



EUDAT

Data Services for Research “The Story”



Per Öster
Director, Research Infrastructures
CSC – IT Center for Science Ltd





Why EUDAT?

THE ISSUE

Research - Collaborations - Infrastructures

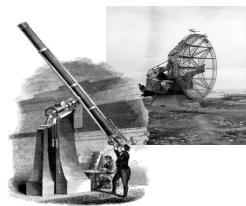
Research Infrastructure trends:

- Internationalisation
- Professionalization
- Increasing dependence on ICT
- The data deluge



European Ris:

- Around 500
- € 100 billion investment





Pan-European RI data challenges requires pan-European solutions

- All Research Infrastructures are facing data challenges

- How 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)





First some history

THE BACKGROUND

History of the EUDAT concept

- Early consortium of 13 European partners (mostly service providers with expertise in HPC, grid computing, networking and data storage)
- EUDAT has its origins in the work of the PARADE (Partnership for Accessing Data in Europe) initiative



- **PARADE White Paper** (September 2009) defining a Strategy for a European Data Infrastructure that should be persistent, multidisciplinary, and based on the need of user communities

Strategy for
a European Data Infrastructure

WHITE PAPER
28.09.2009

Editors

Ernesto Kuhl, C2C

Christof Dierker, C2C

Stefan Thrunert, C2C

Andreas Henschel, C2C

Andreas Henschel, C2C

Frank Weitzberg, C2C

The concept of a shared pan-European infrastructure was then supported and further elaborated by a number of policy and experts bodies:

- **e-IRG and ESFRI:** e-IRG Blue Paper (September 2010) recommends "to identify and promote common (long term) data related services across different RI"
- **High Level Expert Group (HLEG)** report on Scientific Data: "Riding the Wave: How Europe can gain from the rising tide of scientific data" (October 2010) calls for a "Collaborative Data Infrastructure" for scientific data, that supports seamless access, use, reuse, and trust of data

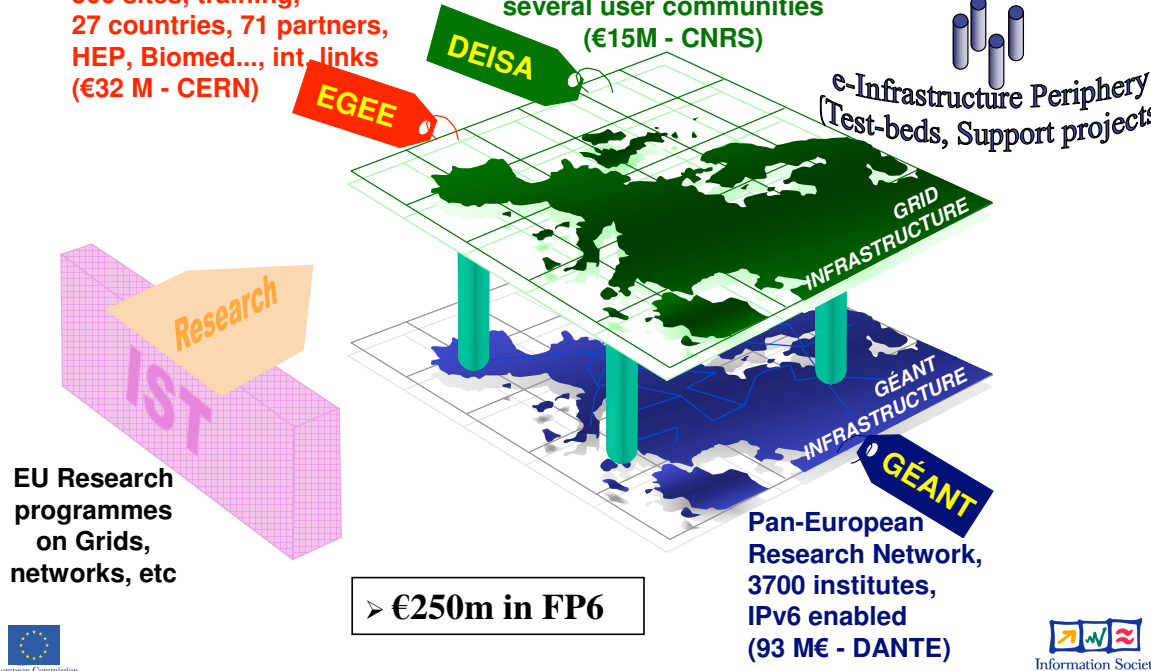
EC e-Infrastructure Vision (~2005)

Strategic building blocks of the e-Infrastructure today

production quality grid,
20000 CPUs, ~5PB storage,
500 sites, training,
27 countries, 71 partners,
HEP, Biomed..., int. links
(€32 M - CERN)

grid of EU supercomputers
networked at Gbps, focus on
global filing system, >120 Tflop/s,
several user communities
(€15M - CNRS)

e-Infrastructure Periphery
(Test-beds, Support projects)



EU Research
programmes
on Grids,
networks, etc

Pan-European
Research Network,
3700 institutes,
IPv6 enabled
(93 M€ - DANTE)

> €250m in FP6



* Slide by European Commission, DG CNECT e-Infrastructure unit Kyriakos Baxevinidis

Building the Generic Data Infrastructure Layer

e-Infrastructures Vision

empower research communities through ubiquitous, trusted and easy access to services for data, computation, communication and collaborative work



Scientific facilities, research communities

Source: European Commission

Evolution

National



EUDAT



Future European Grid

PRACE

GÉANT

~2000

Global

2008

Today

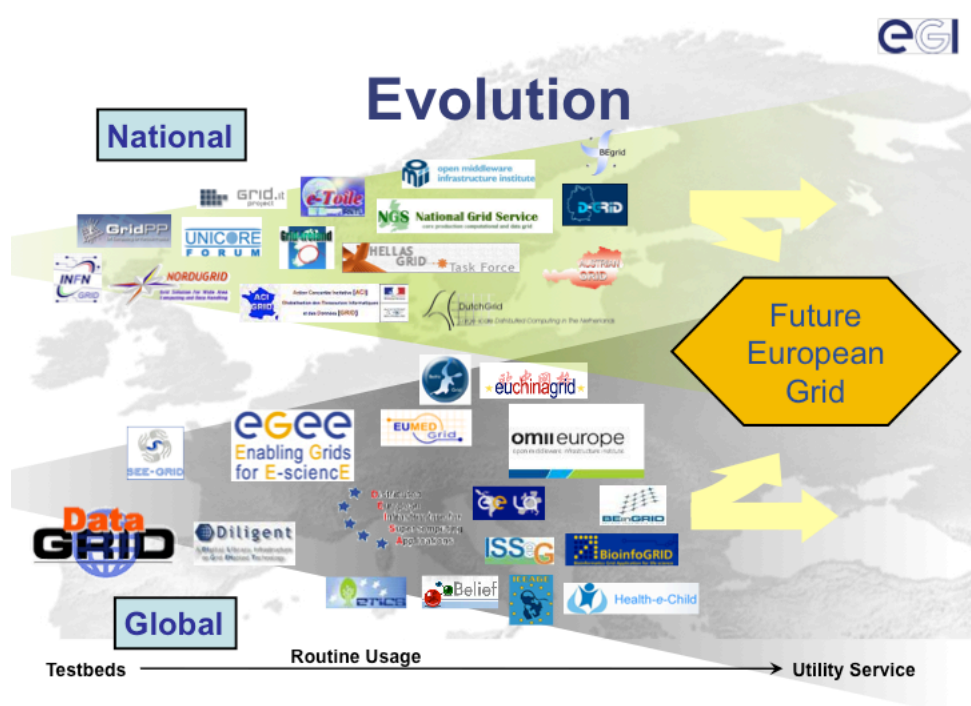
Timeline

Routine Usage

Testbeds

Utility Service





~2000

2008

Today

Timeline

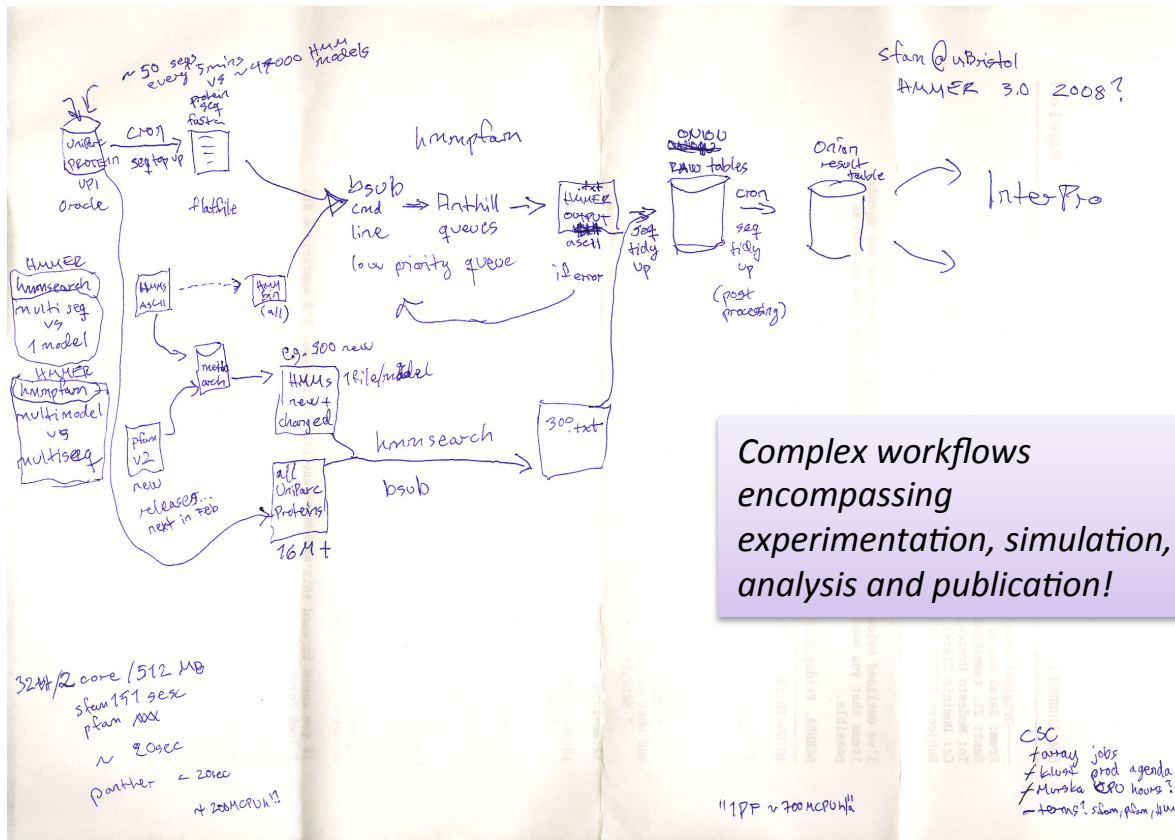




About the big picture, EUDAT and EUDAT 2020

THE ANALYSIS

Complex Collaborations – Complex Workflows



Complex workflows encompassing experimentation, simulation, analysis and publication!



Data is the asset

Research Infrastructures – Where is it going?

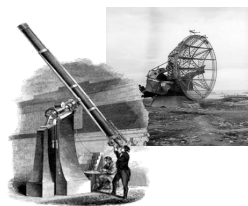
Research Infrastructure trends:

- Internationalisation
- Professionalization
- Increasing dependence on ICT
- The data deluge



European Ris:

- Around 500
- € 100 billion investment



middle age

19th century

20th century

21st century

We need to promote synergy

- The worst case scenario: 500 RI with 500 incompatible self-made ICT and data management solutions
- What can we do to promote collaboration and re-use of e-infrastructure?

EVERY RI NEEDS TO
DEAL
WITH DATA
MANAGEMENT



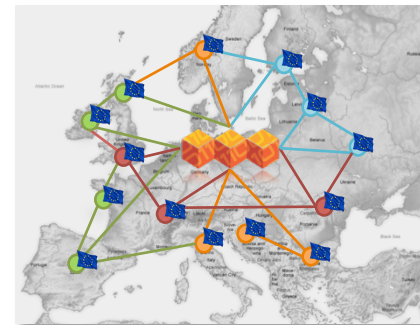
Integrating e-Infra and RIs

- Research Infrastructures → CDI users, partners & stakeholders
- It is not only about developing technical solutions, but also about defining the right partnership model
 - The model must necessarily take into account existing arrangements within pan-European research communities (organisational structure, funding schemes, business models, etc.)



Bridging National and European solutions

- Research and infrastructures are still funded at national levels → need to make sure that the solutions remain interoperable
- EUDAT provides a European gateway to national centers and a European extension to national solutions
- Making national resources more available and visible
 - Access to European resources through national catalogues
 - Making visible valuable national collections through EUDAT



E-Infrastructure Commons

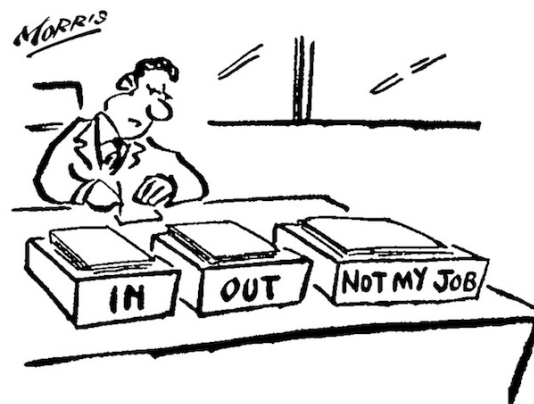


- Think about the users... and all these acronyms!
 - Users should have a “right” to a seamless access to network, data, and computing resources funded by public money
 - It is our role to make it as easy as possible for users → No one should care about e-Infrastructures as such
- Think global!
 - Solutions must also be thought at global level (RDA)
 - Cross-continent collaboration is a must (e.g. NDS, ANDS, etc.)

Roles and Responsibilities

Agreeing on roles and responsibilities

- Between RIs and e-Infrastructures
- Between e-Infrastructures (EUDAT, EGI, PRACE, OpenAIRE, etc.)
- Between national and European e-Infrastructures
- Between funders (EC/National)



Trust is key!

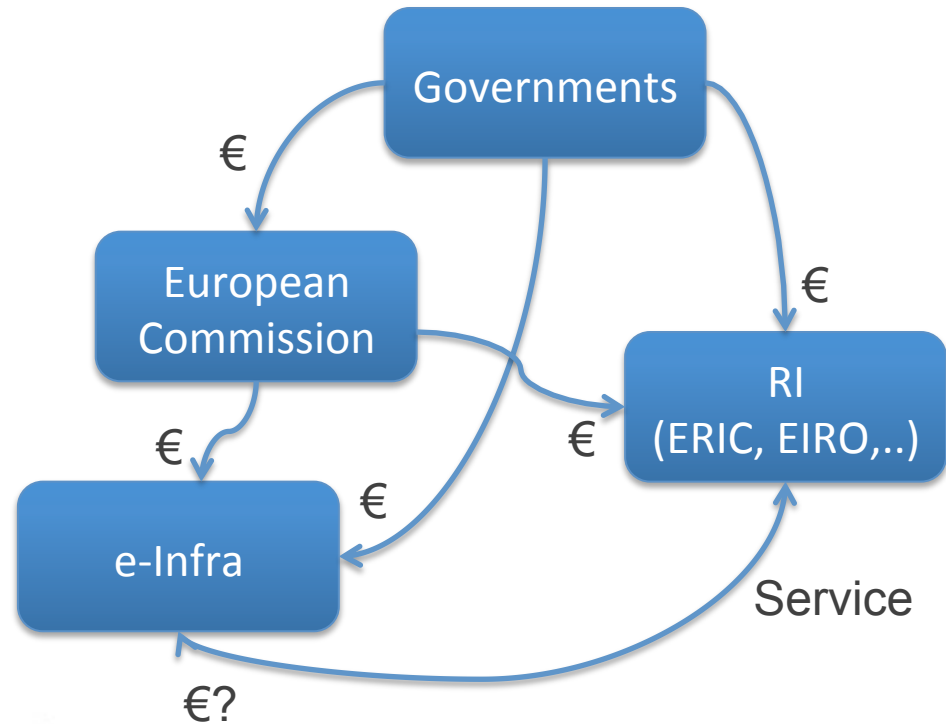


Customers and Flavour of the Currency



Follow the Money

- Funding of research
 - Funding streams
 - Spending restrictions
 - Procurement rules
 - Policy for RI use
 - Policy for e-Infra use



Sustainability?



	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Business planning skills	Initial business plan in place	Market awareness, business modelling	Establishing adaptability to market change	Formulated strategic planning process	Advanced business plans to cover short-term business strategy (3-5 years)
Business operational management skills	Basic facilities management	Appropriate human resources management	Establishing a suitable organisational structure	Settled, effective organisational arrangements	Professional management of facilities and human resources
Business development skills	Initial market research	Accurate market segmentation and development of value proposition	Business modelling including the planning of effective sales, distribution, promotion and brand development	Confirmed development of effective sales approach, systems and methods	Robust and adaptive sales and marketing strategies with achievable targets
Financial management skills	Basic financial management	Appropriate financial systems, including cost control	Provision of quality management measuring information	Professional cost financial management	Advanced financial strategy and facilities with relevant targets
Technical development skills	Minimum skills required to produce the initial service/ product	Higher value depth and breadth of skills/ product	Appropriate management skills for managing complex technical requests, including skills in networking	Manage capacity for site based operational growth	Settle on a framework for managing operational growth, including a number of managing strategy
Legal skills	All the guarantees of national legislation in place	Contracts for customers and suppliers	Processes are in place to address relevant legal issues	Settled management of legal issues in relation to the enterprise	Proactive measures are undertaken to manage relevant contractual and legal issues
Policy awareness	Basic research to understand government, local and other policies relevant to the enterprise	Market needs and their impact on research to maintain current operations of research project	Researcher well established to be involved in policy development	A network of contacts established to monitor relevant policy	Proactive measures are undertaken to manage relevant contractual and legal issues
Governance system	All the major governance systems are developed and operational	The need for appropriate systems of governance because time	Effective governance systems before the organisation starts to scale development	A formal governance system is established and effective in monitoring and leading	Proactive measures of the organisation and effectiveness of the governance system are monitored, showing effectiveness of structure adopted
Organisational structure/interdependence	Initial structure may be vulnerable, not optimal	Early growth may challenge resources of current structure	Structure is flexible enough to permit growth	Settled and effective organisational structure	Proactive measures of the organisation and effectiveness of the governance system are monitored, showing effectiveness of structure adopted

Knowledge Exchange sustainability index



EUDAT and EUDAT2020

THE SOLUTION

Consortium

26 European partners



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



KIT
Karlsruhe Institute of Technology



INGV



BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

epcc



CINECA

INES
Centre Informatique National
de l'Enseignement Supérieur

DKRZ
DEUTSCHES
KLIMARECHENZENTRUM



CSC



Red IRIS

Max-Planck-Institut für Meteorologie

JÜLICH
FORSCHUNGSZENTRUM

rzg
RECHEN-
ZENTRUM
GARCHING

Science & Technology Facilities Council

maatG



umweltbundesamt
ENVIRONMENT AGENCY AUSTRIA



SNIC

Trust-IT Trust-IT Services Ltd
Communicating ICT to markets

UNINETT sigma

SURF SARA

UCL

EUDAT

European Data

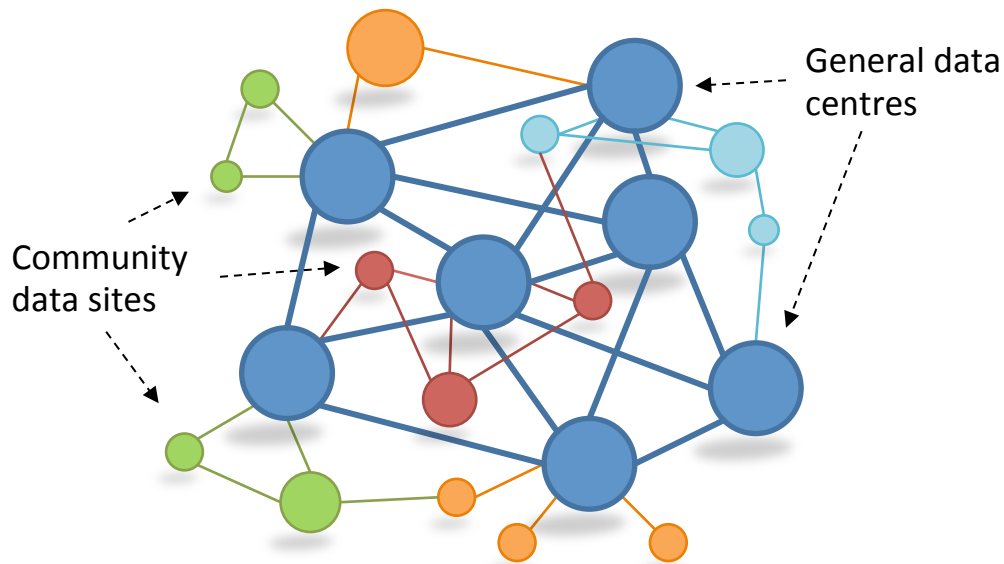


Researchers, citizens,
industry and society...

- Start date: 1st October 2011
- Duration: 36 Months
- Budget: 16.3 M€ (9.3M€ EC)
- EC Call: INFRA-2011-1.2.2
- Consortium: 25 partners from 13 countries
 - National data centers, technology providers, research
- Objectives:
 - Cost-efficient and high-quality CDI
 - Meeting users' needs in flexible and sustainable way
 - Across geographical and disciplinary boundaries

<http://www.eudat.eu>

Sustainable



- Independent and sustainable centres working within a common framework to develop shared services & policies
- EUDAT is about providing solutions in a federated environment
- Partnerships between legal entities relying on OLAs and SLAs

Service-Oriented



- 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 services address the full lifecycle of research data



B2SAFE
Replicate Research Data Safely



B2SHARE
Store and Share Research Data



B2FIND
Find Research Data



B2STAGE
Get Data to Computation



B2DROP
Sync and Exchange Research Data

New Proposal: EUDAT2020

- More services (by user needs)
- Strengthen collaboration between e-Infrastructures and Research Infrastructures
 - Within EUDAT → update plans, partnership models, etc.
 - Joint participations in projects (e.g. RI clusters)
- Foster interoperability between pan-European infrastructures
 - Pilot cross-infrastructure services
 - E-Infrastructure Commons Roadmap
- Bridge national and European solutions
 - Build on success stories and working examples
 - Establish partnership models for national stakeholders
- Make Europe better integrated!
 - Consolidate the infrastructure for academia and industry
 - Establish innovative international collaboration

References and Acknowledgments

- EGI Design Study, prof. Dr. Dieter Kranzmueller
 - Old e-infrastructure evolution slide
- European Commission DG CNECT e-Infrastructure unit
 - 2005 e-Infrastructure vision slide
- Dr. Tommi Nyrönen, Head of Finnish ELIXIR node
 - Napkin with workflow
- Dr. Damien Lecarpentier, EUDAT Project Director
 - Almost all other slides
- Knowledge Exchange Sustainability Index
 - <http://hdl.handle.net/11304/0b51c322-5938-11e4-81ac-dcbd1b51435e>
- Strategy for a European Data Infrastructure (PARADE White Paper)
 - ...



Thank you






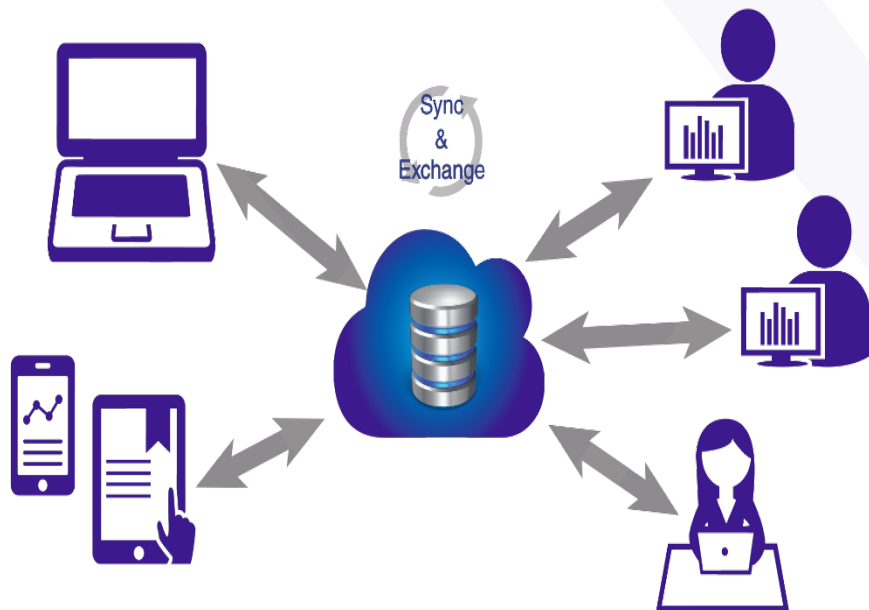
Details of the EUDAT services for the interested

ABOUT THE SERVICES

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:

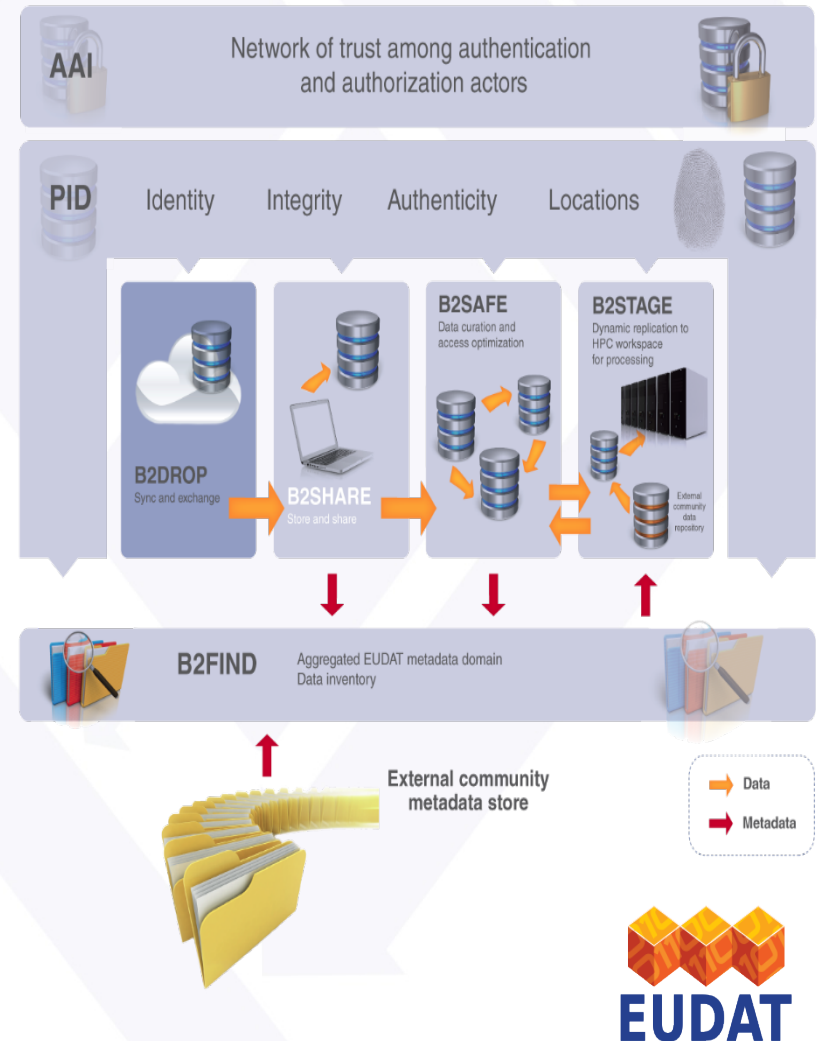
-  **Store and exchange** data with colleagues and team
-  **Synchronize** multiple versions of data
-  Ensure **automatic desktop** synchronization of large files



B2DROP

Sync and Exchange Research Data

b2drop.eudat.eu



B2DROP features

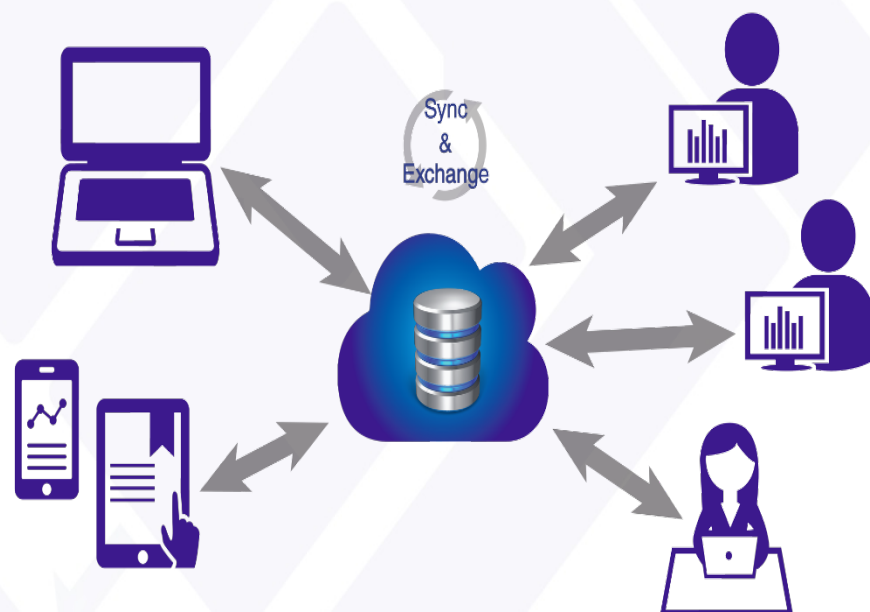


B2DROP

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**



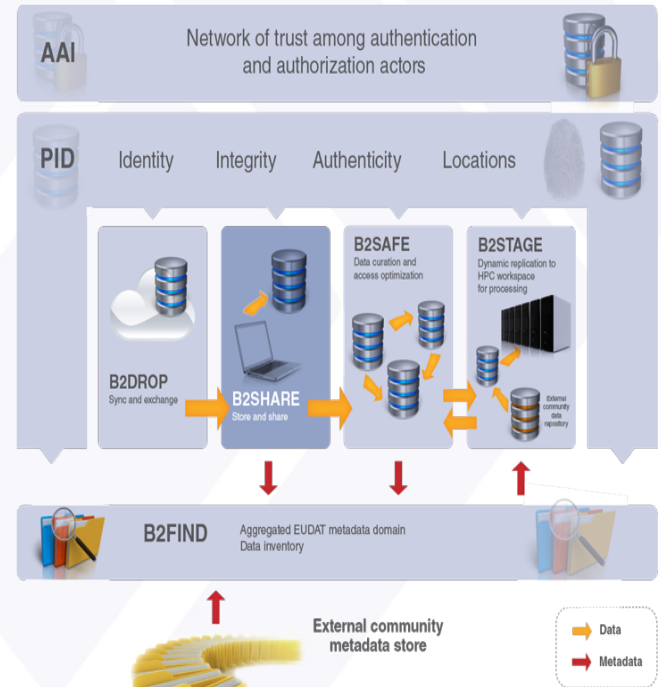
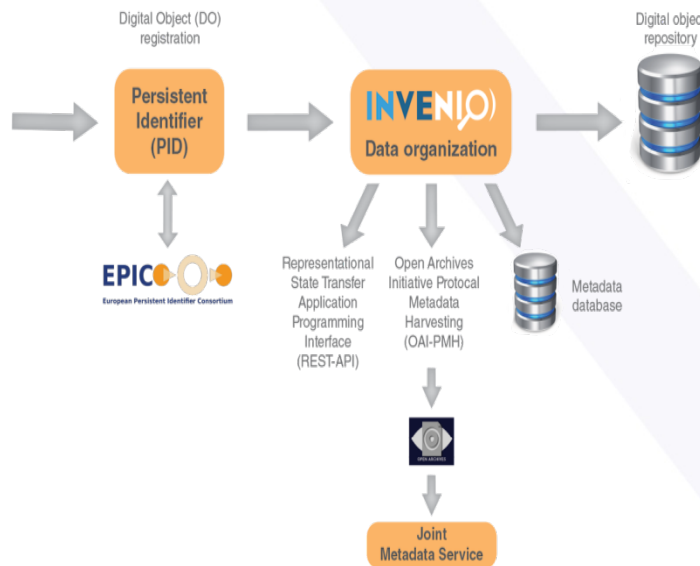
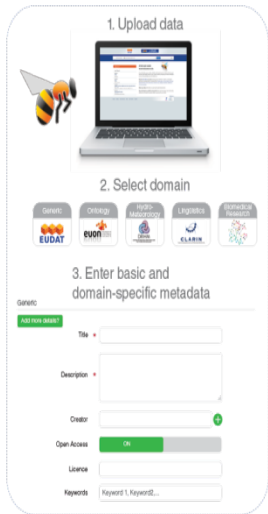
B2SHARE is a **user-friendly, reliable** and **trustworthy** way for researchers, scientific communities and citizen scientists to **store** and **share** small-scale research data from diverse contexts.

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.eudat.eu



B2SHARE features



B2SHARE
Store and Share Research Data





b2share.eudat.eu

- 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
- **data owner defines access policy**
- **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
- **EUDAT user support**
- **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:

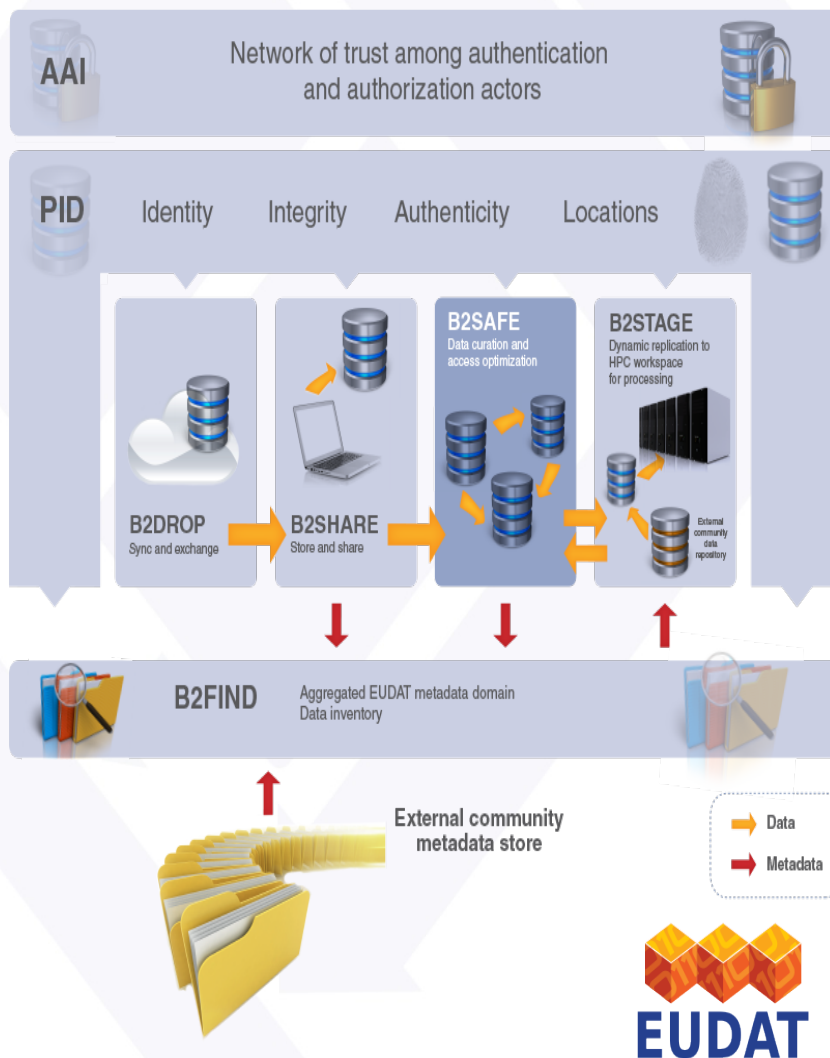
-  Provide an **abstraction layer which virtualizes large-scale data resources**
-  Guard against data loss in long-term **archiving and preservation**
-  **Optimize access** for users from different regions
-  Bring data **closer to powerful computers** for compute-intensive analysis



B2SAFE

Replicate Research Data Safely

eudat.eu/b2safe



B2SAFE features

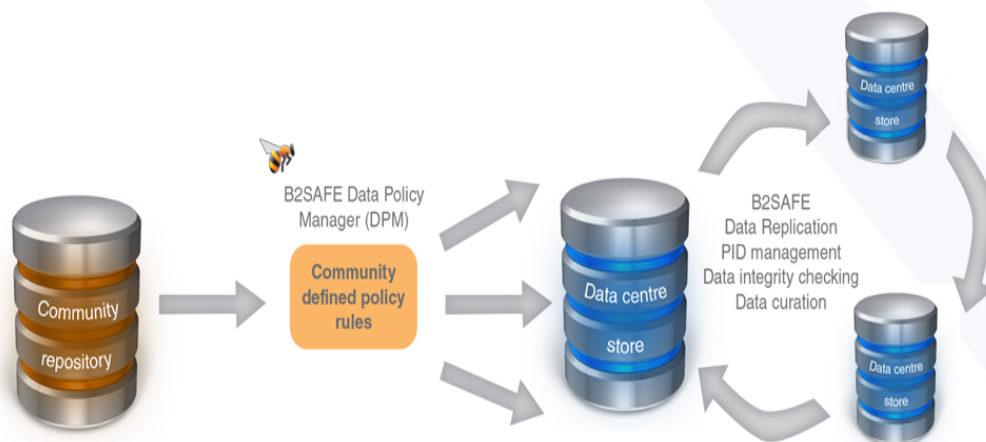


B2SAFE

Replicate Research Data Safely

eudat.eu/b2safe

- 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**



Who benefits



B2SAFE





Replicate Research Data Safely

eudat.eu/b2safe

- repositories lacking the **capacity and / or funding** to offer **reliable storage and access services** over longer periods of time
- repositories **without adequate compute capacity** for **data-intensive computational services** based on their data
- **data producers** who need to be sure that **trusted centres are taking care of their data**
- consumers wishing to access **optimized services on data sources of interest to them**
- consumers who wish to **apply interdisciplinary data-intensive methods** using data collected from **various communities**

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:

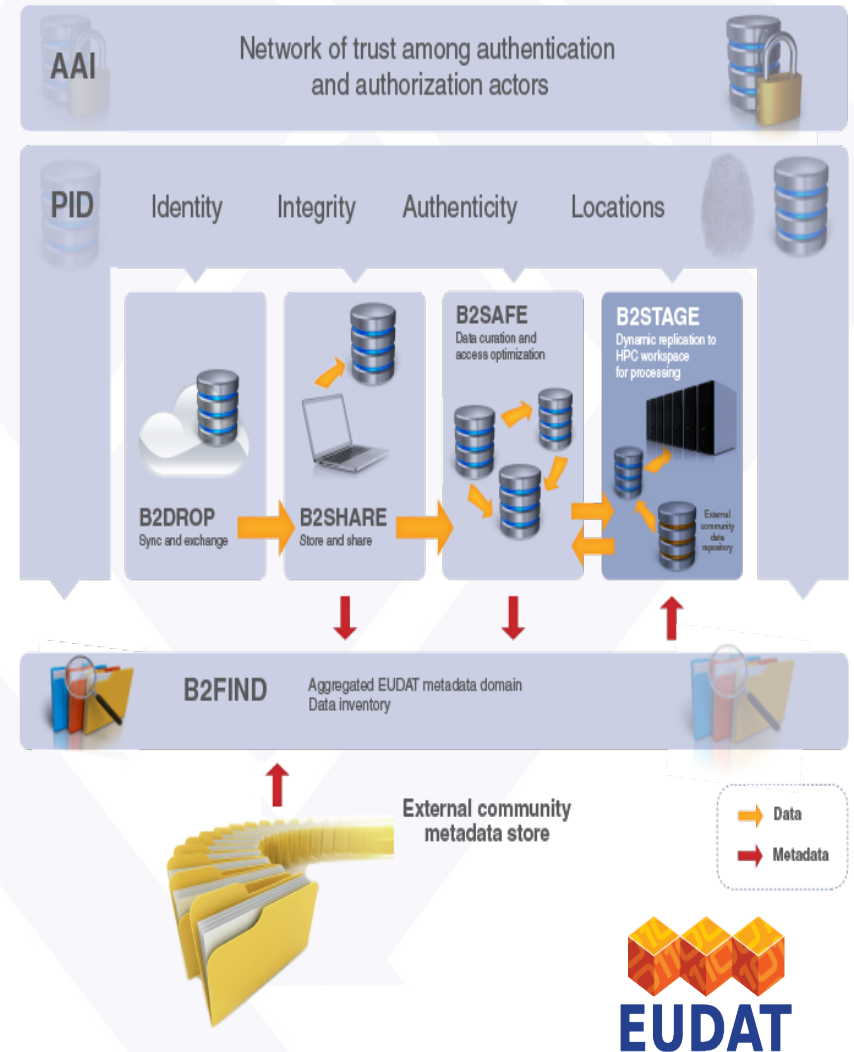
-  **Transfer** large data collections from EUDAT storage facilities to **external HPC facilities for processing**
-  In conjunction with B2SAFE, **replicate community data sets, ingesting them** onto EUDAT storage resources for long-term preservation
-  **Ingest computation results** into the EUDAT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



B2STAGE

Get Data to Computation

eudat.eu/b2stage



B2STAGE features

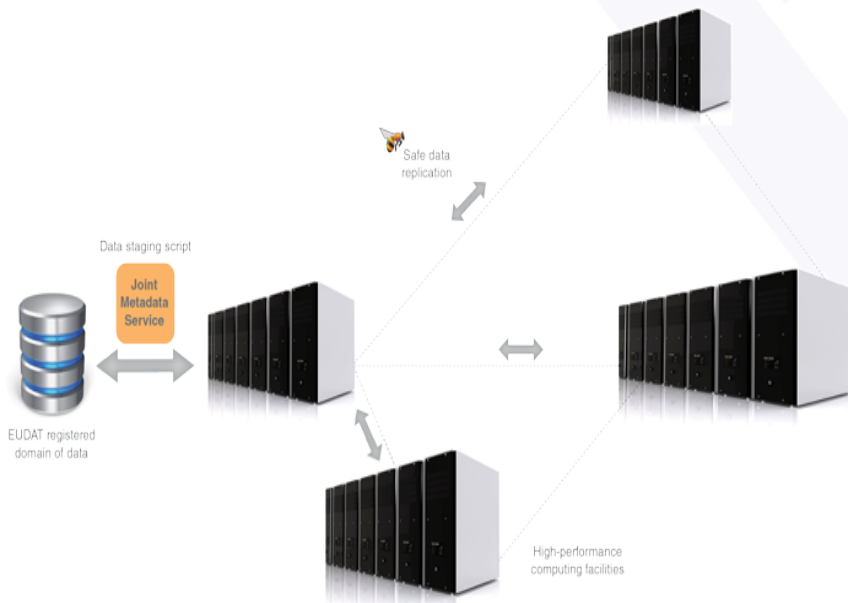
- an extension of the **B2SAFE** and **B2FIND** services, which allow users to store, preserve and find data
- **data-staging script** facilitates staging, ingestion and retrieval of persistent identifier (PID) information of transferred data
- service available **to all registered researchers and interested communities**



B2STAGE
Get Data to Computation


eudat.eu/b2stage

- **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

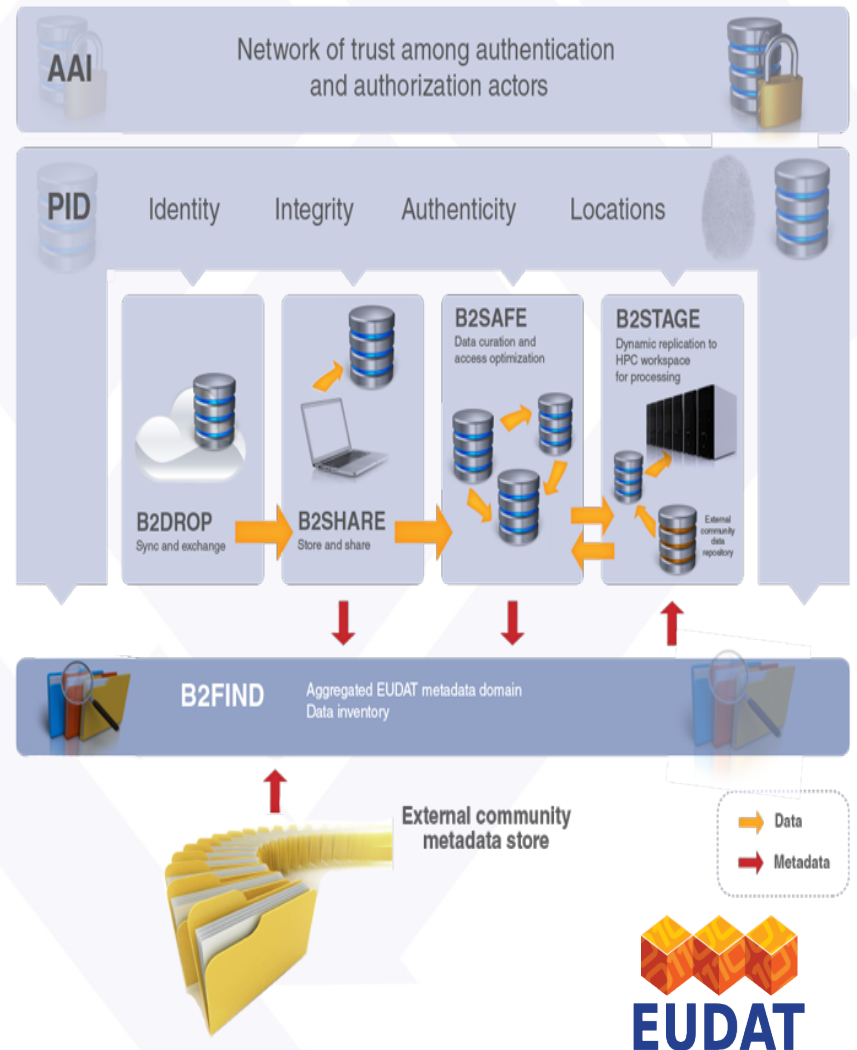
 Browse through collections using **standardized facets**



B2FIND

Find Research Data

b2find.eudat.eu



B2FIND features



B2FIND
Find Research Data

b2find.eudat.eu

- based on a **comprehensive joint catalogue of EUDAT services and external metadata**
- **metadata is mapped onto standardized facets**
- **supports faceted, geospatial and temporal metadata searches**
- allows users **to search and browse datasets via keyword searches**
- 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**
- 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**

