The B2 SERVICE SUITE

www.eudat.eu/services

This work is licensed under the Creative Commons CC-BY 4.0 licence.
Attribution: EUDAT - www.eudat.eu

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065
EUDAT offers a complete set of research data services, expertise and technology solutions to all (European) scientists and researchers.

These shared services and storage resources are distributed across 15 European countries.

Data are safely stored alongside some of Europe’s most powerful supercomputers.
B2 Services

- **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

**AAI**: Network of trust among authentication and authorization actors

**PID**: Identity Integrity Authenticity Locations

**B2FIND**: Aggregated EUDAT metadata domain Data inventory

**External community metadata store**

- Data
- Metadata
B2 SERVICE SUITE is part of EUDAT...

- a pan-European initiative building a sustainable cross-disciplinary and cross-national data infrastructure providing a set of shared services for accessing and preserving research data
- supporting multiple research communities by working closely with them to deliver these technical services as part of the EUDAT Collaborative Data Infrastructure (CDI)
EUDAT offers common data services, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network of 35 European organisations.

Our vision is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure.
Community-Driven Solutions

EUDAT services are designed, built and implemented based on user community requirements.
Why Should I Use it?

- **Outsource** your data management needs...
  - Less effort than building your own infrastructure
  - More time and money for the research
- ... to a group of *respected* and *trusted* data centres...
  - With a national and international remit
  - Have a long past and future
- ... Who have a wealth of experience supporting data management for science.
  - Many of whom support long term curation or are working to support it
Support throughout the lifecycle

**B2DROP**
Sync and Exchange Research Data

**B2FIND**
Find Research Data

**B2STAGE**
Get Data to Computation

**B2SHARE**
Store and Share Research Data

**B2SAFE**
Replicate Research Data Safely

- Creating Data
- Processing Data
- Analysing Data
- Preserving Data
- Giving Access to Data
- Re-Using Data
EUDAT DATA DOMAINS

- **PUBLISHED DATA DOMAIN**
  - Linking Digital Objects to Publications
  - Discovery of Digital Objects

- **REGISTERED DATA DOMAIN**
  - Stage Digital Objects
  - Register Digital Objects

- **WORKSPACE (TEMPORARY - TRANSIENT)**

- **Published**
  - Static data snapshots
  - Long-term persistence and curation
  - Formal metadata & DOIs
  - Links to publications
  - Managed by repository

- **Shared**
  - Active data, subject to change
  - Shared with collaborators
  - Some metadata, some identifiers
  - Managed by community or collaborative group

- **Private**
  - Large numbers of data objects
  - Frequently updated, rapidly changing
  - Locally shared or individually private
  - Managed by individual researchers or labs
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 others.
An ideal solution for researchers and scientists to:

- **Store and exchange** data with colleagues and team members, including research data not finalized for publishing
- **Share data** with fine-grained access controls
- **Synchronize multiple versions** of data across different devices

**Features:**
- 20GB storage per user
- Living objects, so no PIDs
- Versioning and offline use
- Desktop synchronisation
B2SHARE is a user-friendly, reliable and trustworthy way for researchers, scientific communities and scientists to store and share small-scale research data from diverse contexts.
A winning solution for researchers, scientists and communities to:

- **store data safely** at a trusted and certified data centre
- **preserve data** to guarantee long-term persistence
- **control access** and **share data** with colleagues and the world

**Features:**
- metadata management
- permanent PIDs
- Open Access support
B2SAFE

B2SAFE is a robust, safe and highly available service which allows community and departmental repositories to implement data management policies on research data across multiple administrative domains in a trustworthy manner.
The ideal solution for communities with no facility for archival to:

- **replicate research data** into secure data stores
- **archive and preserve** research data in the long-term
- bring data close to **powerful compute resources**
- **co-locate** data with different communities
- benefit from **economies of scale**

**Features:**

- large-scale storage
- robust and highly available
- permanent PIDs
B2SAFE Use Case from CLARIN ERIC
Replication of Linguistic Data

iRODS

SAMQFS

GPFS

HPSS

dCache

DMF

PID

EPIC

16
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.
Facilitating communities to:

- move large amounts of data between data stores and high-performance compute resources
- re-ingest computational results back into EUDAT
- deposit large data sets onto EUDAT resources for long-term preservation

Features:

- high-speed transfer
- reliable and light-weight
- manages permanent PIDs
B2FIND

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

- **seek data objects and collections** using powerful metadata searches
- **catalogue community data** by means of selected metadata
- **browse through multi-disciplinary data collections** filtered by content, provenance and temporal keywords

**Features:**
- simple to use
- standards-based
- comprehensive catalogue
IDI Integration

https://b2access.eudat.eu

Primary Identities

- OpenID IdPs from RIs
- ORCID, ResearchID, Scopus
- ESGF, ENES
- B2ACCESS IdP
- User Profile
- B2SHARE (OAuth 2)
- B2SAFE (X.509)
- B2STAGE (X.509)
- B2DROP (SAML)
- B2HANDLE (SAML)

Social IdP

- Google
- Facebook
- Linkedin

SAML IdP

- eduGAIN
- RIs e.g. CLARIN
- PRACE EGI WLCG

Multi-Protocol Identity Management, LoA support
powered by Unity IDM

OpenID

OAuth 2 authorization server

Access Token

EUDAT federation database

PKI

EUDAT Service Endpoints

Data Project
Coordination Portal

Helpdesk TTS

Site & Service Registry

EUDAT CA

X.509

SAML

B2ACCESS AAI functions

EUDAT federation database

B2ACCESS Integration

Identity & Authorisation

EUDAT
## Roadmap Data Access and Re-Use Service Area

### B2SHARE
- **M12-M18** (Mar 2016 – Aug 2016)
  - Release of B2SHARE 2.0
  - Pilot: integration with B2NOTE
  - Pilot: B2SHARE as EUDAT CDI Metadata Store
  - Integrate with DTR for pilot users
- **M19-M24** (Sep 2016 – Feb 2017)
  - Support for cloud storage services like DropBox and Google
  - Drive
  - Support for more storage back-end solutions, e.g. cloud-storage and object stores
  - Support versioning
  - Support Digital Object Identifiers (DOIs)
- **+M24** (Mar 2017 – Feb 2018)
  - Integration with the EUDAT Generic Execution Framework (GEF)
  - Support for metadata extraction and data exploration via the GEF

### B2DROP
- **M12-M18** (Mar 2016 – Aug 2016)
  - Integrated with B2SHARE
  - Integrated with B2ACCESS
  - Deployment with Puppet and Docker

---

**B2SHARE**
- B2SHARE
- B2SHARE
- B2SHARE

**B2DROP**
- B2DROP
- B2DROP
- B2DROP
## Roadmap Data Access and Re-Use Service Area

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Tasks and Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M12-M18</strong></td>
<td>Mar 2016 – Aug 2016</td>
</tr>
<tr>
<td><strong>M19-M24</strong></td>
<td>Sep 2016 – Feb 2017</td>
</tr>
<tr>
<td><strong>+M24</strong></td>
<td>Mar 2017 – Feb 2018</td>
</tr>
</tbody>
</table>

### B2FIND
- Continued community integration
- Improved user experience
- Resolve granularity issues
- Integration with B2NOTE
- Performance and scalability improvements
- Customisation of GUI for communities
- Prototype of SRU interface
- SRU interface in production
- Integrate with EUDAT CDI Metadata Store
- Integration with Data Type Registry
- Extend harvesting methods (OGC / CSW)
- Improved search functionalities for hierarchical search and taxonomies
- Improved semantic mapping
- Evaluate DTR pilot
- Bring DTR into production
- Integrate DTR with other EUDAT services based on pilot evaluation

### Registry Services
- Pilot instance of the Data Type Registry
- Collaborate with pilot communities on further evaluation and adaptation of the DTR
- Integrate the DTR pilot with B2ACCESS for Federated AAI
- Integration of the DTR with B2SHARE and B2FIND
- Further development of the DTR based on feedback from communities and other EUDAT services
- Integration of the DTR with B2NOTE
- Evaluate DTR pilot
- Bring DTR into production
- Integrate DTR with other EUDAT services based on pilot evaluation
## Roadmap Data Access and Re-Use Service Area

**B2ACCESS**
- B2SHARE integrated with B2ACCESS for authentication
- Integration with Data Type Registry pilot
- Integration with EUDAT CDI common HTTP REST API
- Integration with B2SAFE & DPM
- Integration with B2DROP
- Integration with B2STAGE
- Integration with Data Project Coordination Portal
- Installation packages & distributed setup
- Pilot for XACML based authorisation solution

**B2SHARE**
- Support for integration with external community sites
- Integration with PRACE
- Integration with EGI
- Distributed authorisation

**Complete integration with PRACE**
- Complete integration with EGI

**B2DROP**
- Complete integration with external community sites

**B2STAGE**
- Improvements based on change requests from communities and partners

---

**M12-M18**
Mar 2016 – Aug 2016

**M19-M24**
Sep 2016 – Feb 2017

**+M24**
Mar 2017 – Feb 2018

---
14 generic centres, 15PB committed, 5-10Gb/s per site (potential of > 1000 PB aggregated)
Interactions With Users

- **Data Pilots**
  - 5 biomed and life science
  - 7 earth science, energy and environment
  - 5 physical sciences and engineering
  - 5 Social Science and Humanities

- **Interactions through data project co-ordination portal**
  - Allows user to specify the services (QoS) that they want
  - And request preferences for sites
  - ... but still needs 1 to 1 discussions on many occasions
Making it REALLY useful

- REST API and python library
  - Some services have well supported API (B2DROP (webDAV) and B2SHARE (REST))
  - Other REST APIs under development
EUDAT python library (ongoing)

- A *programmable* and *flexible* python API library to *create and execute workflows* composing different EUDAT services
- Let’s make EUDAT scriptable!

- Enable collaborations with other projects (PRACE, EGI..)
- First draft of the library:
  - find datasets by community names or tags using B2FIND
  - Manage transfers using Globus Online API (and FTS3 later on);
Data search:

```python
from eudat import b2find

# Get “aleph” dataset
datasets = b2find.get_dataset_source(community='aleph')
print('First dataset URI = ' + datasets[0])
```

PID resolution:

```python
from eudat import b2handle

# Get physical URL of the object resolving PID
URL = b2handle.get_url_by_pid("/11100/33ac01fc-6850-11e5-b66e-e41f13eb32b2")
print("Returned URL: " + URL)
```

Data transfer:

```python
from eudat import b2stage

# Initialize the Globus client
globus = b2stage.ClientGlobus(resource_file_path=RESOURCES_FILE, debug=True)
# Perform a data transfer between 2 endpoints (third party transfer)
task_id = globus.transfer(src_endpoint, dst_endpoint, src, dst, recursive=args['-r'])
```
For more info:

- B2DROP: https://b2drop.eudat.eu
- B2SHARE: https://b2share.eudat.eu
- B2SAFE: https://eudat.eu/services/userdoc/b2safe
- B2STAGE: https://eudat.eu/services/userdoc/b2stage
- B2FIND: http://b2find.eudat.eu