EUDAT
Data Services, Tools & Knowledge

Damien Lecarpentier
EUDAT Project Manager
CSC – IT Center for Science Ltd
E-IRG, 11.11.2014

SEVENTH FRAMEWORK
PROGRAMME

EUDAT

EUDAT: A pan-European e-Infrastructure solution for pan-European RI data Challenges

- **All Research Infrastructures are facing data challenges**
    - Where to store the growing amount of data?
    - How to find it?
    - How to make the most of it?

- **Many communities are developing own solutions**
    - This is good…
    - … but we also need to make sure that the solutions remain interoperable

- **EUDAT mission is to fill this gap**
    - Providing a set of services to help RIs managing their growing amount of data
    - Providing these services across communities to ensure minimum level of interoperability
    - EUDAT also help to bring data and computing together (HPC centers core partners)

EUROPEAN DATA INFRASTRUCTURE

EUDAT

2

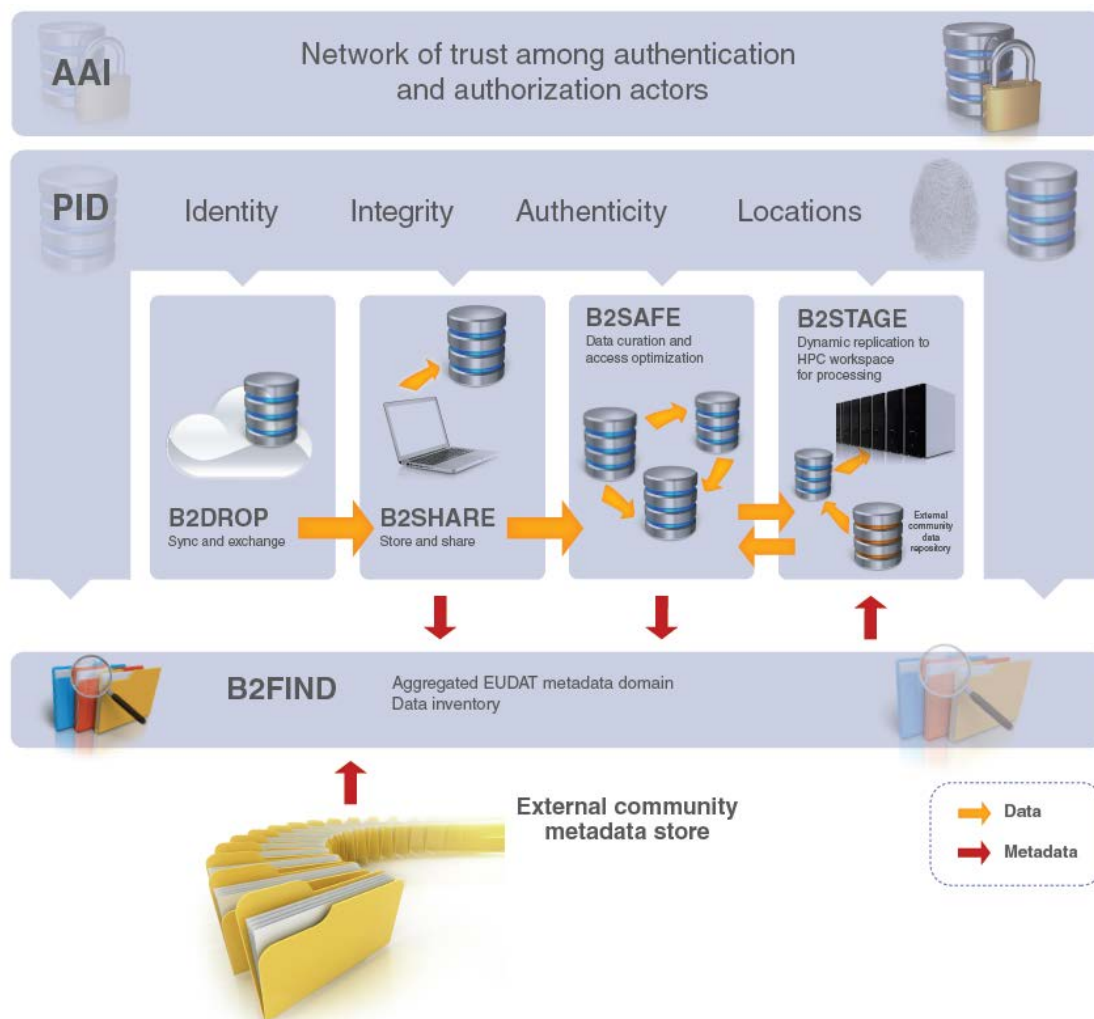# Data Centers and Communities

User Forums + 30 communities

1st User Forum
7-8 March 2012, Barcelona

Services & Resources

Covering both **access** and **deposit**, from **informal data sharing to long-term archiving**, and addressing **identification**, **discoverability** and **computability** of both **long-tail and big data**, EUDAT's services will **address the full lifecycle of research data**

B2DROP is a **secure and trusted** data exchange service for researchers and scientists to keep their research data **synchronized** and up-to-date and to **exchange** with other researchers.
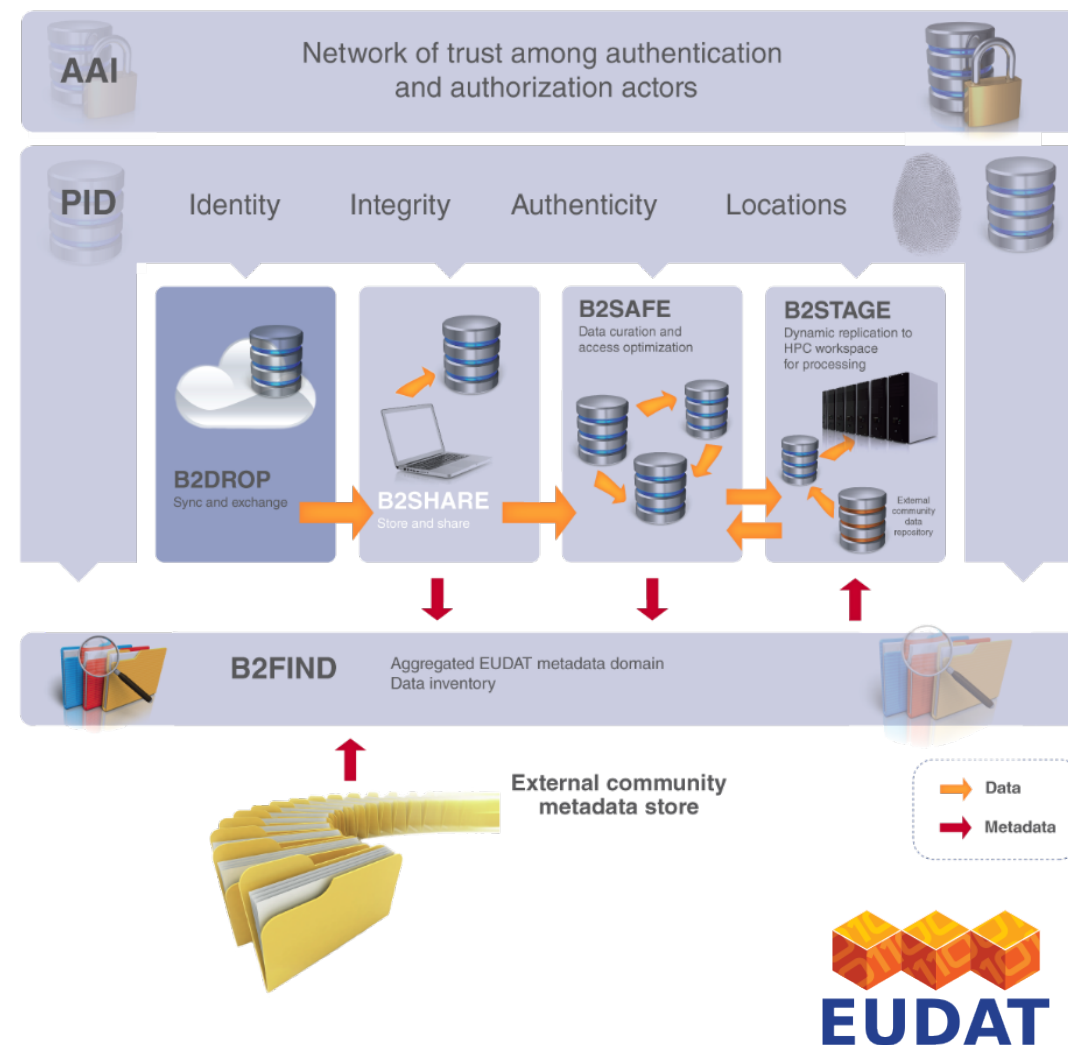
An ideal solution to:

- **Store and exchange** data with colleagues and team

- **Synchronize** multiple versions of data

- Ensure **automatic desktop** synchronization of large files
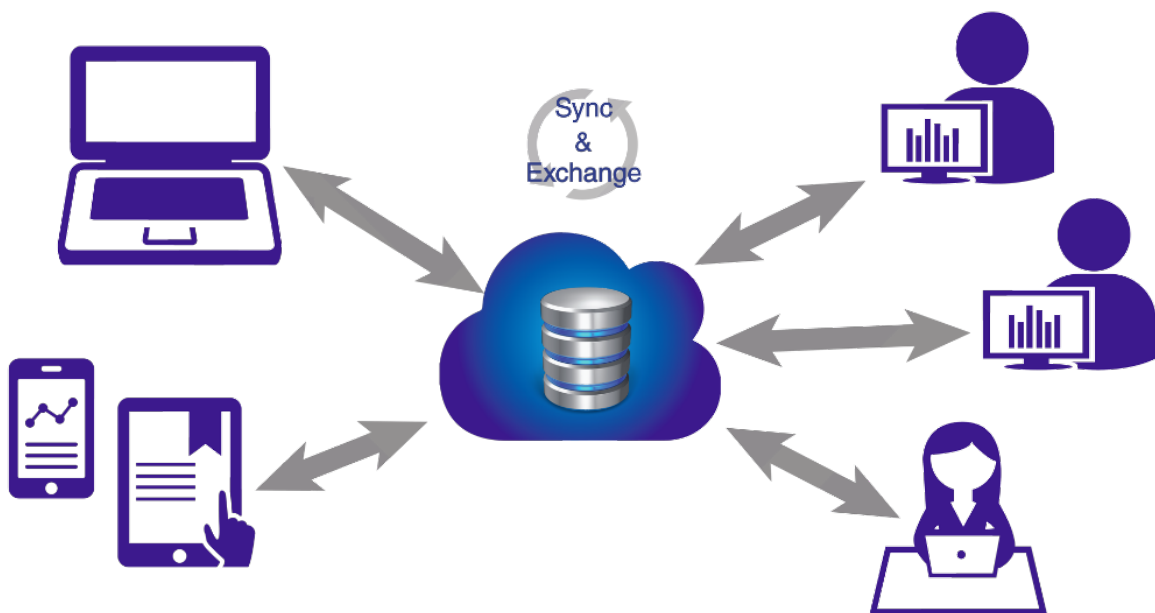
B2DROP features

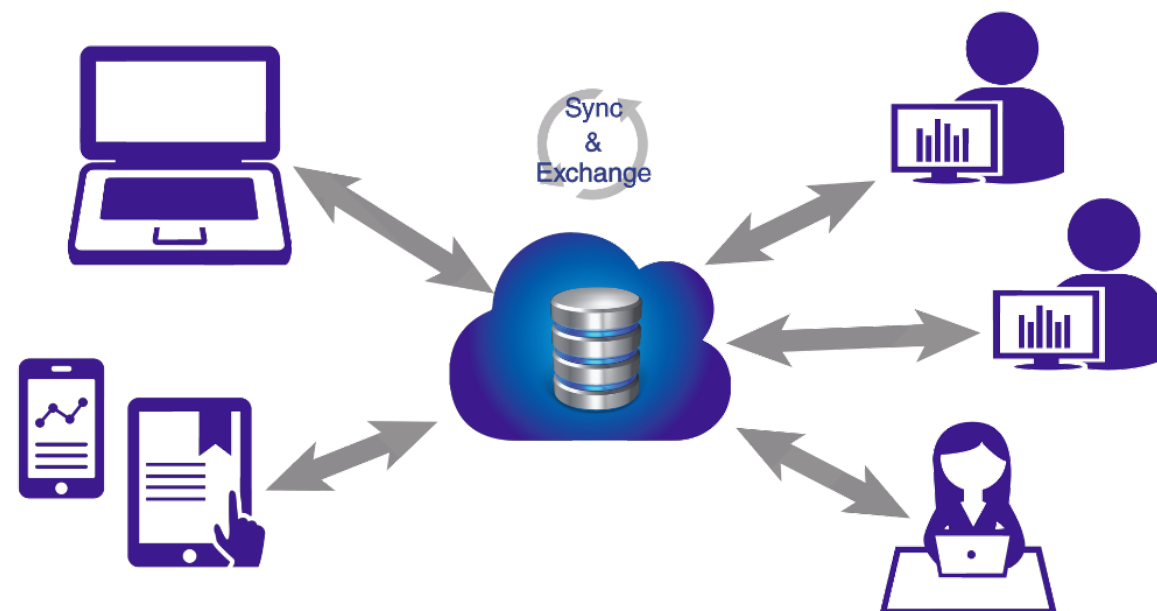B2DROP
Sync and Exchange Research Data

b2drop.eudat.eu

- future **integration with the B2 suite of services** to allow user-friendly data sharing

- users decide **with whom to exchange data, for how long and how**

- **up to 20GB of storage space** for research data

- **access and manage permissions** to files **from any device and any location**

- **simple to use and open to all** researchers, scientists, communities alike to **synchronize and exchange data with one or multiple users**

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.

A winning solution to:

Store: facilitates research data storage

Preserve: guarantees long-term persistence of data

Share: allows data, results or ideas to be shared worldwide

B2SHARE
Store and Share Research Data
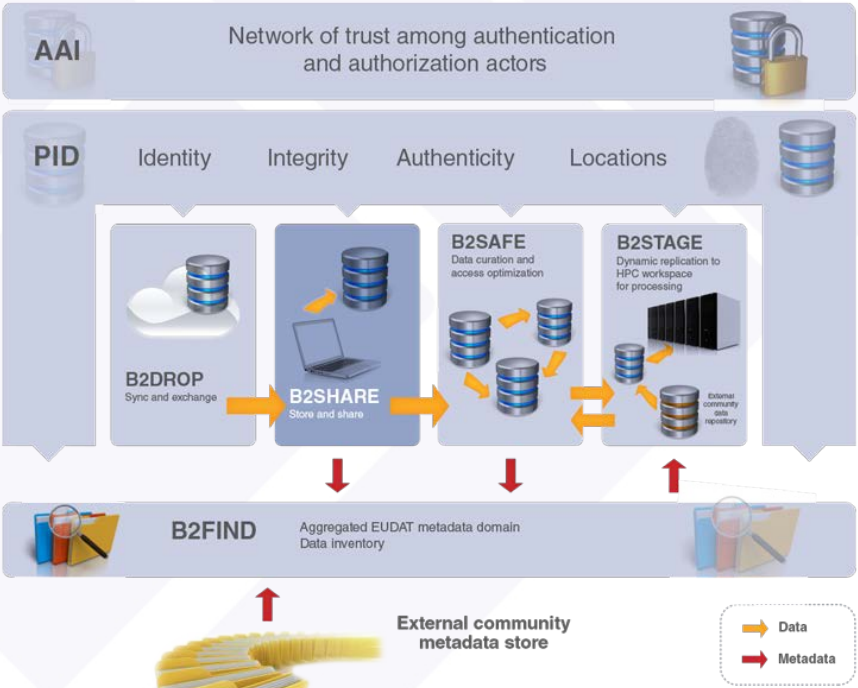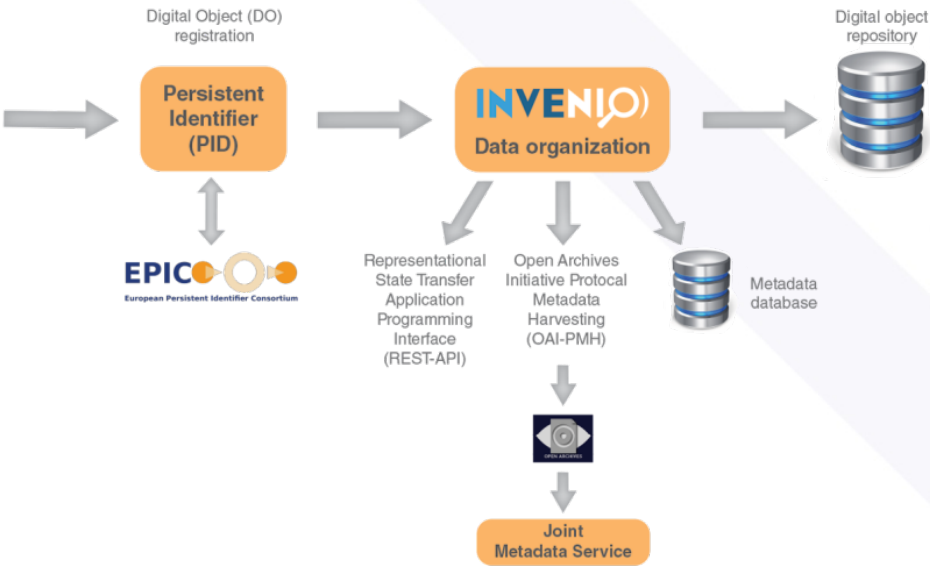
b2share.eudat.eu

A four-click service

B2SHARE
Store and Share Research Data

b2share.eudat.eu

B2SHARE features

- Targets **small-scale research data collected as part of international collaboration and looking for a central repository**

- integrated with the **EUDAT collaborative data infrastructure**

- **free upload and registration** of stable research data

- data assigned a **permanent identifier**, which can be retraced to the data owner

- **community-specific metadata extensions** and **user interfaces**

- **openly accessible** and **harvestable metadata**

- **representational state transfer application programming interface (REST API)** for integration with community sites

- **data integrity ensured by checksum** during data ingest

- **professionally managed storage service –** no need to worry about hardware or network

- **monitoring of availability and use**

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.

A solution to:

- Provide an **abstraction layer which virtualizes large-scale data resources**

- Guard against data loss in long-term **archiving and preservation**

- **Optimize access** for users from different regions

- Bring data **closer to powerful computers** for compute-intensive analysis

B2SAFE
Replicate Research Data Safely

eudat.eu/b2safe

AAI — Network of trust among authentication and authorization actors

PID — Identity   Integrity   Authenticity   Locations

B2DROP Sync and exchange
B2SHARE Store and share
B2SAFE Data curation and access optimization
B2STAGE Dynamic replication to HPC workspace for processing
External community data repository

B2FIND Aggregated EUDAT metadata domain Data inventory

External community metadata store

Data
Metadata

B2SAFE features

- Provides a **European extension to Finnish national data services (e.g. IDA) for replicating and accessing data across European sites**

- based on the execution of **auditable data policy rules** and the use of **persistent identifiers (PIDs)**

- respects the **rights of the data owners** to define the **access rights for their data** and to decide how and when it is made **publicly referenceable**

- data policies are **centrally managed via a Data Policy Manager**, and the policy rules are implemented and enforced by **site-local rule engines**

- able to **aggregate data from different disciplines** into a storage system of **trustworthy and capable data service providers**

- support for **repository packages** (e.g. DSPACE, FEDORA) and a **lightweight HTTP-based solution**

B2SAFE Data Policy Manager (DPM)

Community defined policy rules

Community repository

Data centre store

B2SAFE
Data Replication
PID management
Data integrity checking
Data curation

Data centre store

Data centre store

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.

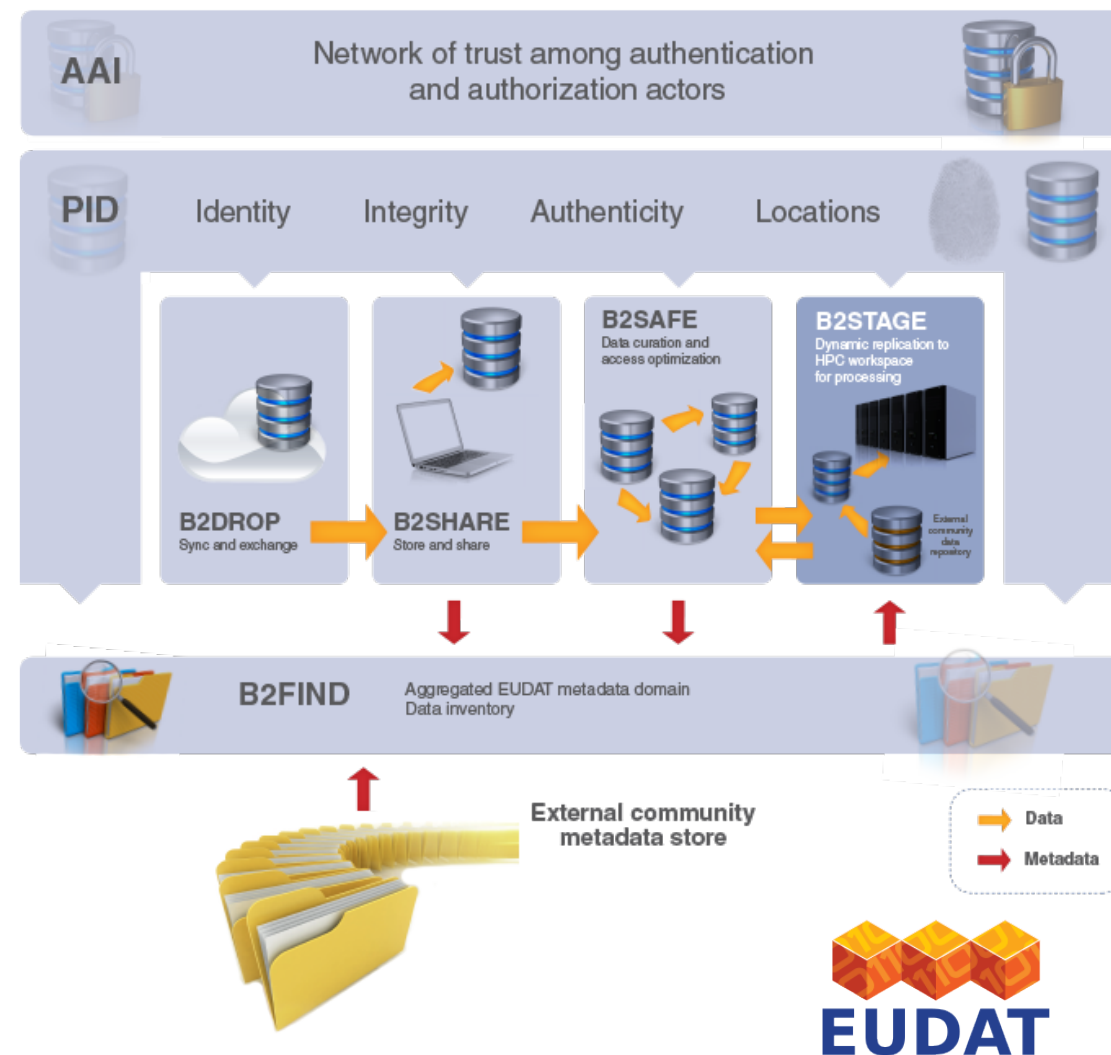**The service allows users to:**

- **Transfer** large data collections from EUDAT storage facilities to **external HPC facilities for processing**

- In conjunction with B2SAFE, **replicate community data sets, ingesting them** onto EUDAT storage resources for long-term preservation

- **Ingest computation results** into the EUDAT infrastructure

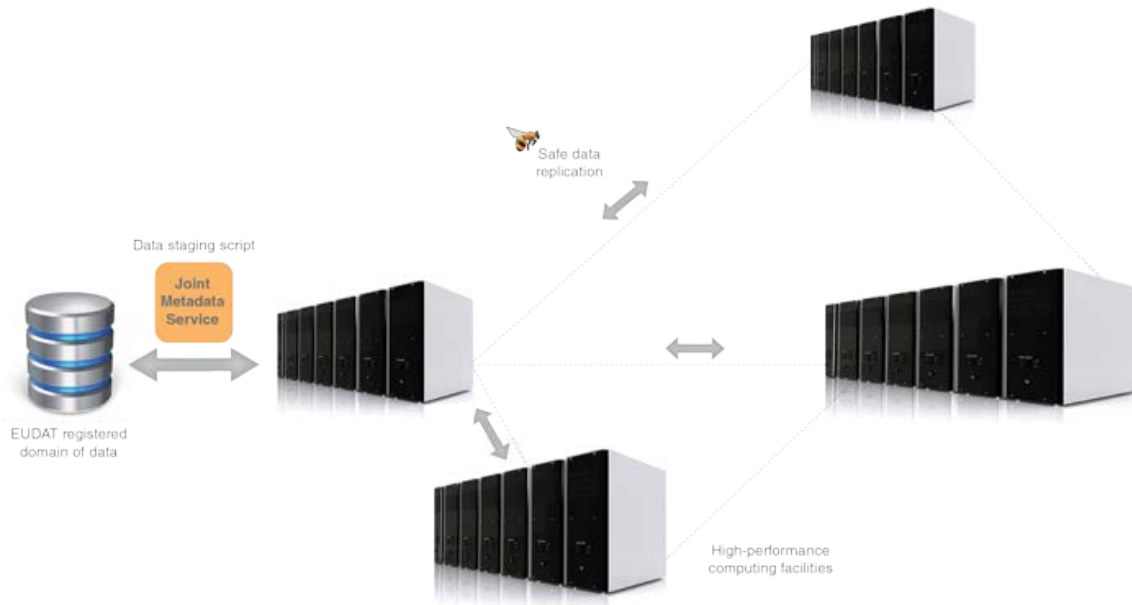- Access data through a **RESTful HTTP interface** (in progress)

B2STAGE
Get Data to Computation

eudat.eu/b2stage

EUDAT

B2STAGE features

- **an extension of the B2SAFE and B2FIND services,** which allow users to store, preserve and find data

- **data-staging script** facilitates staging, ingestion and retrieval of persistent identifier (PID) information of transferred data

- service available **to all registered researchers and interested communities**

- **users negotiate access to remote HPC services in parallel**

- **collaboration with other infrastructures,** such as the European Grid Infrastructure (EGI) and Partnership for Advanced Computing in Europe (PRACE)

- **documentation, educational material and service helpdesk** available to support users

B2FIND is a simple, user-friendly **metadata catalogue of research data collections** stored in EUDAT data centres and other repositories.

A service which allows users to:

- **Find** collections of scientific data quickly and easily, irrespective of their origin, discipline or community

- Get quick **overviews** of available data

- Browse through collections using **standardized facets**

b2find.eudat.eu
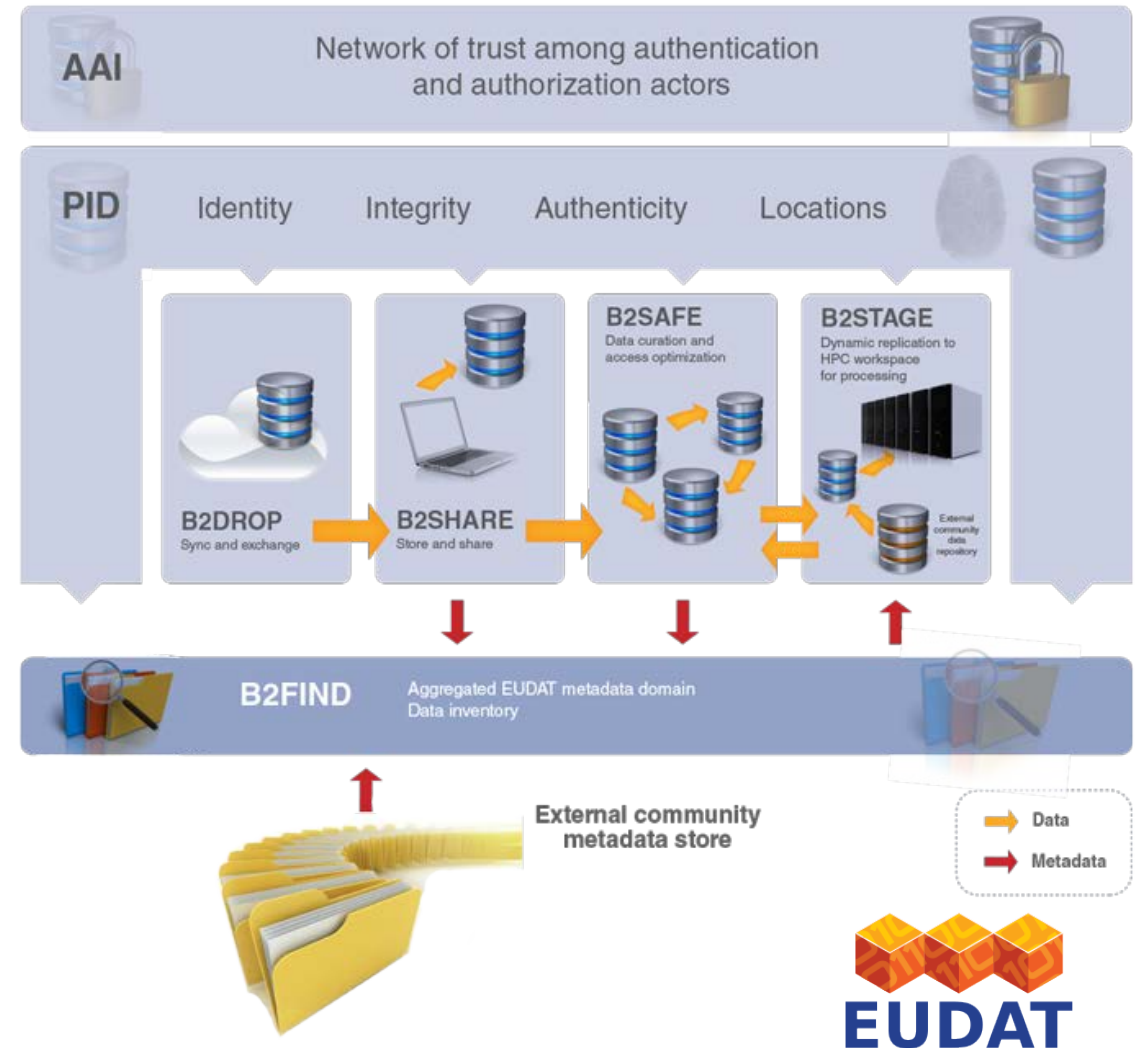
B2FIND features

- **supports faceted, geospatial and temporal metadata searches**

- allows users **to search and browse datasets via keyword searches**

- initially available for **communities in the EUDAT registered domain of data**

- EUDAT will then extend the service **to other interested and reliable data and metadata providers**

- results displayed in **user-friendly format** and **listed in order of relevance**

- access to **the scientific data objects** is given through **references provided in the metadata**

b2find.eudat.eu

Community Outreach & Service take Up

EUDAT H2020 (32)

Factor 6

Start EUDAT (5)
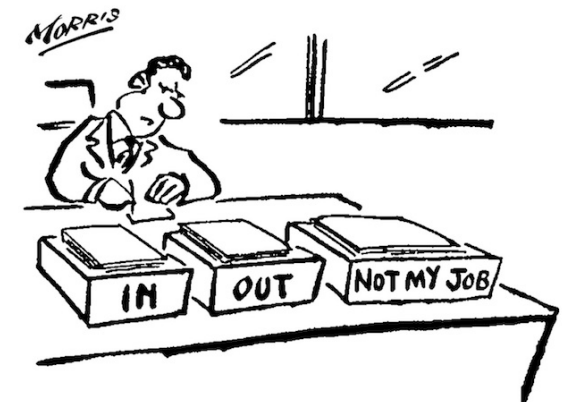
Integrating

Pilots/testing

Interacting

17
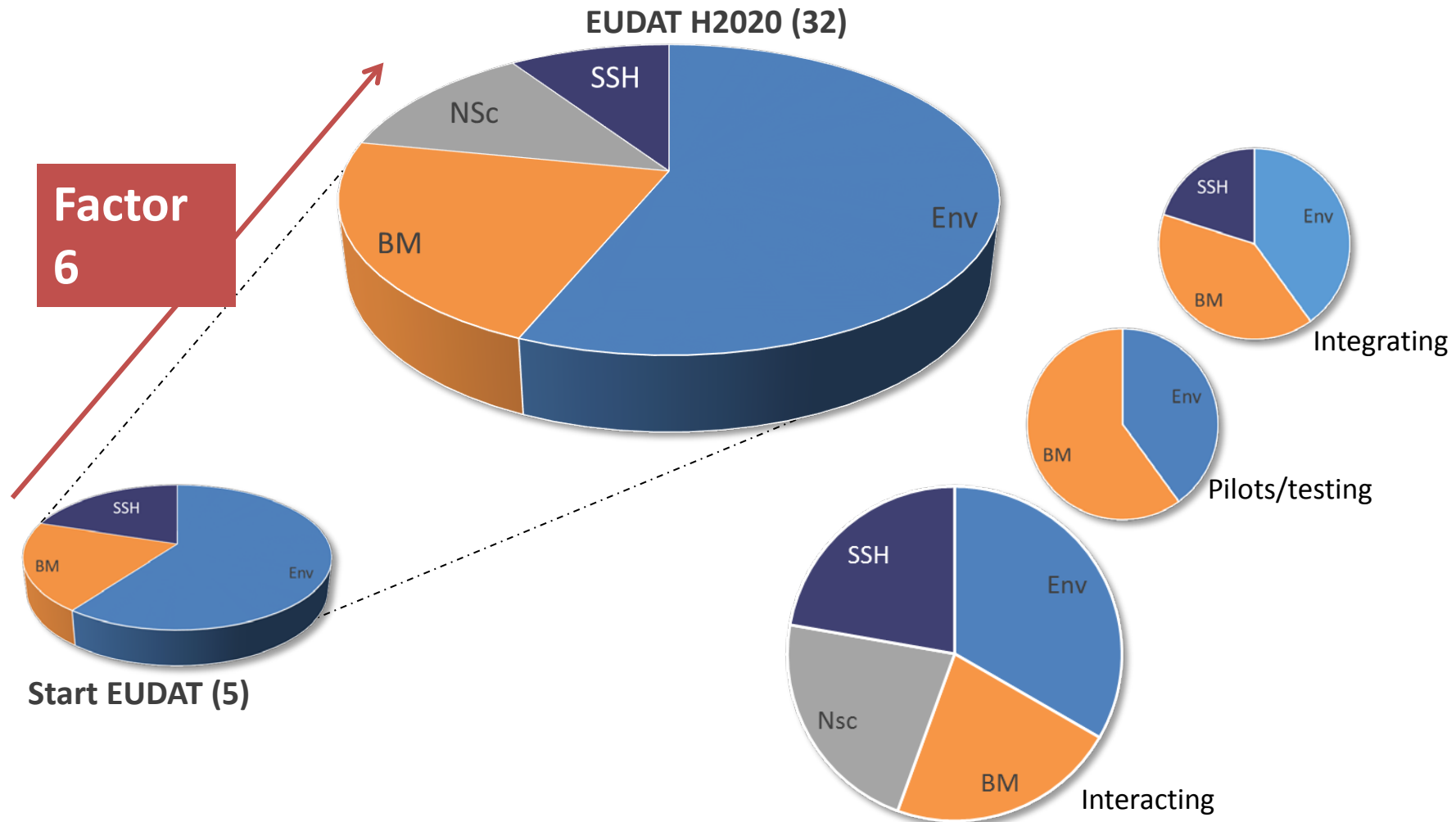
WHAT IS COMING NEXT?

Policies / Data Access and Reuse

- Open Access?

  – Funders: "Yes, absolutely!"

  – Researchers: "Yes, but…"
    - Some data is "sensitive"
    - What about credit and merit?
    - How to find one's way in the legal minefield?

  – What role for e-Infrastructure and service providers?
    - Providing the right tools
    - Training, training, training

Policies / Resource funding & allocation

- ## Business model
  - *Basic* vs *Premium* services
  - *Pre-paid* resources vs *pay-per* use

- ## Access mechanisms
  - *Quality* vs *Market*-based access
  - How to balance contributions?
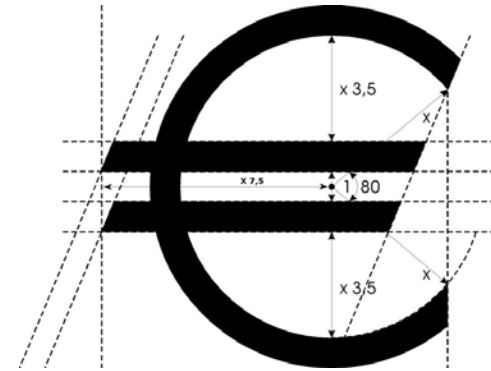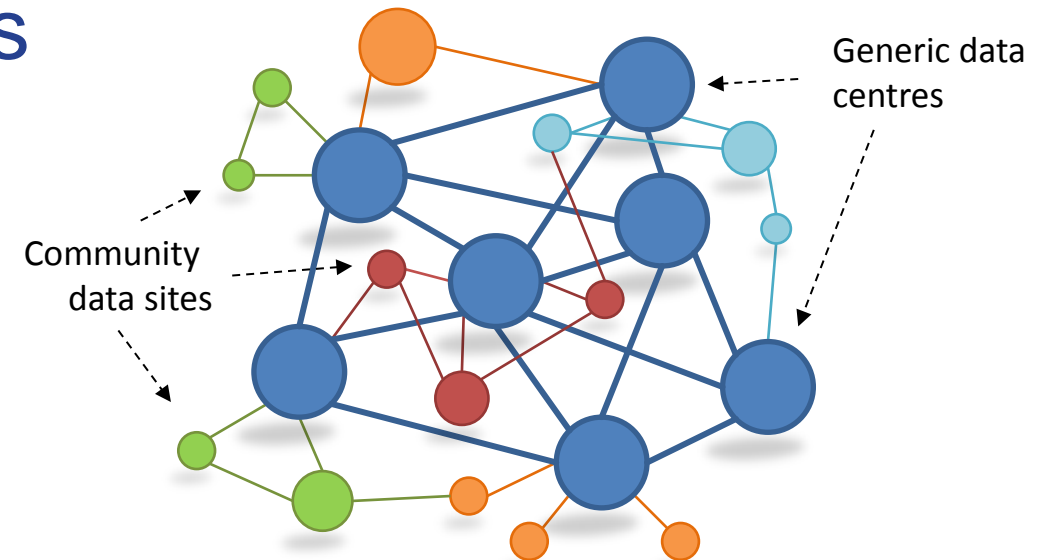
Partnerships!

- ## Research Infrastructures → CDI users

  – Integration of EUDAT components within RIs

  – Governance

  – Business models

- ## EUDAT nodes → CDI providers

  – Operational Level Agreements (OLAs)

  – Service level policies

  – Funding & legal models

Generic data centres

Community data sites
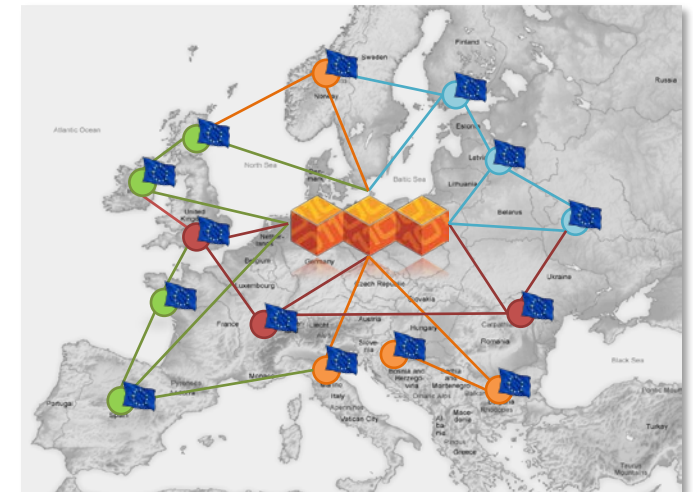
Partnerships!

- ## European E-Infrastructures
  - Cross-Infrastructure services/e-Infra Commons
  - Policies
  - Business models

- ## National initiatives
  - Interoperability → bridging national & European solutions
  - Cross-national collaborations
  - Roles and responsibilities

Thank you