only need to understand the interface exposed by the service in order to use it and should not worry about its internals. Secondly, even though technology is important in the IT environment, we must also consider the other aspects of the interface: users (who, which features, etc) and value (cost model, contracts, etc).

In the PRACE-4IP project, we investigate the possibilities that PRACE outsources services enabling data to be shared or preserved for example. To succeed in this, static (format, volume, ownership, etc) and dynamic (data flow, sharing, evolution over time, etc) dimensions of data must be specified. To facilitate this, data management plan templates have been developed over the previous years and it is now required to fill one by the EC for the H2020 projects. One of the EUDAT 2020 project deliverables [12] is a customised template for the DMP Online [13] produced by the Digital Curation Centre.

To implement interoperability with data services, we need to deal with the multiple layers which define an interface (none of them can be left aside). The following is a list of the layers we will define for each service and use case:

- User identification and roles
- Governance and finance
- Service level agreement
- Use case workflows / requirements
- Data Management Plan
- Service API
- Technical components for the interface
- Infrastructure

3.2 EUDAT data services

EUDAT, a pan-European Data Infrastructure is providing data services and storage capability to researchers and communities for the benefit of scientific research in Europe [14].

As presented in Figure 1 the current data services catalogue aims to cover the research data life cycle. It has been defined and set up using an iterative approach involving both scientific communities and data centres.

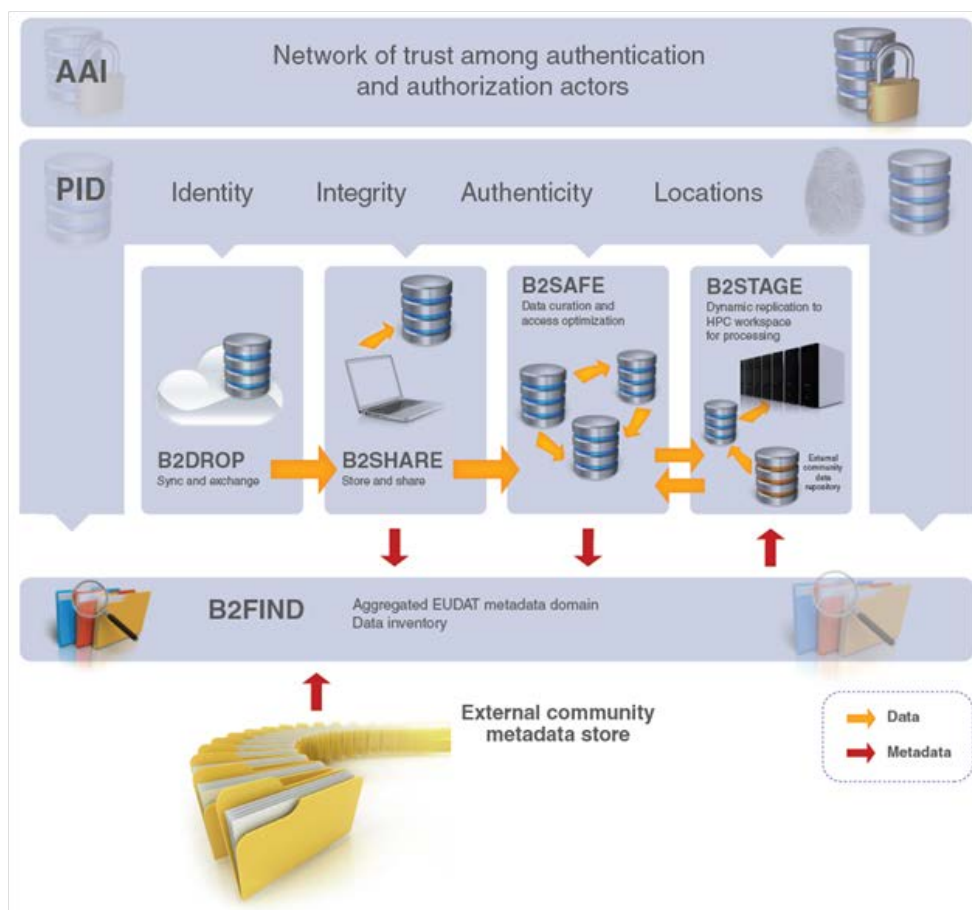


Figure 1: Overview of EUDAT services

Below is a short description of the services, with more details available at [15]:

- B2DROP is a secure and trusted data exchange service for researchers and scientists to keep their research data synchronized and up-to-date which also allows for exchange with other researchers.
- B2SHARE is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store and share small-scale research data from diverse contexts.
- B2SAFE is a robust, safe and highly available service which allows community and departmental repositories to implement data management policies on their research data across multiple administrative domains in a trustworthy manner.
- B2STAGE is a reliable, efficient, light-weight and easy-to-use service to transfer research data sets between EUDAT storage resources and high-performance computing (HPC) workspaces.
- B2FIND is a simple, user-friendly metadata catalogue of research data collections stored in EUDAT data centres and other repositories.

This service catalogue is subject to modifications following the Service Management System processes used in the EUDAT 2020 project. As part of our interactions we will need to react to services evolution, but we will also be able to influence it by providing requirements from the PRACE projects.

It is quite straightforward to work with EUDAT 2020 project members as they have a task which is part of the WP7 'Cross Infrastructure Service' named 'T7.1: Joint Access to Data and HPC services'.

We had a joint PRACE 4IP (WP6 – Task 6.3) and EUDAT 2020 (WP7 – Task 7.1) session on 3 Sep 2015 at the CINECA office (Bologna). We have been able to share knowledge about our infrastructures and agree on an action plan, presented in the following sections.

3.3 Data pilots identification

The PRACE-4IP DoA states that the way to experiment interoperability with other e-infrastructure should be done through joint pilots. This means that we will manage some real life use cases and will be able to implement iteratively technical solutions matching user requirements.

In the following subsections we describe the actions with the goal of identifying and selecting the more relevant pilots.

3.3.1 PRACE calls for project access

11th call:

The 11th Call for Proposals for Project Access opened on 9th February 2015. This call included a pilot call for joint data storage and services with some providers, one of them being EUDAT.

For this first joint call, in order to enable PRACE projects to benefit from EUDAT data services, around one PB storage capacity is available, thus close to 150TB per awarded project. EUDAT guarantees data storage at one or several of the 12 EUDAT sites for at least 24 months after the end of the PRACE grant, free of charge. Some computing resources for analysis purposes may be available.

The technical evaluation criteria are

- Criterion 1: Soundness of the pilot, foreseen impact and adherence to the detailed technical specifications provided (30%);
- Criterion 2: Technical feasibility of the pilot implementation (availability of storage capacity at selected sites, frequency of data access and transmission, nature of data) (30%).
- Criterion 3: Technical excellence and provision of added value to the current infrastructure (e.g. other implementations of existing services, implementation in other scientific disciplines and domains, etc.) (15%);
- Criterion 4: Contribution to Open access (15%);
- Criterion 5: Contribution to expanding EUDAT into new or different research domains and scientific communities (10%)

Unfortunately none of the 4 proposals linked with an EUDAT data pilot has been ranked high enough to be awarded by PRACE. The call was very competitive due to the very limited amount of available CPU hours and thus only strong proposals with high computational requests were successful.

However, while reviewing the proposals, we have been able to gather requirements and identify potential use cases. We can summarise them by the following list:

- Ability to transfer important volume data sets from and to the HPC facility;
- Ability to drop files in a storage space accessible by a workgroup;
- Ability to perform post processing on the data storage environment;
- Publication of some data on a public (or access controlled) repository;
- Data preservation for more than 5 years

12th call:

The 12th Call for Proposals for Project Access opened on 30 September 2015. This call does not explicitly offer access to EUDAT services and storage.

However, the plan is to contact the Principal Investigators of awarded proposals and offer them the equivalent of what was offered by EUDAT in the PRACE 11th call.

Regarding timelines, the allocation decision will be communicated early March 2016. This is the date after which we may be able to identify some pilots.

Future calls

We need to review with PRACE and EUDAT management whether or not we will offer the ability to run data service pilots as part of the future calls. This will depend on multiple factors such as:

- Call timelines vs PRACE 4IP and EUDAT H2020 timelines;
- The available data services and the feedback from previous pilots.

3.3.2 *DECI calls*

The DECI-13 call within the Optional Programme was launched on 14 August 2015 and closed on 21 September.

Access to the successful applications will be awarded for a period of 12 months, beginning 18 January 2016. Resources are available on the following Tier-1 architectures: Cray XC30/Cray XC40, IBM Blue Gene/Q, Intel clusters (various processor and memory configurations) and hybrid systems (clusters with GPGPU accelerators or Xeon Phi Co-processors) made available from Cyprus, Czech Republic, Finland, Greece, Hungary, Ireland, Italy, Norway, Poland, Serbia, Spain, Sweden, the Netherlands, and UK. Applicants need not specify a particular machine or architecture, but if they do, these preferences are taken into account.

The DECI-13 call is issued as a pilot call jointly with EUDAT. Together, they offer the opportunity to make awards of 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.

In total about 115 applications were submitted. Ten out of these requested for access to EUDAT resources.

Currently all the applications are in the process of technical and scientific review. After the review the applications will be ranked according to the feedback received from the scientific review and the resources will be distributed only to those applications which passed their scientific review.

The 10 projects that have asked for EUDAT services are simultaneously being reviewed by the EUDAT panel. However, of these 10 only the projects that will get accepted by both the DECI evaluation panel and the EUDAT evaluation panel will be granted EUDAT services. The access to the EUDAT services will be given either during or after the DECI-13 project period completion in accordance to the request of the applicant.

The plan is to have the complete list of accepted applications within December 2015.

3.3.3 *Alternative sources*

Due to the fact that none of the 11th PRACE call proposals requiring EUDAT services were awarded, we are exploring some alternative sources to identify early pilots.

To start with, all the partners involved in PRACE 4IP Task 6.3 checked if some of their HPC users have requested some data related services. This can be even easier if we can leverage the fact that some of the PRACE partners are also EUDAT partners.

The other potential sources which are being explored are the PRACE preparatory access calls. This could be used to carry on some integration testing on the production environment even if we cannot use full datasets.

3.4 Identified Initial Pilots

3.4.1 *Data drop generic pilot*

This pilot is not from a specific user, but based on requirements gathered from the PRACE 11th call proposals, hence we will call it a 'generic pilot'. Its key objective is to implement a simple use case regarding its features and data. But by tackling all the interoperability layers, we would like to identify as many solutions and roadblocks as possible.

The high level workflow for this pilot is the following:

- PRACE user identifies files he likes to share within a workgroup;
- PRACE user launches an interface;
- The interface uses PRACE user credentials;
- Datasets are transferred and ingested into EUDAT B2DROP;
- PRACE user defines sharing policy using B2DROP interface;
- Authorised persons can search and access the datasets from B2DROP

This fits with the early stages of the data life cycle, usually before a paper has been published by the research team.

3.4.2 *Data publication for UK Turbulence community*

EPCC has received a request from the UK turbulence research community to make datasets generated on their ARCHER system publicly available via a web interface.

The requirements from the research community are quite minimal:

- Ability to upload datasets (and possible metadata) to EPCC data storage area (be it on standard file system or some policy-based service such as iRODS);
- Make uploaded datasets available publicly via web interface;
- Possible additional requirement: search available datasets by metadata

It was decided to use this case as the basis of a pilot as it comes from users (it is a real life use case) and it matches some of the requirements identified in the PRACE 11th call for proposals.

This usually complies with the second stage of the research data life cycle, starting just after papers have been published. The high level solution would be to use the EUDAT B2SHARE service for this purpose.

3.4.3 *Technical pre-requisites for pilots*

We performed a high level design for those pilots and identified that we need to provide a solution for two technical pre-requisites. This will be useful to use any of the EUDAT services from PRACE.

The first one is the user identification. 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 identity federation. We will need to review how those two systems can share enough information to enable proper user access control. This obviously is tightly linked to the work performed for the collaboration around security. The technical solution will need to rely upon proper policies for example.

The other one is data transfer. With the increasing size of research data, efficient solutions for data transfer between the two infrastructures is required on both network and tools dimensions.

Based on the existing PRACE-EUDAT MoU [16], the PRACE network is available to support the data transfers. This means that we will have a state of the art solution to link to the EUDAT nodes hosted by a PRACE partner (for example JUELICH, EPCC, BSC, CINECA, CINES, SURFsara, etc). We will need to assess the available solutions to reach other EUDAT nodes.

Regarding data transfer tools, GridFTP is used by both infrastructures and would be the basis of the initial solution. The objective will be to wrap it into a packaged solution hiding the complexity to the end user.

3.5 Proposed Planning for Pilots

The current high level planning is based on the assumption that some pilots will come out of the 13th DECI call. The awarded projects will be notified early January 2016 for an allocation start on 18/01/2016, meaning that data pilots would kick off in the first quarter of 2016.

Consequently, we would like to progress with the two initial projects described in the previous paragraph from November 2015 up to February 2016. This would provide important information and ease the DECI call related pilots to take off.

3.6 Strengthening Collaboration for Sustainability

The first pilots are due to progress with some of the core domains of interoperability. They will prove, based on real life use cases that it is possible and beneficial for users to use EUDAT data services from PRACE.

However, we have already identified some additional topics we cannot ignore which will require to build a stronger collaboration. To go further than some pilots and offer a sustainable interoperability, we will need to plan or deliver some solution in those domains.

3.6.1 Governance and contractual aspects

This might be the most complex topic, but it is crucial to reach some agreements between PRACE and EUDAT if we want to offer researchers a sustainable solution in which they will have trust and confidence over years.

Each of the infrastructures is defining its own organisation and governance that will go further than the current projects (PRACE 4IP, EUDAT 2020). This will include a contractual and cost model to access the services. So how to ensure a user will see this as one integrated solution, with one point of contact and a unique service level agreement?

We will raise this to management attention and as we will setup some pilots, propose some solutions.

3.6.2 *User training*

It appears that there is a lack of understanding of some of the best practices about data management and data life cycle. This quite often results in data producers asking for storage space only, but not being able to plan further (metadata, licensing, data curation, data preservation etc).

We plan to work with EUDAT and PRACE training teams on the best way to offer some training modules around those areas, maybe including this in some existing training cycle.

3.6.3 *User support*

Both infrastructures have their own user support process and toolset. Should we want to offer a seamless integration, we will need to update the processes to ensure a user support request can be handled with a single point of contact.

3.6.4 *Synchronise services evolution*

PRACE and EUDAT services are still evolving on both functional and technical aspects. One of the challenges of interoperability will be to ensure that the deployment of one new service release doesn't disrupt the availability of the whole solution.

So we need to have a process in place to guarantee that in case any change impacts the service external interface that it is managed (plan, review design, test, deploy).

This will also include emergency changes, security related for example.

3.6.5 *Scientific workflows*

This topic has been identified in the PRACE-4IP DoA. With both PRACE and EUDAT willing to offer service to researchers, we should make sure they can be accessed from workflow engines.

To understand what this means, let's take the example of a quite simple and generic scientific workflow. The tasks would be:

- Get data from some instruments and sensors
- Pre process raw data to clean and normalize them
- Identify and store the input data
- Extract some input data and run a simulation
- Compare the results with previous simulations and validate them
- Annotate, archive and publish the results

We can see that some of those tasks could be performed using PRACE (run a simulation) or EUDAT (transfer data, archive data, publish data ...) services.

The scientist can launch all the computer based tasks manually but to repeatedly perform the experiment, he could automate this by using a workflow engine. In this case PRACE and EUDAT services should offer a machine independent and standardized interface to access them.

4 Other Potential Collaboration

The initial focus for our collaboration has been, as described in the previous chapters, the existing security groups and EUDAT for the data services. But we have planned over the coming months to liaise with some other organisations and identify how we could collaborate with them.

4.1 Collaboration with CoEs

Recently, eight new Centers of Excellence (CoEs) for computing applications have been selected for support by the EC Horizon 2020 program [18]. Table 1 lists the eight CoEs.

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	Center of Excellence for Global Systems Science

Table 1 List of Centres of Excellence

In the end of September 2015 the European eXtreme Data and Computing Initiative (EXDCI), a support action co-ordinated by PRACE and ETP4HPC, hosted a meeting in Rome for CoEs and FET projects related to HPC [19]. PRACE presented there the facilities available for the CoEs. Half a percent of PRACE resources is available for CoE applications and from T6.3, support is available for data management and other interoperable requirements the CoEs may have.

We are in contact also with other PRACE-4IP work packages about their collaborative activities with CoEs, which also may lead to activities for our task. WP4 is preparing a 'CoE-Training-Questionnaire' to gather training requirements and WP7 is writing a document "Applications Areas Supported by New Centres of Excellence" to identify which codes they

will have to focus on. WP6 also will contribute to the EXDCI activities for the support of the CoEs.

As the CoEs have just started their activities, no specific support questions have yet been received. In the coming months this task will contact the different CoEs if they need active support.

The workshop between PRACE, CoE and FET HPC projects that is being organized at the PRACE Days 16 will be the occasion to further integrate the cooperation.

4.2 Collaboration with XSEDE and RIKEN

During PRACE-3IP Project PRACE and XSEDE have collaborated as described in the deliverable 6.1.3. [17].

At Super Computing SC15 (15 to 20 November 2015) a meeting between PRACE and XSEDE was organised. As a result PRACE and XSEDE agreed to work on the preparation of joint allocations. This would include the definition of areas like scientific steering committee, peer review systems alignment, MoU and just return mechanism.

The plan is to establish a small team by end of this year with members from both XSEDE and PRACE to be able to work on the preparation starting next year.

In addition, PRACE, XSEDE and RIKEN would prepare a call for enhancing interoperability between the infrastructures. This would start with collecting communities' requirements. Then a working session is planned for next year either during PRACEdays16, XSEDE conference in Miami or a RIKEN event in autumn. The expected outcome is to launch a joint call comparable to the PRACE-XSEDE one.