

EUDAT

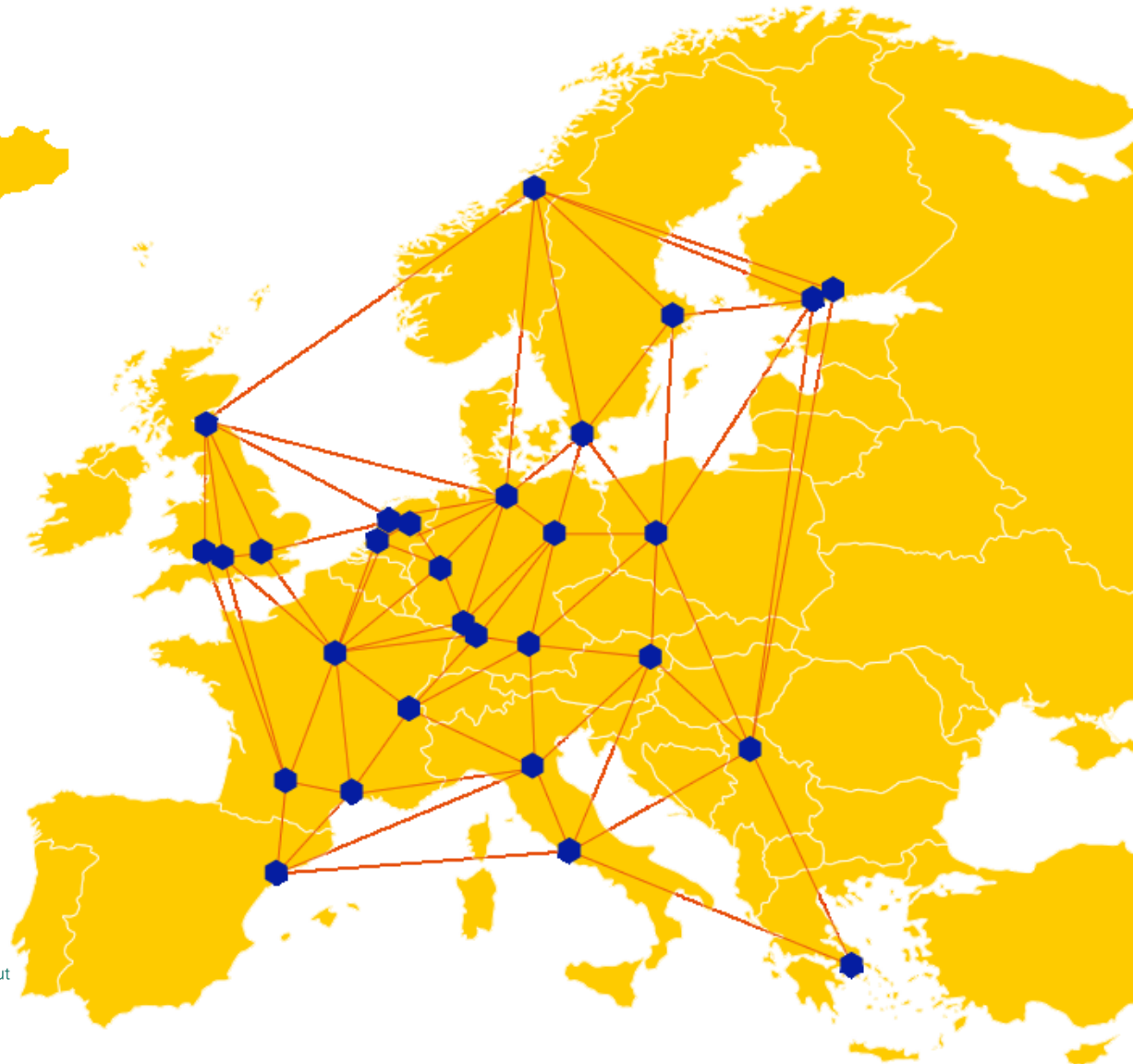
“Tieteelliset lehdet ja tutkimusdata 2017-02-01”

Ari Lukkarinen, CSC - IT Center for Science

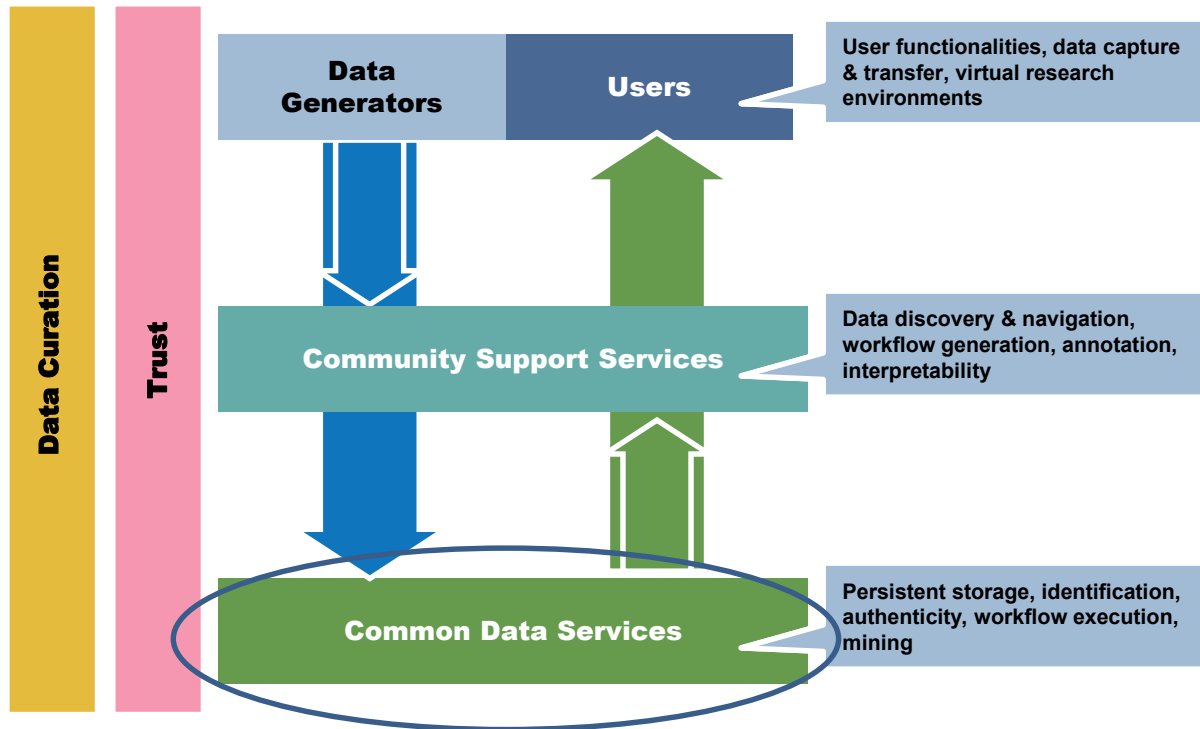
EUDAT2020 - 33 Partners



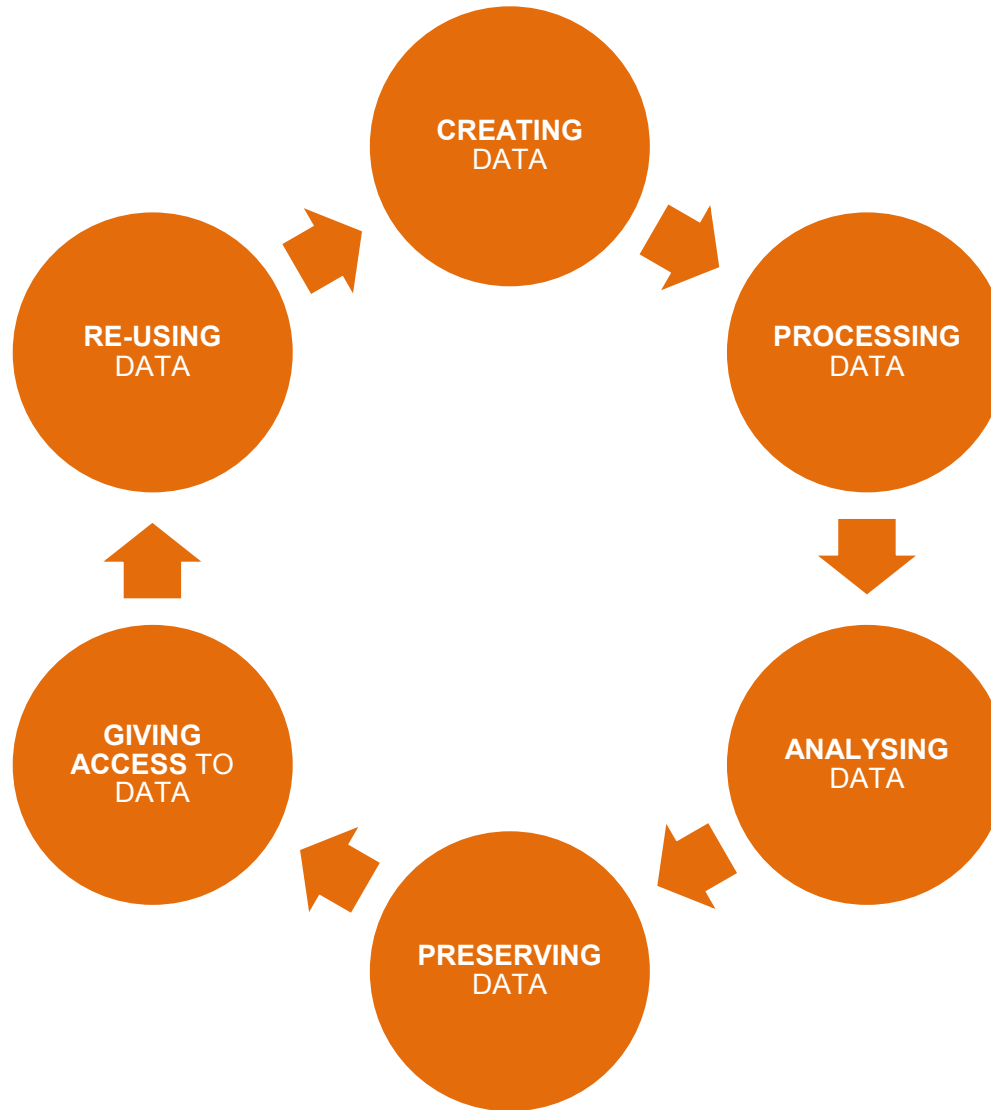
e-Science Data Factory



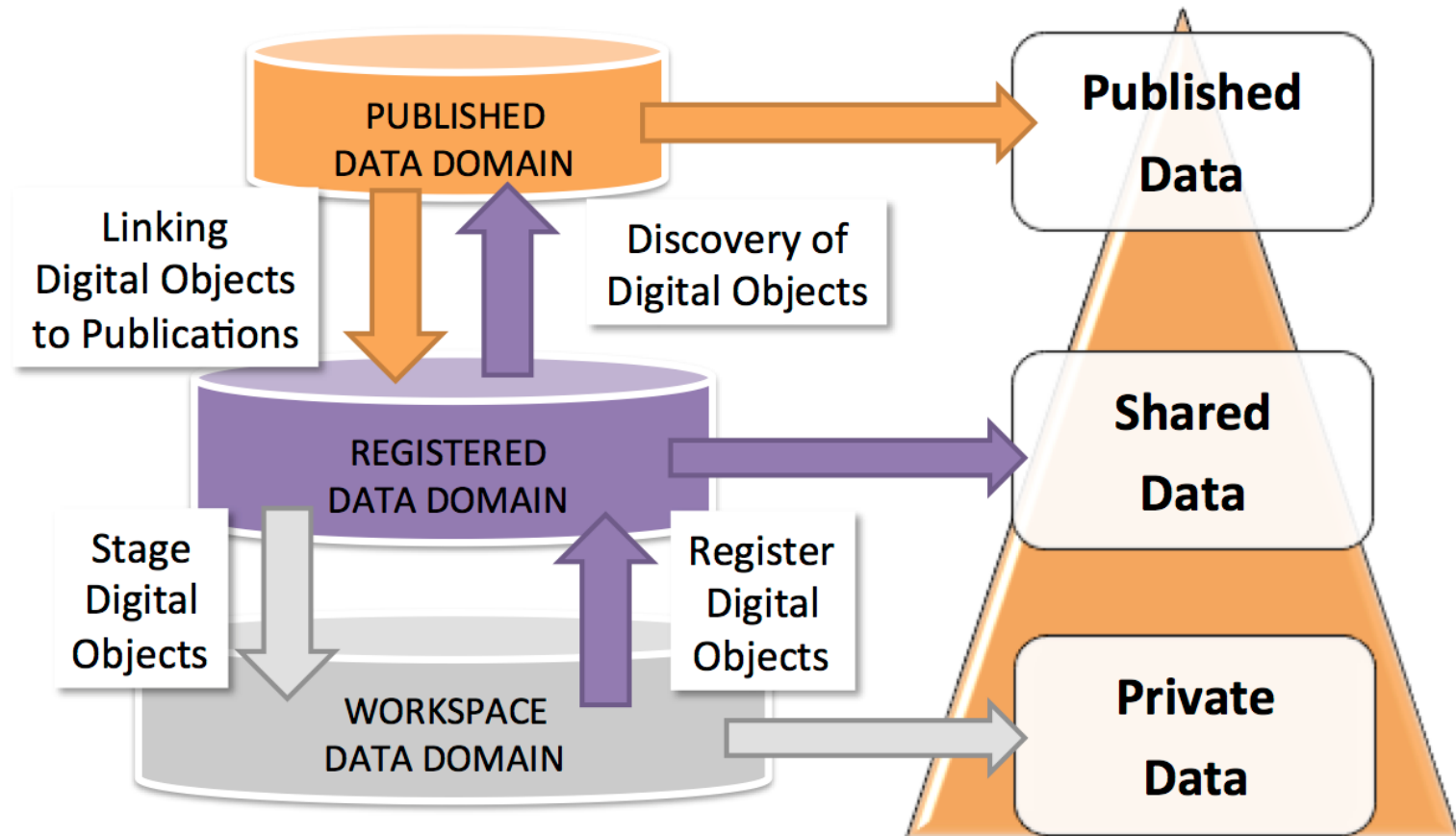
Collaborative Data Infrastructure



Research Data Lifecycle



EUDAT Data Domain



EUDAT Data Domain modeled on the ANDS¹ Data Curation Continuum

Building solutions with the communities



- Common Language Resources and Technology Infrastructure (CLARIN)



- European Network for Earth System Modelling (ENES)



- Distributed infrastructure for life-science information (ELIXIR)



- European Plate Observing System (EPOS) - Solid Earth sciences Research Infrastructure



- Integrated Carbon Observation System (ICOS) to quantify & understand greenhouse gas balance

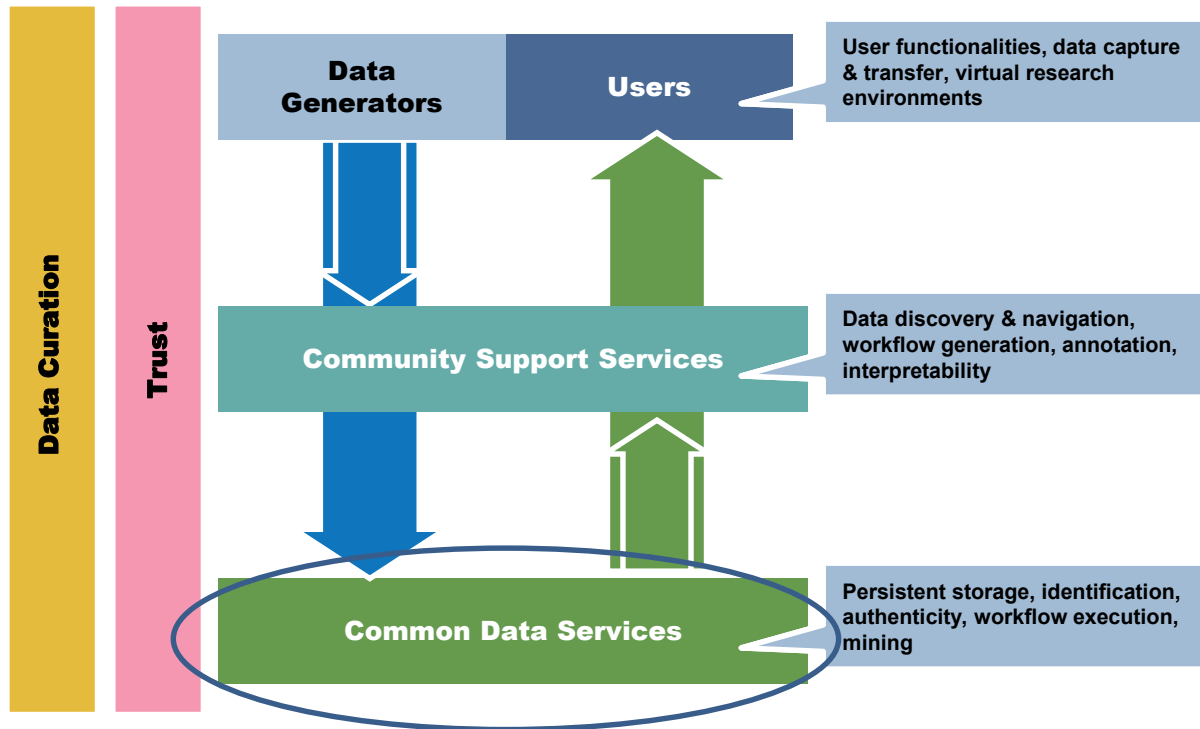


- Long-Term Ecosystem Research (LTER) in Europe

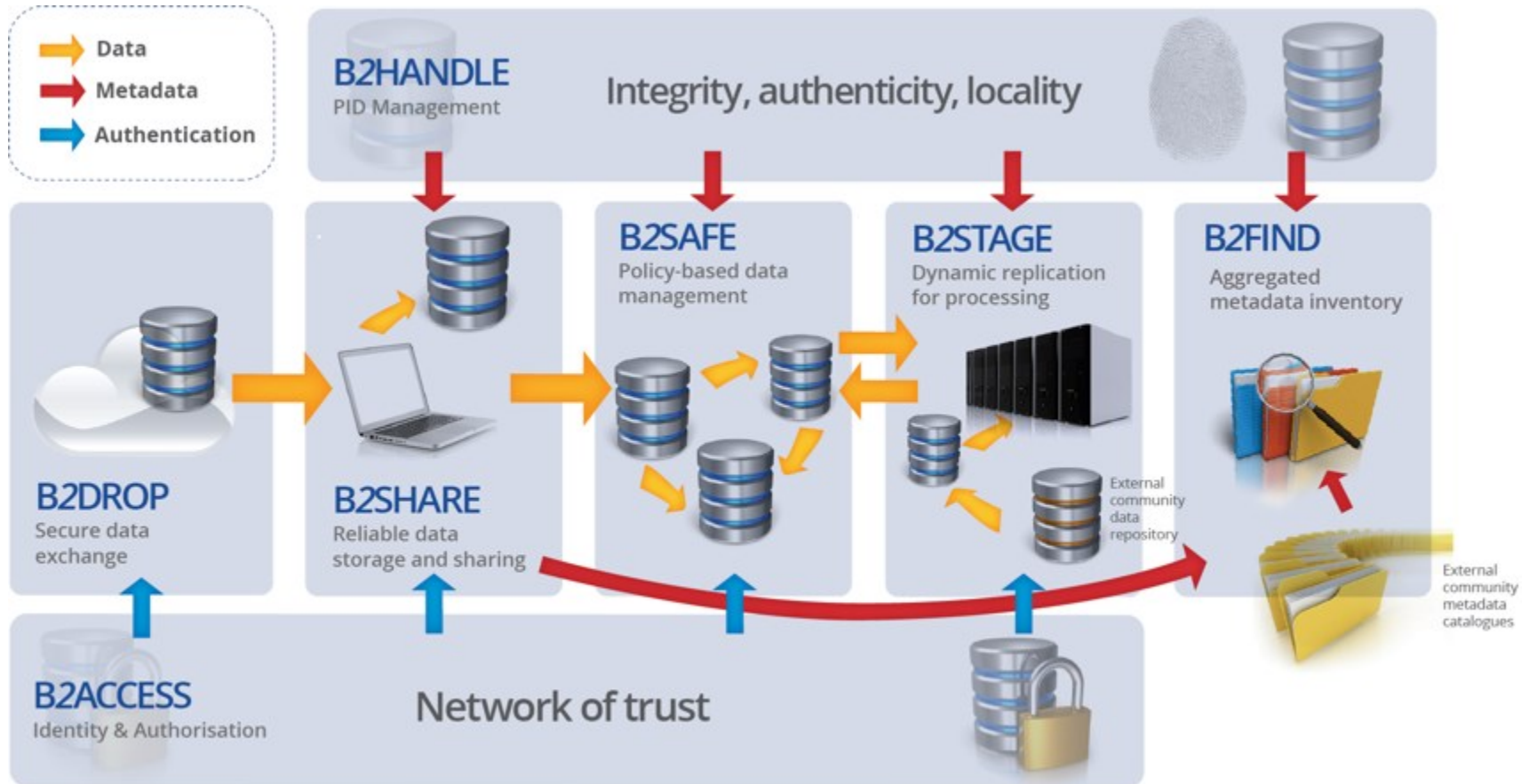


For more information visit - <https://eudat.eu/use-cases>

Collaborative Data Infrastructure



What is the EUDAT Service offer?



B2FIND
Find Research Data



B2STAGE
Get Data to Computation



B2HANDLE
Register your Research Data



B2SHARE
Store and Share Research Data



B2SAFE
Replicate Research Data Safely



B2DROP
Sync and Exchange Research Data



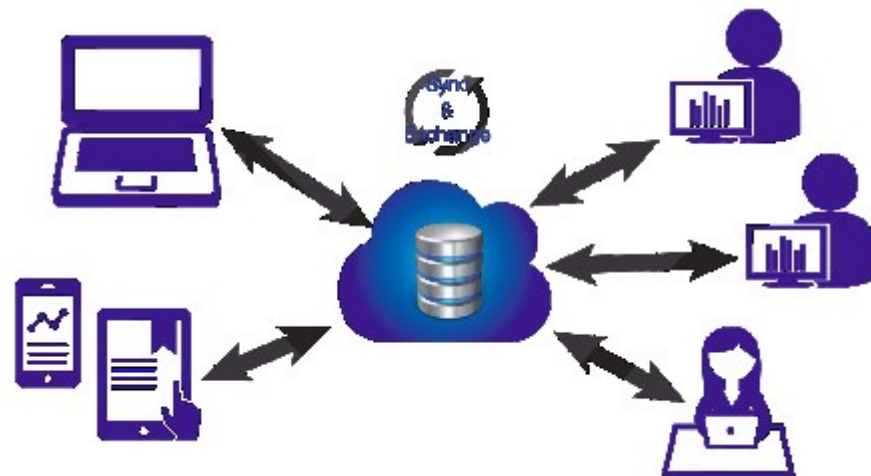
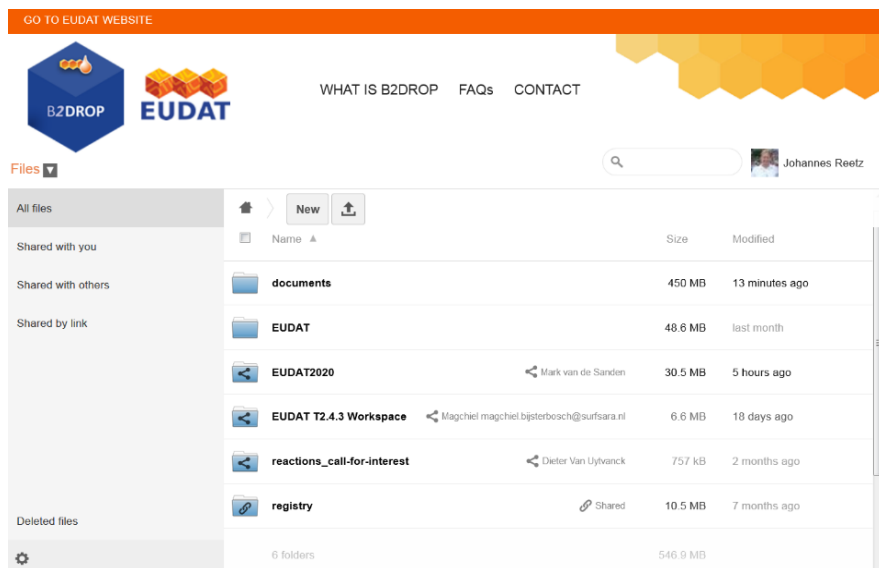
B2ACCESS
Identity & Authorisation



B2NOTE

DATA TYPE REGISTRY

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



EUDAT2020

- **Cloud Storage Federation**, collaboration with GEANT in OpenCloudMesh
- B2DROP as a **workspace** area to **computing facilities**
- **Integration** with EUDAT CDI (e.g. B2SHARE)

Who

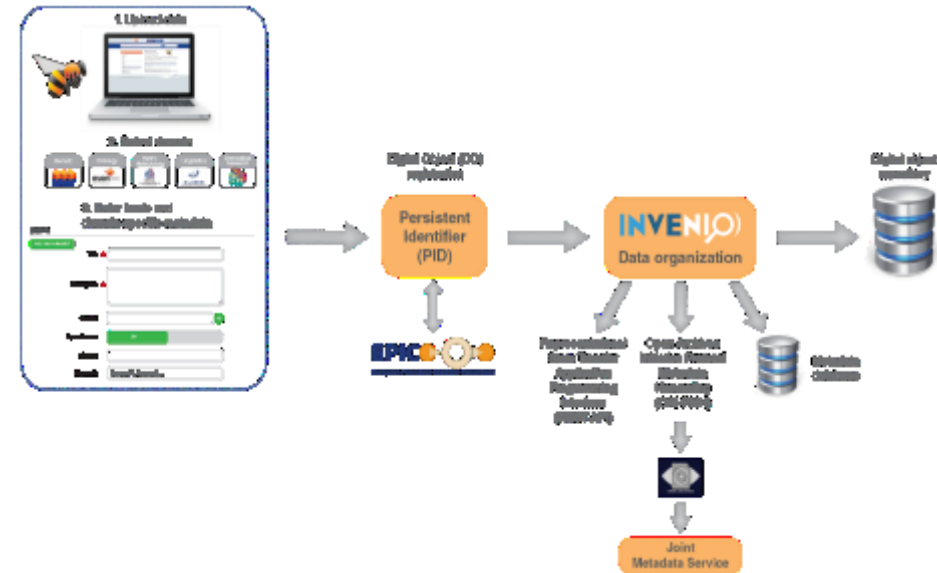
- Small to Medium Teams

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
- Make known to wider community



Select a domain or project

Linguistics CLARIN	Herbarium GHA	Biodiversity Biodiversity	Generic EUDAT	Ontology EUDAT	RNA RNA	Hydro-Meteorology DRIW	Biomedical Research Biomedical Research
-----------------------	------------------	------------------------------	------------------	-------------------	------------	---------------------------	--

Generic

Title * Title of the resource

Description *

Creator author

Open Access ☒

Embargo Till

Licence

Keywords keyword1, keyword2, ...

Contact Email contact email

Discipline None selected ▼

EUDAT2020

- Focus on **authorization**
- Data **versioning** and **annotation**
- Extended HTTP Restful API interface
- Easy installable **software package**
- Integration with other services**

Who

- Community Data Managers
- 'Sophisticated' Organisations

What

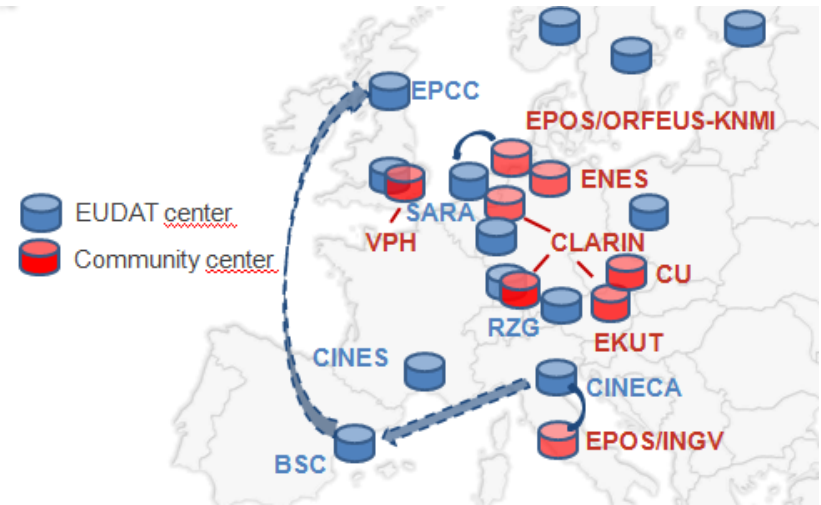
- Provide an abstraction layer which virtualizes large-scale data resources
- Optimize access** for users from different regions
- Bring data **closer to powerful computers**

Why

- Performance
- Replication between trusted sites
- Data Preservation



B2SAFE
Replicate Research Data Safely



EUDAT2020

- Support **metadata**
- Optimize and **extend** policies to support **data curation** and **provenance**
- Support **authorization** on basis of **community access rules**
- Integration with other EUDAT services

Who

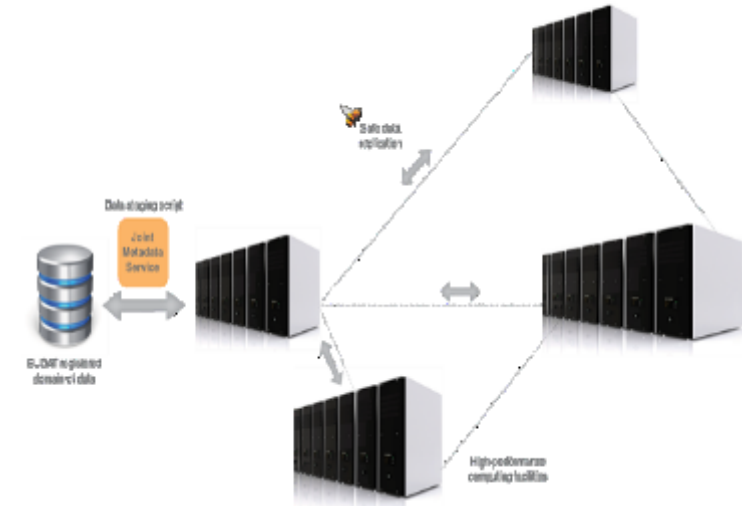
- Users and Communities with Significant Computational Needs

What

- Transfer** large data collections from EUDAT storages to **external HPC facilities for processing**
- Copy large data sets, ingesting them** onto EUDAT storage resources

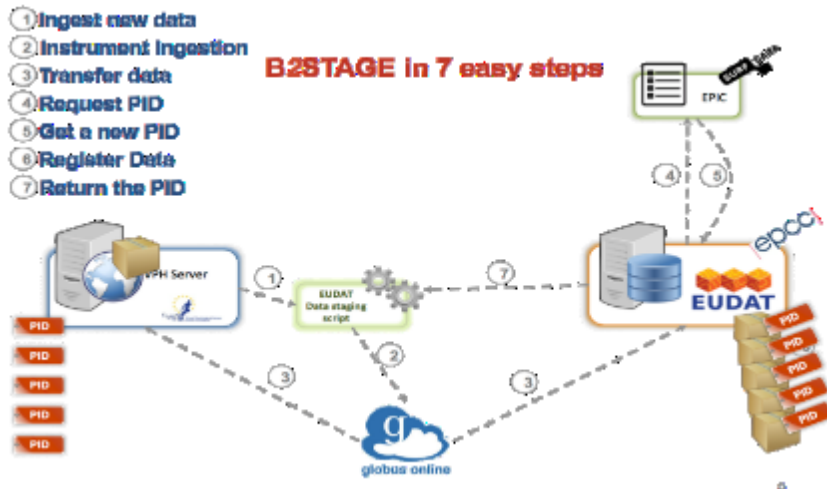
Why

- Integration/Collaboration with PRACE & EGI
- Simplify Data Transfer



EUDAT2020

- Further develop **HTTP** to a **mature interface** and extend functionality to **metadata**
- Extend EUDAT **client API** library to other **B2 services** (e.g. B2SHARE, B2FIND, PID)
- Further integration with B2ACCESS












- **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



Search communities...

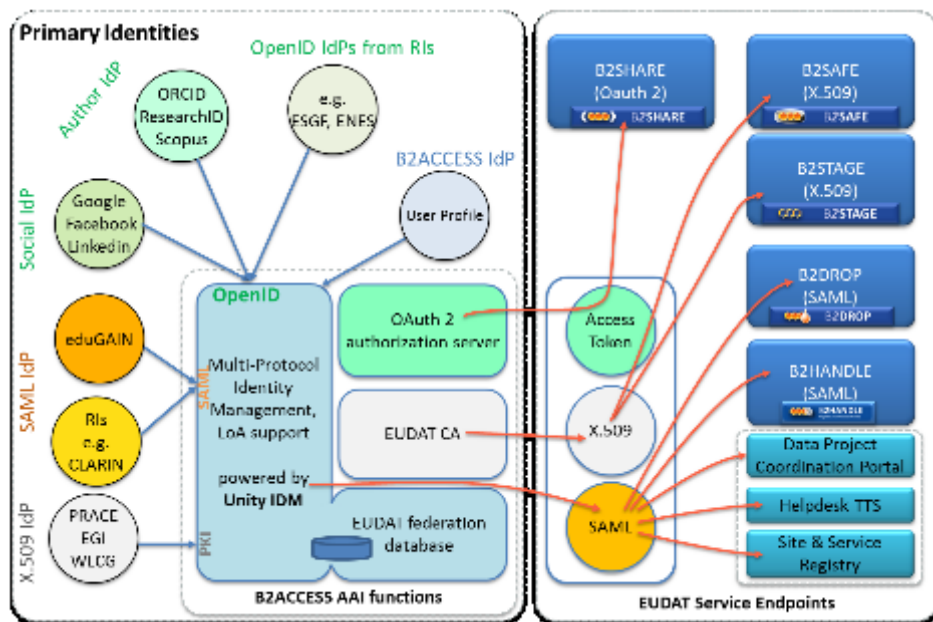
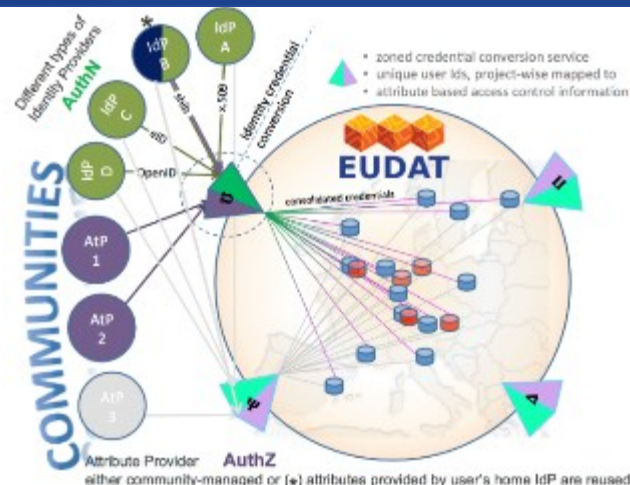
13 communities found Order by: Name Ascending

 <p>ALEPH ALEPH was a particle physics experiment installed on the Large... 185 Datasets</p>	 <p>B2SHARE EUDAT offers B2SHARE - a user-friendly, secure, robust, reliable and... 182 Datasets</p>	 <p>CESSDA CESSDA is an umbrella organisation for the European national data archives.... 529 Datasets</p>
 <p>CLARIN The Common Language Resources and Technology Infrastructure CLARIN project is... 137451 Datasets</p>	 <p>DataCite DataCite is a not-for-profit organisation formed in London on 1 December... 241180 Datasets</p>	 <p>EARLINET EARLINET, the European Aerosol Research Lidar Network, (www.earlinet.org) is... 5 Datasets</p>
 <p>GBIF The Global Biodiversity Information Facility (GBIF) was established by... 13814 Datasets</p>	 <p>ENES The European Network for Earth System modelling (ENES) provides information... 830 Datasets</p>	 <p>NARCIS NARCIS provides access to scientific information, including (open access)... 175442 Datasets</p>





EUDAT2020

- Harvesting of **metadata** stored in **B2SAFE**
- **Community customizations**
- **Annotation** of datasets
- Further assess **RDF** and **Linked Data**
- Further assess **scalability** and **performance**

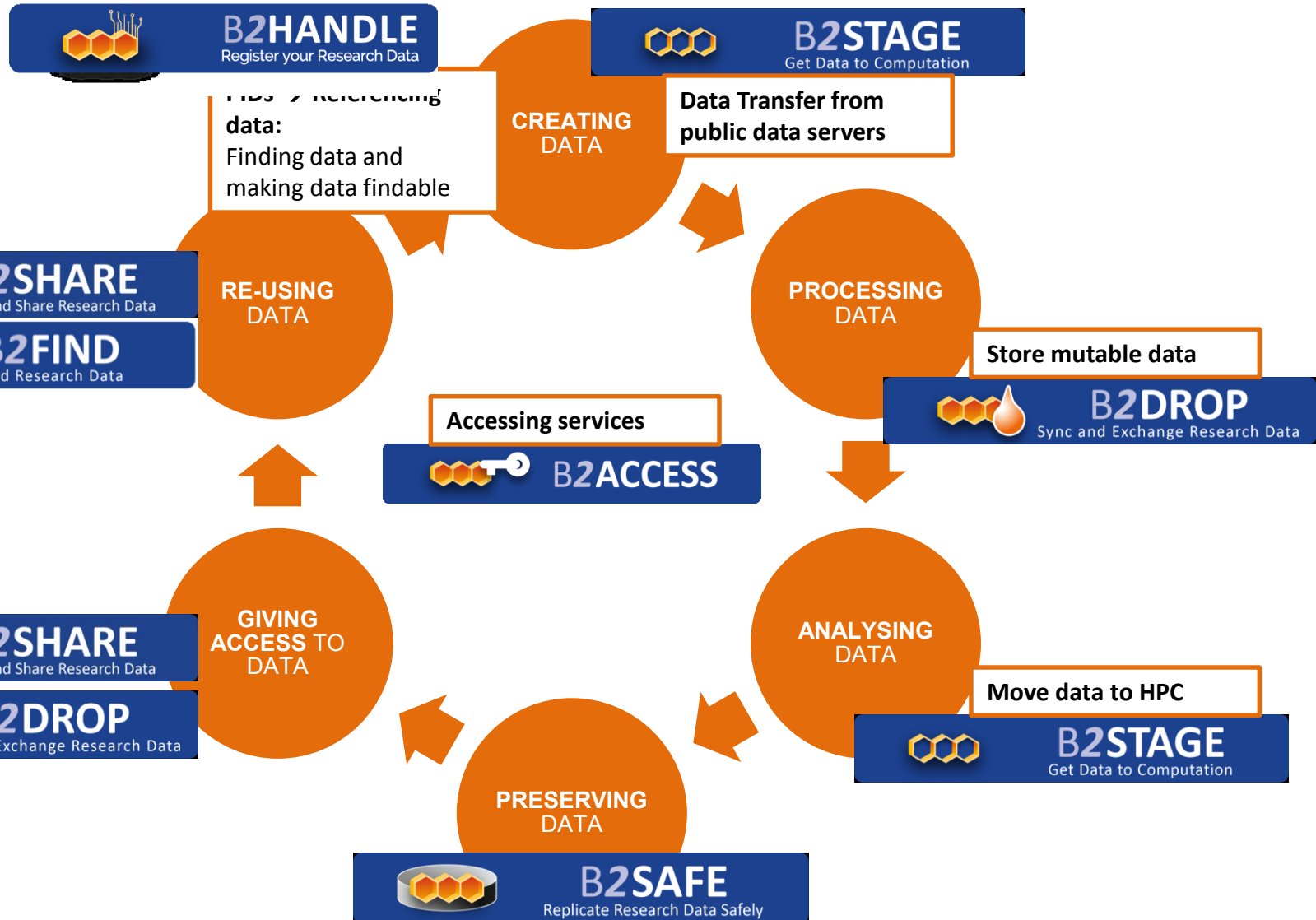
- **Who**
 - Anyone wanting to use the B2 Services
- **What**
 - Complies with **community ownerships** and **access rights**, basis of trust
 - Credential **conversion approach** (e.g. SAML, OpenID, X.509, Username/password)
 - Identity provider for **citizen scientists**
- **Why**
 - Use your own ID in federated environment

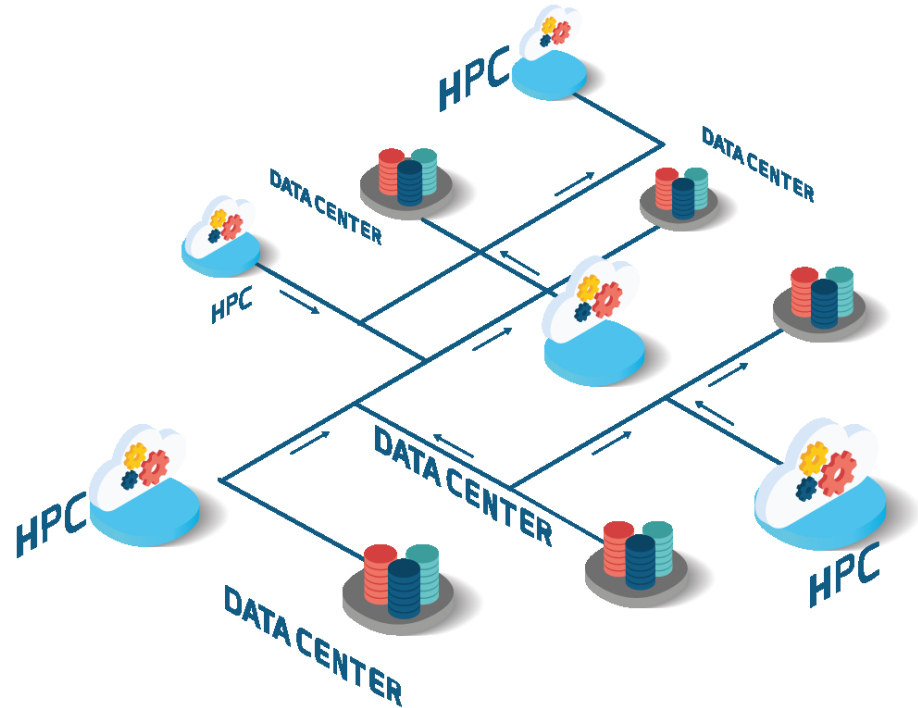
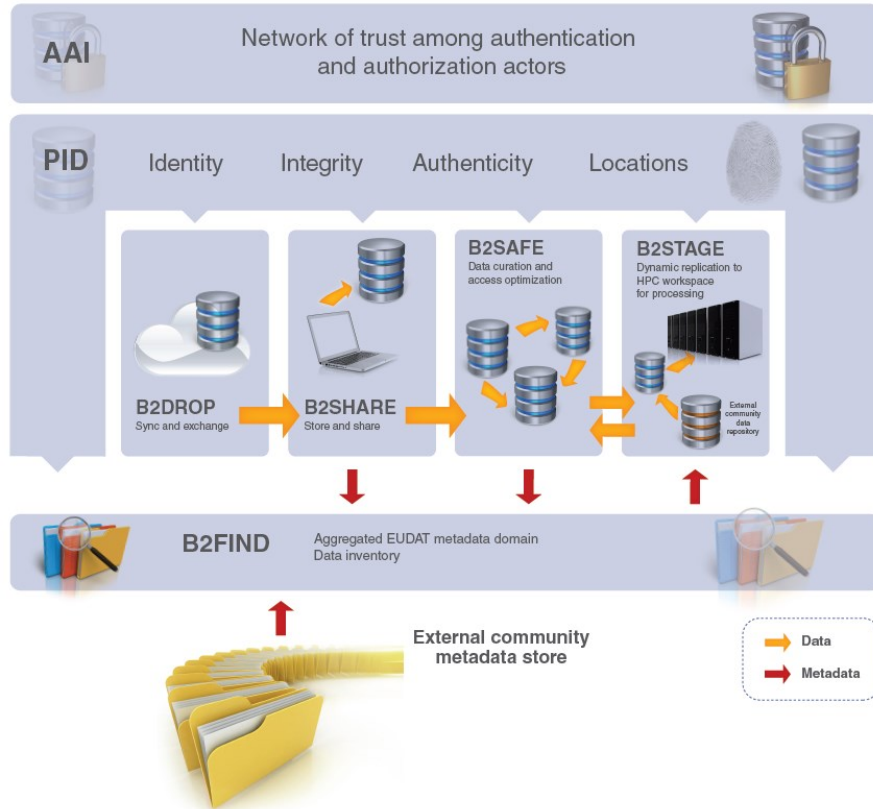


EUDAT2020

- Integration with operational and B2 services
 - B2SHARE ✓
 - B2SAFE ✓
 - B2DROP 
 - DPM ✓
 - B2STAGE ✓
 - CREG ✓
 - HTTP API 
 - GRIDFTP ✓
- Integration with community IdP domains and portal environments 
- Enabling access via eduGAIN ✓ Social IdPs ✓
- Focus on **authorization** 

Research Data Lifecycle





- Instead of individual services, all services are capable of working together
- Common methods to access the services (API)
- Standardized IT service maintenance practises

Palveluväylä

- ”Palveluväylä on tiedonvälityskerros, joka määrittää miten tietoja ja palveluja välitetään eri tietojärjestelmien välillä.”
- ”Palveluväylä on muutakin kuin tekniikkaa, sillä sen tarkoituksena on:
 - mahdollistaa palvelujen ja tietovarantojen yhtenäinen kokonaisuus
 - mahdollistaa standardoitu, turvallinen ja hallittu ympäristö
 - madaltaa tiedonvaihdon kynnyksiä
 - mahdollistaa uudenlaisia toiminta- ja toteutusmalleja.”

EUDAT and ATT ?

- Some services are almost identical
- National vs. European focus
- University/researcher vs. scientific community
- Small set of services vs. wider service set
- Different choice of persistent identifiers
- Are services interoperable ?
- What service should I use, as a researcher ?
 - National scope, ATT
 - European project (or legal entity), EUDAT

Scientific Journals ?

- Reliable storage space.
 - For research data (large amount of data) or only small space for published data ?
- Persistent identifiers
 - National library provides URN service used by ATT
 - EUDAT uses EPIC handle identifiers and can also provide DOIs
 - CSC is a member of DataCite
 - We are in a process of creating a policy how to get DOIs from from CSC.
- Some EUDAT services are free to use
 - B2Find [to find data]
 - B2Share [to publish data]

10 year commitment

A landmark agreement sustaining the pan European Collaborative Data Infrastructure for the next 10 years

3rd October 2016

16 major European research organisations, data and computing centres signed an agreement to sustain the EUDAT – pan European collaborative data infrastructure for the next 10 years. The organisations stand together behind a long term sustainability plan and commit to develop, maintain and deploy pan-European research data services and to promote harmonization of research data management practices across centres.



What does this mean?

Service providers, both generic and thematic, and research communities have joined forces as part of a common framework for developing an interoperable layer of common data services. Known as the EUDAT Collaborative Data Infrastructure (CDI), this is essentially **a European e-infrastructure of integrated data services and resources** to support research. This infrastructure and its services have been developed in close collaboration with **over 50 research communities** spanning across many different scientific disciplines and involved at all stage of the design process. The establishment of the EUDAT CDI is timely with the imminent realization of the European Open Science Cloud^[1] which aims to offer open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines.

What benefits will this bring to researchers?

Researchers can rely on **innovative data services** to support their research collaboration and data management. Additionally they benefit from a common service management framework delivered by CDI service providers and the connection between sites. These 16 organizations^[2] from 11 countries with a further 4^[3] in the process of joining, are European leaders in the deployment of data management and data infrastructure services. By joining the EUDAT CDI they are able to better **serve their local customers across borders and support cross-national collaborations**.

European Open Science Cloud

