

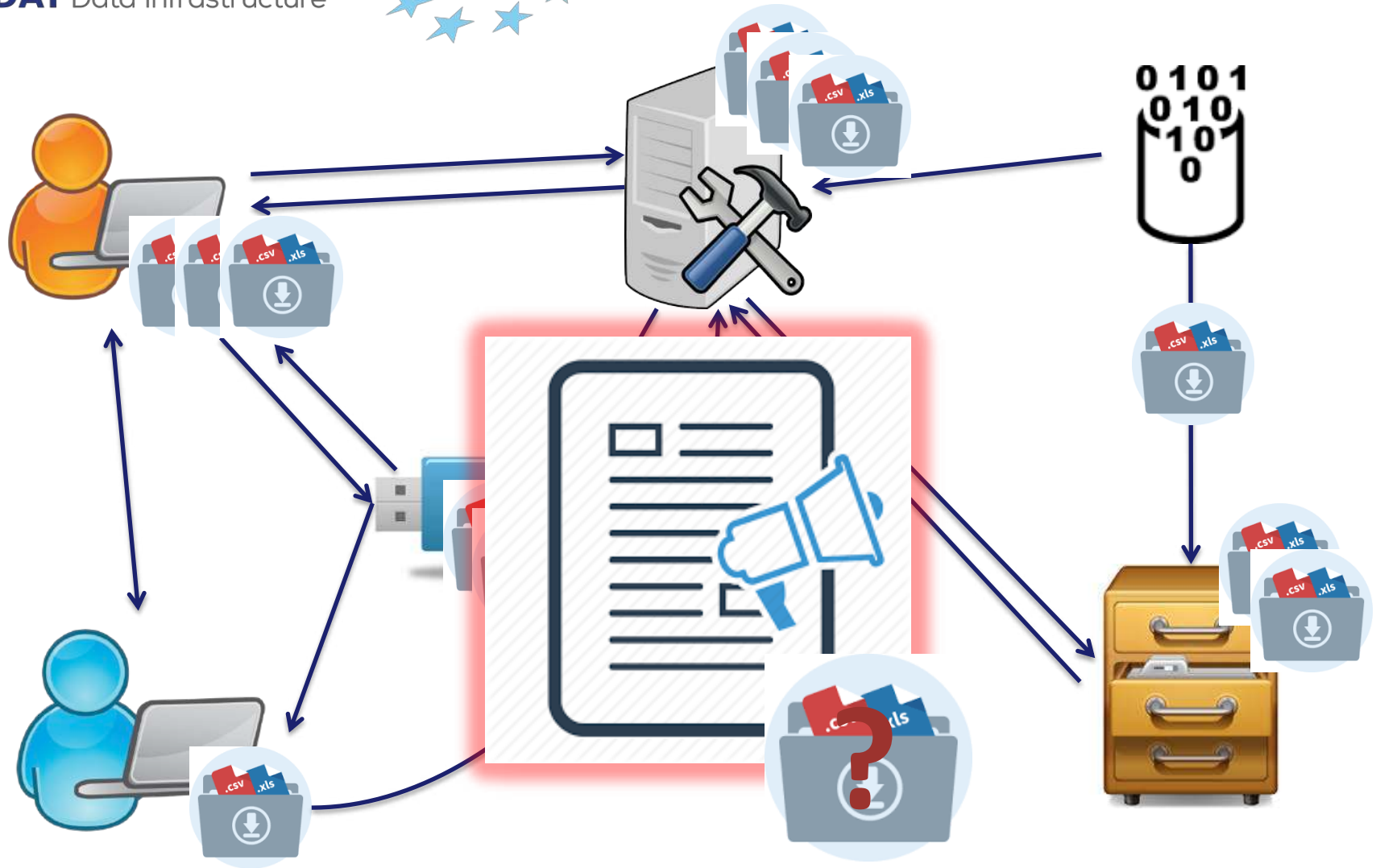


## Data Sharing and Publishing

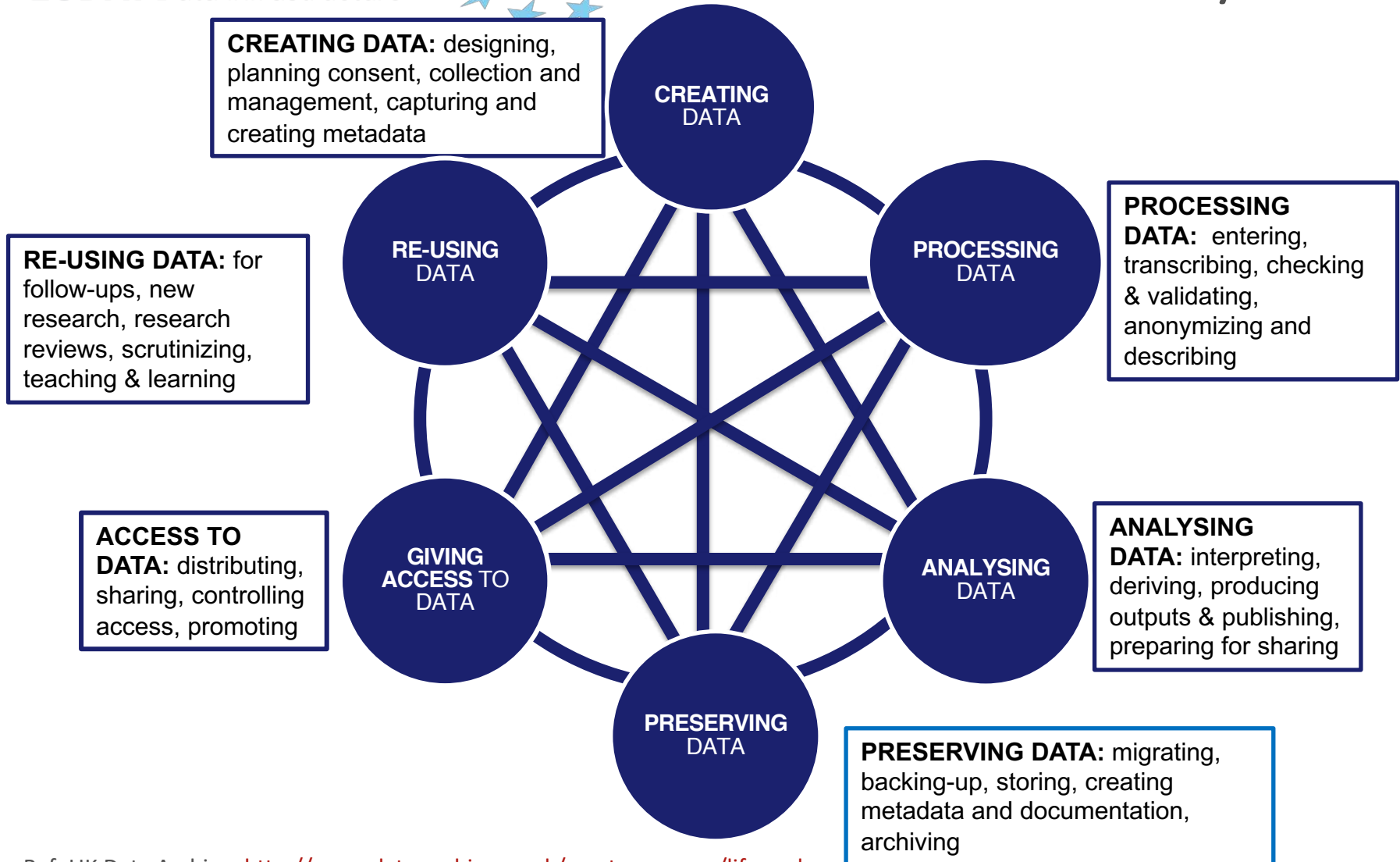
*Sara Ramezani*    *SURFSARA*

---

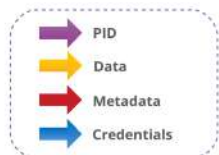
# Research Data- Where is the challenge?



# The Data Life Cycle



Ref: UK Data Archive: <http://www.data-archive.ac.uk/create-manage/life-cycle>



Data discovery



Data access & sharing



Data management & preservation



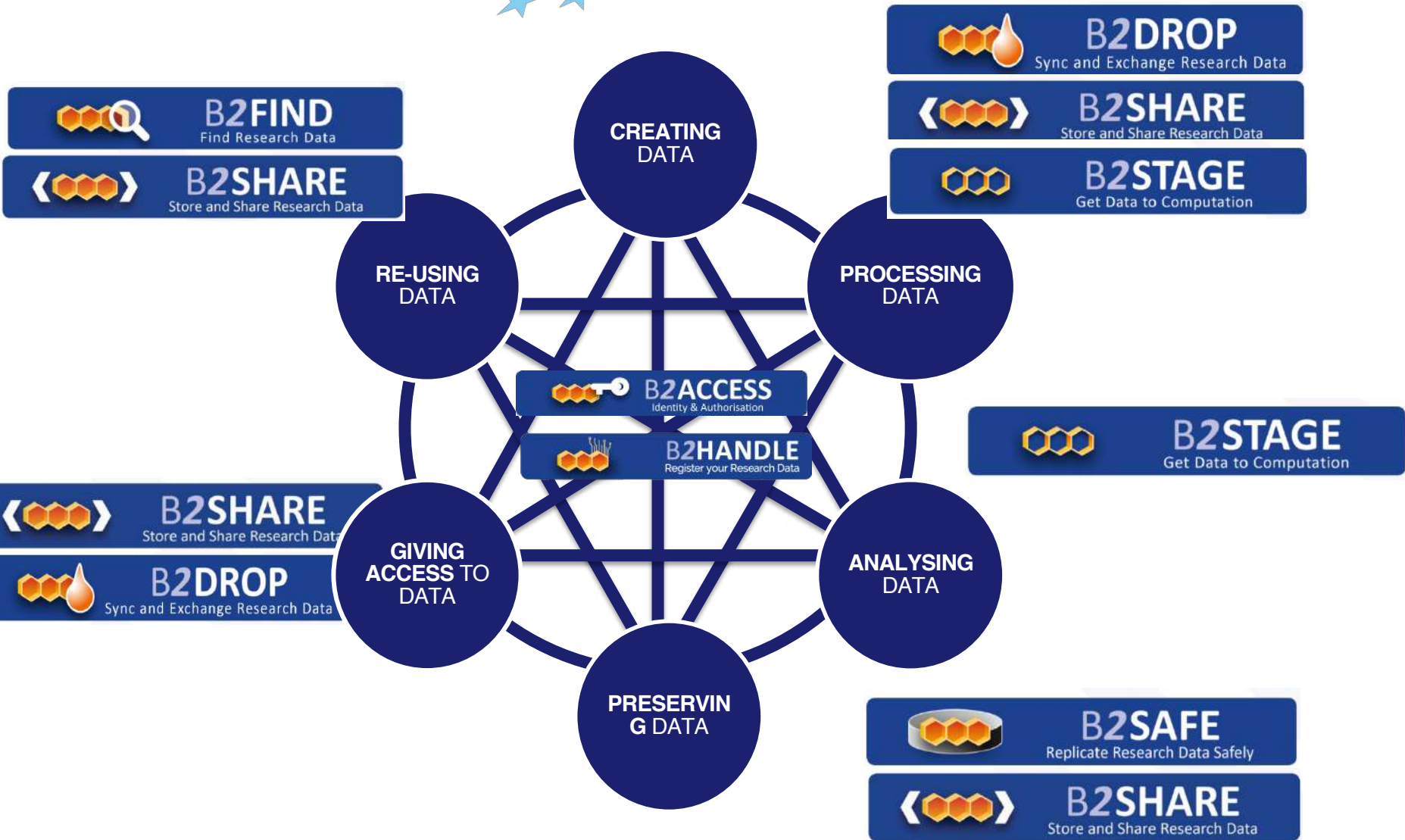
User management



**EUDAT**  
Domain

External domain





## The FAIR Data Principles

- **Findable**
  - F1. (Meta)data are assigned a globally unique and persistent identifier
  - F2. Data are described with rich metadata (defined by R1 below)
  - F3. Metadata clearly and explicitly include the identifier of the data they describe
  - F4. (Meta)data are registered or indexed in a searchable resource
- **Accessible**
  - A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
    - A1.1 The protocol is open, free, and universally implementable
    - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
  - A2. Metadata are accessible, even when the data are no longer available
- **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
- **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

- B2DROP
- B2HANDLE
- B2SAFE
- B2STAGE
- B2SHARE
- B2NOTE
- B2FIND

## ■ How EUDAT services link to data lifecycle

- How EUDAT services support the FAIR principles
  - Helping scientists to generate FAIR data

Who

Anyone wanting to use the B2 Services

What

Complies with **community ownerships** and **access rights**, basis of trust

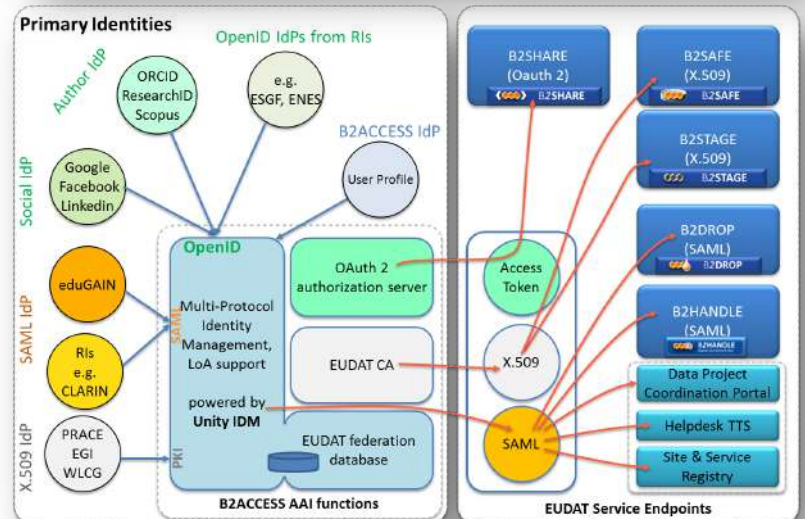
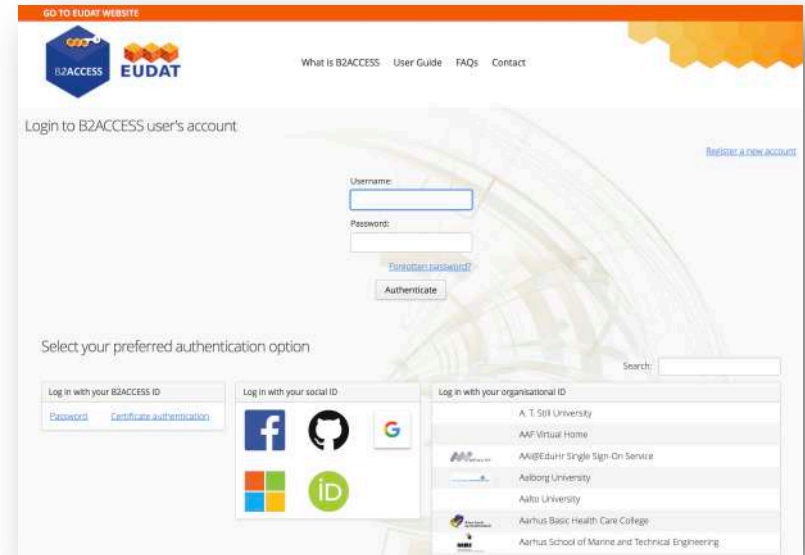
Credential **conversion approach** (e.g. SAML, OpenID, X.509, Username/password)

Identity provider for **citizen scientists**

Why

Use your own ID in federated environment

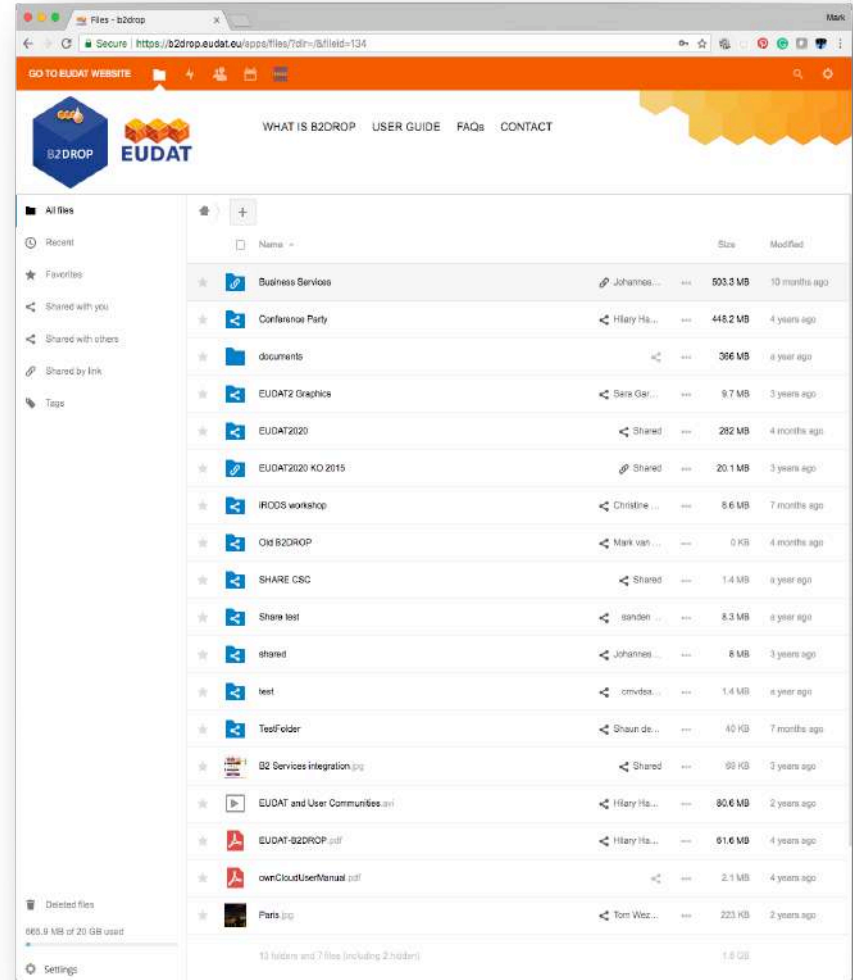
<https://b2access.eudat.eu/>





- 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

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



## Who

- Small to Medium Teams

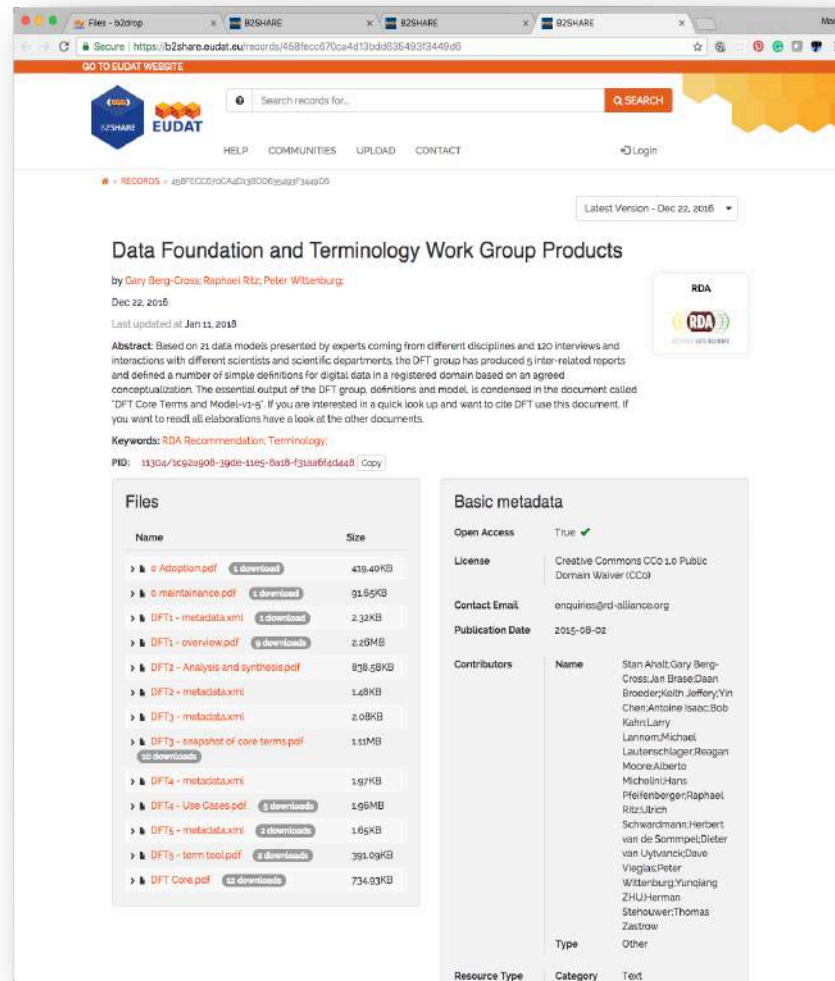
## What

- 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

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



The screenshot shows the B2SHARE interface for a specific record. The record title is "Data Foundation and Terminology Work Group Products" by Gary Berg-Cross, Raphael Ritz, and Peter Wittenburg, dated Dec 22, 2016. The abstract describes the work of the DFT group in producing scientific definitions and model. The interface includes a search bar, navigation links (HELP, COMMUNITIES, UPLOAD, CONTACT), and a list of files for download. The files list includes items like "Adoption.pdf", "maintenance.pdf", and "DFT1 - metadata.xml". A "Basic metadata" section on the right shows details like "Open Access: True", "License: Creative Commons CC0 1.0 Public Domain Waiver (CC0)", and a list of contributors.

Name	Size
Adoption.pdf	418.49KB
maintenance.pdf	91.55KB
DFT1 - metadata.xml	2.32KB
DFT1 - overview.pdf	2.05MB
DFT2 - Analysis and synthesis.pdf	826.26KB
DFT2 - metadata.xml	1.48KB
DFT3 - metadata.xml	2.08KB
DFT3 - snapshot of core terms.pdf	1.11MB
DFT4 - metadata.xml	1.97KB
DFT4 - Use Cases.pdf	1.95MB
DFT5 - metadata.xml	1.65KB
DFT5 - term tool.pdf	391.09KB
DFT Core.pdf	734.93KB

Basic metadata																									
Open Access	True <input checked="" type="checkbox"/>																								
License	Creative Commons CC0 1.0 Public Domain Waiver (CC0)																								
Contact Email	enquiries@rd-alliance.org																								
Publication Date	2015-08-02																								
Contributors	<table border="1"> <thead> <tr> <th>Name</th> <th>Contributor</th> </tr> </thead> <tbody> <tr> <td>Stan Ahalt</td> <td>Gary Berg-Cross</td> </tr> <tr> <td>Jan Brase</td> <td>Dan Broeder</td> </tr> <tr> <td>Colin Jeffrey</td> <td>Yin Chen</td> </tr> <tr> <td>Antoine Isaac</td> <td>Bob Kahn</td> </tr> <tr> <td>Larry Lennom</td> <td>Michael Lauterschlager</td> </tr> <tr> <td>Reagan Moore</td> <td>Alberto Micholin</td> </tr> <tr> <td>Hans Pfeilenberger</td> <td>Raphael Ritz</td> </tr> <tr> <td>Ilch Schwachmann</td> <td>Herbert van de Sompele</td> </tr> <tr> <td>Dieter van Lyfvenick</td> <td>Dave Vieglac</td> </tr> <tr> <td>Peter Wittenburg</td> <td>Yunqiang ZHU</td> </tr> <tr> <td>Herman Stehouwer</td> <td>Thomas Zastrow</td> </tr> </tbody> </table>	Name	Contributor	Stan Ahalt	Gary Berg-Cross	Jan Brase	Dan Broeder	Colin Jeffrey	Yin Chen	Antoine Isaac	Bob Kahn	Larry Lennom	Michael Lauterschlager	Reagan Moore	Alberto Micholin	Hans Pfeilenberger	Raphael Ritz	Ilch Schwachmann	Herbert van de Sompele	Dieter van Lyfvenick	Dave Vieglac	Peter Wittenburg	Yunqiang ZHU	Herman Stehouwer	Thomas Zastrow
Name	Contributor																								
Stan Ahalt	Gary Berg-Cross																								
Jan Brase	Dan Broeder																								
Colin Jeffrey	Yin Chen																								
Antoine Isaac	Bob Kahn																								
Larry Lennom	Michael Lauterschlager																								
Reagan Moore	Alberto Micholin																								
Hans Pfeilenberger	Raphael Ritz																								
Ilch Schwachmann	Herbert van de Sompele																								
Dieter van Lyfvenick	Dave Vieglac																								
Peter Wittenburg	Yunqiang ZHU																								
Herman Stehouwer	Thomas Zastrow																								
Type	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

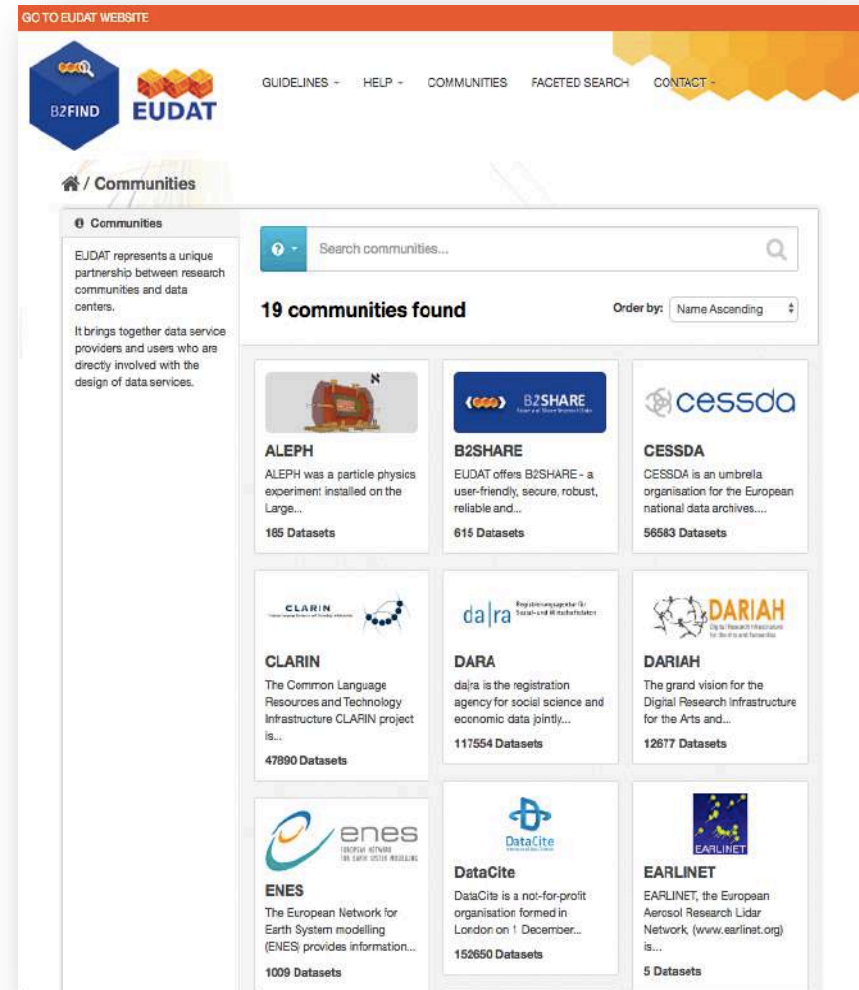
- Browse through collections using **standardized facets**

Why

- Unique collection

- Ease of Searching

<http://b2find.eudat.eu/>



GO TO EUDAT WEBSITE

B2FIND EUDAT

GUIDELINES - HELP - COMMUNITIES - FACETED SEARCH - CONTACT -

Home / Communities

Communities

EUDAT represents a unique partnership between research communities and data centers. It brings together data service providers and users who are directly involved with the design of data services.

Search communities...

19 communities found Order by: Name Ascending

Community	Description	Datasets
ALEPH	ALEPH was a particle physics experiment: installed on the Large...	165 Datasets
B2SHARE	EUDAT offers B2SHARE - a user-friendly, secure, robust, reliable and...	615 Datasets
CESSDA	CESSDA is an umbrella organisation for the European national data archives...	56583 Datasets
CLARIN	The Common Language Resources and Technology Infrastructure CLARIN project is...	47890 Datasets
DARA	da ra is the registration agency for social science and economic data jointly...	117554 Datasets
DARIAH	The grand vision for the Digital Research Infrastructure for the Arts and...	12677 Datasets
ENES	The European Network for Earth System modelling (ENES) provides information...	1009 Datasets
DataCite	DataCite is a not-for-profit organisation formed in London on 1 December...	152650 Datasets
EARLINET	EARLINET, the European Aerosol Research Lidar Network, (www.earlinet.org) is...	5 Datasets



## Who

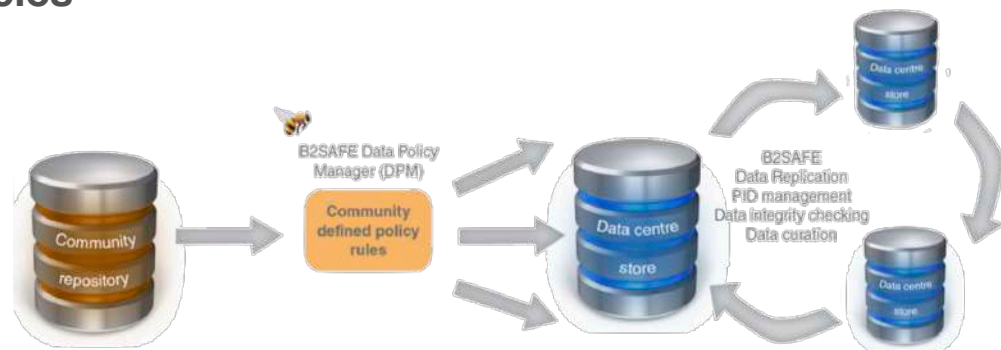
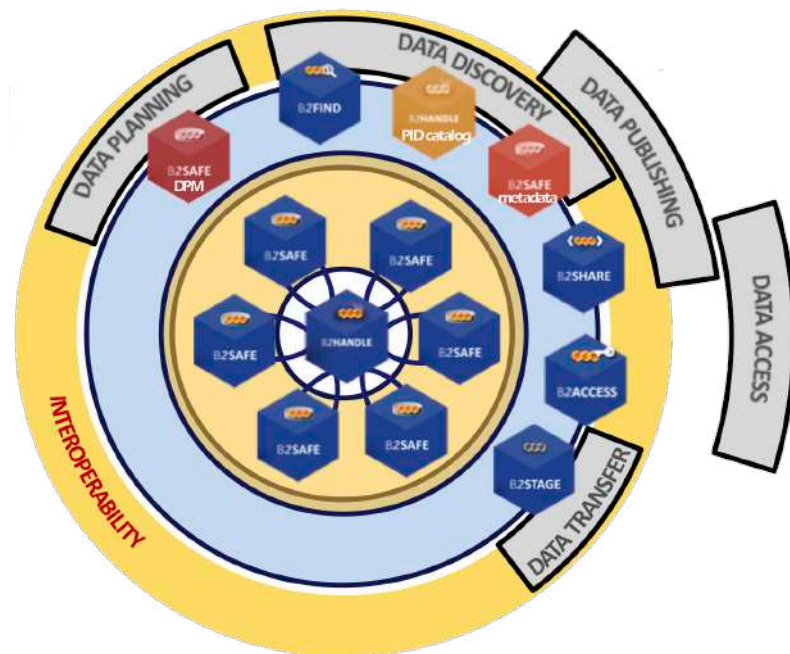
- Community Data Managers
- 'Sophisticated' Organizations

## 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

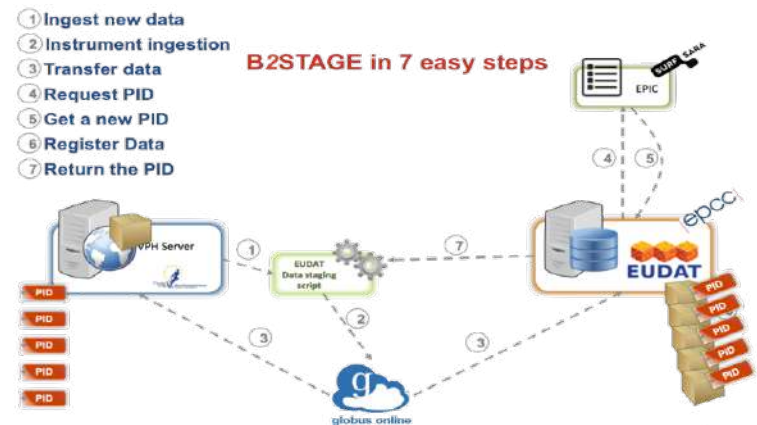
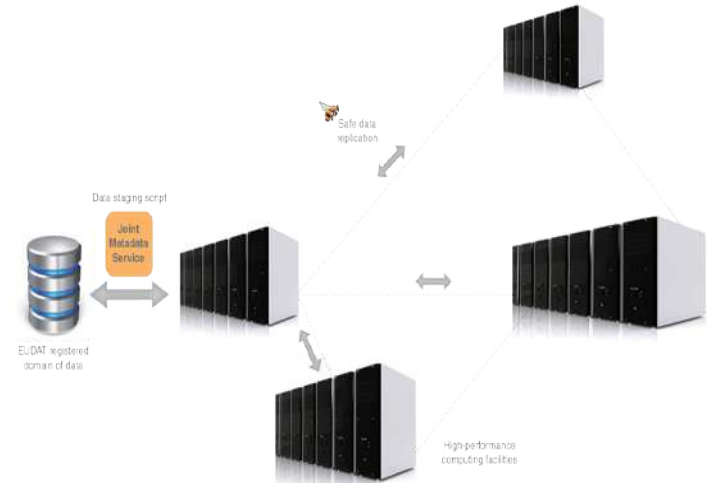


- Who

  - 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



<http://petstore.swagger.io/?url=https://b2stage.cineca.it/api/specs&docExpansion=none> - /

Who

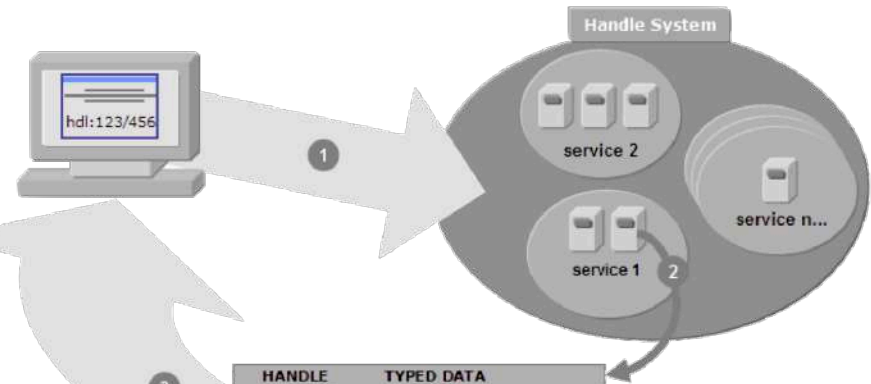
- Groups or communities who want to make their data **referenceable**, improving data management tasks

What

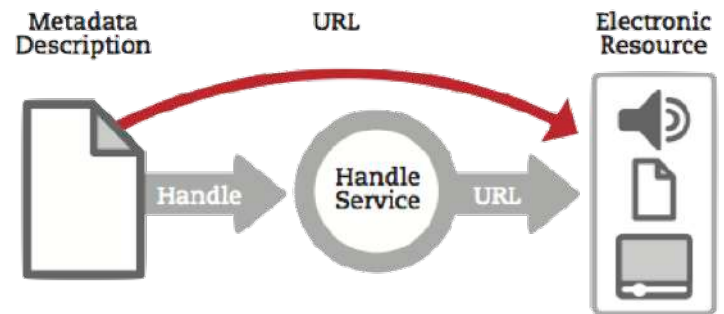
- Follows **policies** to register data and make it **long term referenceable**
- Reliability through mutual **PID mirroring**
- Provides **abstraction layer** between a globally **unique persistent identifier** and **physical location** of data objects
- PIDs **global resolvable**

Why

- Simple integration
- Technology Agnostic



HANDLE	TYPED DATA
123/456	http://www.bell.edu/doc.pdf http://www.bell.edu/music.mp3 <xml> <auth>J.Doe</auth> <date>050211</date> <...> </xml> 0100101011010110110101001
123/457	http://library.edu/doc.html
.....	.....





## Who

- Anyone

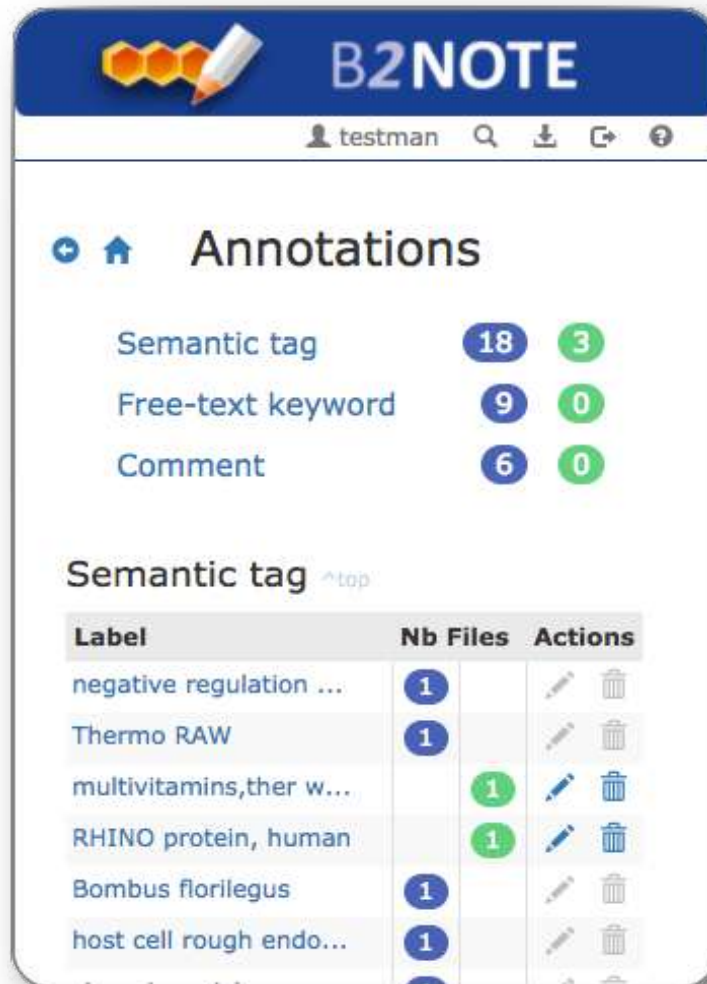
## What

- Enrich** data with Semantic tag, Free-text keyword or comment **without changing the data record**
- Share** annotations
- Manage** annotations
- Integrate with data repositories**

## Why

- Retrieve and aggregate** heterogeneous files from distributed sources on basis of annotations

<https://b2note.eudat.eu/>















**B2NOTE**

testman

### Annotations

Semantic tag	18	3
Free-text keyword	9	0
Comment	6	0

#### Semantic tag <sup>^top</sup>

Label	Nb Files	Actions
negative regulation ...	1	 
Thermo RAW	1	 
multivitamins,ther w...	1	 
RHINO protein, human	1	 
Bombus florilegus	1	 
host cell rough endo...	1	 

# Hands-on!

- B2DROP
- B2DROP -> B2SHARE publication workflow
- B2FIND -> B2SHARE discovery and download
- B2DROP -> CLARIN Switchboard example



# The CLARIN Community



- CLARIN is the European Research Infrastructure for Language Resources and Technology. It makes digital language resources and language technology available to researchers from all disciplines but especially the Humanities and Social Sciences
  - CLARIN is a partner in EOSC-hub and
    - Potentially contributes a huge amount of metadata for language resources and services to B2FIND.
    - Provides thematic service(s) : CLARIN Metadata Infrastructure
      - Virtual Language Observatory (VLO)
      - Virtual Collection Registry (VCR)
      - Language Resource Switchboard
  - These services should be integrated with other EOSC-hub services e.g.
    - CLARIN metadata visible in B2FIND
    - Language Resource Switchboard integrated in B2FIND and B2DROP
-



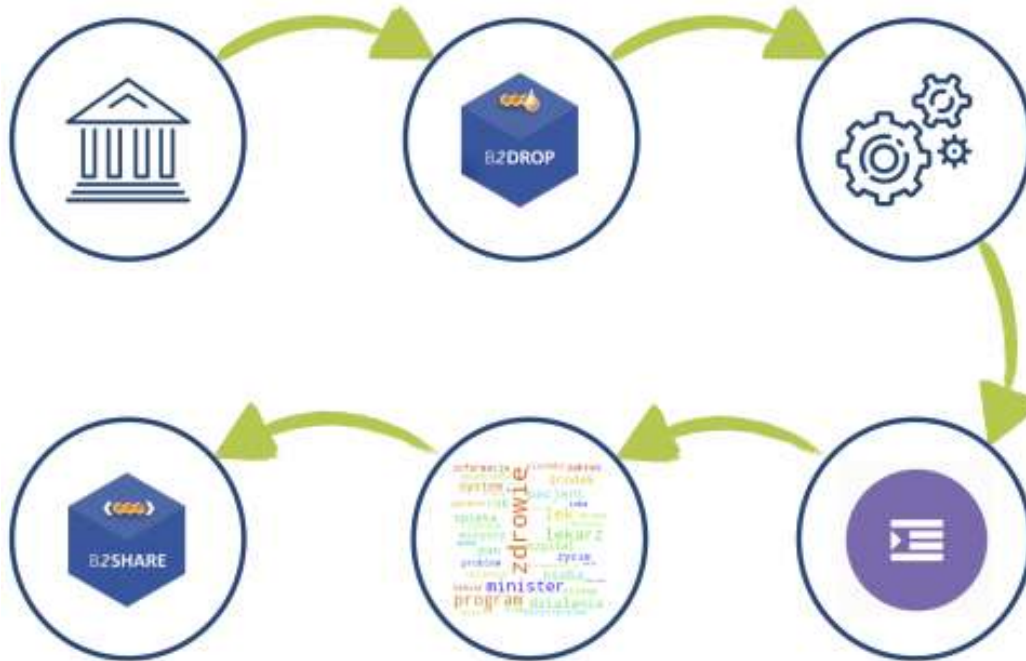
Collaborative  
Data Infrastructure



# The CLARIN Community

- CLARIN is the European Research Infrastructure for Language Resources and Technology. It makes digital language resources and language technology available to researchers from all disciplines but especially the Humanities and Social Sciences
- CLARIN is a partner in EOSC-hub and
  - Potentially contributes a huge amount of metadata for language resources and services to B2FIND.
  - Provides thematic service(s) : CLARIN Metadata Infrastructure
    - Virtual Language Observatory (VLO)
    - Virtual Collection Registry (VCR)
    - Language Resource Switchboard
- These services should be integrated with other EOSC-hub services e.g.
  - CLARIN metadata visible in B2FIND
  - Language Resource Switchboard integrated in B2FIND and B2DROP

- First step in CLARIN/EOSC-hub services integration
- Extensive demo available at <https://www.clarin.eu/eosc>





Thanks to:  
Narges Zarrabi  
SURFsara and EUDAT colleagues

**Please provide your feedback about the training session:**

**[https://www.surveymonkey.com/r/EOSC-hub week 01](https://www.surveymonkey.com/r/EOSC-hub_week_01)**