Coupling HPC and Data services together for large scale use cases

DI4R Conference - Krakow

28th September 2016

Giuseppe Fiameni
g.fiameni@cineca.it

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

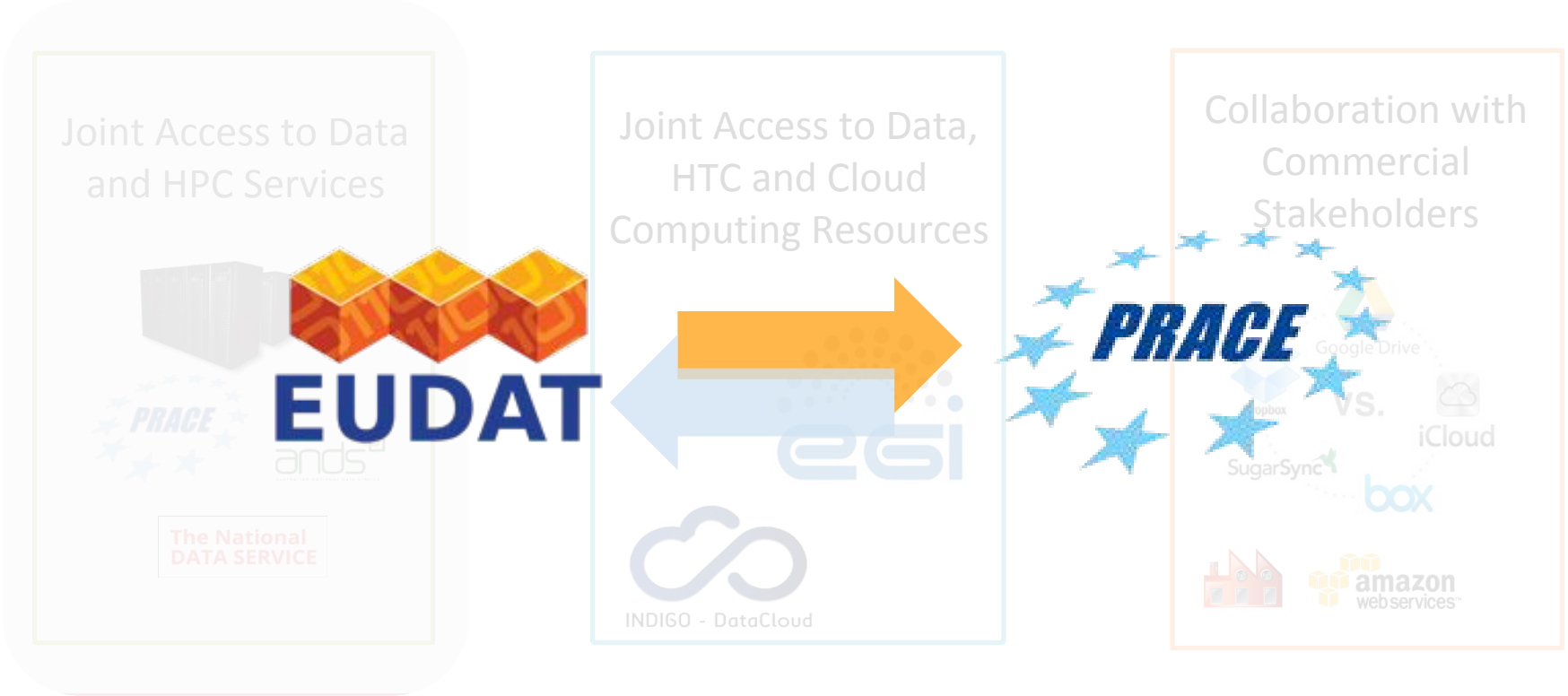
Background

- The amount of data produced in any particular discipline is starting to exceed the ability to manage them individually, leading to the development of a new analytic field named ***High Performance Data Analytics*** where data and computing services need to interact closely with each other.
- A large variety of services already exists but e-Infrastructures have evolved along different dimensions creating separate offerings for computing and data.
- Scientific communities do not access low level computing and data services directly, but rather work with portals and workflows to perform complex tasks.

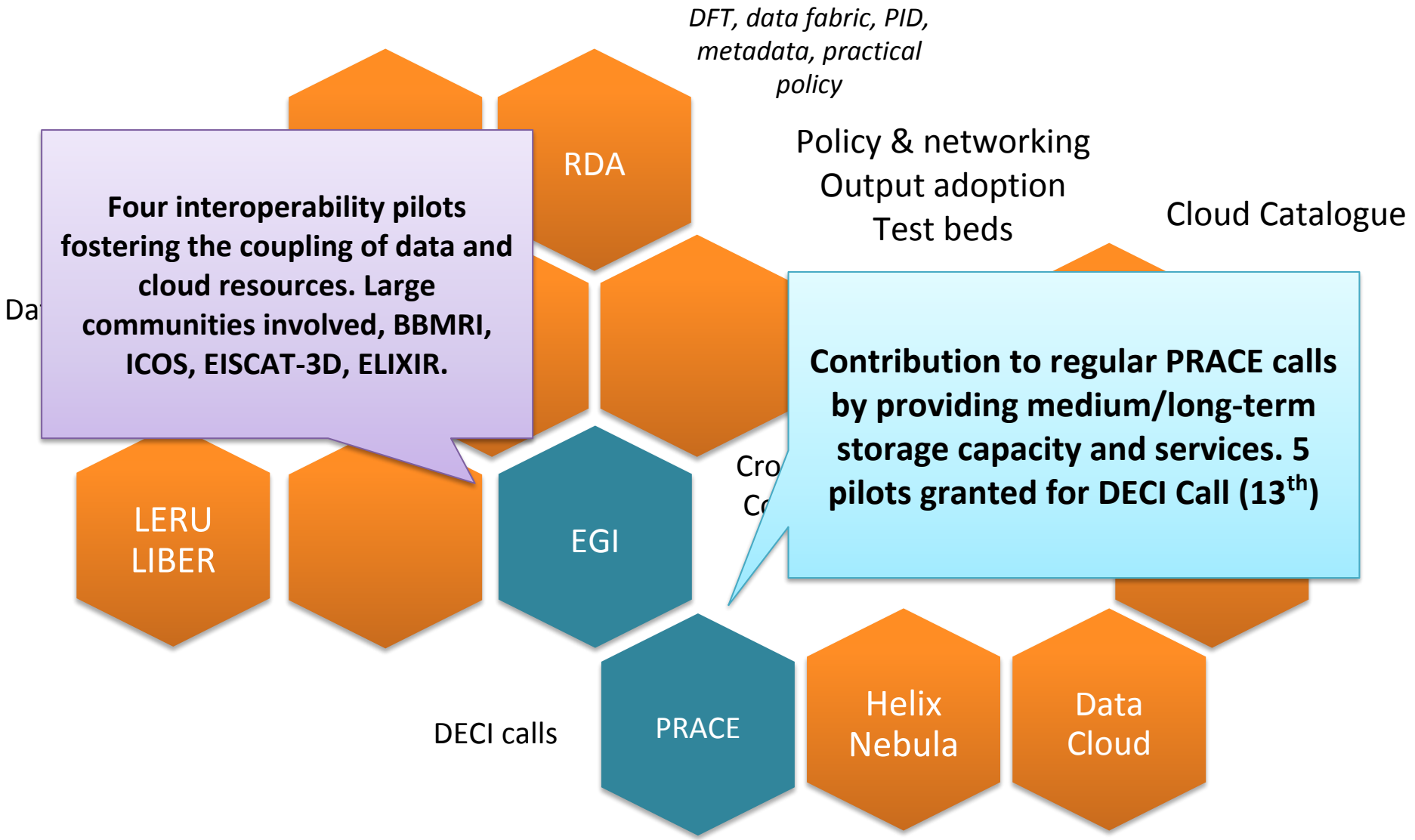
Need for interoperability

- Ensure the interoperability of EUDAT with other public and private e-Infrastructures, lowering technical and policy barriers **by piloting concrete use cases with user communities**
- Provide European researchers and industries with seamless access to data and computing resources for cross-utilization use cases
- Implement the **Open Science** vision where resources of any kind and size are accessible without any technical barrier.

Interoperability



E-Infrastructure Collaboration



How?

- **Joint Open Calls for proposals**
  - EUDAT offering data services and resources through regular PRACE calls
  - Review process is transparent to users
- **Joint training activities**
  - PRACE project investigators involved into EUDAT Data Management webinars and courses
- **Continuous technical discussion and developments of new components**
  - Definition of the EUDAT Workspace area
  - Synchronization of authentication credentials for single sign-on
  - EUDAT clients as part of the PRACE Common Production Environment

Areas of e-Infras harmonization

- **Technical**
  - Cross-utilization use cases, e.g. data transfer, workflow execution, data discoverability and provenance (PID), federated AAI, etc.
  - Combination of respective services catalogue
- **Policy**
  - Harmonization of access policy fostering the uptake of services on the long-term
- **Operational**
  - Harmonization and cross-fertilization of operational tools, technologies, practices and policies
  - *Security for Collaboration among Infrastructures* (https://www.eugridpma.org/sci/) Security collaboration : WISE (https://wise-community.org/)

PRACE Research Infrastructure

- PRACE – ***the Partnership for Advanced Computing in Europe*** – Research Infrastructure enables high impact European scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society.

- PRACE seeks to realize this mission through world class computing and data management resources and services open to all European public research through a peer review process.

B2 Service Suite

B2DROP
Sync and Exchange Research Data

B2SHARE
Store and Share Research Data

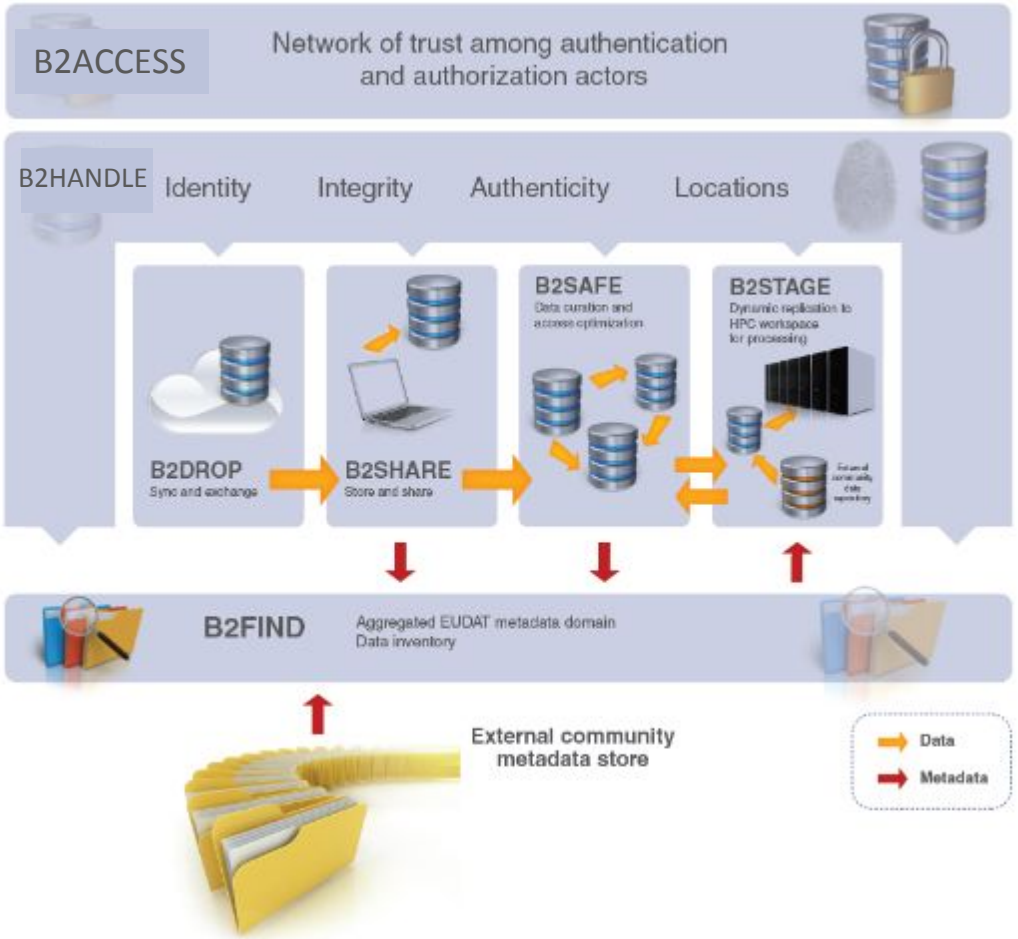
B2SAFE
Replicate Research Data Safely

B2STAGE
Get Data to Computation

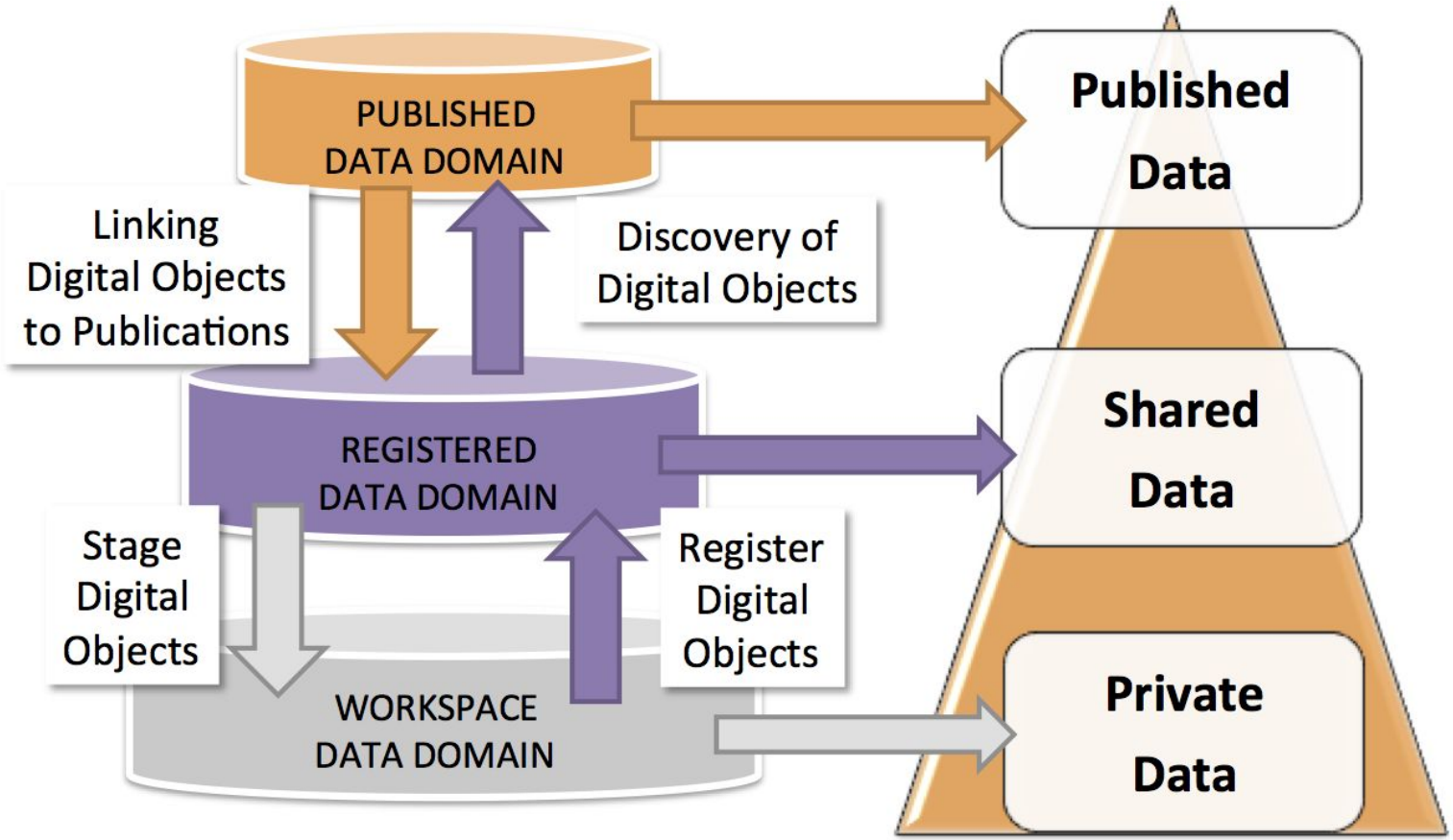
B2FIND
Find Research Data

B2HANDLE
Register your Research Data

B2ACCESS
Identity & Authorisation



CDI Data Domain



EUDAT Data Domain modeled on the ANDS[1] Data Curation Continiuum

1. Australian National Data Service organization – www.ands.org.au

Some Facts about the collaboration

- **5 pilots out of 10 granted resulting from the 13th PRACE DECI Call**
  - 20% of all applicants requested to access EUDAT data services
  - ~350TB of storage space
- **Different scientific fields**
  - Engineering, Material Science, Astrophysics, Earth Science
- **Relevant requirements**
  - Temporary store of large collection -> Workspace area
  - Sharing of intermediate results, simulations input -> B2DROP
  - Deposit of relevant result for publication -> B2SHARE

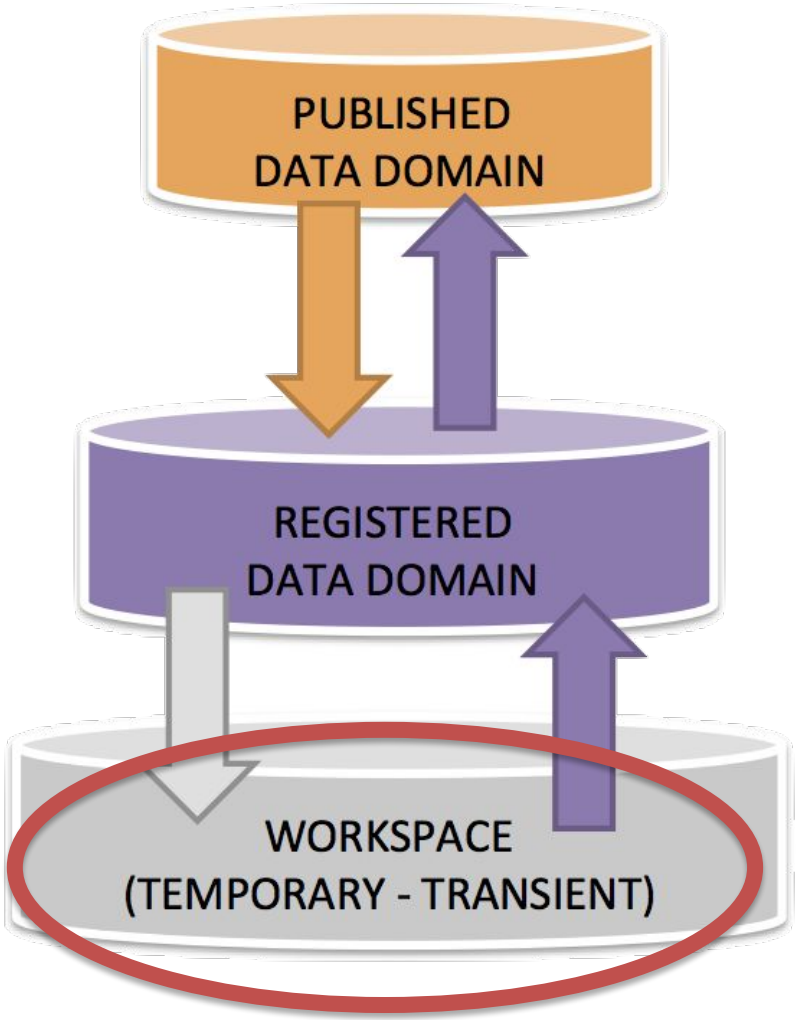
Some numbers

Code	Project Name	Field	Country	Data requirements during the PRACE	Data requirements in TB after the PRACE project	Duration	EUDAT Site	PRACE Site
HybTurb3D	Hybrid 3D simulations of turbulence and kinetic instabilities at ion scales in the expanding solar wind	Astro Sciences	IT	140 TB	140 TB	24 m	CINECA	SurfSARA
MULTINANO	Multiscale simulations of nanoparticle suspensions	Engineering	IT	30 TB	30 TB	24 m	CINECA	MPDCF
CHARTERED	Charge transfer dynamics by time dependent density functional theory (CHARTERED)	Materials Science	SE	30TB	20TB	24 m	KTH/PDC	IT4I
HiResClimate	High Resolution EC-Earth Simulations	Earth Sciences	IE	150TB	150TB	12 m	EPCC	KTH
AFiD	Effect of rotation and surface roughness on heat transport in turbulent flow	Engineering	NL	11TB	10TB for 24 months 1TB for long-term storage	10TB for 24 m 1TB for long-term storage	SURFsara	EPCC

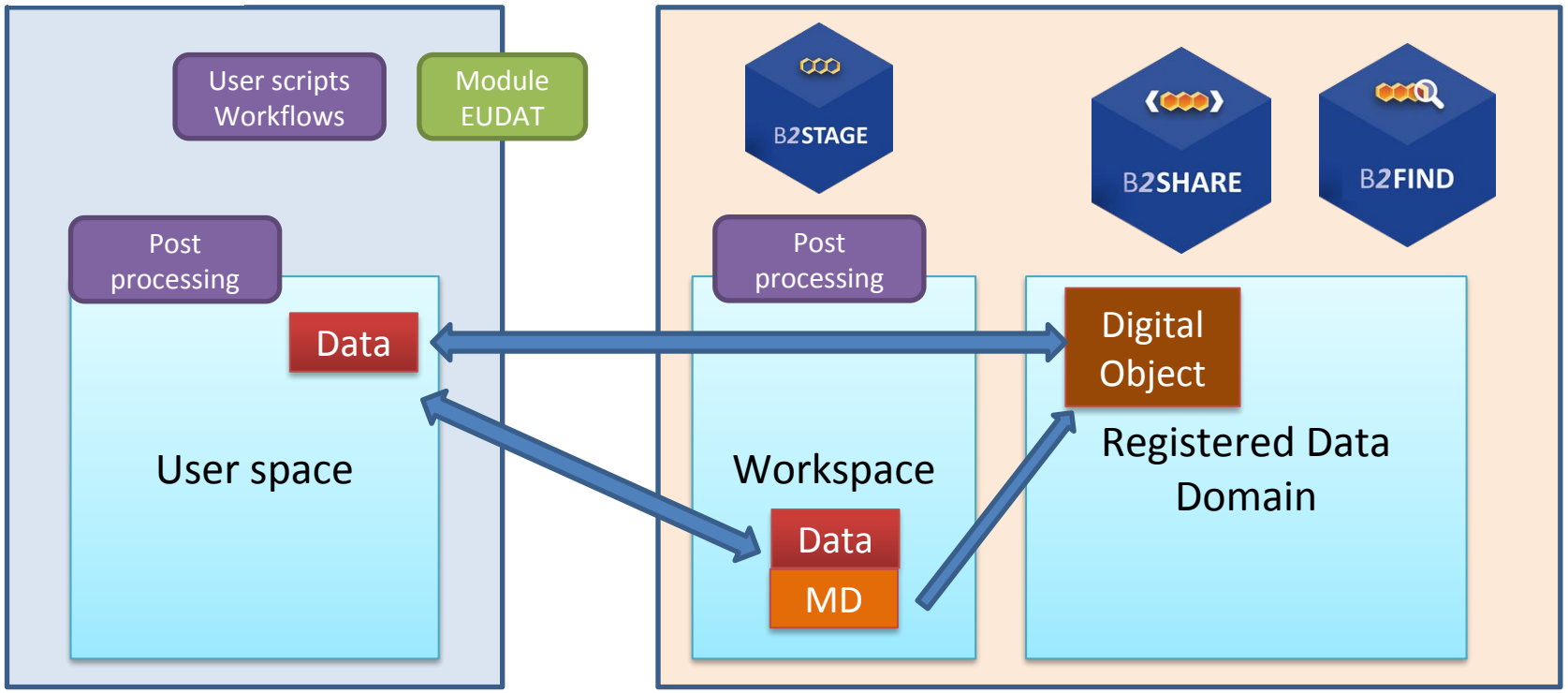
WORKSPACE

Requirements

- Communities and users (e.g. PRACE) want deposit area for digital entities from computing simulations
- Connect to existing, community specific access services
- Support multiple protocols: GridFTP, Webdav, POSIX (full, light, like)
- Integrated within the CDI domain and services



WORKSPACE


User scripts
Workflows
Module
EUDAT
B2STAGE
B2SHARE
B2FIND
Post
processing
Post
processing
Data
Digital
Object
User space
Workspace
Registered Data
Domain
Data
MD

EUDAT

Challenges

- Synchronizing people, funds, and resources of projects which have their own implementation plan is challenging

- PRACE does not target RIs but rather individual researchers or research groups

- Transferring large amounts of data across internet is difficult
  - Large archives maintained close to computational power, at least for HPC applications

For more info:

https://b2drop.eudat.eu
https://eudat.eu/services/userdoc/b2drop

https://b2share.eudat.eu
https://eudat.eu/services/userdoc/b2share

https://eudat.eu/services/userdoc/b2safe

https://eudat.eu/services/userdoc/b2stage

http://b2find.eudat.eu
https://eudat.eu/services/userdoc/b2find

http://b2access.eudat.eu

EUDAT

Many thanks for your attention!

www.eudat.eu