

EUDAT - Open Data Services for Research

Per Öster

05.03.2015



CSC-IT CENTER FOR SCIENCE

CSC at a Glance

- Founded in 1971 as a technical support unit for Univac 1108
- Connected Finland to the Internet in 1988
- Reorganized as a company, CSC – Scientific Computing Ltd. in 1993
- All shares to the Ministry of Education and Culture of Finland in 1997
- Operates on a non-profit principle
- Facilities in Espoo, close to Otaniemi campus (of 15,000 students and 16,000 technology professionals) and Kajaani (data center in North-east Finland)
- Staff >250
- Turnover 2015 ~34 million euros





EUDAT

Open Data Services for Research



Per Öster
Director, Research Infrastructures
CSC – IT Center for Science Ltd
per.oster@csc.fi
ECMWF, 05.03.2015

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 offers a pan-European solution

- Providing a set of generic services to help RIs managing their growing amount of data
- Providing these services across communities to ensure minimum level of interoperability
- Linking community specific repositories to the largest European scientific data and HPC centers → Collaborative Data Infrastructure (CDI)



Data Centers and Communities

26 European partners



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



KIT
Karlsruhe Institute of Technology



BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

epcc

DKRZ
DEUTSCHES
KLIMARECHENZENTRUM



CINECA

INES
Centre Informatique National
de l'Enseignement Supérieur



Red IRIS

Max-Planck-Institut für Meteorologie

CSC

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 Jigma

SURF SARA

UCL



User Forums + 30 Communities



GBIF



1st User Forum
7-8 March 2012, Barcelona



VERGE



European Integrated Data Archive



DRIHM
DISTRIBUTED RESEARCH INFRASTRUCTURE
FOR HYDRO-METEOROLOGY



DARIAH
Digital Research Infrastructure
for the Arts and Humanities



diva
Digital Infrastructure
for Visualisation and
Analysis



Geo-Seas



LTER europe



SCIDIP-ES
SCIENCE DATA INFRASTRUCTURE FOR PRESERVATION - EARTH SCIENCE



pandata europe



WLCG
Worldwide LHC Computing Grid



i marine



ICOS
integrated carbon observation system



CESSDA
Council of European Social Science Data Archives



ISIS
Science & Technology Facilities Council



elixir



ECRIN
EUROPEAN CLINICAL RESEARCH INFRASTRUCTURES NETWORK



aginfra
A data infrastructure for agriculture



MAPER



EUROVO
AIDA
Astronomical Infrastructure for Data Access



EMBRC
EUROPEAN MARINE BIOLOGICAL RESOURCE CENTRE

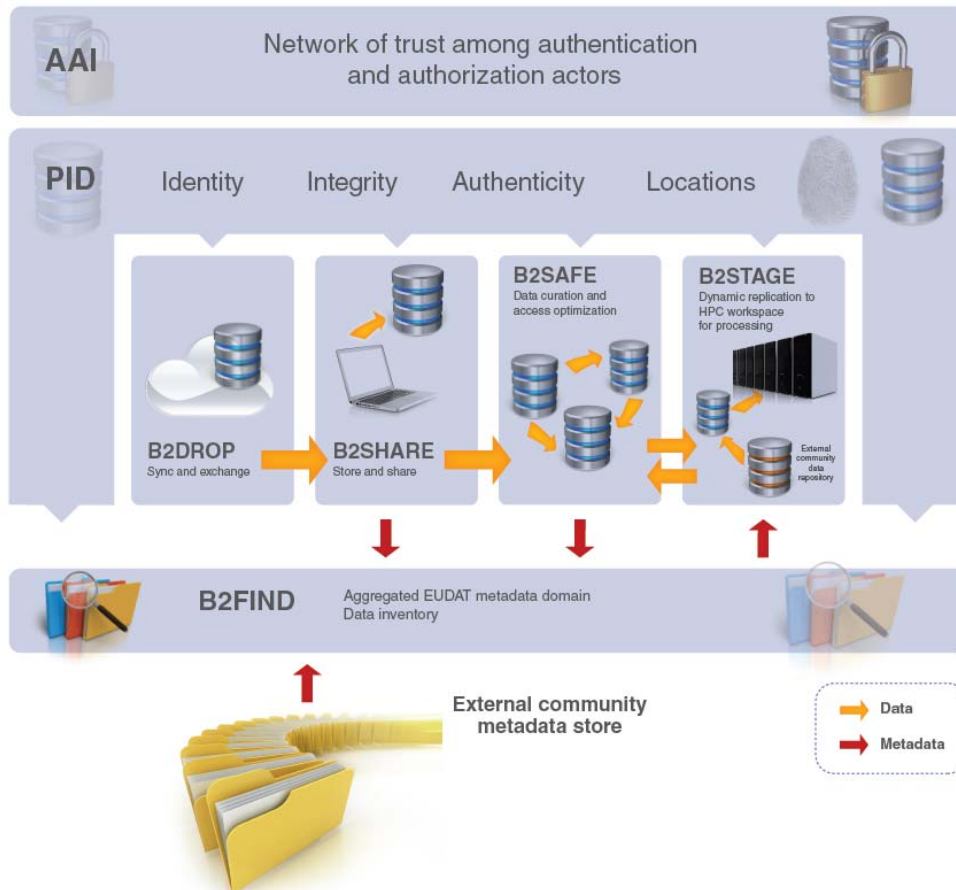


IAGOS



EUDAT




Services & Resources

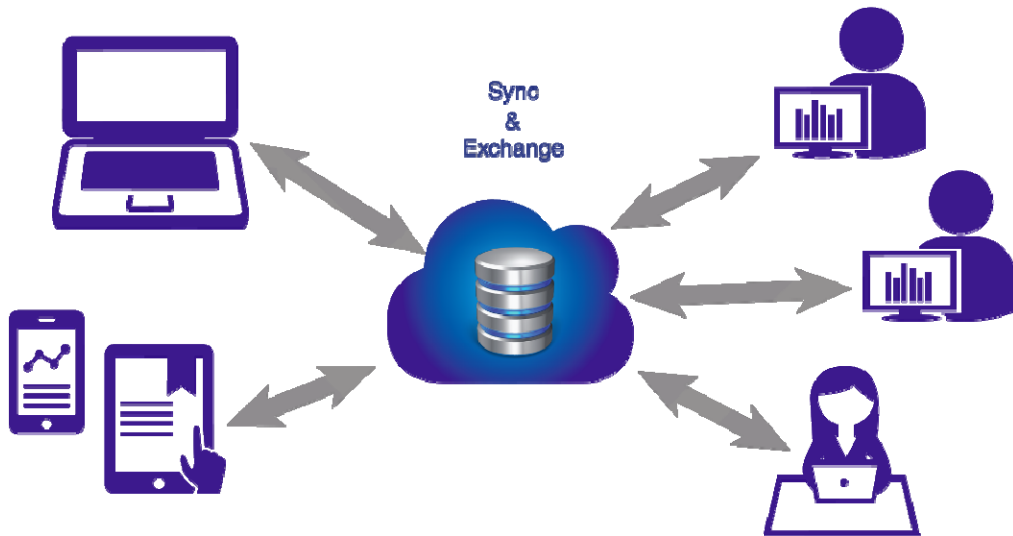


Covering both **access** and **deposit**, from **informal data sharing** to **long-term archiving**, and addressing **identification**, **discoverability** and **computability** of both **long-tail** and **big data**, EUDAT's services will **address the full lifecycle of research data**

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.

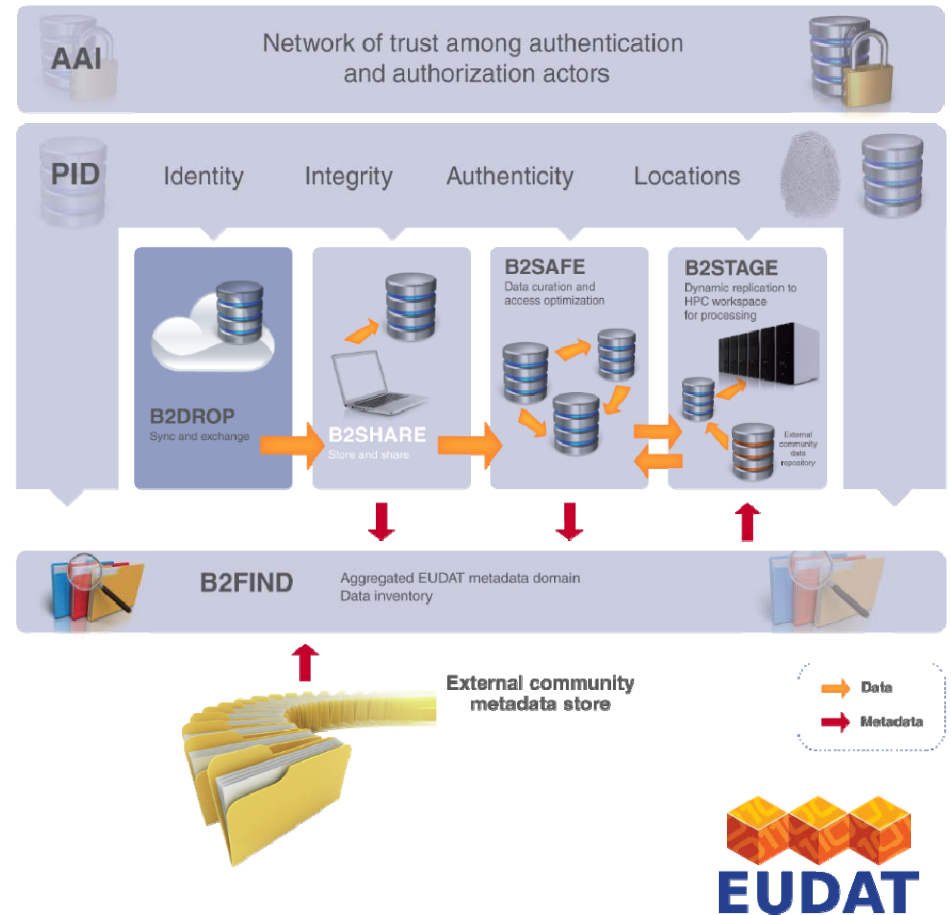
An ideal solution to:

-  **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






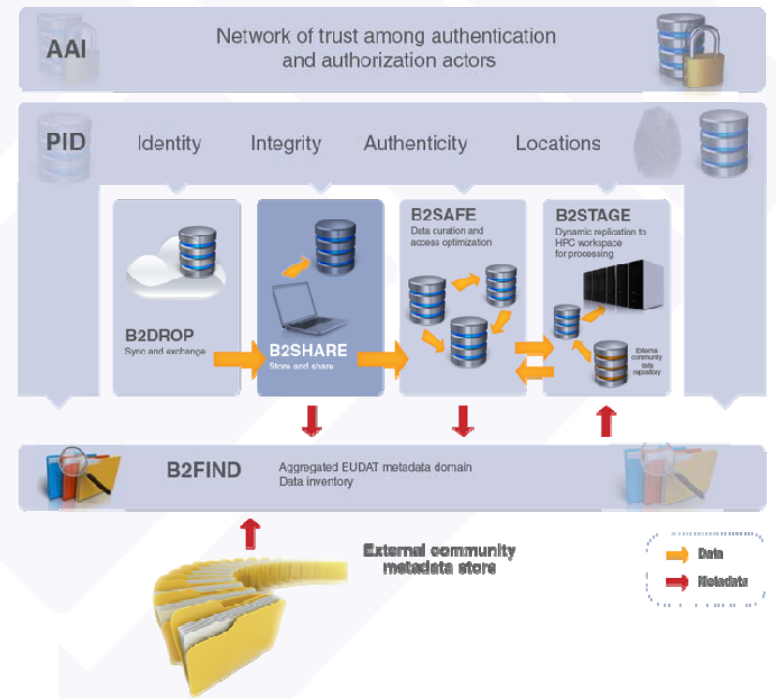
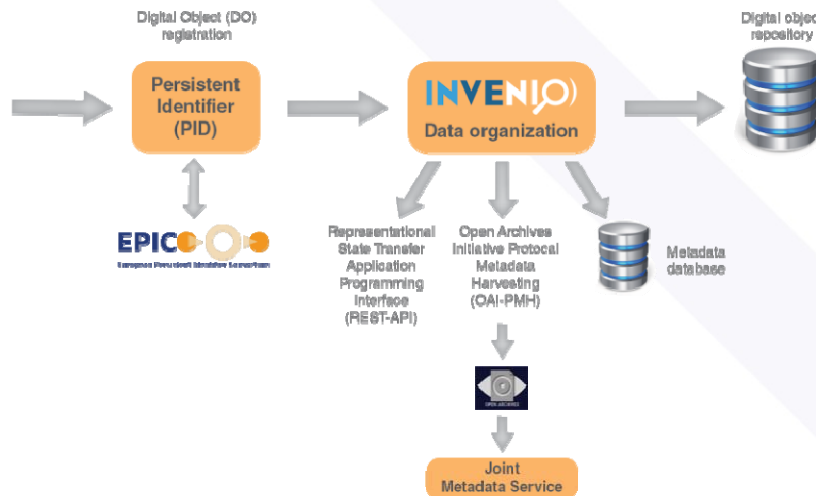
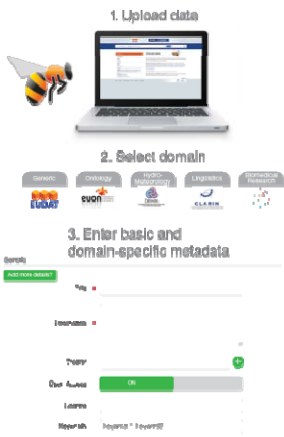
B2SHARE is a **user-friendly, reliable** and **trustworthy** way for researchers, scientific communities and citizen scientists to **store** and **share** small-scale research data from diverse contexts.



b2share.eudat.eu

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



Step 03

Information

Comments

Reviews

Usage statistics

EUDAT Presentation FOT-NET

Damien Lecarpentier

16 December 2014

<https://trng-b2share.eudat.eu>

Abstract: Presentation of EUDAT services at FOT NET workshop in Barcelona

Search 139 records fo

You are logged in as Damien.

Deposit >>>

Latest Deposits

2014-10-23
**PARADE, Strategy for a Eu
Paper**
by Kimmo Koski[...]
Strategy for a European Data Infrac

2014-10-21
Knowledge Exchange Sust
Output from the Knowledge Excha

2014-10-10
**MHD run for a single clust
ENZO-MHD at z=0.**
by Franco Vazza
HiFi files for 3D dataset for a clust

2014-10-09
ENZO-AMR data for cluste
by Franco Vazza
HiFi monolithic 256*3 files for a sim

Files ▾

Name	Date	Size	
EUDAT-FOTNET-16.12.14.pptx	16 Dec 2014	10.5 MB	Download

Export

Export as [BibTeX](#), [MARC](#), [MARCXML](#), [DC](#), [EndNote](#), [NLM](#), [RefWorks](#)

Metadata

PID:	11304/61e61764-8503-11e4-ac7e-860aa0063d1f
Publication:	https://trng-b2share.eudat.eu
Publication Date:	2014-12-16
Licence:	CC-BY
Uploaded by:	damien.lecarpentier@csc.fi
Domain:	generic
Checksum:	d51306c2b65e3f63ebc456f935d711fe28d627cf9a d1bee9f1206ae83d79b7c3

Rate this document:



(Not yet reviewed)

[Report abuse](#)

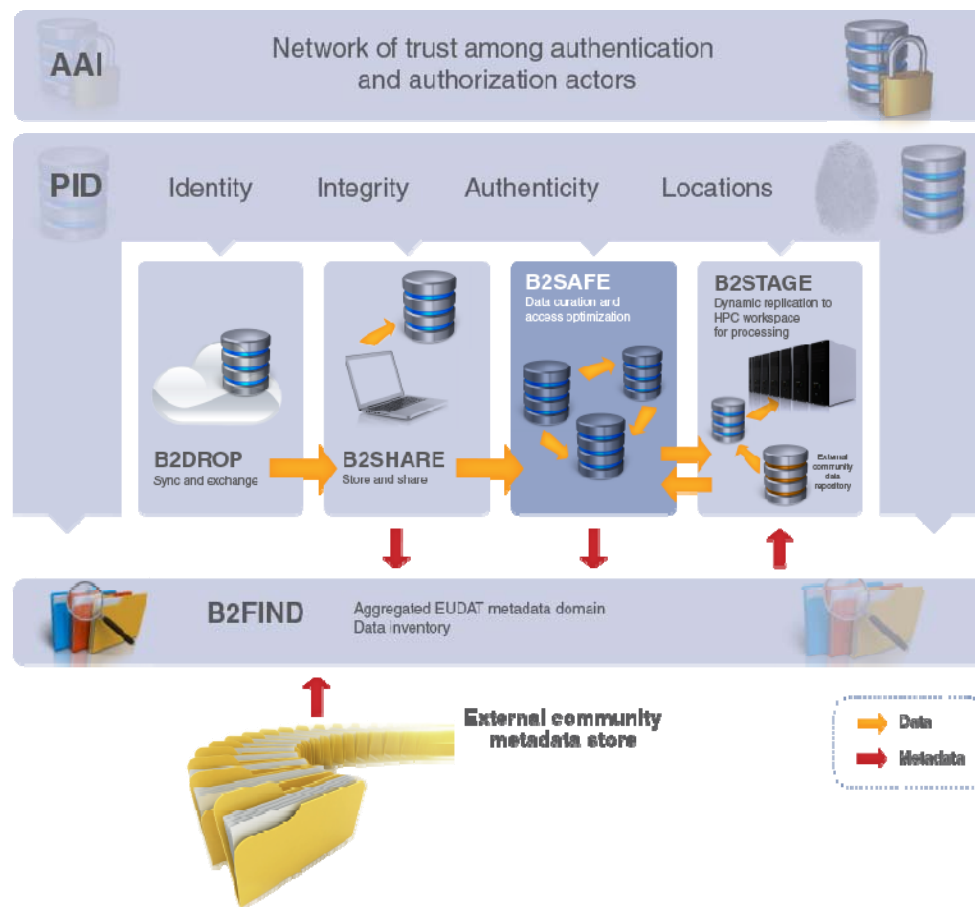
B2SAFE is a **robust, safe and highly available service** which allows community and departmental repositories to **implement data management policies on their research data** across multiple administrative domains in a trustworthy manner.

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



eudat.eu/b2safe



B2SAFE Features

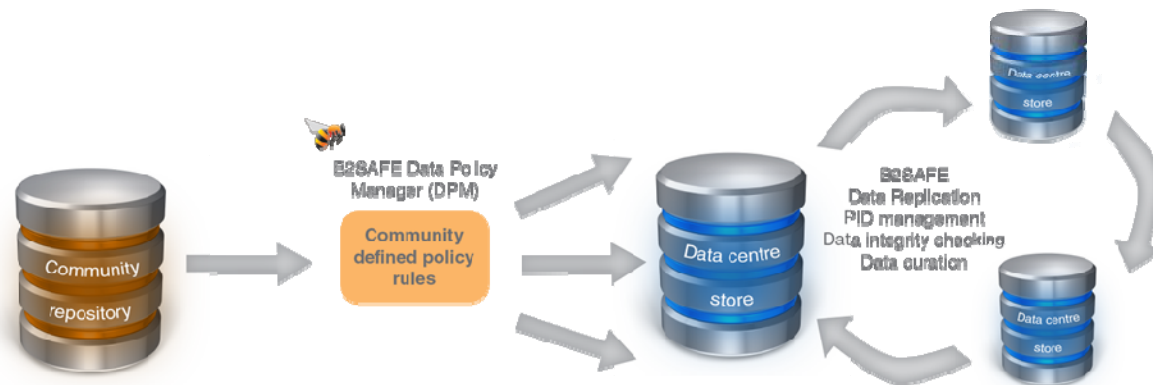


B2SAFE

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**



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:

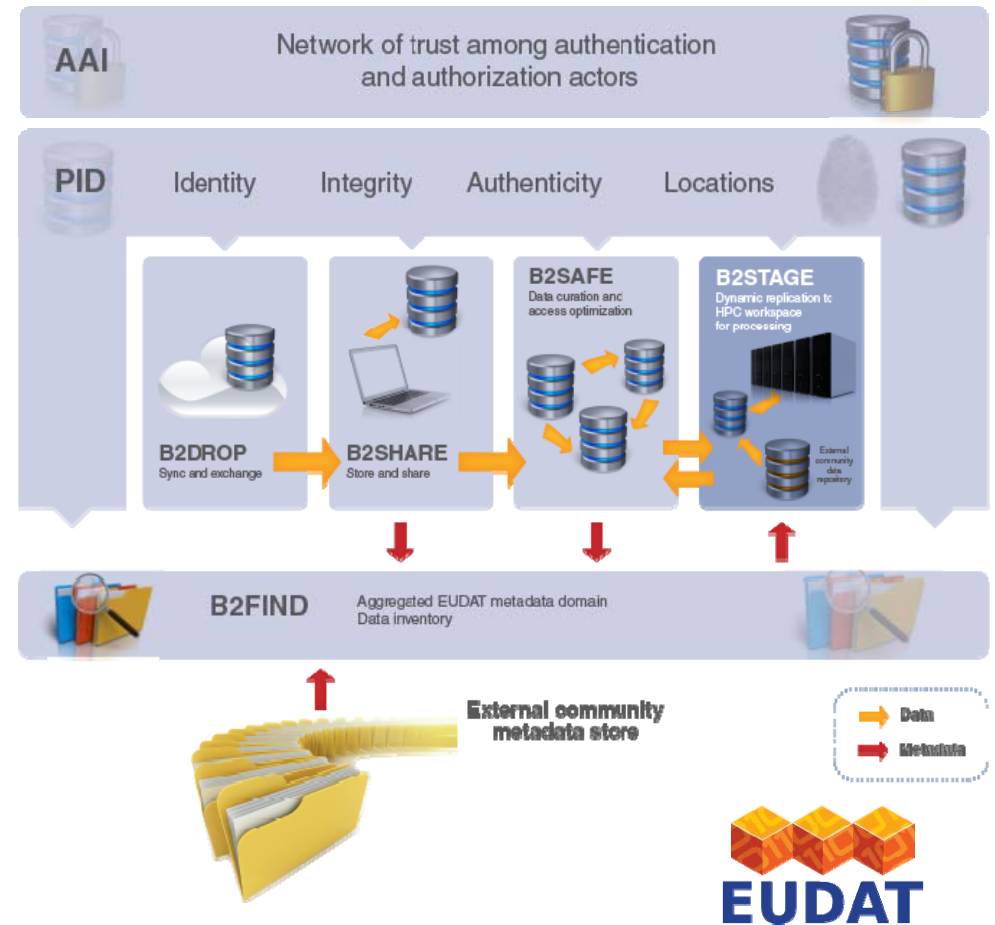
-  **Transfer** large data collections from EUDAT storage facilities to **external HPC facilities for processing**
-  In conjunction with B2SAFE, **replicate community data sets, ingesting them** onto EUDAT storage resources for long-term preservation
-  **Ingest computation results** into the EUDAT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



B2STAGE




Get Data to Computation

eudat.eu/b2stage



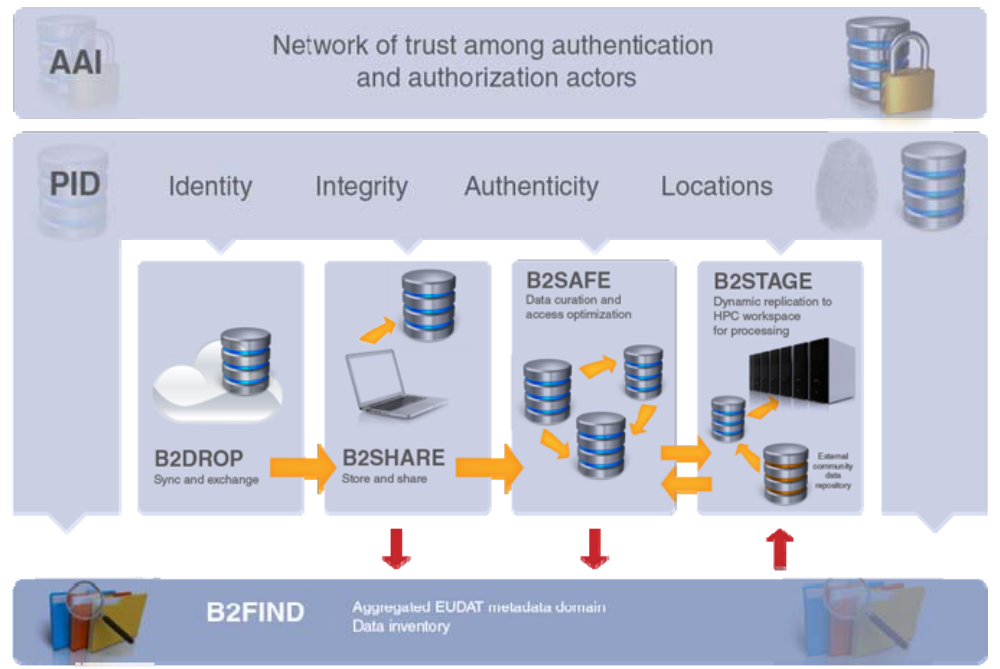
B2FIND is a simple, user-friendly **metadata catalogue of research data collections** stored in EUDAT data centres and other repositories.

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



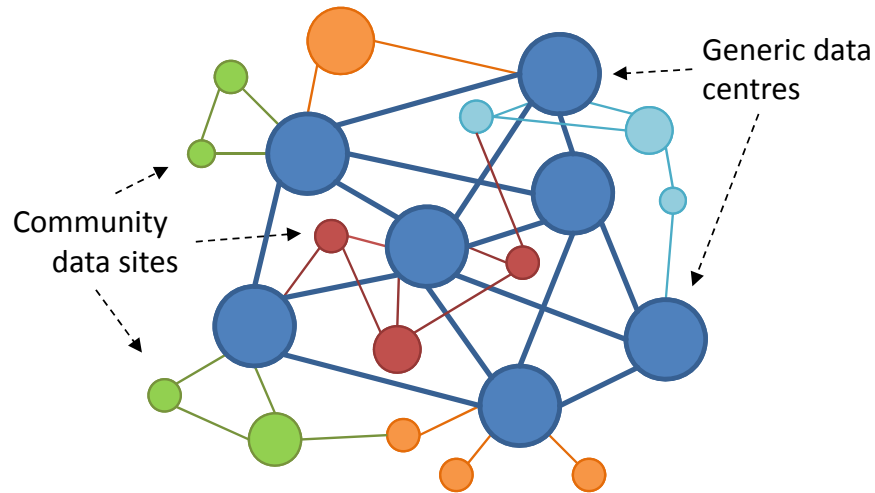
B2FIND
Find Research Data

b2find.eudat.eu

The screenshot shows a web interface for a dataset. On the left, there is a map of the Tyrrhenian Sea region with a location pin. Below the map are social media sharing options for GBIF, Google+, Twitter, and Facebook. The main content area displays the dataset title 'Foraminifera abundance of sediment core MD81-LC07' and its description: 'Hayes, Angela (2008): Foraminifera abundance of sediment core MD81-LC07, doi:10.1594/PANGAEA.407822'. Under 'Data and Resources', it states 'This dataset has no data' and includes a search box with 'Globigerina bulloides' entered. An 'Additional Info' table provides metadata details.

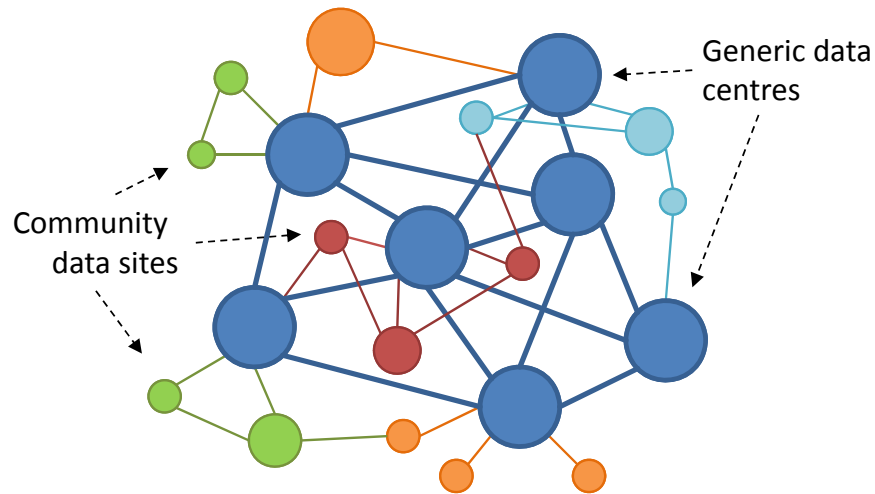
Field	Value
Source	http://dx.doi.org/10.1594/PANGAEA.407822
Author	Hayes, Angela (2008)
Version	68c7252d8134ebb8a19ce88a84e98ade
Discipline	Species
GeographicCoverage	Informationnotprovided
MetaDataAccess	http://metadata.gbif.org/catalogue/OAIHandler?verb=GetRecord&metadataPrefix=eml&identifier=oa:metadata.gbif.org/eml/portal/2290.xml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

A Federated and Distributed CDI



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

A Federated and Distributed CDI



- **Using EUDAT services:** finding and accessing data, for instance, or storing smaller data sets by interacting with one of the CDI public front-end services

vs

- **Joining the CDI:** implies a tighter integration with at least one of the EUDAT centre → partnership between legal entities relying on OLAs and SLAs



Need an EUDAT Specific Offering?

- **Storage capacities** located at selected centers across Europe, based on clear SLAs
 - Replication based on policy rules defined by the customer
 - Possibility to use large-scale computing power close to the data
 - Service based on clear SLAs with hosting centers
- **(Open) Data Sharing platform** tailored for specific needs (researchers, citizen scientists, etc.)
 - B2DROP, B2SHARE – plus extensions
- Dissemination and **better discoverability and reusability** of data sets
 - B2FIND – for data hosted both inside and outside of EUDAT



Conclusion

- **Pilot existing services**

- B2DROP, B2SHARE, and B2FIND available immediately through the Web with free access at the point of use
 - Pilots within the EUDAT 2020 project can include customization of the service (e.g. B2SHARE extension, B2FIND mapping, etc.)
- B2SAFE and B2STAGE requires selection of hosting sites and agreement on resource allocation
 - Limited free resources available for piloting service through the EUDAT 2020 project

- **Contact EUDAT to discuss long term strategy and partnership on data management**

- Data management and preservation policies and requirements
- Business models and partnership agreements (pre-paid resources, pay-per use, etc.)



Contacts

- Project Manager
 - Damien Lecarpentier (damien.lecarpentier@csc.fi)
- Scientific Coordinator
 - Peter Wittenburg (peter.wittenburg@mpi.nl)
- Web
 - <http://www.eudat.eu>
- Twitter
 - @Eudat_eu