EUDAT Collaborative Data Infrastructure

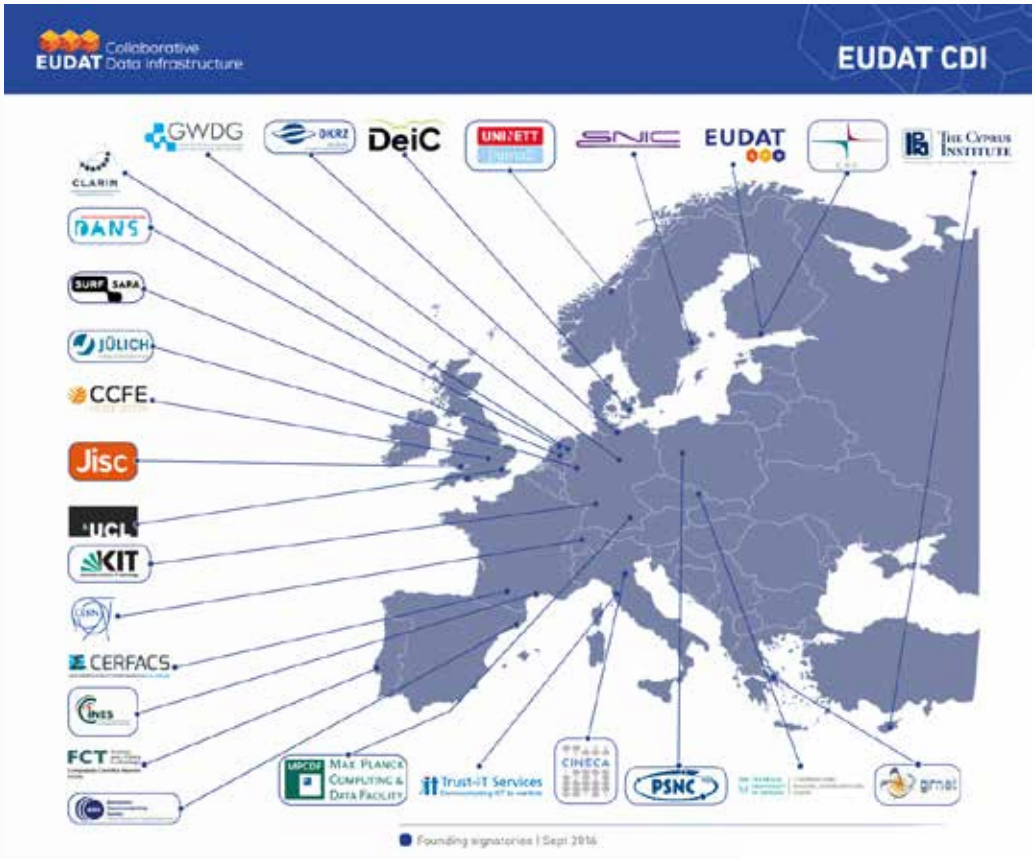
Services for Research Data Management

About EUDAT

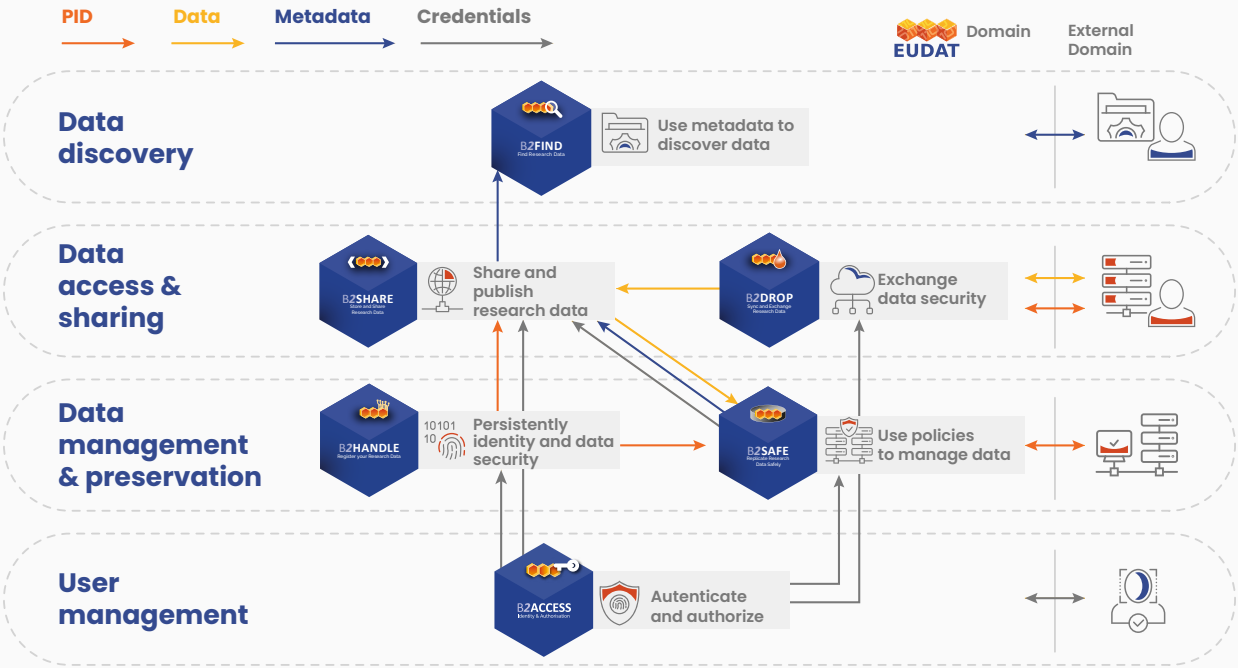
EUDAT Collaborative Data Infrastructure

The EUDAT Collaborative Data Infrastructure (or EUDAT CDI) is one of the largest infrastructures of integrated data services and resources supporting research in Europe. It is sustained by a network of more than 20 European research organisations, data and computing centres that, in September 2016, have signed an agreement to maintain the EUDAT CDI for the next 10 years, and in 2018 have supported the establishment of the limited liability company, EUDAT Ltd.

This infrastructure and its services are continuously being developed in close collaboration with over 50 research communities spanning across many different scientific disciplines.



EUDAT Services



EUDAT has developed a service stack that forms the Collaborative Data Infrastructure (CDI). The services are as follows:

B2DROP
Sync and share research data

B2SHARE
Store and publish research data

B2FIND
Find research data

B2SAFE
Keep research data safe via data management policies

In addition, a set of EUDAT core operational services, essential for the management of the CDI

B2ACCESS
(identity and authorisation), easy-to-use and secure Authentication and Authorisation platform

B2HANDLE
(persistent identification management), a service to register persistent identifiers called "handles" to data objects and retrieve data objects via these identifiers, serving a purpose similar to DOIs for papers

Most of these components are described in more detail in the following sections.

Sync and share research data

B2DROP is a low-barrier, user-friendly and trustworthy storage environment which allows users to synchronise their active data across different desktops and to easily share this data with peers. EUDAT offers a free public basic instance for any researcher. For communities and organisations a premium service is offered on the public instance. Communities and organisations can also request customised instances

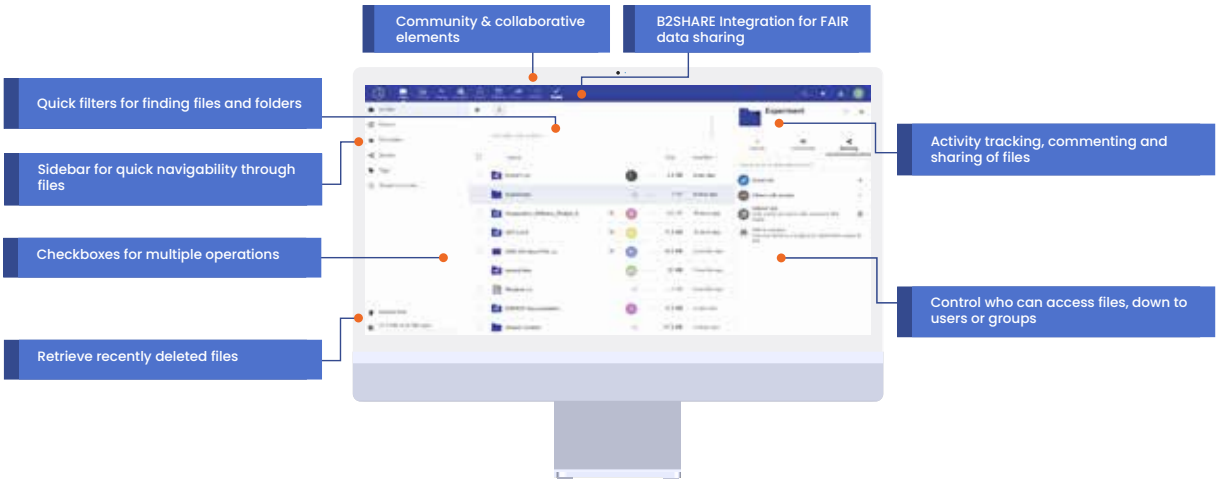
Features

- Default quota of 20GBs (Basic), high quotas optional (Premium)
- Access via Web GUI, desktop clients and Webdav
- Multiple versions of files are kept
- Enabled apps: Contacts, Calendar, Tasks, Circles (social communities)
- Sharing within B2DROP, across different instances (via OCM-API) and via links
- Publishing of datasets to B2SHARE
- Integration with CLARIN Language Resource Switchboard (Basic)
- Integration with other community services optional (Premium)
- Group management (Premium)
- OnlyOffice (Premium)

Technologies
- NextCloud

Standards
- OpenCloudMesh API for Federated Cloud Sharing

Dependencies
- B2ACCESS



How it works

The service is intended for the long-tail and still volatile data which can change and are still subject to active research e.g. drafts of research papers. Therefore, B2DROP offers versioning of all ingested files but does not attach persistent identifiers to them.

B2DROP is hosted at the Jülich Supercomputing Centre, which guarantees that your research data stays in Europe. Daily backups of all files in B2DROP are taken and kept on tap. B2DROP is based on Nextcloud. Data is encrypted on transmission through the exclusive use of the https protocol for data transfer. Each B2DROP user is allowed 20GB of storage.

B2DROP offers an intuitive user-interface via the web. In addition to web access, users can mount B2DROP as a drive on their desktop machines via WebDAV, or use a desktop client which also allows offline synchronisation.

Packages

B2DROP Basic For Researchers
Capacity: 20GB
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. B2DROP is an ideal solution to store and share data with colleagues and team members, synchronise multiple versions of data, ensure automatic desktop synchronisation of large files. Only self registration is needed and can be accessed online. You can share files with local users, or across other B2DROP and ownCloud/Nextcloud instances. You can do synchronisation using desktop clients or WebDAV mounts. Additionally, you can publish data to B2SHARE.

B2DROP Premium For Researchers
Capacity: 100-2000GB
The basic B2DROP service has a per user quota limit of 20GB. Researchers which have a need for higher quotas, higher storage capacity, specific requirements and/or (Nextcloud) applications enabled can request this via the B2DROP Premium service.

B2DROP For Data Managers
Capacity: 1-10TB
Quota per user: 100-500GB
The public B2DROP service has a per user quota limit of 20GB. Communities, organisations or principal research investigators which have a need for higher quotas, higher storage capacity, specific requirements and/or (Nextcloud) applications enabled can submit a support request to request this. Depending on the requirements and/or requested storage capacity the B2DROP service is offered as a paid service. It includes all the features of the basic service except with higher quotas.

B2DROP For Communities and Organisations
Capacity: 1-10TB
Quota per user: 100-500GB
The public B2DROP service is provided on a fair-share basis. Communities and/or organisations with specific requirements which can not be support via the public fair-share service or or need higher level of security or access control can request consultancy in setting up and deploying a local B2DROP instance or can request a B2DROP instance which is hosted at one of the EUDAT partners. Features are negotiable.

How to get access?

For researchers needing 20GB or less, use B2DROP for free now:

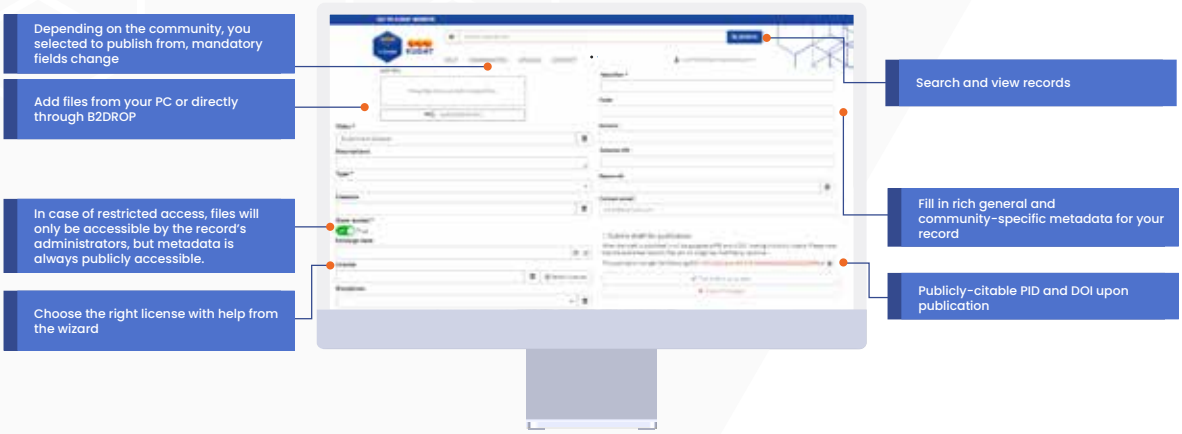
b2drop.eudat.eu

For higher needs from individual researchers, data managers and communities and organisations, contact us to set-up your dedicated instance:

eudat.eu/contact-support-request

Store and publish research data

B2SHARE is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store, publish and share research data in a FAIR way. B2SHARE is a solution that facilitates research data storage, guarantees long-term persistence of data and allows data, results or ideas to be shared worldwide. B2SHARE supports community domains with metadata extensions, access rules and publishing workflows. EUDAT offers communities and organisations customised instances and/or access to repositories supporting large datasets.



Features

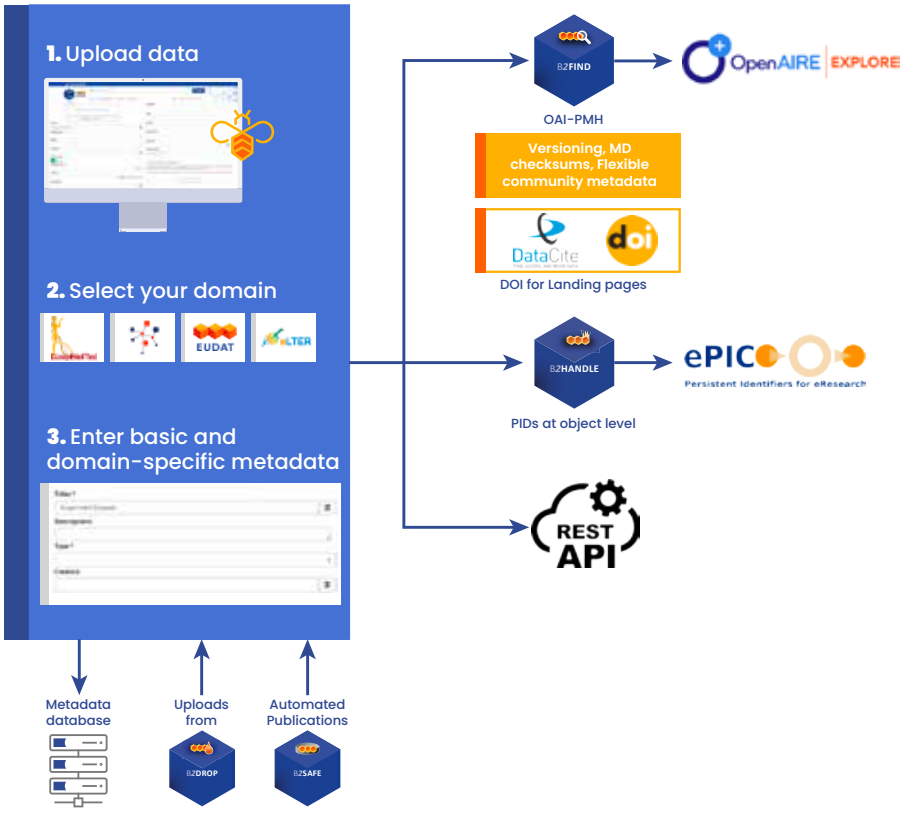
- Support of metadata descriptions via the EUDAT metadata schema
- Registers DOIs for datasets and Handle PIDs for data objects
- Supports versioning
- Harvested by B2FIND and OpenAIRE explorer
- Direct upload from B2DROP
- Accessible via a Web GUI and an API to support automatic publishing workflows
- Supports community domains
- Allows communities to define metadata extensions, access rules and publishing workflows
- Annotate research data through the B2NOTE extension

Technologies
Invenio

Dependencies
B2ACCESS

How it works

B2SHARE is a service for what's known as "long tail of data". This is a key component of research but is often under-exploited and even at risk when only stored on separate machines and storage devices such as notebooks, desktops or departmental servers. Any kind of stable research data which you want to preserve and share in a safe environment can be deposited in B2SHARE.



The research data can have various formats - papers, spreadsheets, audio-visual media, and more. The only requirement is that the source or purpose of the data is scientific.

The data is uploaded from your local or departmental storage device as digital objects which are then automatically assigned with persistent identifiers, ensuring long-lasting access and reference to the data.

The data is deposited in Finland at the Finnish IT Center for Science. More precisely it is stored on servers in Kajaani, a newly-built and environmentally-friendly data centre using the most advanced and modern data centre technology. EUDAT retains the right of archiving, i.e. creating replicas at trusted centres to take care of long-term persistence.

Packages

B2SHARE Basic For Researchers (free)
B2SHARE is a free-to-use public service supporting cross disciplinary research, self-registration for any scientists and researchers, free upload of stable research data, data will receive a DOI which can be used for citing, data objects will receive a B2HANDLE PID, access policy is defined by the data owner, metadata in openly accessible and made discoverable via B2FIND, Direct uploads from B2DROP, data integrity is ensured by checksums which are calculated during data ingest.

B2SHARE Premium For Data Managers
The B2SHARE - Premium service is the same as the basic service, but provides researchers the option to upload larger data files and datasets (>20GB) and the customization of metadata extensions.

B2SHARE For Communities And Organisations
B2SHARE for communities and organisations offers the option for dedicated and branded B2SHARE instances which can be customised to community and organisational needs. This supports customisation to the metadata schemas, community domains and access rights.

B2SHARE For Large Datasets
EUDAT is a network service providers across Europe specialising in large ICT infrastructure supporting research. To researchers and/or communities seeking to publish and/or make large datasets accessible in a FAIR way, EUDAT can offer access to data repositories supporting large datasets (>50TB). These repositories are made available through the EUDAT partners.

How to get access?

Use B2SHARE Basic now

b2share.eudat.eu

For communities wanting to publish using their own metadata scheme fields, contact us

eudat.eu/contact-support-request

Find research data

B2FIND is an interdisciplinary discovery portal for research outputs that allows free term search. Results may be narrowed down using several facets, including spatial and temporal search options.

B2FIND is the EUDAT metadata indexing service and provides a discovery portal which allows users to find data collections within an international and inter-disciplinary scope. It is based on a comprehensive metadata catalogue of research data collections stored in EUDAT data centres and community repositories. Harmonisation of the metadata descriptions collected from heterogeneous sources enables not only the presentation in a consistent form but also the faceted search across scientific domain boundaries. The service is geared for communities and other providers of research data who need to publish and give visibility to their metadata and individual researchers who need to search data from everywhere, and see data in the context with an across community approach.

Features

- Harmonisation of the metadata descriptions via the EUDAT Core metadata schema
- Harvesting of repositories via different protocols (e.g. OAI-PMH, CSW, Rest-APIs)
- Faceted search via 12 facets (including geospatial, temporal search options), additional free text search
- Metadata aggregation from community repositories, multiple metadata standards are supported
- Harvested by OpenAIRE explorer

Technologies

- CKAN

Supported harvesting protocols

- OAI-PMH
- REST-API
- CSW

Dependencies

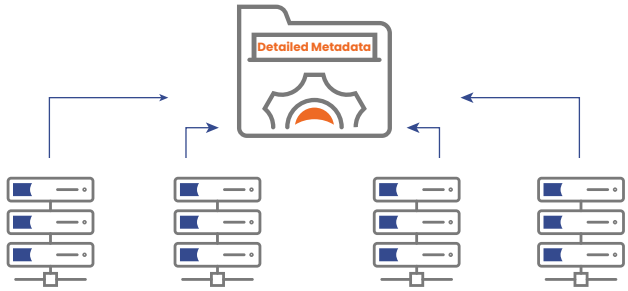
- B2SHARE

Supported metadata standards:

- Dublin Core
- Datacite
- OpenAire
- EUDAT Core
- INSPIRE
- ISO 19139/19139, FGDC
- community specific standards

How it works

Metadata that are made searchable in B2FIND are harvested from metadata providers, such as using the standard OAI-PMH interface. The community itself decides which metadata are made available to EUDAT and how their metadata elements are mapped to specific metadata elements for being represented as facets within the search portal. A sophisticated framework ensures that metadata from various providers are harvested regularly to display complete and up-to-date information. This framework also provides the translation from the community´s metadata elements to the EUDAT Core Metadata Schema, which is based on generic schemas like DataCite and OpenAire. Close communication with the community´s representatives is key for making research data findable and thus supporting FAIR data principles.



The B2FIND service can be utilised in two ways: (1) via a graphical, web-based tool, available at b2find.eudat.eu, which is designed to be self-explanatory, and (2) using a command line tool based on the API interface to the portal.

The following two prerequisites must be fulfilled in order to publish and make your metadata visible in B2FIND: (1) a service for providing and transferring the metadata, and (2) the definition of a community specific mapping. The B2FIND entry page lists all communities that provide metadata under 'Communities'. In order to publish metadata, please follow the "Guidelines for Providing Metadata for EUDAT-B2FIND" (v), where supported harvesting protocols and metadata schemas are described in detail.

Packages

B2FIND For Researchers (Free)
B2FIND is a discovery service based on metadata steadily harvested from research data collections from EUDAT data centres and other community repositories. The service offers faceted browsing and it allows in particular to discover data that is stored through the B2SHARE service.

B2FIND For Data Providers
To enlarge the discovery of existing research output, data repository owners can make their research data collections stored in existing data repositories harvestable and discoverable via the public B2FIND service. To ease the process of harvesting data repositories, owners can assess the B2FIND guidelines for data providers.

How to get access?

Access and find research through the service now

b2find.eudat.eu

For Data Providers, request activation through the form:

eudat.eu/contact-support-request

Keep research data safe via data management policies

B2SAFE is a robust 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. It offers an abstraction layer of large scale, heterogeneous data storages, guards against data loss in long-term archiving, allows optimized access for users (e.g. from different regions), and brings data closer to facilities for compute-intensive analysis.

Features

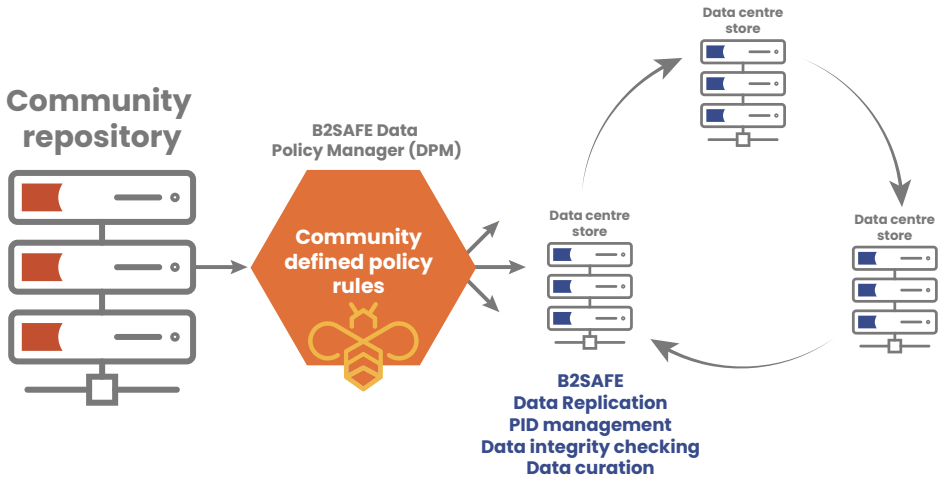
- Support for data management policies (e.g. registration of PIDs, cross-site replication, data integrity checks)
- Support for policies customised to community and organisational needs
- Support for less frequently used archival data, but can also support active data
- Support for large scale storage resources (e.g up to PB-scale)
- A single namespace across heterogeneous storages
- Supports integration with different kind of storage systems (e.g. Tape based HSM, POSIX filesystems, Object storage)
- Access via GridFTP, Webdav, iRODS commands
- Service offered by a network of EUDAT service providers

Technologies
iRODS

Dependencies
B2HANDLE

How it works

Tomorrow's research data will be a domain of registered data objects and collections where persistent identifiers (PIDs) identify the data objects and collections, while information associated with the PIDs allows the integrity and authenticity of the data to be checked. Furthermore, data management operations will be governed by formalised policy rules that will also form the basis of any audits to assess quality. To meet this vision, B2SAFE is designed to be based on the execution of auditable policy rules and the use of PIDs, as offered by the EPIC Handle service.



This will increase trust in data and services in the anonymous world of data re-usage and repurposing. Suitable technology that allows the execution of sequences of policy rules will be applied, and the EUDAT data federation will be built on trust agreements, which ensure that ownership rights to the data reside with the originators. To make replicas accessible via community-defined portals, service providers will need to host community-based software respecting all access permissions.

Packages

B2SAFE – For Archival Data

B2SAFE can provide a higher level of abstraction to different underlying Data Archival mechanisms, in order to maintain Data Objects metadata (including also PIDs) and integrating with external tools. This offer targets data which is not frequently used and needs to be stored for long term and offers the option to replicate data across different sites for safe keeping.

B2SAFE – For Active Data
As a standard default mechanism B2SAFE can allow users to simply store data directly in the iRODS File System (iRODS FS), rather than the Object Storage or the Data Archival (as previously described). B2SAFE can create and assign PIDs to data objects, and store metadata associated to files and directories that have been saved to the Posix FS, on users behalf. This offer targets active data which is frequently used.

How to get access?

If you are a Repository Service Provider (Thematic Service Provider) you can do the following:

Request to use B2SAFE

eudat.eu/contact-support-request

Join the EUDAT CDI as a provider

eudat.eu/eudat-cdi/joining

Any community and departmental data repositories that have a proper repository infrastructure supporting PIDs and metadata describing the properties and context of the data being replicated can participate in B2SAFE. The replication service currently relies on iRODS but also supports other federation technologies. Participating repositories can choose to use iRODS and link it with their local repository to enable a tight integration to the EUDAT infrastructure, or use other easy-to-integrate federation approaches and client libraries.

EUDAT will help interested community centres to set up and use this service by running training courses, and providing support for the necessary adaptation work, including offering a service help desk.

B2ACCESS

Identity & authorisation

B2ACCESS is a federated cross-infrastructure authorisation and authentication proxy for user identification and community-defined access control enforcement. It allows users to authenticate themselves using a variety of credentials providing federated access and single-sign-on to services and service providers in a trusted way. B2ACCESS offers communities and service providers an AARC compliant AAI proxy ready to be integrated within the EOSC AAI Federation.

Features

- Compliant to the AARC Blueprint and REFEDS Sirtfi
- Supports group management
- Supports authorisation via group membership
- Supports authentication e.g. via eduGAIN, ORCID and Social Identities
- Support for EUDAT B2ACCESS local accounts and user authentication by X.509 certificates
- Support for IdP and service integration via SAML, OAuth and OIDC

Technologies

- Unity Idm

Standards for Service and IdP integration

- SAML
- OAuth2/OIDC

How it works

B2ACCESS allows EUDAT users to authenticate themselves using a variety of credentials. The following log-in options are supported:

- User's home organisation identity provider
- Social account (e.g. Google, Microsoft Live and Facebook)
- B2ACCESS ID

The user's home organisation has the highest level of assurance and is the preferred way of authentication. If a user's home organisation is not available in the EUDAT B2ACCESS system, a social account or B2ACCESS ID are the alternatives. However, since these have a lower level of assurance, access to the EUDAT services and resources can be limited for users authenticated under these options.

The figure shows the interaction of different methods of authentication with the B2ACCESS system. The diagram shows external identity providers (IdPs) authenticating to the B2 services in EUDAT with B2ACCESS providing the Authentication/Authorisation Infrastructure (AAI). Thus, users can use their existing identities to identify themselves to the EUDAT services. The organisational identity will be provided by a user's home organisation. In order to support a wide range of IdPs we have joined the EduGain federation

How to get access?

To request the use of B2ACCESS as your identity and access solution, reach out through the form below:

eudat.eu/support-request

B2HANDLE

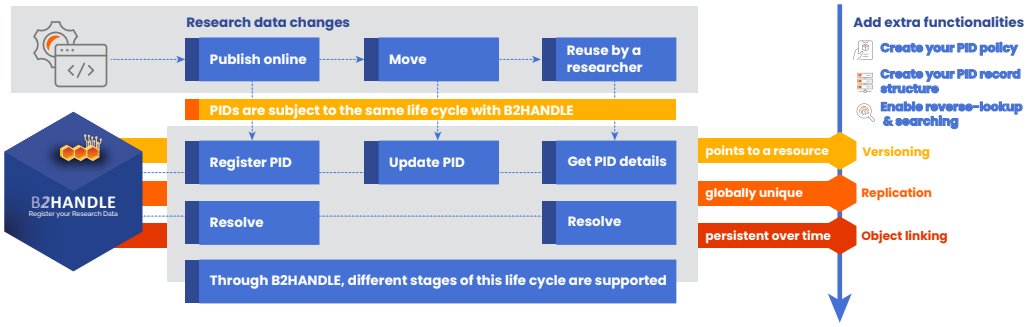
Register your research data with a persistent identifier

B2HANDLE is the distributed service for storing, managing and accessing persistent identifiers (PIDs) and essential metadata (PID records) as well as managing PID namespaces. The implementation of the service relies on the DONA/Handle persistent identifier solution. B2HANDLE can be used by middleware applications, end-user tools and other service to reliably identify data objects over longer timespans and through changes in object location or ownership. The B2HANDLE service encompasses management of identifier namespaces (Handle prefixes), establishment of policies and business workflows, operation of Handle servers and technical services, and a user-friendly Python library for general interaction with Handle servers and EUDAT-specific extensions. B2HANDLE is mostly transparent to end-users, shielding them from the complexity of infrastructure details. B2HANDLE supports a dedicated Handle record structure (a PID profile) for the safe data management within an infrastructure with a given topology.

Features

- Globally resolvable identifiers via the Global Handle Network (DONA, hdl.handle.net)
- Communities and organisations can obtain Handle Prefixes (ePIC, 21.######) for their own use
- PIDs can be hosted at EUDAT service providers
- B2HANDLE operates as a federation of EUDAT service providers based on policies
- PIDs are mirrored across multiple providers for high resolution and high availability
- Handle records can be customised to community and organisational needs (for example include checksums and timestamps to ensure authenticity of the data objects)
- Support for reverse look-ups of Handle records
- REST API for easy registration and minting of PIDs

Technologies
- Handle (handle.net)



B2HANDLE is EUDAT's main Persistent Identifier (PID) service. B2HANDLE is a distributed service, designed to contribute to data persistence by maintaining opaque, globally unique persistent identifiers. PIDs are used to reliably identify and cite data objects throughout their lifecycle and they are thus a vital part of long-term data management. Moreover, data can be directly retrieved by PIDs and corresponding key metadata can be stored together with them in the so-called PID entry.

B2HANDLE is based on the Handle System, which is a reliable, redundant and scalable system built on an open architecture and designed to allow end-users to manage PIDs with ease. The B2HANDLE service encompasses management of identifier namespaces (Handle prefixes), establishment of policies and business workflows, operation of Handle servers and technical services, and a user-friendly Python library for general interaction with Handle servers and EUDAT-specific extensions.

In order to access a data object stored in EUDAT, an associated PID is needed. B2HANDLE enables EUDAT services and user communities to assign PIDs to different kinds of managed objects stored in the EUDAT CDI. The advantage of the service is that the underlying technology is mostly transparent to the end-user (users need not know the complexity of infrastructure details).

B2HANDLE service resolution is based on the Handle System which offers a very reliable resolution service. It forwards the user to the current location of the object.

One of the key advantages of B2HANDLE is the B2HANDLE Python library. It is a client library to enable easy interaction with Handle services using the native REST interface offered by the Handle system. The library offers the main functionalities to create, update and delete PIDs as well as advanced functionality such as searching over Handles using an additional search servlet and managing multiple location entries per PID. The library is available for all services that want to enable PIDs.

In the EUDAT ecosystem, EUDAT services make use of B2HANDLE to enable data access, guarantee long lasting references to data and facilitate data publishing. B2SAFE and B2SHARE use the service to create and manage PIDs for their hosted data objects, whereas B2FIND and B2STAGE use the resolving mechanism of B2HANDLE to retrieve objects and to refer to them.

Packages

B2HANDLE For Researchers
PIDs are normally automatically minted at and within data management and data repository systems, for example via the EUDAT B2SHARE data repository service. Therefore, researchers in need of a Persistent Identifier for a data set can publish the data set via the B2SHARE service.

B2HANDLE *For Data Managers*
B2HANDLE is the distributed service for minting, storing, managing, and accessing persistent identifiers (PIDs) and essential metadata (PID records) as well as managing PID namespaces. The B2HANDLE service allows Data Managers and Data Providers the registration and minting PIDs which can be integrated within Data Repositories and other Data Management platforms.

Prefix
PIDs are comparable to the ISBN numbers assigned to books. A PID consists of a Prefix and a Suffix. The Prefix identifies the naming authority of the PIDs identified by the Prefix. The prefixes acquired via EUDAT are ePIC Handle Prefixes and are globally resolvable via the Handle.net Global Resolving Network. This offer does not include the PID Hosting service, the Prefixes need to be configured in a Handle service separately. If you also require PID hosting, please select the PID Hosting offer.

PID Hosting
B2HANDLE is the distributed service for minting, storing, managing, and accessing persistent identifiers (PIDs) and essential metadata (PID records) as well as managing PID namespaces. The B2HANDLE PID Hosting allows Data Managers and Data Providers the registration and minting PIDs which can be integrated within Data Repositories and other Data Management platforms.

How to get access?

For researchers, use B2HANDLE though B2SHARE now!

b2share.eudat.eu

For other packages such as for data managers, prefixes or PID hosting:

eudat.eu/support-request

How to proceed?

For community data managers and system administrators:
There are two modes of integration with EUDAT:

1. "use" EUDAT: an EUDAT datacentre ingests the community data and ensures its storage and replication. EUDAT will also expose several web-interfaces and APIs to its services, which will allow a looser, but still feature-rich connection with EUDAT.
2. "join" EUDAT: the data centre/institute becomes a part of the EUDAT network. To this end it is necessary to adopt the EUDAT Collaborative Data Infrastructure (CDI).

In both cases, the community will benefit from access to EUDAT's primary services – B2SAFE for safe replication, B2ACCESS for user authentication and authorisation, B2FIND, the metadata harvesting and cataloguing service, B2SHARE for publishing, storing and sharing of smaller data sets, and B2DROP for storing, synchronising and exchanging data with colleagues and team members.

Joining EUDAT is more demanding, but it allows to benefit from the complete set of EUDAT data centre features such as: choice of data centres to federate with by means of B2SAFE; use of an own PID service within the EUDAT domain; optimised data transfer within the EUDAT CDI; and tighter control of data management policies (replication, authorisation) through direct management of core services at a site.

Joining EUDAT requires the installation and configuration of a minimum software stack (the EUDAT CDI). Furthermore, it requires the community to dedicate some resources (storage, compute and man-power) to the project.

In practice, three packages are required: iRODS deployed at the site; access to a system for PIDs (or B2HANDLE); and (optionally) deployment of the B2STAGE DSI package.

Configuration requires: access to disk space and connection to other iRODS installations in EUDAT; monitoring configuration; user-access configuration; and connection to the metadata catalog.

For end-users: The EUDAT services B2ACCESS, B2SHARE, B2FIND and B2DROP all offer interfaces for end users.

Ready to talk? Get in touch with us now!

eudat.eu/contact-support-request

EUDAT