



# EUDAT

## Data Services, Tools & Knowledge



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

- Where to store the growing amount of data?
- How to find it?
- How to make the most of it?



- This is good...
- ... but we also need to make sure that the solutions remain interoperable



- 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)



# Data Centers and Communities

26 European  
partners



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



**KIT**  
Karlsruhe Institute of Technology



**Barcelona  
Supercomputing  
Center**  
Centro Nacional de Supercomputación

**epcc**



**DKRZ**  
DEUTSCHES  
KLIMARECHENZENTRUM



**Red IRIS**



**CSC**



Max-Planck-Institut  
für Meteorologie

**JÜLICH**  
FORSCHUNGSZENTRUM

**rzg**  
RECHEN-  
ZENTRUM  
GARCHING



Science & Technology  
Facilities Council

**maatG**



**umweltbundesamt**  
ENVIRONMENT AGENCY AUSTRIA



**SNIC**



Trust-IT Services Ltd  
Communicating ICT to markets

**UNINETT**  
sigma

**SURF SARA**

**UCL**



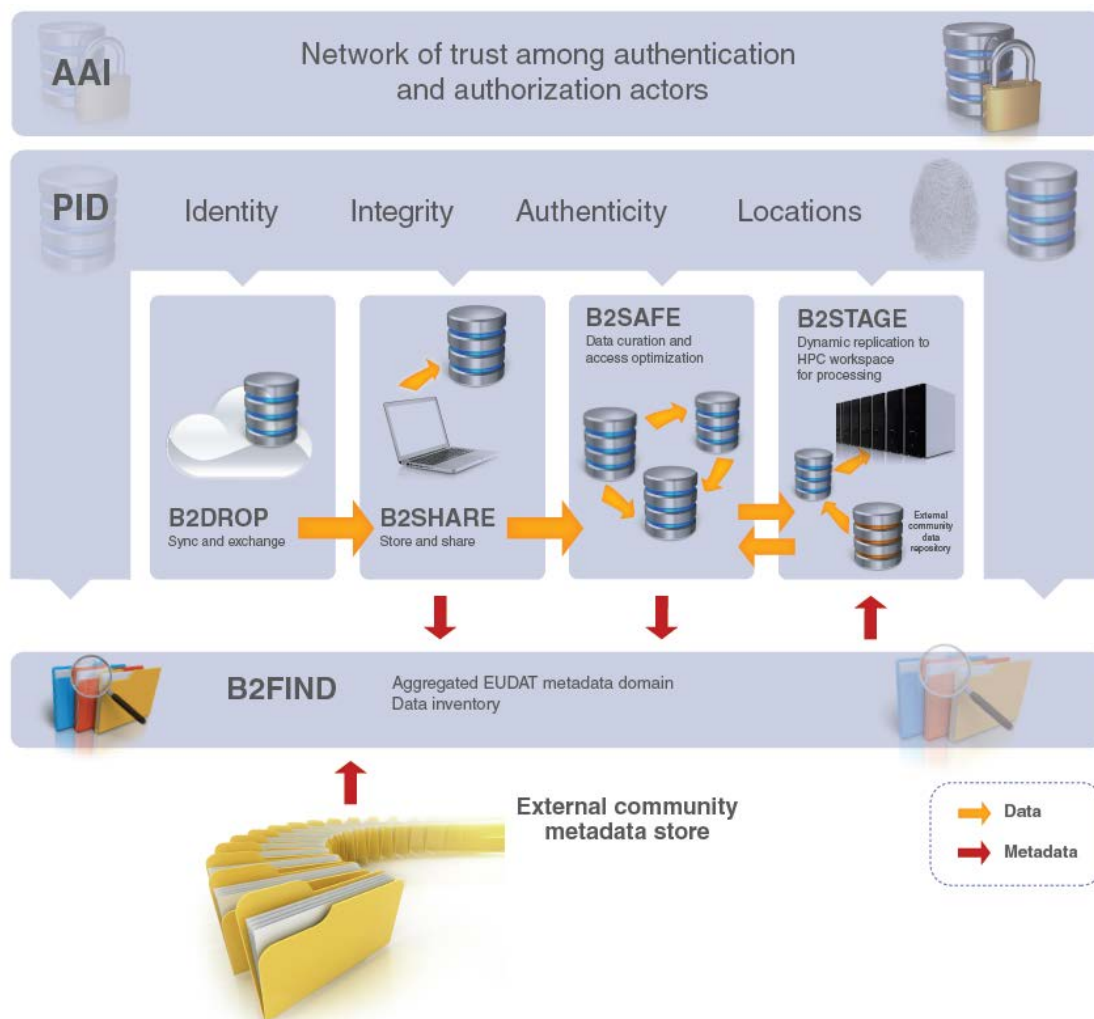
# User Forums + 30 communities



1<sup>st</sup> User Forum  
7-8 March 2012, Barcelona






# Services & Resources

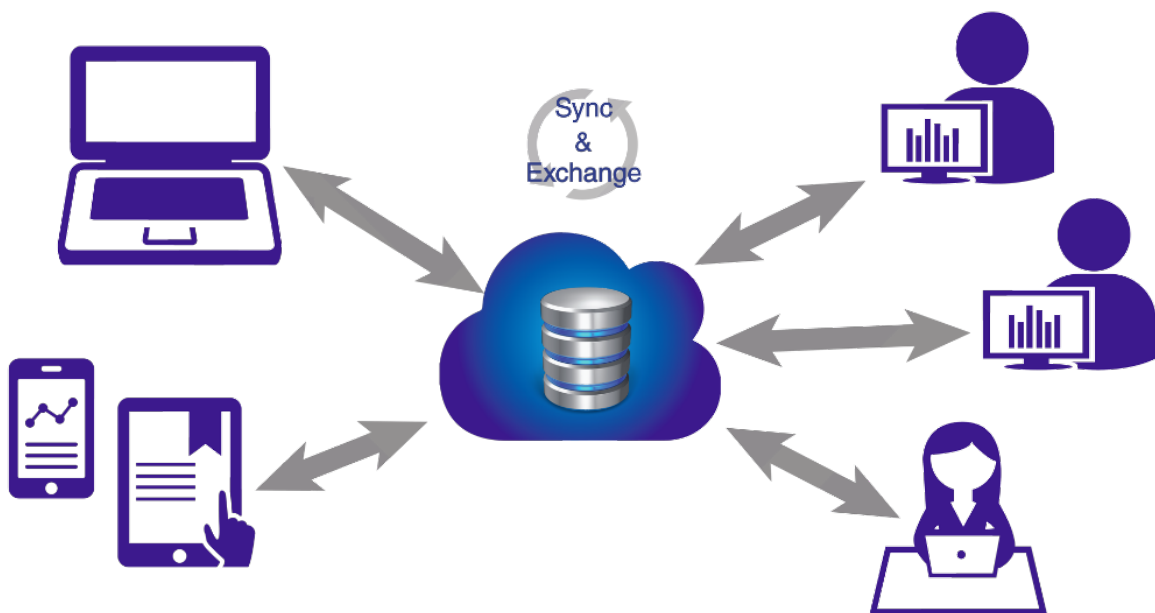


Covering both **access** and **deposit**, from **informal data sharing** to **long-term archiving**, and addressing **identification**, **discoverability** and **computability** of both **long-tail** and **big data**, EUDAT's services will address the full lifecycle of research data

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:

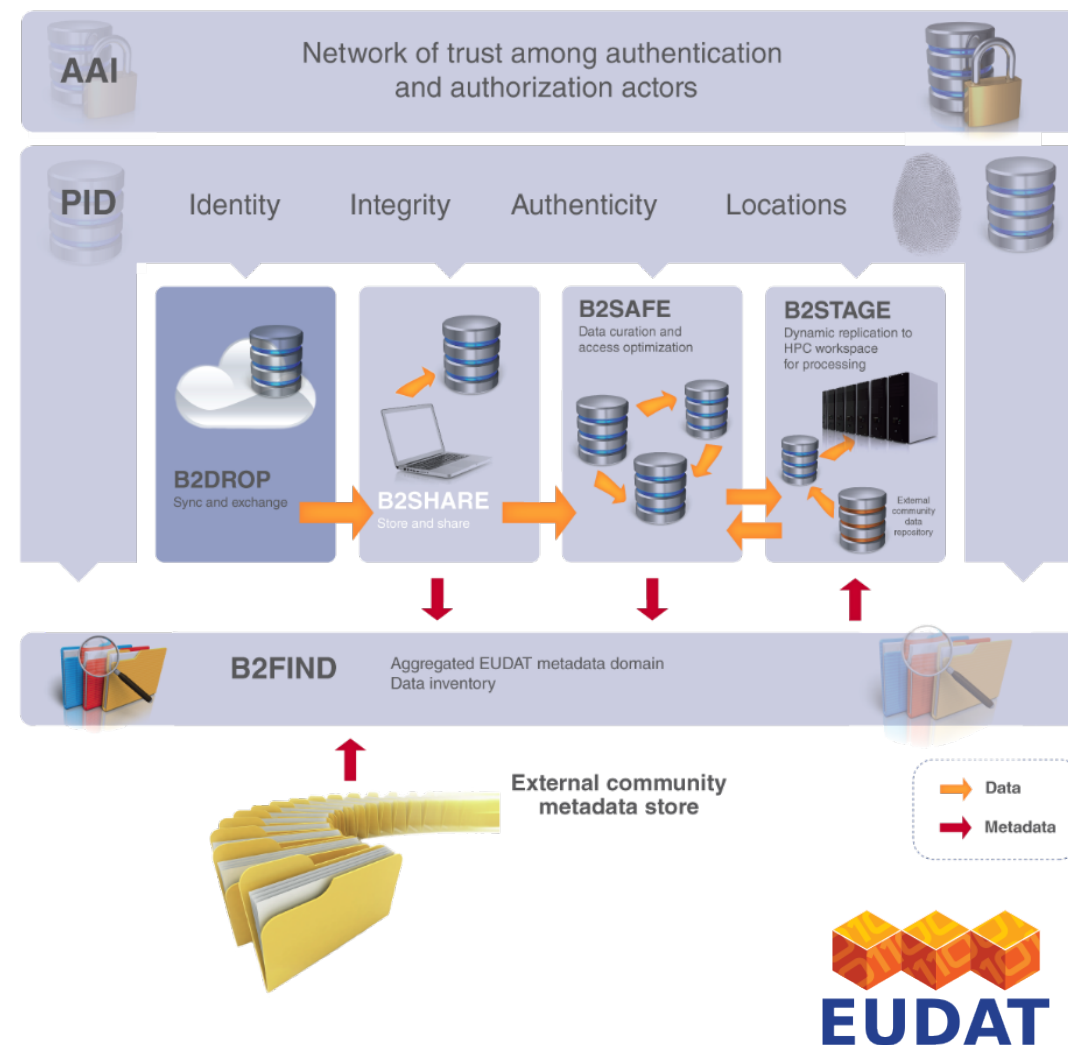
-  **Store and exchange** data with colleagues and team
-  **Synchronize** multiple versions of data
-  Ensure **automatic desktop** synchronization of large files



# B2DROP

Sync and Exchange Research Data

[b2drop.eudat.eu](http://b2drop.eudat.eu)



**EUDAT**

# B2DROP features

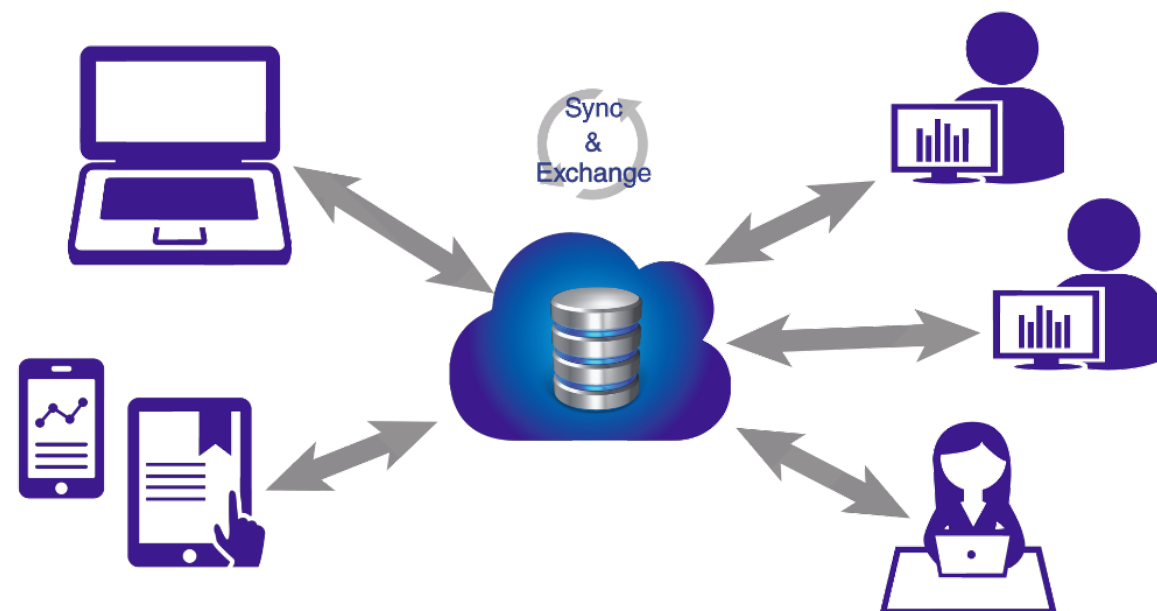
- future **integration with the B2 suite of services** to allow user-friendly data sharing
- users decide **with whom to exchange data, for how long and how**
- **up to 20GB of storage space** for research data
- **access and manage permissions** to files from any device and any location
- **simple to use and open to all** researchers, scientists, communities alike to **synchronize and exchange data with one or multiple users**



## B2DROP

Sync and Exchange Research Data




[b2drop.eudat.eu](http://b2drop.eudat.eu)





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](https://b2share.eudat.eu)

1. Upload data

2. Select domain

3. Enter basic and domain-specific metadata

Generic

ADD NEW DATA

Title

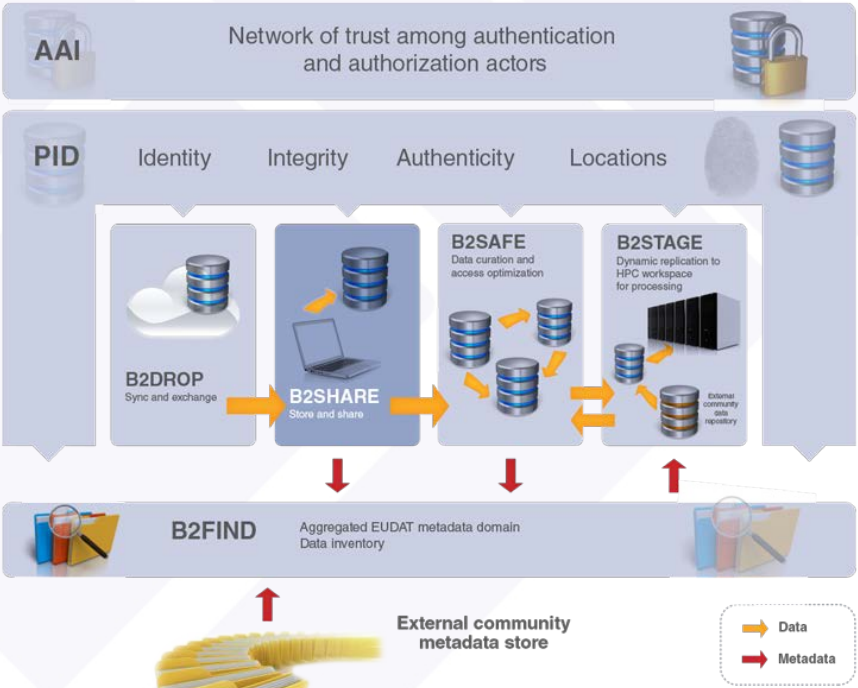
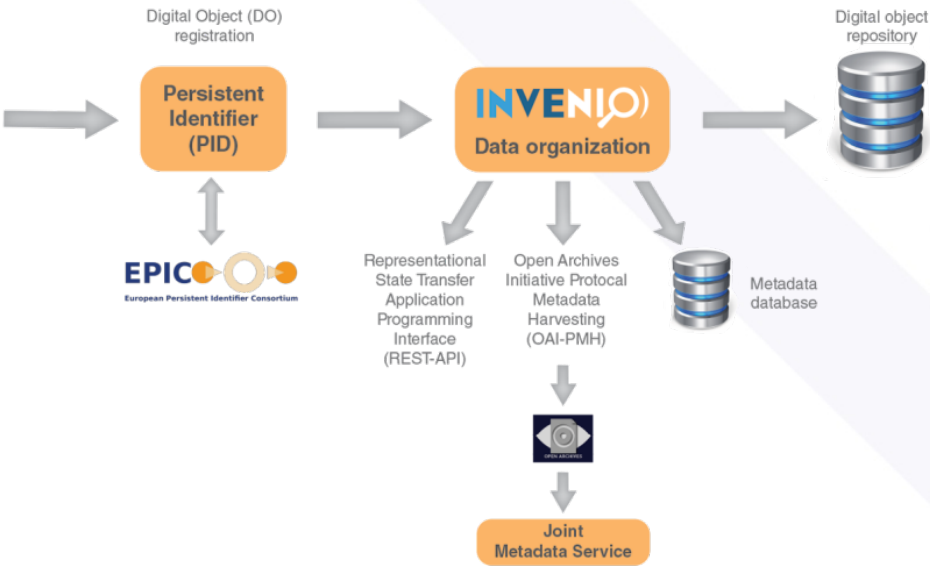
Description

Creator

Open Access

License

Keywords





# A four-click service



**B2SHARE**  
Store and Share Research Data

[b2share.eudat.eu](https://b2share.eudat.eu)



**B2SHARE**  
Store and Share Research Data

Search 139 records for



Search

You are logged in as Damien.

Deposit >>>

## Latest Deposits

2014-10-23  
**PARADE, Strategy for a European Data Infrastructure White Paper**  
by Kimmo Koski[...]  
Strategy for a European Data Infrastructure White Paper by PARADE, Partn ...

2014-10-21  
**Knowledge Exchange Sustainability Index**  
Output from the Knowledge Exchange workshop: Sustainable Business Models ...

2014-10-10  
**MHD run for a single cluster simulated with 640^3 cells with ENZO-MHD at z=0.**  
by Franco Vazza  
Hdf5 files for 3D dataset for a cluster run using a uniform grid box wit ...

2014-10-09  
**ENZO-AMR data for cluster E1**  
by Franco Vazza  
Hdf5 monolithic 256^3 files for a simulated galaxy cluster at z=0, using ...

## STORE AND SHARE YOUR RESEARCH DATA



A user-friendly, secure, robust, reliable and trusted service to share and store your research data *adding value to your research data by assigning Persistent Identifiers to ensure long-lasting access and reference.*

Deposit and release your data via the generic interface or select a community extension including specific metadata fields. Releasing your data implies that your deposited data can be referred to, therefore any changes should be reflected in new data uploads.

Share your data with others in a safe and trusted environment.

Do you belong to a scientific community? Brand and create your own community collection with specific metadata fields customized for your field.

## Step 01

Drag and drop files here

Select files

Stop upload

## Add basic details

Filename

EUDAT-DEL-WP5-DS 2 2-EUDAT Ea  
Services.pdf

Generic

Add more details?

Title

Description

Creator

Open Access

ON

Licence

Publisher

Publication Date

Tags

Linguistics

Language Code

Country/Region

Resource Type

Project Name

Quality

\* Indicates required field

## Step 02

Select a domain

Generic



Ontology



## Step 04

Deposit

# B2SHARE features



**B2SHARE**  
Store and Share Research Data

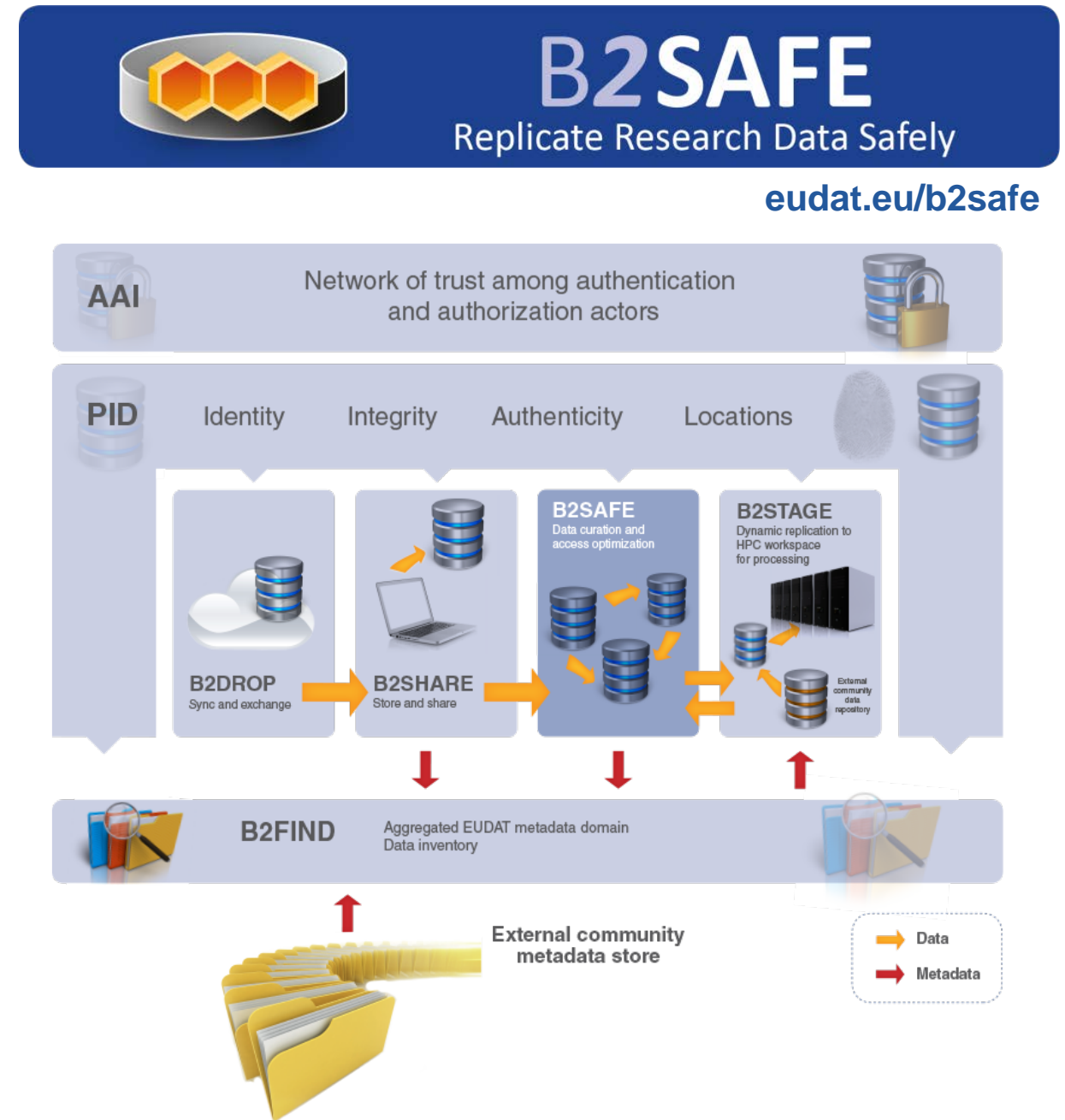
[b2share.eudat.eu](http://b2share.eudat.eu)

- 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.

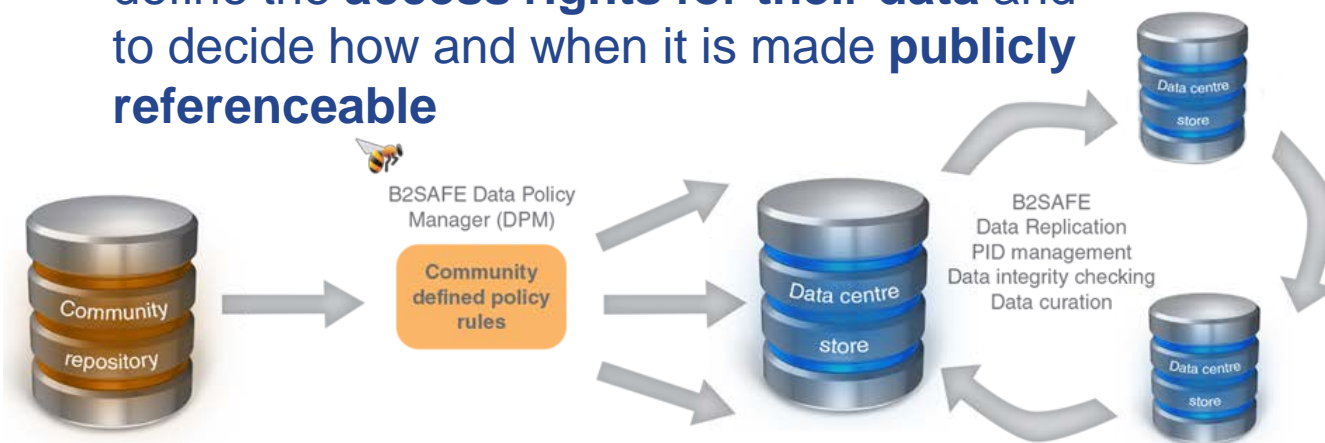
**A solution to:**

- 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 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**



## B2SAFE

Replicate Research Data Safely

[eudat.eu/b2safe](http://eudat.eu/b2safe)

- 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**



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:

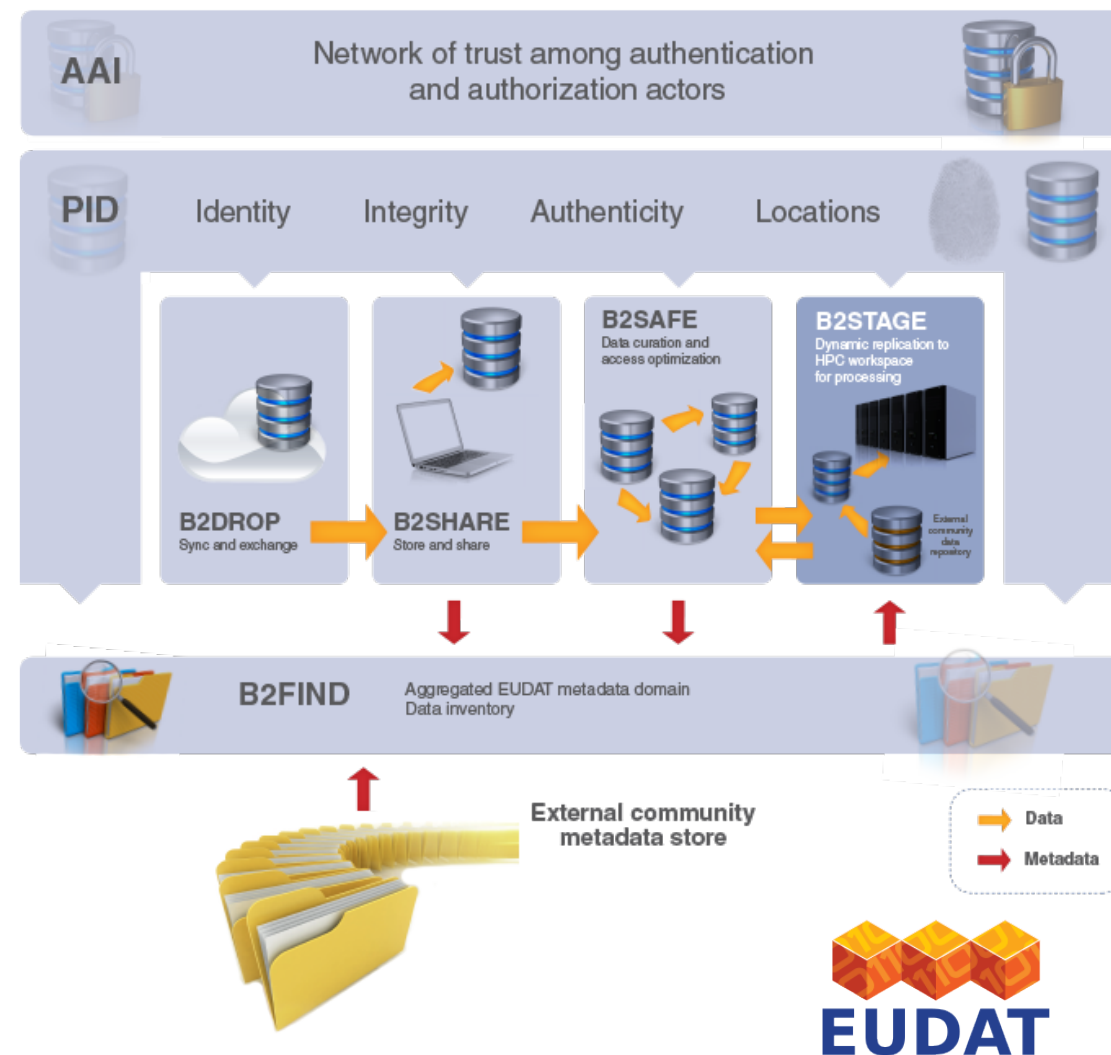
- Transfer large data collections from EUDAT storage facilities to **external HPC facilities for processing**
- In conjunction with B2SAFE, **replicate community data sets, ingesting them** onto EUDAT storage resources for long-term preservation
- Ingest computation results** into the EUDAT infrastructure
- Access data through a **RESTful HTTP interface** (in progress)



# B2STAGE

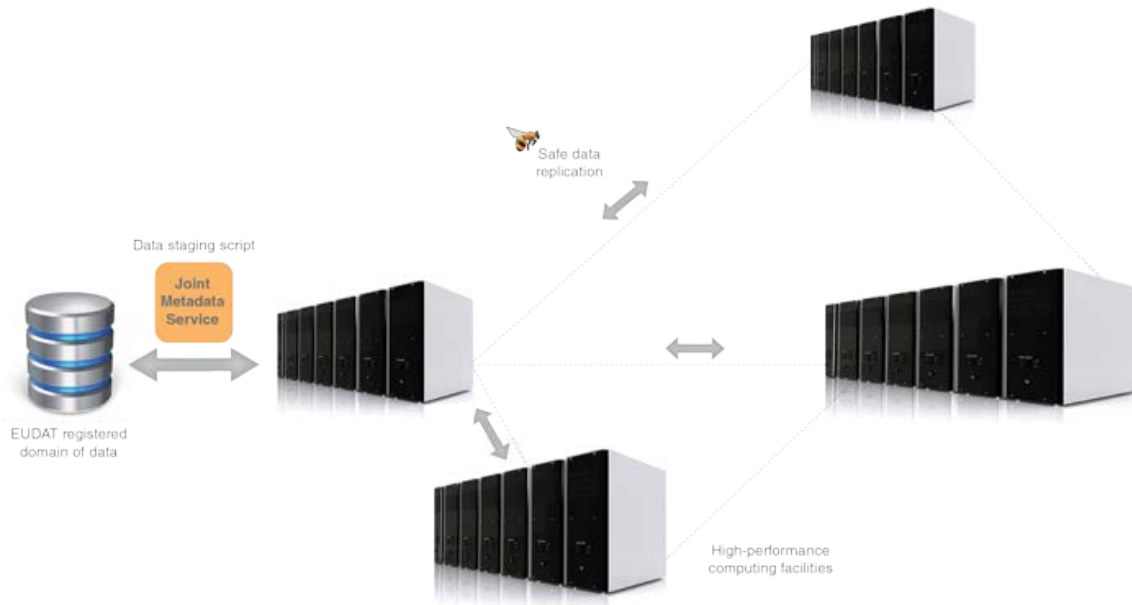
Get Data to Computation

[eudat.eu/b2stage](http://eudat.eu/b2stage)



# 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**



**B2STAGE**  
Get Data to Computation

[eudat.eu/b2stage](http://eudat.eu/b2stage)

- **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

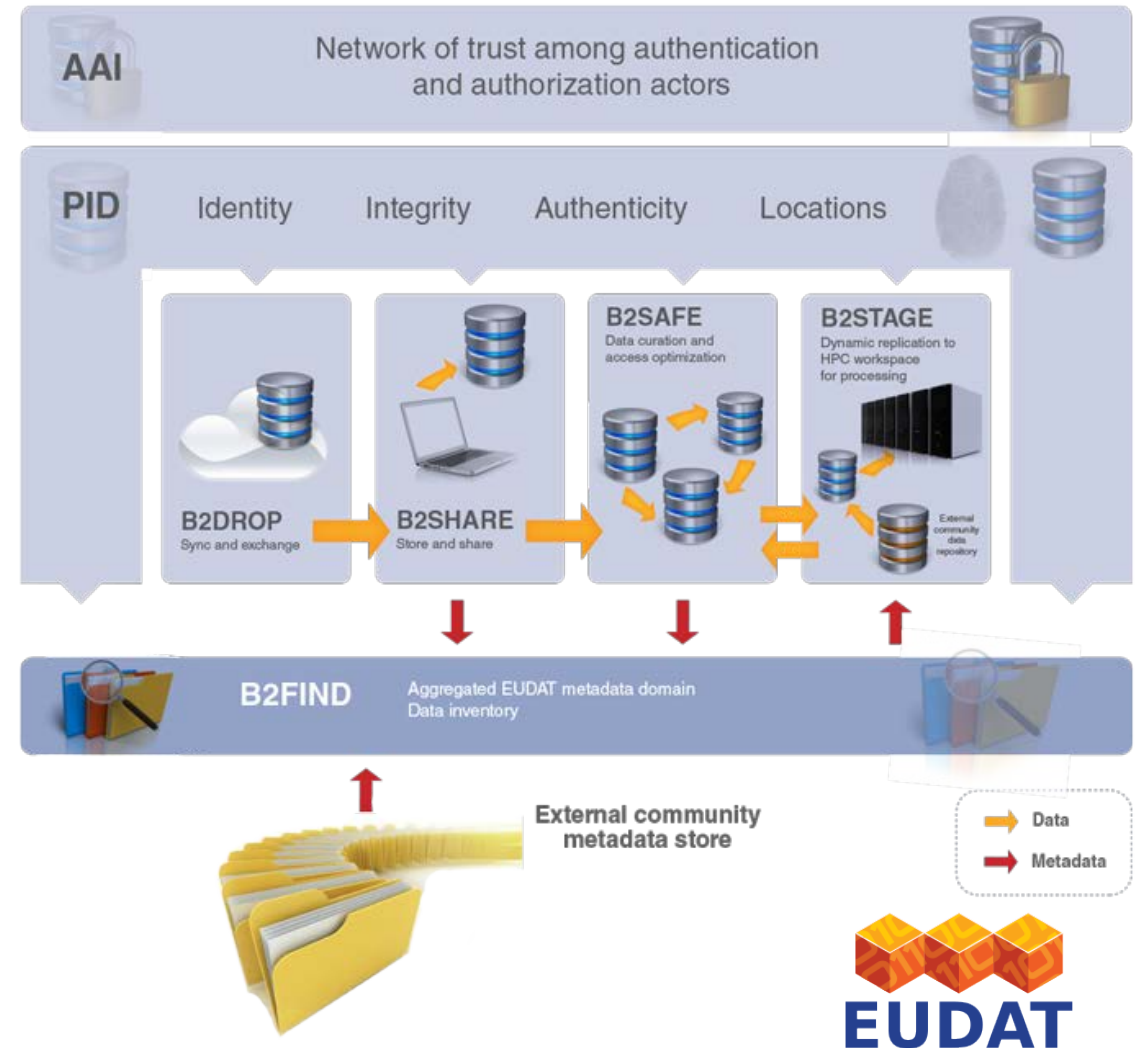
B2FIND is a simple, user-friendly **metadata catalogue of research data collections** stored in EUDAT data centres and other repositories.

**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](http://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**  
Find Research Data

[b2find.eudat.eu](http://b2find.eudat.eu)

Home / Datasets / Foraminifera abundance of sediment core MD81-LC07

Dataset extent



Map data © OpenStreetMap contributors  
Tiles by MapQuest

Dataset

Related

### Foraminifera abundance of sediment core MD81-LC07

Hayes, Angela (2008): Foraminifera abundance of sediment core MD81-LC07, doi:10.1594/PANGAEA.407822

Data and Resources

This dataset has no data

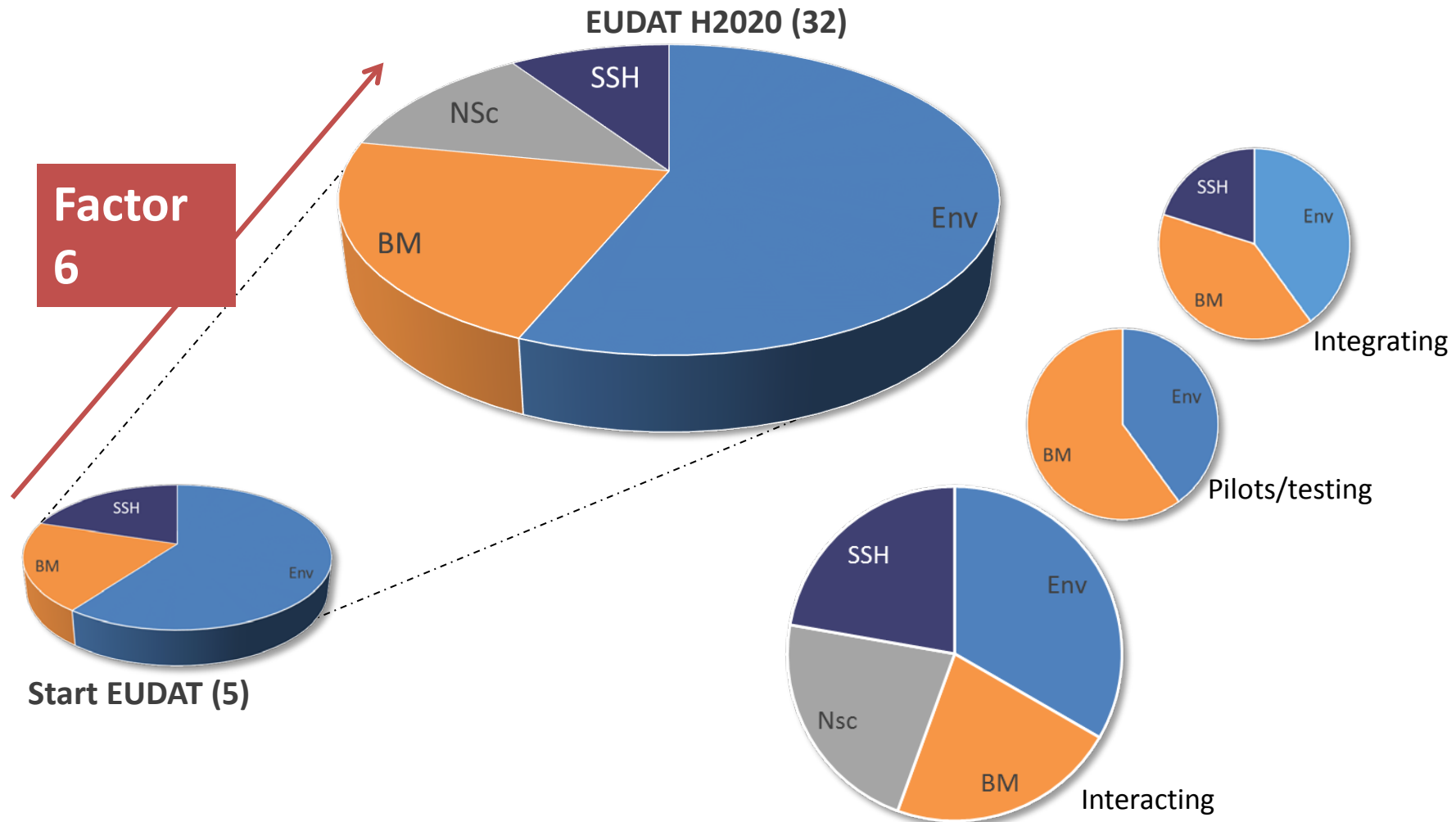
Globigerina bulloides

Additional Info

Field	Value
Source	<a href="http://dx.doi.org/10.1594/PANGAEA.407822">http://dx.doi.org/10.1594/PANGAEA.407822</a>
Author	Hayes, Angela (2008)
Version	68c7252d8134ebb8a19ce88a84e98ade
Discipline	Species
GeographicCoverage	Informationnotprovided
MetadataAccess	<a href="http://metadata.gbif.org/catalogue/OAIHandler?verb=GetRecord&amp;metadataPrefix=eml&amp;identifier=oai:metadata.gbif.org:eml/portal/2290.xml">http://metadata.gbif.org/catalogue/OAIHandler?verb=GetRecord&amp;metadataPrefix=eml&amp;identifier=oai:metadata.gbif.org:eml/portal/2290.xml</a>
Origin	PANGAEA - Publishing Network for Geoscientific and Environmental Data
PublicationYear	2008
TemporalCoverage:BeginDate	1995-01-01
TemporalCoverage:EndDate	1995-01-01



# Community Outreach & Service take Up

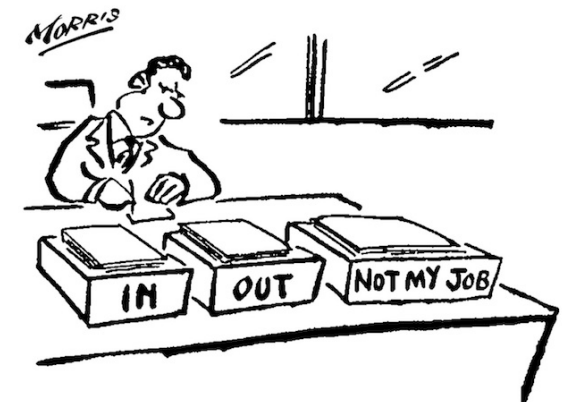




WHAT IS COMING NEXT?

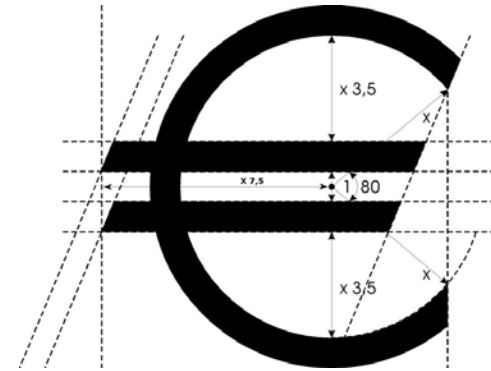
# Policies / Data Access and Reuse

- 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!

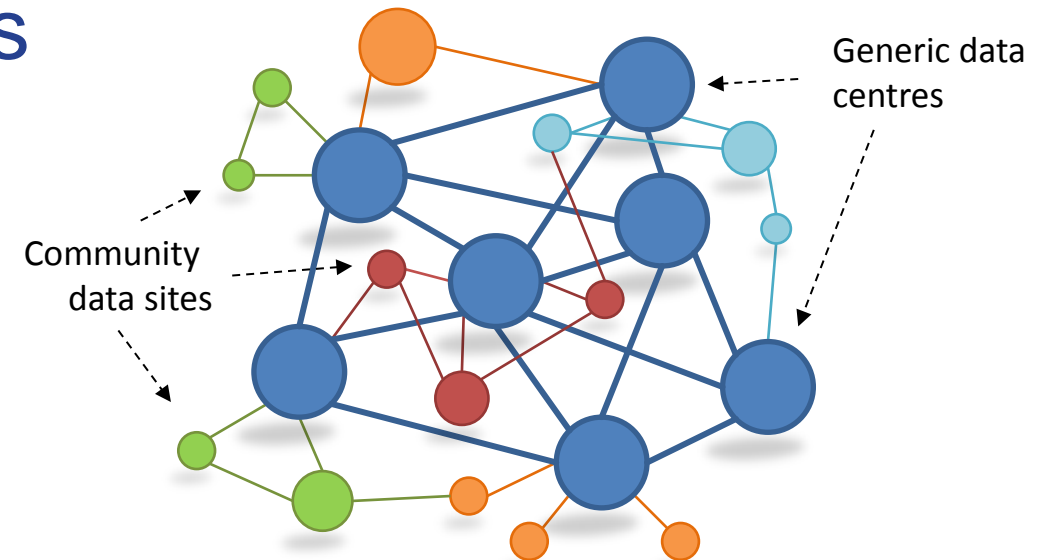
- Research Infrastructures → CDI users

- Integration of EUDAT components within RIs
- Governance
- Business models



- EUDAT nodes → CDI providers

- Operational Level Agreements (OLAs)
- Service level policies
- Funding & legal models



# Partnerships!

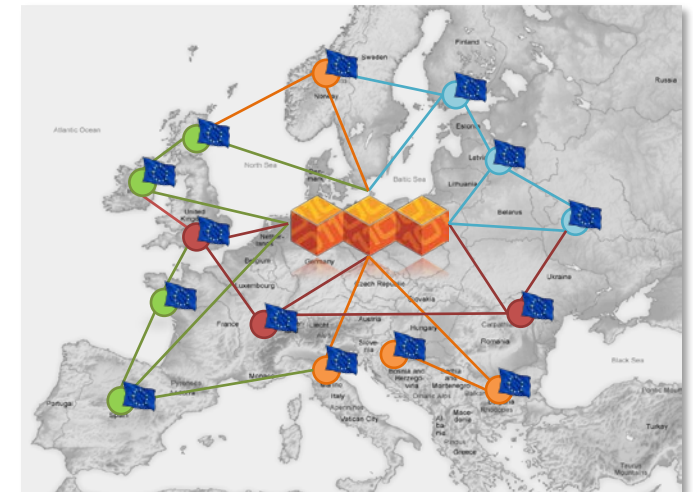
- European E-Infrastructures

- Cross-Infrastructure services/e-Infra Commons
- Policies
- Business models



- National initiatives

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





Thank you