



**CompBioMed**

A Centre of Excellence in Computational Biomedicine

**Webinar series**

# Webinar #9

## EUDAT Services for FAIR Data Management

**27 June 2019**

**Welcome!**



**Presenter:**  
**Dr Narges Zarrabi (SURFsara)**



**Moderator:**  
**Ben Czaja (UvA)**

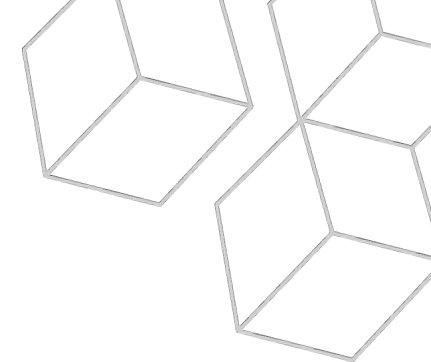
In collaboration with:



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 675451

The webinar series is run  
in collaboration with:





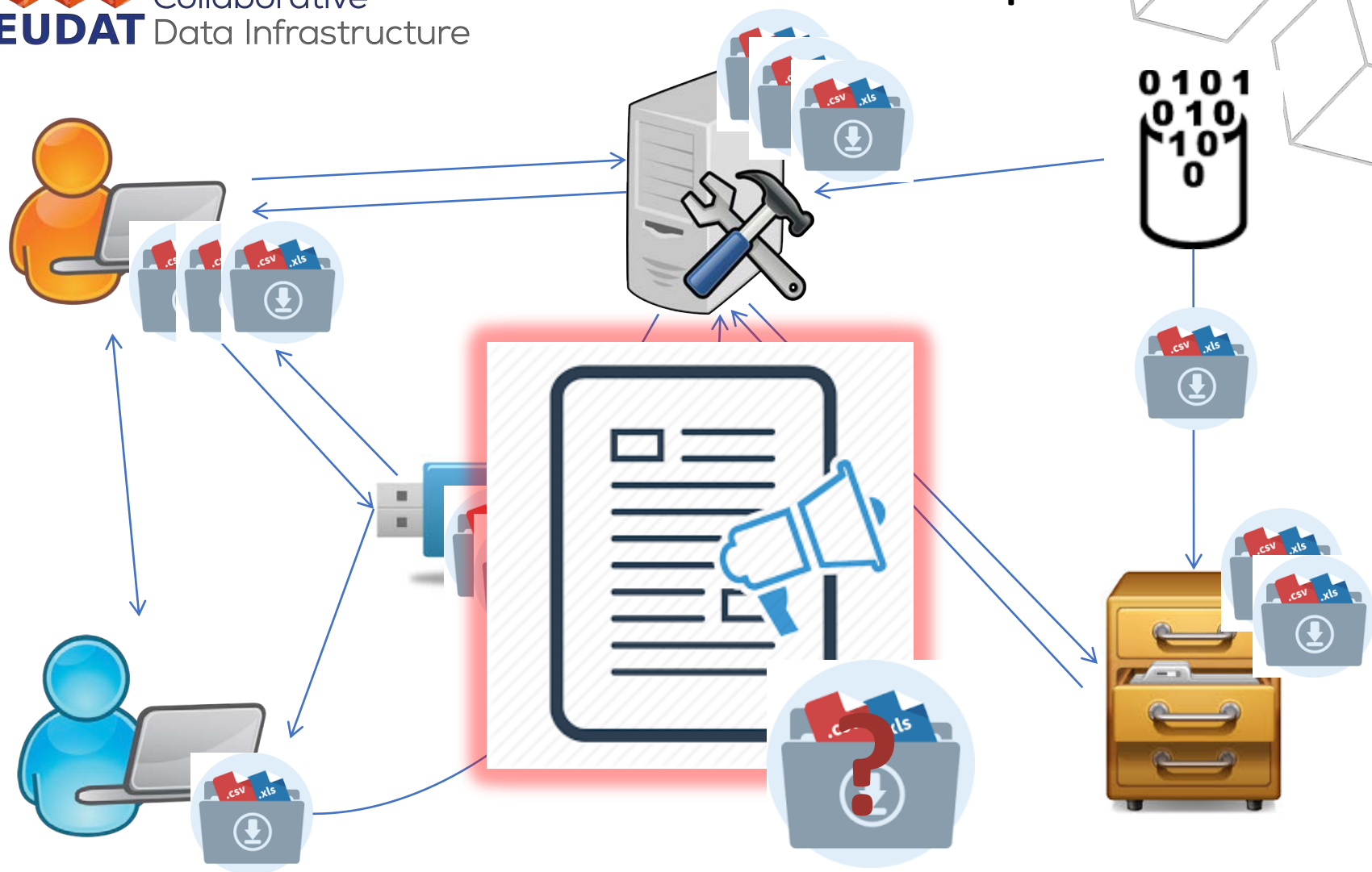
## Part 1:

- Data management requirements of research communities (10')
- Overview of B2Services for FAIR data management (20')
  - B2DROP, B2SAFE, B2SHARE, B2STAGE, B2HANDLE, B2FIND, B2ACCESS, B2NOTE...
  - Integration between B2Services

## Part 2:

- Example data pipelines and workflows (Live demo) (25') –
    - Safe data replication with B2SAFE (CompBioMed use case)
    - Data sharing and publishing workflow
    - Data discovery and download workflow
  - Q&A (5')
-

# Data- Where is the problem?



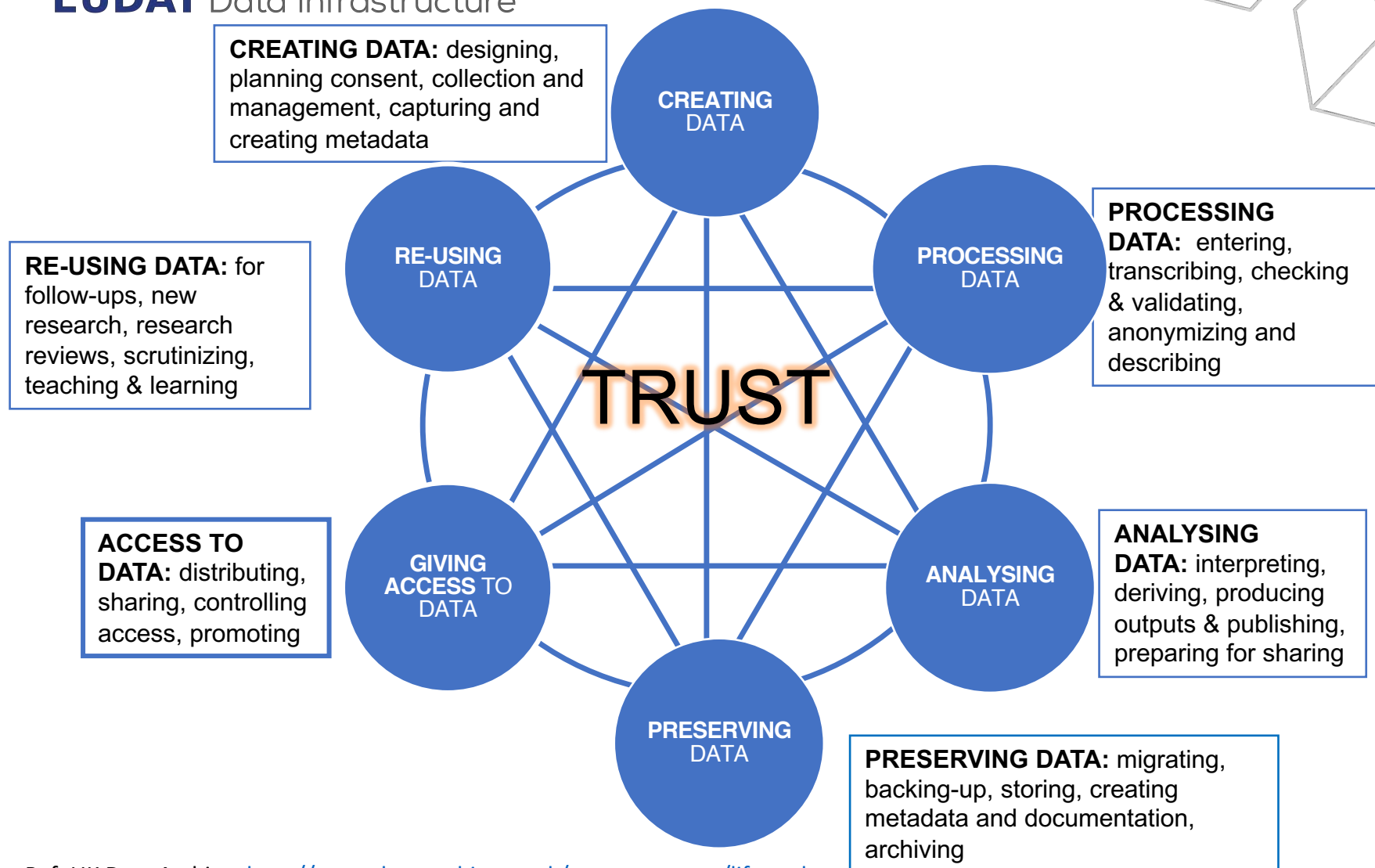
# Data requirements of research communities

- **More efficient data access, sharing and transfer**
    - *Intensive data-sharing and transfer*
    - *Restricted data-sharing and transfer*
  - **Preserving research data**
    - *Storage, backup and archiving large data, synchronizing data over distributed places*
    - *data provenance*
  - **Accessible research Data**
    - *Making data accessible to research communities, PIDs*
    - *Publishing data with domain specific metadata*
    - *Linking published data to processed and raw data*
  - **Findable research data**
    - *A major challenges scientific communities is to discover data from research data collections and repositories*
-



Collaborative  
Data Infrastructure

# Research data life cycle



# What is... FAIR ?

## Findable:

- F1.** (meta)data are assigned a globally unique and persistent identifier;
- F2.** data are described with rich metadata;
- F3.** metadata clearly and explicitly include the identifier of the data it describes;
- F4.** (meta)data are registered or indexed in a searchable resource;

## Interoperable:

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles;
- I3.** (meta)data include qualified references to other (meta)data;

## Accessible:

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol;
  - A1.1** the protocol is open, free, and universally implementable;
  - A1.2.** the protocol allows for an authentication and authorization procedure, where necessary;
- A2.** metadata are accessible, even when the data are no longer available;

## Reusable:

- R1.** meta(data) are richly described with a plurality of accurate and relevant attributes;
  - R1.1.** (meta)data are released with a clear and accessible data usage license;
  - R1.2.** (meta)data are associated with detailed provenance;
  - R1.3.** (meta)data meet domain-relevant community standards;

- **EUDAT B2Service Suite**

- **B2ACCESS**

- **B2DROP**

- **B2HANDLE**

- **B2SAFE**

- **B2STAGE**

- **B2SHARE**

- **B2FIND**

- **B2NOTE**

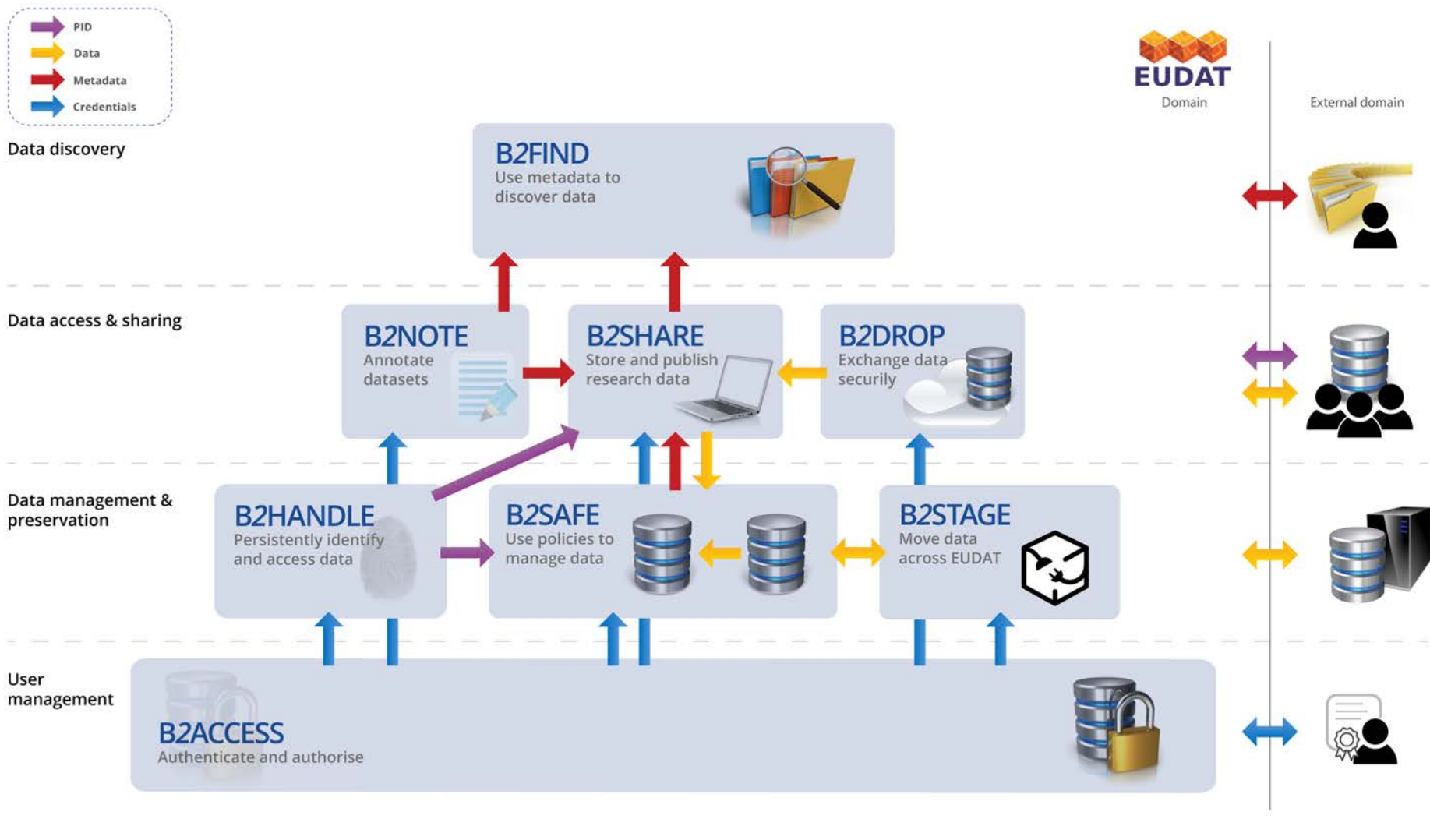
- **How EUDAT services link to the research data lifecycle**

- **How EUDAT services support the FAIR principles**

- **EUDAT contact & support:** <https://eudat.eu/support-request>

---







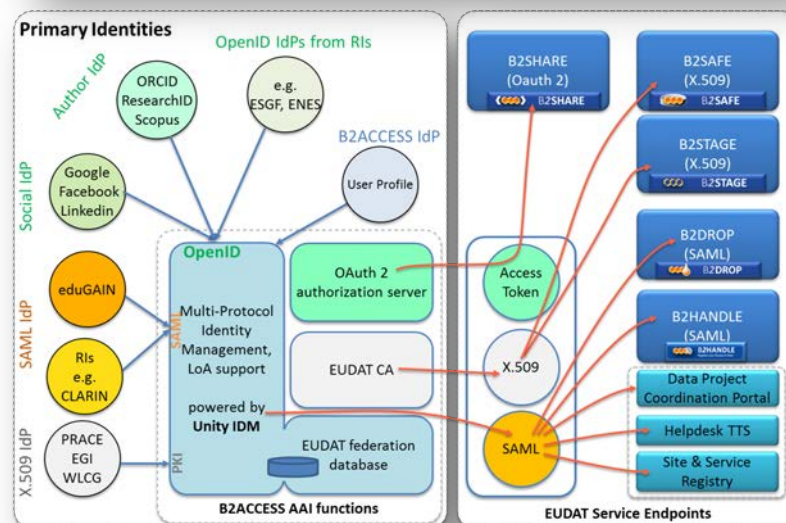


- Who

  - Anyone wanting to use the B2 Services
- What

  - Complies with **community ownerships** and **access rights**, basis of trust
  - Credential **conversion approach** (e.g. SAML, OpenID, X.509, Username/password)
  - Identity provider for **citizen scientists**
- Why

  - Use your own ID in federated environment







## ■ Create an account in B2ACCESS

- Go to: <https://b2access.eudat.eu/>
- Click on: Register a new account
- Create B2ACCESS user account (username only)
- Fill in the required information and click Submit

GO TO EUDAT WEBSITE

What is B2ACCESS User Guide FAQs Contact

Login to B2ACCESS user's account

[Register a new account](#)

Username:

Password:

[Forgotten password?](#)




Authenticate



Select your preferred authentication option

Log in with your B2ACCESS ID

[Password](#) [Certificate authentication](#)

Log in with your social ID

Log in with your organisational ID

A. T. Still University

AAF Virtual Home

AAI@EduHr Single Sign-On Service

Aalborg University

Aalto University

Aarhus Basic Health Care College

Aarhus School of Marine and Technical Engineering

Select a registration form to fill

[OAuth 2.0 Client Registration \(only for service integration\)](#)

[Create B2ACCESS user account \(username only\)](#)

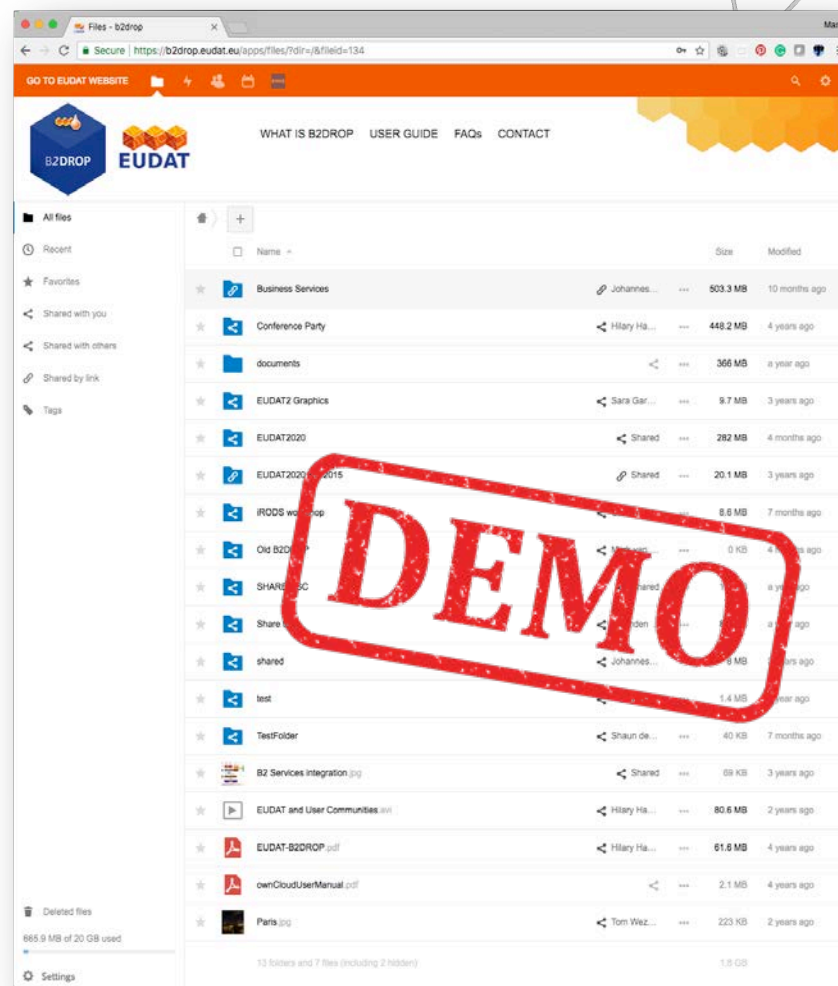
[Create B2ACCESS user account \(certificate + optional username\)](#)

[SDC Replication Managers Registration](#)

Close



- Who
  - Citizen scientists and small teams
- What
  - Store and exchange data
  - Synchronize multiple versions
  - Ensure automatic desktop synchronization
- Why
  - Ease of Use
  - Trusted European Service





## Who

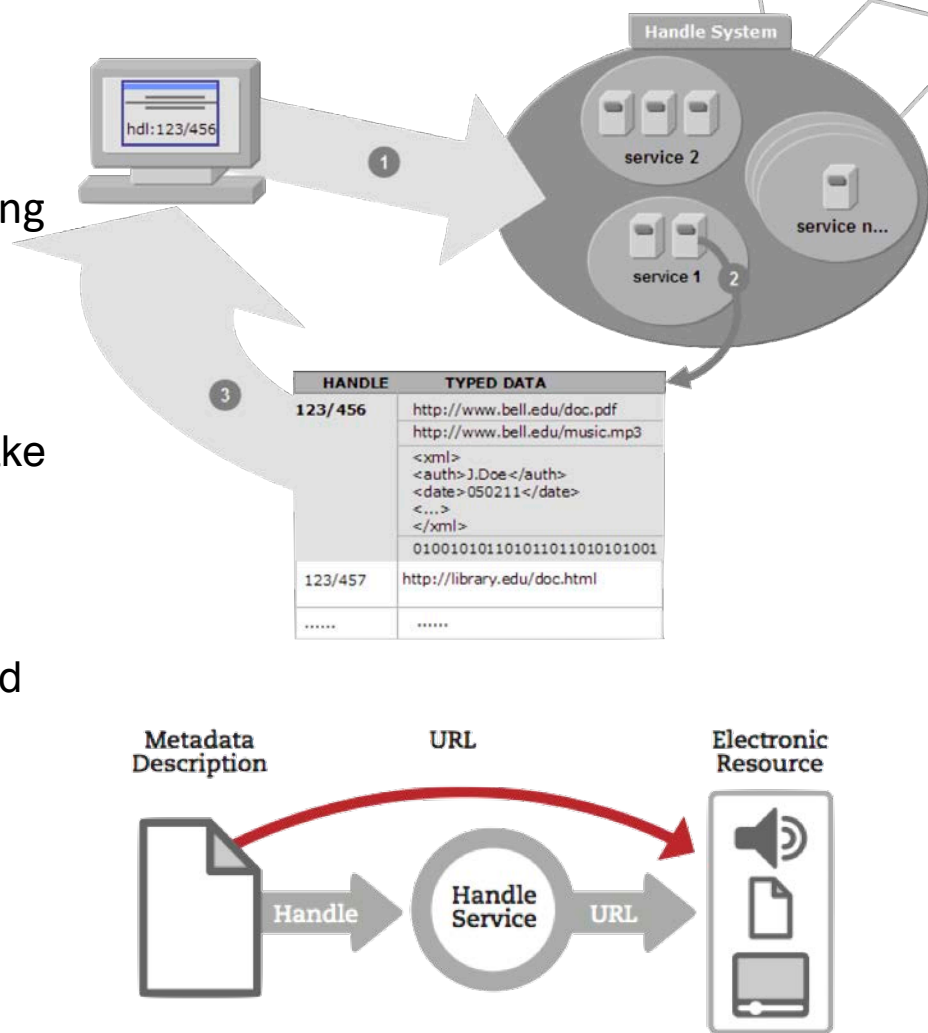
- Groups or communities who want to make their data **referenceable**, improving **data management tasks**

## What

- Follows **policies** to register data and make it **long term referenceable**
- Reliability through mutual **PID mirroring**
- Provides **abstraction layer** between a globally **unique persistent identifier** and **physical location** of data objects
- PIDs **global resolvable**

## Why

- Simple integration
- Technology Agnostic



## Who

- Small to Medium Teams

## What

- Store** data (incl. software) and add domain meta data
- Share** registered research data worldwide
- Preserve** (small-scale) research data for long-term

## Why

- Register Data for Publications (FAIR)
- Make known to wider community

The screenshot shows the B2SHARE website interface. At the top, there's a navigation bar with 'HELP', 'COMMUNITIES', 'UPLOAD', and 'CONTACT'. Below this, a search bar and a 'GO TO EUDAT WEBSITE' link are visible. The main content area displays a record titled 'Data Foundation and Terminology Work Group Products' by Gary Berg-Cross, Raphael Ritz, and Peter Wittenburg. The record includes an abstract, keywords, and a list of files. A large red 'DEMO' stamp is overlaid on the file list.

Name	Size	Access	True
Adoption of DFT Core Terms.pdf	4.1 KB	1 download	
maintaining the DFT Core Terms.pdf	9.1 KB	1 download	
DFT1 - overview.pdf	2.32 KB	1 download	
DFT1 - overview.pdf	2.32 KB	1 download	
DFT2 - Analysis and synthesis.pdf	836.56 KB	1 download	
DFT2 - metadata.xml	1.48 KB	1 download	
DFT3 - metadata.xml	2.08 KB	1 download	
DFT3 - snapshot of core terms.pdf	1.11 MB	1 download	
DFT4 - metadata.xml	1.97 KB	1 download	
DFT4 - Use Cases.pdf	1.95 MB	1 download	
DFT5 - metadata.xml	1.65 KB	1 download	
DFT5 - term tool.pdf	391.09 KB	1 download	
DFT Core.pdf	734.93 KB	12 downloads	

**Basic metadata**

**Contributors**

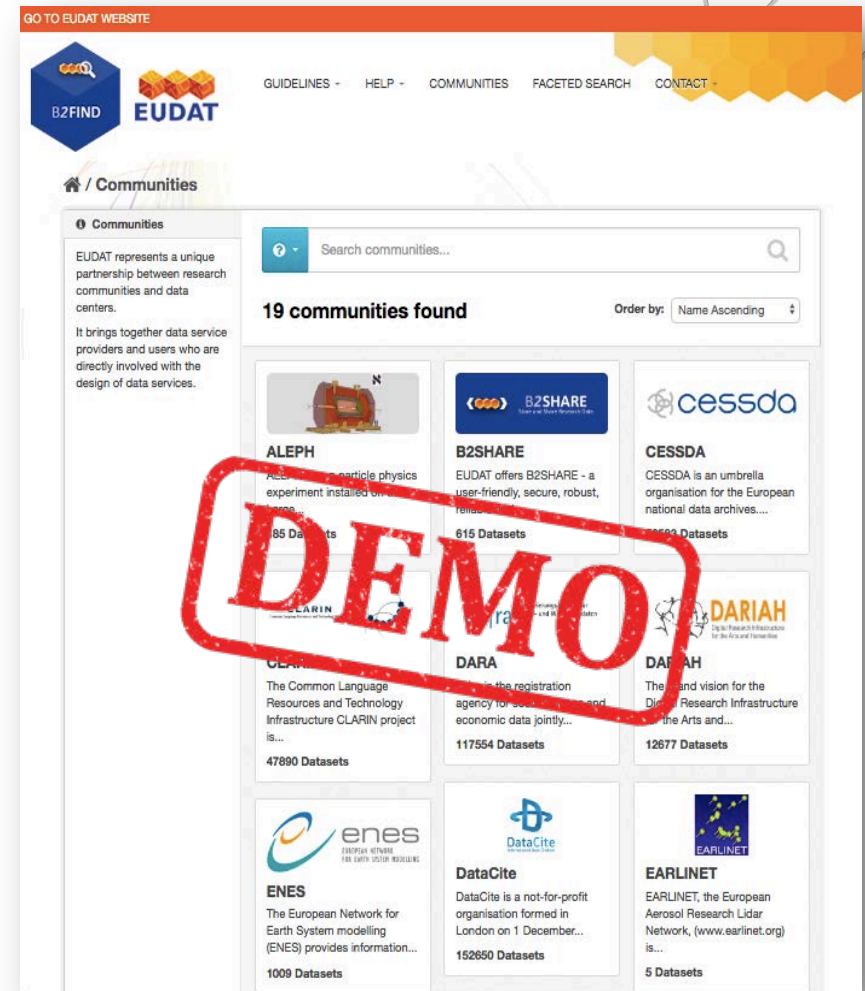
Name	Type
Stan Ahalt	Other
Gary Berg-Cross	Other
Jan Berg	Other
Dieter van de Sompele	Other
Jeffery Vin	Other
Chris Boone	Other
Isaac Bob	Other
Kahn, Larry	Other
Lannom, Michael	Other
Lautenschlager, Reagan	Other
Moore, Alberto	Other
Micheli, Hans	Other
Pfeifenberger, Raphael	Other
Ritz, Ulrich	Other
Schwardmann, Herbert	Other
van de Sompele, Dieter	Other
van Uytvenck, Dave	Other
Vieglas, Peter	Other
Wittenburg, Yunqiang	Other
Zhu, Herman	Other
Stehouwer, Thomas	Other
Zastrow	Other

**Resource Type** Category Text





- **Who**
  - Anyone
- **What**
  - **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**
- **Why**
  - Unique collection
  - Ease of Searching







## Who

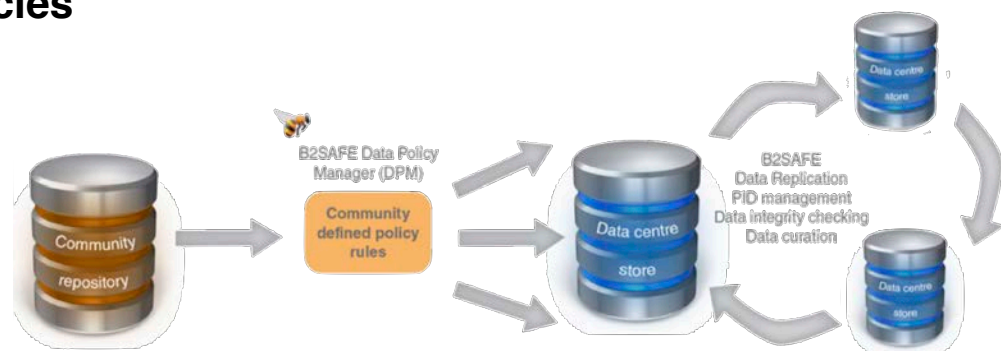
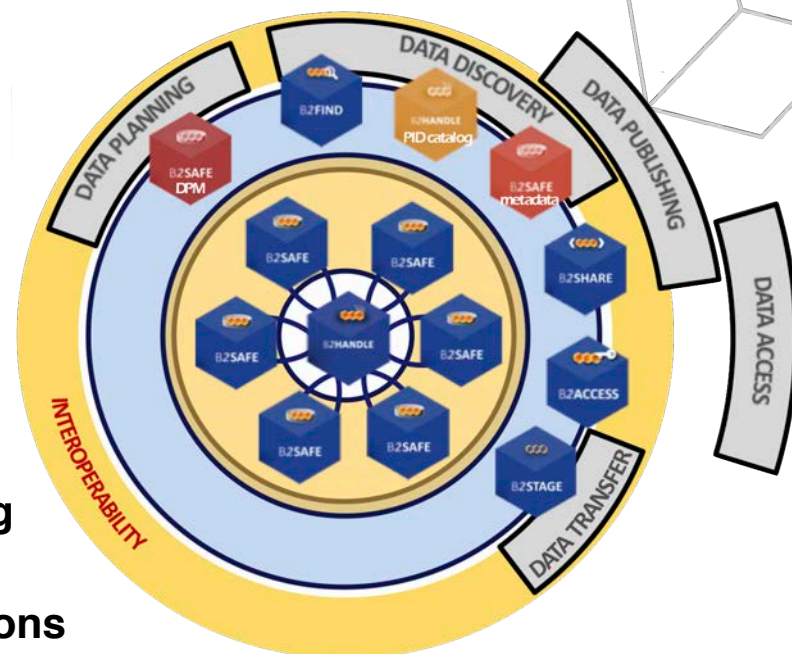
- Community Data Managers
- 'Sophisticated' Organizations

## What

- 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** and to **computing** resources
- Data management on basis of **policies**

## Why

- Performance
- Replication between trusted sites
- Data Preservation





## Who

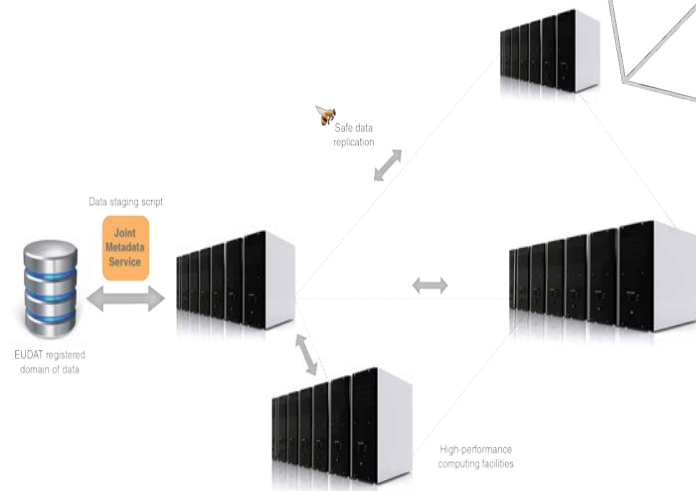
- Users and Communities who want to interact with EUDAT CDI services

## What

- Provide a **common access** layer to B2 services
- Copy large data sets, ingesting them** onto EUDAT data services
- Enables **data transfer** for large data collections from EUDAT storages to **external HPC facilities for processing**

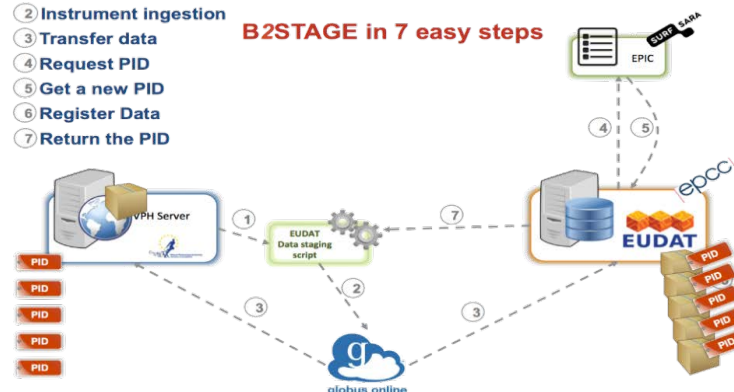
## Why

- Support data transfers between PRACE and EGI
- Simplify data transfers

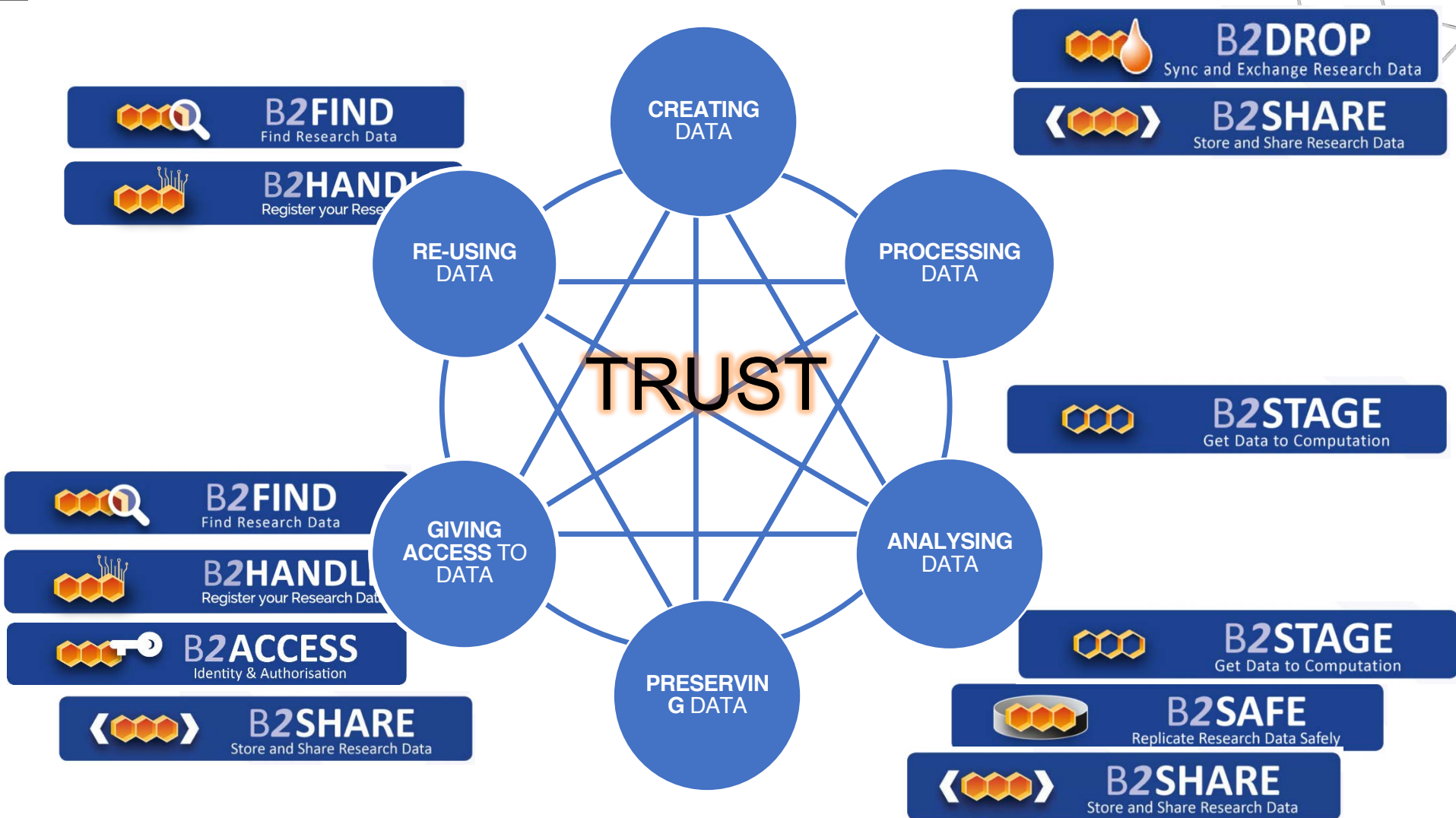


- 1 Ingest new data
- 2 Instrument ingestion
- 3 Transfer data
- 4 Request PID
- 5 Get a new PID
- 6 Register Data
- 7 Return the PID

### B2STAGE in 7 easy steps



Service Component	Development status	Version	Release Level	TRL level	Remark
B2SAFE-CORE	Production	4.1.0	Stable	9	
B2SAFE-DPM	Production	1.2.0	Stable	8	
B2SAFE-METADATA	Proof-of-Concept		Alpha	3	Local metadata store to manage structural metadata. No release defined in GitHub
B2SHARE	Production	2.1.0	Stable	9	
B2DROP	Production	12.0.4	Stable	9	B2DROP version is based on Nextcloud version
B2DROP-B2SHARE bridge	Production	1.0.0	Stable	8	
B2STAGE-GridFTP	Production	1.9.0	Stable	8	
B2STAGE-HTTP	Production	1.0.0	Stable	8	
B2HANDLE	Production	8.1.0	Stable	9	B2HANDLE version is based on Handle version.
B2HANDLE library	Production	1.1.1	Stable	8	
B2ACCESS	Production	1.9.6	Stable	9	B2ACCESS version is based on Unity-IDM version
B2FIND	Production	2.3.2	Stable	9	
B2NOTE	Production	1.0.0	Stable	8	
GEF	Pilot		Beta	6	
DATA DISTRIBUTION	Proof-of-Concept		Alpha	3	
WORKSPACE	Proof-of-Concept	0.4	Alpha	4	Prototype of the HTTP API for workspaces has been released.



## At **EUDAT** we are working to make our services **FAIR**

### EUDAT & FINDABLE

- 📦 **B2FIND**: multi-disciplinary metadata catalogue
- 📦 **B2HANDLE**: policy-based prefix & PID management
- 📦 **B2SHARE**: research data repository
- 📦 **B2SAFE**: policy-driven data management

### EUDAT & ACCESSIBLE

- 📦 **B2STAGE**: data staging service
- 📦 **B2SHARE**: research data repository
- 📦 **B2SAFE**: policy-driven data management
- 📦 **B2NOTE**: research data annotation

### EUDAT & INTEROPERABLE

- 📦 **B2HANDLE**: policy-based prefix & PID management
- 📦 **B2STAGE**: data staging service
- 📦 **B2SHARE**: research data repository
- 📦 **B2SAFE**: policy-driven data management
- 📦 **B2FIND**: multi-disciplinary metadata catalogue

### EUDAT & REUSABLE

- 📦 **B2SHARE**: research data repository
- 📦 **B2SAFE**: policy-driven data management
- 📦 **B2NOTE**: research data annotation



Engage	Deploy	Use
<i>For Community Decision-Makers &amp; Data Managers</i>	<i>For Systems and Support Engineers</i>	<i>For Researchers and End-Users</i>
<ul style="list-style-type: none"> <li>EUDAT Primer</li> </ul>	<ul style="list-style-type: none"> <li>EUDAT Primer</li> </ul>	<ul style="list-style-type: none"> <li>EUDAT Primer</li> </ul>
Services	Services	Services
<ul style="list-style-type: none"> <li>B2FIND</li> <li>B2STAGE</li> <li>B2SAFE <ul style="list-style-type: none"> <li>What is B2SAFE</li> <li>Using B2SAFE</li> <li>Joining B2SAFE</li> </ul> </li> <li>B2HANDLE</li> <li>B2SHARE</li> <li>B2DROP</li> <li>Use B2HOST</li> <li>Join B2HOST</li> </ul>	<ul style="list-style-type: none"> <li>B2FIND Integration</li> <li>B2STAGE Administration</li> <li>B2SAFE Configuration <ul style="list-style-type: none"> <li>iRODS Deployment <ul style="list-style-type: none"> <li>MPI-PL, SURFSara, RZG iRODS Zone Federation</li> <li>The dCache to iRODS connection at SURFSara</li> </ul> </li> </ul> </li> <li>B2HANDLE for Communities</li> <li>B2SHARE Deployment</li> <li>B2ACCESS Management</li> <li>B2ACCESS Service Integration</li> </ul>	<ul style="list-style-type: none"> <li>B2FIND Usage</li> <li>B2STAGE</li> <li>B2SHARE <ul style="list-style-type: none"> <li>B2SHARE Usage</li> <li>B2SHARE API</li> </ul> </li> <li>B2DROP <ul style="list-style-type: none"> <li>Publish from B2DROP to B2SHARE</li> </ul> </li> <li>B2HANDLE for end-users</li> <li>B2ACCESS Usage</li> </ul>
Tools	Tools	Tools
<ul style="list-style-type: none"> <li>License Selector</li> <li>Monitoring information</li> </ul>	<ul style="list-style-type: none"> <li>Monitoring for Operators</li> <li>Resource Coordination Tool</li> <li>Site and Service Registry Administration</li> </ul>	<ul style="list-style-type: none"> <li>License Selector</li> <li>Monitoring information</li> </ul>

- Total 33 documents maintained and revised
- 3 levels of documentation:
  - Engage: for Community decision-makers and data managers
  - Deploy: for system and support engineers
  - Use: for researchers and end users
- Participation from community experts





## EUDAT and the research data lifecycle


The EUDAT offer:


### The EUDAT B2services suite overview


### How to use & deploy the EUDAT services


 B2DROP, the EUDAT's  
Personal Cloud Storage  
Service

 How to share and  
store research data us-  
ing B2SHARE

 Finding data objects  
and collections through  
a web discovery portal:  
B2FIND

 Implementing data  
management policies  
trustworthy manner:  
B2SAFE

 Shifting large  
amounts of data with  
B2STAGE

 The Authentication  
and Authorization plat-  
form: B2ACCESS

### How to manage Persistent Identifiers: B2HANDLE

### All you need to know about copyright, sui generis database and personal data

### How to put the FAIR principles into practice

### About metadata

### Research Data Management

- Total of 14 training modules developed and maintained
- Hands-on training environments for:
  - B2SAFE
  - B2SHARE
  - B2FIND
  - B2HANDLE
  - B2NOTE



# Demo data pipelines and workflows

### Safe data replication with B2SAFE

- CompBioMed is a European commission H2020 funded Centre of Excellence
- Focus on the use and development of computational methods for biomedical applications.
- Data-intensive research
- More than 40 international and associate partners



**Safe data replication and large data transfer is one of the major requirements within the CompBioMed community**

<https://www.compbiomed.eu/>

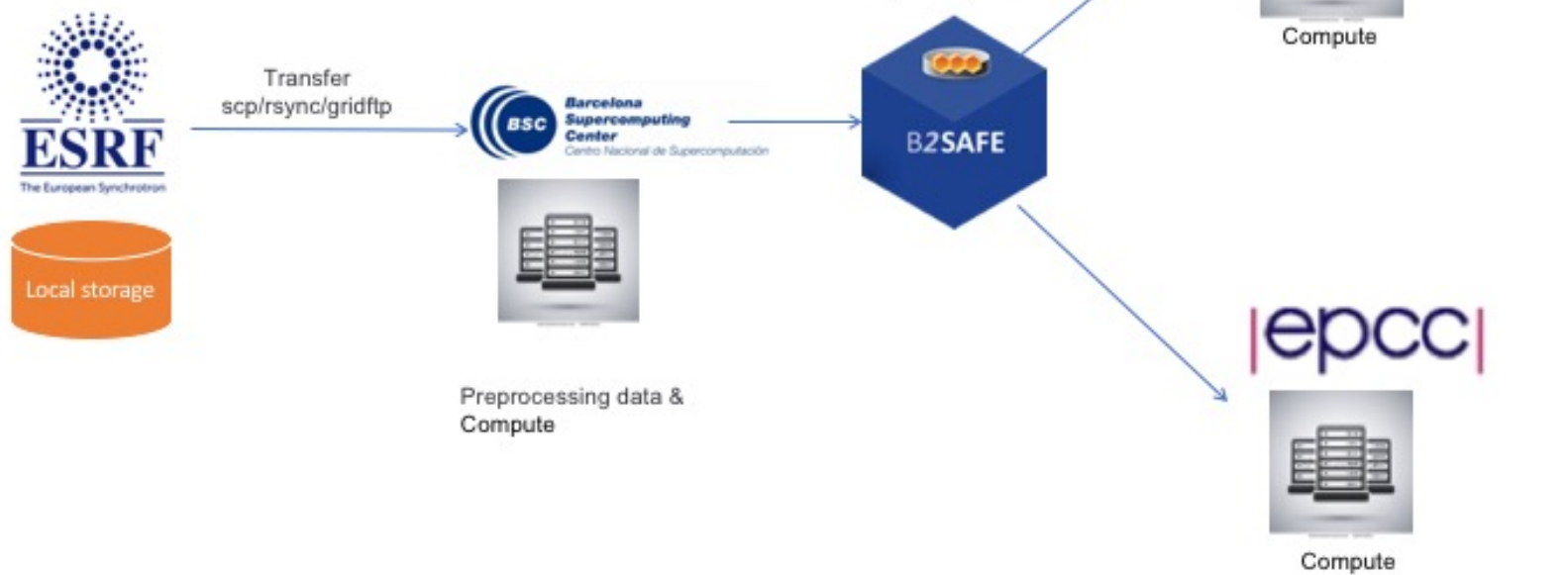
---

## Resources

**Service:** EUDAT B2SAFE service

**HPC Centers:** BSC, SURFsara, EPCC

**Resources:** allocation of at least 24 TB storage at each of the HPC centers



- B2DROP, B2SHARE, B2FIND
- Data publication workflow (B2DROP-B2SHARE integration)
- Data discovery and download (B2FIND-B2SHARE integration)

### Question:

Have you been able to create an account in B2ACCESS?

If yes, try to log into:

<https://b2drop.eudat.eu/>



## Q&A

To pose a question, you can write your question  
in the “Questions” tab

Or

Send an email to: [narges.zarrabi@surfsara.nl](mailto:narges.zarrabi@surfsara.nl)





## Thank you for participating!

## ...don't forget to fill in our feedback questionnaire...

Visit the CompBioMed website ([www.compbiomed.eu/training](http://www.compbiomed.eu/training))  
for a full recording of this and other webinars,  
to download the slides  
and to keep updated on our upcoming trainings

