



**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

D6.6

Final Report on Collaborations with e-Infrastructures and CoEs

Final

Version: 1.0
Author(s): Stéphane Coutin, CINES
Date: 20.04.2017

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: < D6.6 >	
	Deliverable Nature: Report	
	Dissemination Level: PU*	Contractual Date of Delivery: 30 / April / 2017
		Actual Date of Delivery: 30 / April / 2017
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: Final Report on Collaborations with e-Infrastructures and CoEs	
	ID: D6.6	
	Version: <1.0>	Status: <i>Final</i>
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2010	
	File(s): D6.6.docx	
Authorship	Written by:	Stéphane Coutin, CINES
	Contributors:	Giovanni Morelli, CINECA Ralph Niederberger, JUELICH Vincent Ribailier, IDRIS Hans Van Piggelen, SURFsara
	Reviewed by:	Manuel Fiolhais, UC-LCA Veronica Teodor, JUELICH
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.5	03/ April /2017	Draft	Initial
0.6	04/ April /2017	Draft	Adding missing links
0.7	18/ April /2017	Draft	Revision
1.0	20/ April /2017	Final version	Sent for approval to PMO

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, 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-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

© 2017 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	ii
Table of Contents	iii
List of Figures	iv
List of Tables.....	iv
References and Applicable Documents	iv
List of Acronyms and Abbreviations.....	v
List of Project Partner Acronyms.....	vii
Executive Summary	1
1 Introduction	2
2 Partners	3
2.1 Active collaborations	3
2.1.1 <i>EGI</i>	3
2.1.2 <i>EUDAT</i>	3
2.1.3 <i>GÉANT</i>	4
2.1.4 <i>Human Brain Project</i>	4
2.2 Early contacts and potential collaborations	4
2.2.1 <i>Centres of Excellence (CoEs)</i>	5
2.2.2 <i>XSEDE / RIST</i>	6
3 Collaboration on the security domain	6
3.1 Collaboration with e-Infrastructures and NRENs - WISE.....	6
3.2 Incident Handling – Trusted Introducer (TI) programme	7
3.3 Federated Authentication and Authorisation	7
4 Collaboration on the data management domain	8
4.1 Leveraging EUDAT services.....	8
4.2 Leveraging security collaboration.....	9
4.3 Pilots identification – DECI-13 and DECI-14 calls.....	9
4.3.1 <i>DECI-13</i>	9
4.3.2 <i>DECI-14</i>	10
4.4 MULTINANO data pilot	11
4.5 HybTurb3D data pilot.....	11
4.6 HiResClimate data pilot.....	11
4.7 AFiD data pilot	12
4.7.1 <i>Project and actors</i>	12
4.7.2 <i>Current status</i>	12
4.7.3 <i>Next steps</i>	13
4.8 CHARTERED data pilot	13
4.8.1 <i>Project and actors</i>	13
4.8.2 <i>Current status</i>	13
4.8.3 <i>Next steps</i>	15
4.9 Collaboration with EUDAT on the training domain	15
5 Annex.....	16

5.1	AFiD data pilot – data management plan.....	16
5.2	CHARTERED data pilot – data management plan	18

List of Figures

Figure 1:	Data Flow diagram for AFiD data pilot.....	12
Figure 2:	Data Flow diagram for CHARTERED data pilot	14

List of Tables

Table 1	List of Centres of Excellence	5
Table 2:	DECI-13 call data pilots.....	10

References and Applicable Documents

- [1] EGI <https://www.egi.eu/>
- [2] EUDAT project www.eudat.eu
- [3] GEANT Project <http://www.geant.org/>
- [4] Human Brain Project <https://www.humanbrainproject.eu/>
- [5] Centres of Excellence <http://ec.europa.eu/programmes/horizon2020/en/news/eight-new-centres-excellence-computing-applications>
- [6] XSEDE <https://www.xsede.org/>
- [7] RIST <http://www.tokyo.rist.or.jp/rist/en/>
- [8] WISE <https://wise-community.org/>
- [9] eduGAIN <http://services.geant.net/edugain/Pages/Home.aspx>
- [10] EUDAT service descriptions www.eudat.eu/services
- [11] Data Management Plan <https://www.eudat.eu/d22-data-management-plan>
- [12] PRACE-4IP D6.5: http://www.prace-ri.eu/IMG/pdf/D6.5_4ip.pdf
- [13] PRACE-3IP D6.1.3: http://www.prace-ri.eu/IMG/pdf/D6.1.3_3ip.pdf
- [14] AARC Blueprint architecture <https://aarc-project.eu/aarc-draft-blueprint-architecture-available-for-comments/>
- [15] EUGRIDPMA 39th meeting <https://www.eugridpma.org/meetings/2017-01/>
- [16] SNCTFI <https://wiki.geant.org/display/AARC/Snctfi>
- [17] WISE SCIV2 workgroup <https://wiki.geant.org/display/WISE/SCIV2-WG>
- [18] Scalable Negotiator for a Community Trust Framework in Federated Infrastructures, <https://wiki.geant.org/display/AARC/Snctfi>
- [19] HybTurb3D data pilot published files : <https://b2share.eudat.eu/records/a58135af9c9d429f92c15ce88bdfdd55>
- [20] EUDAT data management training : <https://www.eudat.eu/training/research-data-management>

List of Acronyms and Abbreviations

AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research and Collaboration
aisbl	Association International 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
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
EUDAT	EUropean DATa project
EUGridPMA	European Policy Management Authority for Grid Authentication in e-Science
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
ISC	International Supercomputer Conference
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
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
RIST	Research Organization for Information Science and Technology (Japan)
SC	Super Computing
SCI	Security for Collaborating Infrastructures
SP	Service Provider
TB	Technical Board (group of Work Package leaders)
TB	Tera (= 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.
TERENA	Trans-European Research and Education Networking Association
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
WISE	WISE Information Security for collaborating E-infrastructures
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	Cyprus Research and Educational Foundation, 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
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)
IT4Innovations	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA, IT4Innovations National Supercomputing Center, Czech Republic
IUCC	INTER UNIVERSITY COMPUTATION CENTRE, Israel
JKU	Institut fuer Graphische und Parallele Datenverarbeitung der Johannes Kepler Universitaet Linz, Austria
JUELICH	Forschungszentrum Juelich GmbH, Germany
KIFÜ	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	Stichting Academisch Rekencentrum Amsterdam as part of the SURF Foundation
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)
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 some ongoing collaboration with the major European e-Infrastructure (EGI [1], EUDAT [2], GÉANT [3], Human Brain Project [4]). There are also some ongoing preliminary discussions with the Centres of Excellence (CoE) [5] and with worldwide organisations (XSEDE [6] and RIST [7]).

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) [17]. The 4th workshop, attended by representatives from EGI, EUDAT, GÉANT, NRENs, PRACE and also included US participation, was held from 27 to 29 March 2017 in Amsterdam. A Steering Committee coordinates the follow-up activities of this active collaboration, organised in working groups for specific topics. PRACE actively participates 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 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 [9]. This task follows from activities in this area, especially those of the AARC project. 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, we collaborate with EUDAT, which offers a service catalogue [10] complementing PRACE in order to fulfil user needs during the full research data life cycle. The requirements gathering and future pilot identification is done essentially through a call for data pilots included in PRACE and DECI calls. We launched four data pilots as a result of DECI-13 call and might expect two more from DECI-14 call.

As the first data pilots have been running for nearly a year, we have implemented some integration in the areas of data transfer, data storage and data sharing. We have used the Data Management Plan template [11] to ease requirements gathering and will recommend using it systematically. We have also identified the need for cross PRACE and EUDAT trainings, some have been delivered to support data pilots. This collaboration increases now with EUDAT delivering some sessions in PRACE training events and vice versa.

1 Introduction

One crucial direction given for the H2020 programme 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).

In this document, we aim to provide a status about ongoing collaboration activities at the end of PRACE-4IP project. The goal is to implement the collaboration as much as possible on the basis of user driven pilots.

In Section 2, we provide an overview of the PRACE collaboration on the operational domain. This is presented on a per partner basis, gathering on one hand those who we are actively collaborating with and on the other hand those we expect to work with in the near future. The following two sections will give a deeper dive into the key collaboration domains.

In Section 3 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.

Section 4 focuses on a vertical collaboration, the interoperability with EUDAT data services. We begin with a presentation of the added value 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.

2 Partners

In this chapter we identify the organisations PRACE is having or is planning to have collaboration for operations with. For each of them, we identify formal agreements and we provide a list of the collaboration areas. Further details are provided unless the area is more precisely described in the following sections.

2.1 Active collaborations

The partners appearing in this paragraph are those who are actively collaborating with PRACE and/or have signed a Memorandum of Understanding (MoU) with PRACE.

2.1.1 *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.

Today, security related activities are active (see section 3), the others have not taken off yet.

2.1.2 *EUDAT*

The first official agreement between PRACE and EUDAT is the MoU that was signed off in March 2014. As a lot of data centres are involved in both EUDAT and PRACE, and assuming some data used by or produced by simulations ran on PRACE HPC would benefit from EUDAT data services, the MoU allowed the usage of PRACE network to transfer data.

With the H2020 projects (EUDAT 2020 and PRACE-4IP) coming, the objective has been revised to offer seamless data service integration between the two infrastructures. To support this, a new MoU has been signed off in June 2016. It covers, on the top of 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. While running the data pilots, the need for training was clearly identified and a fruitful collaboration started (see subsection 4.9).

EUDAT is also involved in the cross organisation security workgroups (see section 3).

The last identified collaboration area came out of a requirement from PRACE-4IP WP3. There is a need for a PRACE service explicitly requested by the European Commission 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). To create such a data or document repository, a possibility is to create a brand new PRACE service. However, we identified that the EUDAT B2SHARE service was a potential off the shelf solution. After checking that the B2SHARE features meet our requirements, it was agreed to use it. This will be implemented as part of PRACE-5IP project.

2.1.3 GÉANT

The collaboration with GÉANT has a long history since the PRACE dedicated network starting in the DEISA project already in 2006 has been fully based on the GÉANT footprint. This network has been overtaken by PRACE after the end of DEISA and is operational now for more than 10 years.

In 2013 GÉANT has set up the International User Advisory Committee (IUAC) since their services must be adjusted to user's needs continuously. Several European research projects appointed staff members to this committee. PRACE as one of those projects, being highly dependent on massive high speed communication exchanges between supercomputers and data storage providers, as well as end users, via national and European research networks and being explicitly interested in using secure communication paths and authentication & authorisation services on a European and global basis has been part of this committee from the beginning. Several meeting took place in the meantime. PRACE staff members also supported GÉANT in an EU consultation meeting with the EU in 2016.

Already by mid of 2014 further tight collaborations between GÉANT and PRACE have been set up to design the new PRACE network layout. Because of diverse reasons, the upgrade had to be delayed. The activities gathered their way again, after setting up the PRACE-GÉANT Memorandum of Understanding in March 2016.

After having discussed different technical alternatives, cost scenarios, management issues, and contractual dependencies, at the end of 2016 the migration phase from the old dedicated physical network infrastructure to the new PRACE-VPN environment could be started. Close collaborations between both projects have led to a fast set up of the new infrastructure. Currently 7 PRACE partners have migrated to the new infrastructure and due to the tight collaborations a smooth migration path could be defined by connecting both infrastructures (old physical and new virtual). It is anticipated that the full migration will be completed at end of April 2017.

In parallel in March 2017 a further PRACE-GÉANT collaboration meeting took taken place, where further collaboration opportunities have been discussed. Here collaborations on different layers, from management layer down to operations, IT security, and network have been anticipated. It was agreed to get together in further meetings to elaborate all the relevant details.

2.1.4 Human Brain Project

PRACE-HBP MoU was signed in November 2014. The main content agreement focuses on using the PRACE network.

Main HBP HPC infrastructure partners (BSC, CINECA, CSCS, and JUELICH) are also PRACE partners. Therefore usage of the PRACE network infrastructure seems obvious. Close collaboration took place when migrating the partners to the new PRACE-VPN.

2.2 Early contacts and potential collaborations

This second list includes those partners we have been in contact with in order to identify potential collaboration areas. However, despite the fact we hope seeing them taking off soon, no activity has started yet.

2.2.1 Centres of Excellence (CoEs)

Eight new Centres of Excellence (CoEs) for computing applications have been selected for support by the EC Horizon 2020 programme. Since the beginning of PRACE-4IP WP6 team had several meetings with them (as reported in PRACE-4IP D6.5 [12]) in order to understand their requirements and to give them access to the relevant PRACE services. You will find the list of CoEs in the following table.

Coordinator		Acronym	Title
Organisation	Country		
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 KLIMARECHENZENTR UM GMBH	DE	ESiWAC E	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

We have been regularly checking with other PRACE-4IP work packages about their collaborative activities with CoEs, which also may lead to activities for our task.

But so far, no requirement has been raised to us by any CoE for operational activities. Talking to some CoEs, it appears that they are focusing first in activities like code optimisation, and that the need to run those codes on an operational mode will appear in a second time.

This is why we keep committed to support the CoEs in the PRACE-5IP project.

2.2.2 XSEDE / RIST

Research teams become more and more globally distributed. To support their computational needs, PRACE, XSEDE and RIST, three of the major players in the worldwide HPC ecosystem, work to develop some collaboration. For example, during PRACE-3IP project PRACE and XSEDE have collaborated as described in the deliverable 6.1.3 [13].

At Supercomputing Conference SC16 a meeting between PRACE, RIST and XSEDE senior management was organised on 15 November 2016. During this, the three parties reviewed a draft MoU. Its sign-off is planned for mid-2017.

This will serve as a basis 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.

It is worth mentioning that we have already some ongoing security collaboration, XSEDE experts participating in WISE (see section 3).

3 Collaboration on the security domain

3.1 Collaboration with e-Infrastructures and NRENs - WISE

Beginning in 2013 staff members of PRACE joined a working group looking into security for collaborating infrastructures (SCI). This SCI working group, compound by members of the EU projects EGI, PRACE, EUDAT, HBP, and several American initiatives, met several times. Here a trust framework has been defined, which has been revised and complemented continuously over the years. In 2015 this collaboration led to a joined meeting with GÉANT and TERENA working groups in Barcelona, where the WISE community (WISE Information Security for collaborating E-Infrastructures) was founded. PRACE staff members have been participating actively in this community, temporarily providing two participants of PRACE as work group leaders. Currently the WISE community has set up four working groups “Updating the SCI framework”, “Security Training and Awareness”, “Risk assessment”, and “Security in Big and Open Data”. All this workgroups are highly relevant to PRACE, so that this collaboration will help defining and upgrading the PRACE security policy, daily secure operations, and risk assessment.

The community is further on very active having monthly steering committee meetings, presenting at TNC2016 and in the Di4R WS in Krakow.

It just had its 4th work group meeting in Amsterdam at end of March 2017. During 2 days three out of the four working groups progressed with some deliverables that will be directly usable by PRACE, either for internal activities or for collaboration.

The “Risk assessment” has produced a spreadsheet based on best practices presented by the participants. A first version will be finalised over the coming months with some guidelines, keeping in mind the objective to be generic and simple. Members of the workgroup agreed to use it for some of the ongoing collaboration, and to use feedback as an input to improvement.

The “Security Training and Awareness” workgroup has listed the type of trainees based on their skills and roles. For each population, the main needs and the existing trainings are identified.

The “Updating the SCI framework” workgroup has finalised a version 2 of the document. This framework, alongside with a self-assessment spreadsheet, will be presented to NRENS during the next GÉANT conference (end of May 2017).

3.2 Incident Handling – Trusted Introducer (TI) programme

PRACE has set up a computer security incident response teams (CSIRTs) to deal with network security incidents since several years. Obviously security handling cannot be done in a locked environment. The task of PRACE is to provide supercomputing resources throughout Europe and beyond, which leads to the fact, that interaction with the outside world is indispensable.

Building mutual trust between security teams is very important because incidents often originate from outside the network that is affected.

The GÉANT Trusted Introducer (TI) service forms the backbone of infrastructure services and acts as a clearinghouse for all computer security incident response teams (CSIRTs), building a ‘web of trust’ between CSIRTs by listing known teams, and accrediting and certifying teams according to their demonstrated and checked level of maturity.

CSIRTs accredited and/or certified by GÉANT are allowed to participate in closed meetings where sensitive and confidential information about incidents and threats is exchanged. Furthermore vital members-only services are provided enabling those teams to interact more efficiently and effectively.

Therefore in 2016 the security team of PRACE asked PRACE Management Board (MB) to approve the process of being accredited to the TI programme of GÉANT. In the meantime PRACE MB voted to start the accreditation process. Due to this vote, the PRACE security team is preparing the first steps for the accreditation process currently.

3.3 Federated Authentication and Authorisation

During the past year, PRACE has been following closely the different initiatives in the Federated and Authorisation field. It is essential for PRACE to understand the directions that are taken in the Identity federation world. Indeed, if the standards and technologies that are currently being designed are well-understood and anticipated by PRACE, it will offer to the project both a great opportunity to ease interoperability of the PRACE research infrastructure with other international projects and also an opportunity to increase significantly the overall security level offered to its users.

PRACE has in particular manifested a great interest in the work performed by the AARC project which has released in June 2016 the first draft of the “Blue print architecture”. This document [14] proposes a high-level AAI architecture suited for e-Infrastructures and built on top of the eduGAIN [9] robust and well-established identity federation.

PRACE had the opportunity to attend the 39th EUGridPMA meeting [15] that was held in Florence between 30 January 2017 and 1 February 2017. This meeting has been for PRACE a great opportunity to exchange ideas on the different technologies addressed in the “Blue print architecture” and to define how the current PRACE security model could evolve towards a more coupled integration with the Identity federation world. Among the different topics that

have been addressed during this meeting, the most relevant for PRACE are the bridge technologies like the new IGTF to eduGAIN bridge, the token translation services and the attribute enrichment tools.

PRACE is also involved in several security working groups. In the last months, PRACE has mainly contributed in the Scnfti working group [16] and in SCIV2 [17]. The Scnfti working group is elaborating a trust framework aimed at helping relying parties to coordinate a policy allowing them to represent a coherent view to the rest of the world. This framework is described in the Scalable Negotiator for a Community Trust Framework in Federated Infrastructures document [18]. PRACE has participated to the elaboration of this document by attending two Scnfti videoconferences in February and March 2017. SCIV2 is the follow-up of SCI (Security for Collaboration among Infrastructures) and defines best practices, trust and policy standards for collaboration with the aim of managing cross-infrastructure operational security risks. PRACE has attended the different videoconferences during May and July 2016 where the elaboration of the version 2 document has been started.

4 Collaboration on the data management domain

4.1 Leveraging EUDAT services

In addition to the internal PRACE services, we are developing collaboration with other e-Infrastructures to increase our service portfolio and offer PRACE users some further data management services. The initial collaboration on this topic is done with EUDAT, implementing pilots selected within the PRACE/DECI user projects.

Compared to PRACE, which is oriented to provide HPC resources, EUDAT provides a collection of services oriented to Data Management [10]. In this context, collaborations for projects can take advantage of the computing power offered by PRACE in the stage of production and of the long-term data preservation by EUDAT in post-production. In this collaboration, the most widely used EUDAT service was B2STAGE especially because it is easy to use and has reduced setup times. Even B2SHARE was used to explore the possibilities offered by this service for sharing set of experimental data and post production results.

The first step of those pilots is to dig into the user requirements to understand the nature of the data and the expected use cases. We use the Data Management Plan template to drive this and make sure that we will explore the full data life cycle. It is also important to offer research data management training as we noticed different maturity levels within the users.

To fulfil the requirements we have identified some existing EUDAT services:

- data transfer,
- data and metadata publication and search,
- data preservation

but also described new ones:

- workspace storage,
- 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 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.

So, we want to leverage the work done in the collaboration for the security activities ('WISE, AARC project, EuGridPMA ...).

The area we focused on 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 identity federation.

We implemented a change in the PRACE LDAP schema, adding a `deisaUserProfile` attribute. This attribute with value "EUDAT" is provided to users who indicated consent that their user 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.

4.3 Pilots identification – DECI-13 and DECI-14 calls

The DECI-13 and DECI-14 calls are 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.

4.3.1 DECI-13

The DECI-13 call within the Optional Programme was launched on 14 August 2015 and closed on 21 September 2015. Access to the successful applications were awarded for a period of 12 months, beginning 18 January 2016.

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

The 10 projects that have asked for EUDAT services have been reviewed by the EUDAT panel. As a result of DECI and EUDAT review processes, five DECI awarded projects have been accepted as data pilots. The table below summarises them:

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 access to the EUDAT services has been given during the DECI-13 project period completion and for an additional two years period.

Those data pilots will be described in the following chapters.

4.3.2 DECI-14

The DECI-14 call within the Optional Programme was launched on 14 October 2016 and closed on 21 November 2016. Access to the successful applications will be awarded for a period of 12 months, beginning 3 April 2016.

Three out of the applications requested for access to EUDAT resources. After the DECI selection process only two candidates remain. One of them is the continuation of data pilot that ran as part of DECI-13 call.

The assessment by the EUDAT panel is ongoing and decision will be provided by the end of April 2017.

4.4 MULTINANO data pilot

The MULTINANO project has been awarded as part of the DECI-13 call. This project is in the field of Engineering, focusing on “Multiscale simulations of nanoparticle suspensions”. The PI is Prof. Pietro ASINARI. The PRACE execution site is RZG (Germany) and the project started in April 2016.

The data requirements from EUDAT pilot is 30 TB during and after the PRACE project. As requested, the EUDAT site is CINECA (Italy).

The goal of the project was to create an archive for the data produced in molecular dynamics simulations for the study of materials. To allow the community to rapidly achieve the data transfer phase, the enabling team has been providing support to establish the transfer channel via GridFTP (a particular case of B2STAGE service) among RZG and CINECA endpoints. The project is in production and at present about 1TB of data has been transferred.

Data are stored on plain storage and the next step is to complete data transfer.

4.5 HybTurb3D data pilot

The HybTurb3D project has been awarded as part of the DECI-13 call. This project is in the field of Astronomy, focusing on “Hybrid 3D simulations of turbulence and kinetic instabilities at ion scales in the expanding solar wind”. The PRACE execution site is SURFsara and project started in April 2016.

The data requirements from EUDAT pilot is 140 TB during and after the PRACE project. As requested, the EUDAT site is CINECA (Italy).

The first requirement relates to output data storage. At the end of PRACE project, this community needs to store data from high resolution three dimensional simulations of solar wind turbulence. Data are in HDF5 format. As for MULTINANO project, with enabling team support, users were able to transfer their data via GridFTP.

The project is in production and at present about 1TB of data has been transferred.

In addition, there is a requirement to publish some of the data. For this reason the users have also used the B2SHARE service in order to explore the possibility of sharing the experimental data. In particular, they published a collection of files in HDF5 format [19].

4.6 HiResClimate data pilot

The HiResClimate project has been awarded as part of the DECI-13 call. This project is in the field of Earth Science, focusing on “High Resolution EC-Earth Simulations”. The PRACE execution site is KTH and project started in April 2016.

The data requirements from EUDAT pilot is 150 TB during and for 12 months after the PRACE project. As requested, the EUDAT site is EPCC (UK).

The expectation from the project team is to transfer the simulation outputs into the ESGF (Earth System Grid Foundation) node hosted in ICHEC. However, there is a temporary lack of available storage at this destination. Consequently, the request to EUDAT was to use storage as a buffer for about one year, data being removed after final transfer to ICHEC.

This data pilot has not progressed as the simulations have not been performed, due to code issues.

4.7 AFiD data pilot

4.7.1 Project and actors

The AFiD project has been awarded as part of the DECI-13 call. This project is in the field of Engineering, focusing on “Effect of rotation and surface roughness on heat transport in turbulent flow”. The PI is Richard STEVENS from University of Twente (NL). The PRACE execution site is EPCC and project started in April 2016.

The data requirements from EUDAT pilot is 11 TB during and after the PRACE project. As requested, the EUDAT site is SURFsara (NL).

4.7.2 Current status

For this data pilot, we have used the data management template to gather user requirements. It has been completed iteratively along some regular calls. The current version is available in annex 5.1, knowing that this is a living document that will certainly be completed in the future, for example to define long term preservation.

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

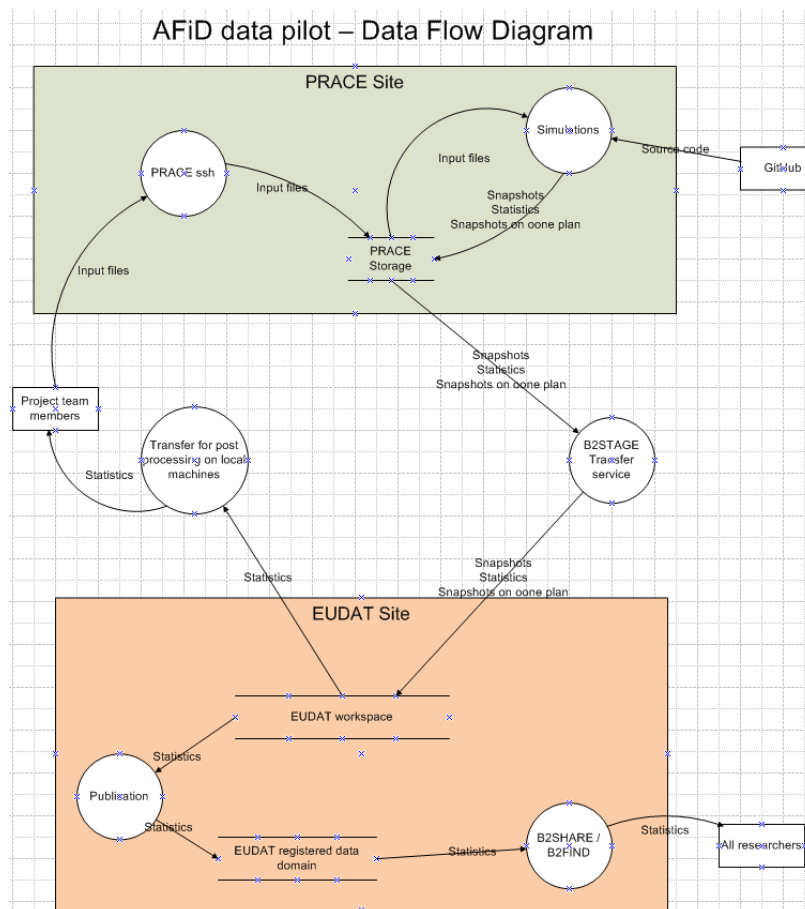


Figure 1: Data Flow diagram for AFiD data pilot

To start with we have 2 logins per person: the PRACE login and the SURFsara login. The 2 team members also have an EPCC account.

EPCC and SURFsara are connected to the PRACE network. GridFTP is installed on both sites. Globus Online can be used.

For data transfer we use GridFTP as it is the solution working behind EUDAT B2STAGE. For data transfer between EUDAT and local computers, users will use their SURFsara account to access the data on the EUDAT workspace. They can transfer data using scp or gridftp for example.

Currently all data is stored at EPCC and is being transferred to the SURFsara Data Archive. All data transfers will be completed in July 2017.

4.7.3 *Next steps*

In order to publish the long-tail data, first post-processing is done at the SURFsara compute infrastructure. The resulting ‘result/statistical’ data will be curated to allow sustainable preservation using B2SHARE. A specific metadata schema will be defined and metadata will be gathered using the produced data. Persistent identifiers will be attached.

Optionally the snapshot data sets can be preserved in the B2SAFE instance at SURFsara.

For all data holds that file format conversion is not required for long-term preservation and publication as the currently used HDF5 format is a suitable format for this purpose.

4.8 CHARTERED data pilot

4.8.1 *Project and actors*

The CHARTERED project has been awarded as part of the DECI-13 call. This project is in the field of Materials Science, focusing on “Charge transfer dynamics by time dependent density functional theory”. The PI is Prof. Biplab SANYAL from Uppsala University (Sweden). The PRACE execution site is IT4Innovations (Czech Republic) and project started in April 2016.

The data requirements from EUDAT pilot is 20 TB during and after the PRACE project. As requested, the EUDAT site is PDC/KTH (Sweden).

4.8.2 *Current status*

For this data pilot, we have used the data management template to gather user requirements. It has been completed iteratively along some regular calls. The current version is available in annex 5.2, knowing that this is a living document. For the CHARTERED data pilot we will need to complete the data sharing paragraph.

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

CHARTERED data pilot – Data Flow Diagram

20/6/16

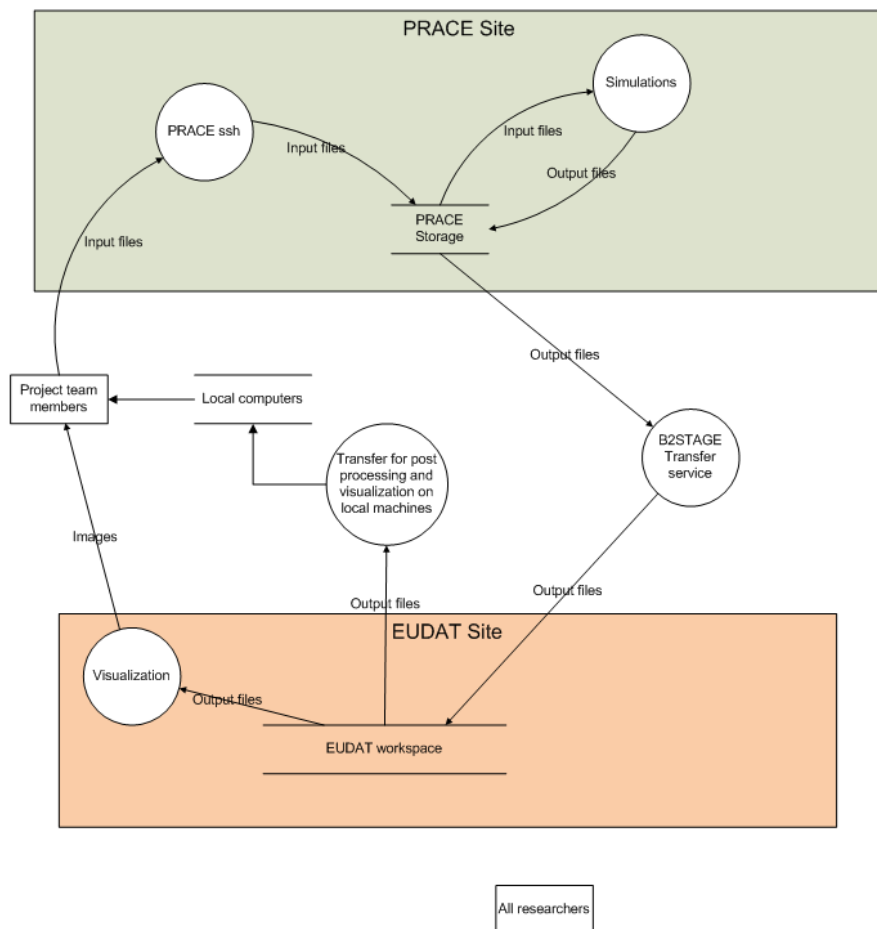


Figure 2: Data Flow diagram for CHARTERED data pilot

To start with, we have 2 logins per person: the PRACE login and the PDC/KTH login. The 5 team members have already a PRACE account associated with the project and a PDC/KTH account.

As KTH/PDC is not connected to the PRACE network, we use the public network between IT4Innovations and PDC/KTH. IT4Innovations is connected to public Internet via redundant 40 GB/s connection on top of which is also an IPsec 1Gb/s tunnel provided to securely connect to PRACE network. Over both connections, standard transfer tools like SSH/SCP, GSI-SSH/GSI-SCP and GridFTP are supported. All these transfer tools provide encrypted file transfer and thus securing the contents transferred between the IT4Innovations and KTH/PDC sites.

B2STAGE is installed in PDC including up-to-date gridftp and relevant certificates.

For data transfer we use GRIDFTP as it is the solution working behind EUDAT B2STAGE. Most of the simulation outputs have been transferred to EUDAT site, target for completion being end of April 2017.

Between EUDAT and local computers, GridFTP is used. This will be used as some of the post processing is performed at user's location.

4.8.3 *Next steps*

The CHARTERED project has applied for a continuation project as part of the DECI-14 call, requesting a data pilot also.

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.

4.9 Collaboration with EUDAT on the training domain

As we are working on the PRACE EUDAT data pilots, it appears that PRACE users are usually small groups (a project and not a community) who produce large volume of data. They are domain experts, but would need to be trained to data management best practices.

To facilitate this we established a working group involving PRACE and EUDAT training managers, meeting via video conference.

We started by delivering informally some research data management training using some of the EUDAT material [20] to PRACE project staff (during a face to face meeting at EPCC) and to data pilots users. The objective was to present research data challenges, the European Commission recommendation and the required services along the data life cycle.

The feedback has been really positive, and we think this type of training should be the starting point. We also invited them to the EUDAT/OpenAire webinar on the same topic.

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 advertised on each other mailing list.

Finally, we have agreed to include some PRACE content into EUDAT trainings and vice versa. The current agenda for this is as follows:

- During PRACE 2017 Spring School joint event with VI-SEEM, Cyprus, there will be a EUDAT delivered 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 ISC 2017 conference, PRACE organises the 3rd in the series of the education and training related workshops (22 June 2017). A EUDAT training team member will participate at the event.
- During 1st EUDAT & Research Data Management Summer School (July 2017), there will be several session delivered by PRACE team about HPC data management. The detailed program is under construction).

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, from the CHARTERED data pilot.

5 Annex

5.1 AFiD data pilot – data management plan

Data Collection

What data will you collect or create?

Data collected are results of fluid simulation. Plan is to have about 60 simulations.

The project will produce the following data set types:

Data set type	Size	Format	Comments
Snapshots	Big volume. Each snapshot is 2Gb to 50 Gb Multiple snapshots per simulation	HDF5	.Taken during the simulation, they are useful to the project team. They can be used to restart simulations
Results / statistics	Smaller Less than 1GB file, 1 per simulation + some log files 1000 files for the project	HDF5 Data stored in 2D fields.	They are easier to use
Snapshots on one plan	1000 times smaller but thousands of of them	HDF5	

In addition to those data sets, some elements are seen as data objects:

- The source code

How will the data be collected or created?

Simulations:

All data set types are created as results of the simulations ran on HPC as part of the DECI project.

The data are organized with the following directories structure tree:

Post processing:

Most of the post processing is done on the statistical data. They are downloaded on local computers by researchers using gridftp.

Documentation and Metadata

What documentation and metadata will accompany the data?

There are some description metadata, that are sufficient for the project team, but not for other persons.

For each system, important information will be used to name files and folders structure.
Some metadata are included into the files themselves, following HDF5 standards.

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 EPCC (PRACE site):

Results will be created on the /work space and can remain here during the project duration.

The data are copied to RDF.

We assume quotas won't be an issue based on the volume forecast.

In SURFsara (EUDAT site):

The storage is a disk + tape solution (based on HSM). This is what we will use for the EUDAT workspace.

In addition, some checkpoint/restart data will be stored in SURFsara outside of EUDAT.

How will you manage access and security?

In EPCC: Permissions are associated to the PRACE login in accordance to EPCC rules.

In SURFsara: based on the EUDAT services rules.

Selection and Preservation

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

Data set type	Preservation duration	Comments
Snapshots	5 to 10 years	Preserving them a few years would save time in case of further analysis.
Results / statistics	Long term / permanent	
Snapshots on one plan	To be defined	This would depend upon usage
Source code	Long term	

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

To be completed. We will explore possibility to use B2SAFE service for long term preservation.

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	Comments
Snapshots	Limited to project team	To be defined	Problem with the description required to make them public
Results / statistics	Limited to project team	Public	
Snapshots on one plan	Limited to project team	???	
Source code	Public	Public	Already on github : https://github.com/PhysicsOfFluids/AFiD

Note that there is already collaboration between AFiD and SURFsara for data publication (www.afid.eu).

Some early tests should happen later this year with B2SHARE, not necessarily with the current project data.

5.2 CHARTERED 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	Comments
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	