



Using EUDAT services to replicate, store, share, and find cultural heritage data

...in PSNC... and beyond...

Maciej Brzeźniak, Norbert Meyer, PSNC
Damien Lecarpentier, CSC





Plan and purpose of presentation



- ❑ **Summarize the work done and planned in EUDAT an DCH-RP project**
 - How DCH sector can use e-Infrastructures?

- ❑ **Undersdand how selected EUDAT services can be used in DCH domain?**
 - How EUDAT CDI architecture can be integrated with domain-specific services?
 - Overview of current services (Simple Store, Safe Replication)
 - Presentation of integration work in DCH-RP

 - *Discussion on possible future extensions / services*



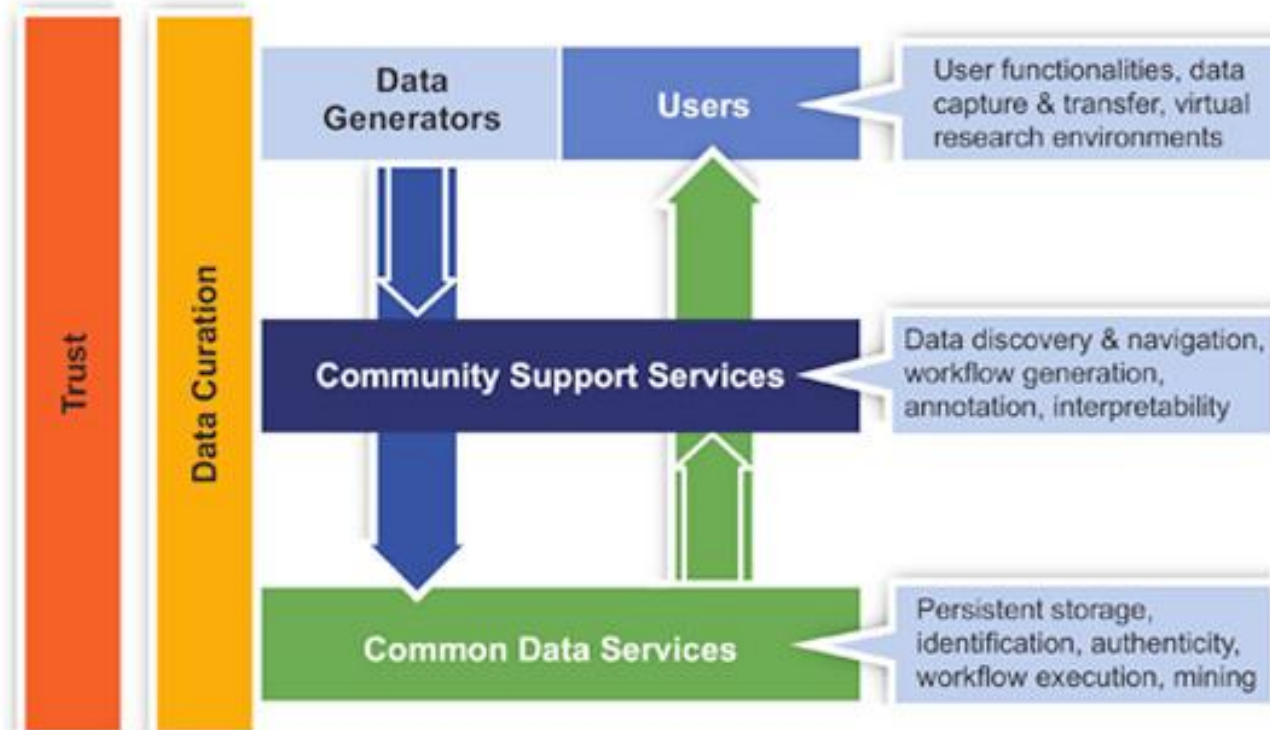
Some aspects of EUDAT



□ EUDAT – European Data Infrastructure

- **Vision:** to support a Collaborative Data Infrastructure
- **Aims:**
 - Provide a **sustainable platform** of **technologies, tools, services driven by user needs**
 - **Engage users in defining/shaping a platform for shared services**
- Support **data-intensive, multi-disciplinary research:**
 - Humanities and Social Science: CLARIN
 - But also: earth science (ENES, Earth system modelling; EPOS: European Plane Observing System), ecology (LifeWatch), Virtual Physiological Human (VPH)
- Deliver **common low-level services** that are required to provide the level of interoperation and trust of data
- Ensure that the data **infrastructure is robust/scalable** (able to address ‚data tsunami‘)
- **Build community/domain-specific services** on top of the common services with participation of users

The Collaborative Data Infrastructure: A framework for the future



Source: High Level Expert Group on Scientific Data, *Riding the wave*, 2010.

CDI layers vs tools and services in DCH

Integration challenges...

High-level services:

EUDAT simple store



eCulture Science Gateway



dLibra dArceo dLab



Invenio



Another community solution



Low-level services:

Local LTS:
long-term storage



EUDAT storage

- Safe Replication



Cloud storage

- S3
- CDMI



Grid storage

- FTS
- GridFTP





CDI layers vs tools and services in DCH Integration challenges...



High-level services:

EUDAT simple store



eCulture Science Gateway



dLibra dArceo dLab



Invenio



Another community solution



DCH-RP <-> EUDAT proof of concept: Do they go together?

Low-level services:

Local LTS:
long-term storage



EUDAT storage
• Safe Replication



Cloud storage

- S3
- CDMI



Grid storage

- FTS
- GridFTP





EUDAT: services for DCH domain



□ Services covered in the presentationis

- **Simple Store service:**
 - enable researchers and scientists to upload, store and share data
 - designated for the „long tail of data“
- **Safe Replication:**
 - Allow communities to replicate data to selected EUDAT data centers
 - Automated replication (iRODS), PID registration (EPIC)
- **Data Staging:**
 - Staging data from user community premises/systems (iRODS)
 - to computing systems, e.g. PRACE's HPC centres (GridFTP, FTS)
- **Metadata service:**
 - Joint metadata domain for all EUDAT data centres
 - Searchable catalogue covering all data stored within EUDAT
- **AAI:**
 - Provide a solution for a working AAI system in a federated

□ Simple Storage Service:

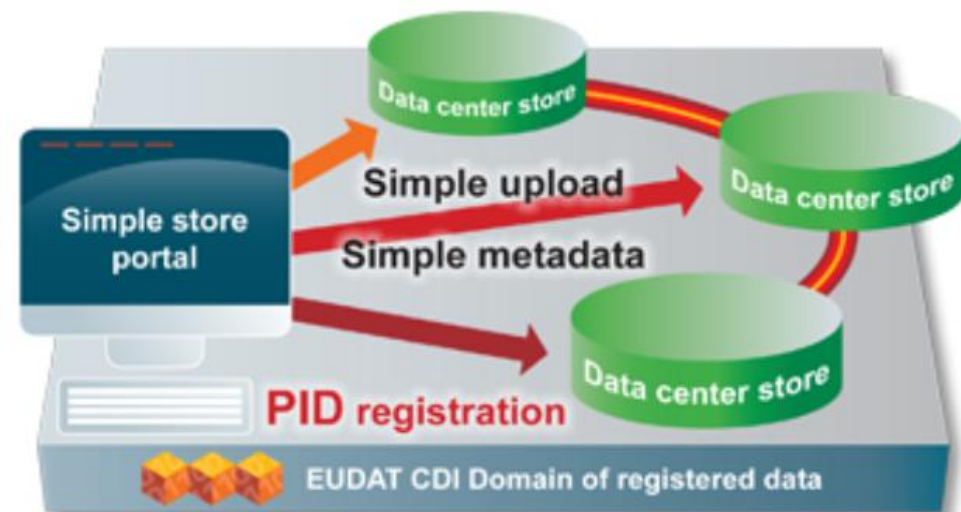
- Address the issues of small user groups and individual users
- Provides solution for "long tail data": often stored on laptops and departmental servers

□ Functionality:

- allowing registered users to **upload typical** data objects into the EUDAT store
- enabling users to **share** such **objects** and collections with other researchers.
- lets utilising other EUDAT services
 - Safe Replication
 - PIDs
 - etc.
- May be integrated with AAI

□ More:

<http://www.eudat.eu/simple-store>





EUDAT for DCH Simple Storage Service



□ Simple Storage Service internals:

- Referred also as Researcher Data Store
- Based on Invenio:
 - Developed by CERN
 - <http://invenio-software.org/>
- Storage backend:
 - Disk
 - iRODS (EUDAT safe replication)
- Front-end:
 - Developed a new submission portal to invenio





EUDAT for DCH Simple Storage Service



EUDAT B2Share

https://trng-b2share.eudat.eu

2nd eudat conference

B2SHARE Search Deposit Help maciekb

Search 0 records for Search



Image courtesy of Frisno Boström

Durable Storage of Research Data

The EUDAT BE2Share (ex-SimpleStore) service provides a quick and easy mechanism for preservation of research data.

In future, the SimpleStore will work with other EUDAT services to ensure data is stored reliably and can be easily searched and retrieved.

The BE2Share is currently in an **early testing phase**. Users are encouraged to sign up and test the service, but please be aware that it is **not ready for production use**. In particular, note that data submitted to the service in its current state is not stored reliably and will be removed at some stage. Also note that all submissions are made public; there is currently no support for restricting access to files.

Please contact livenson@kth.se with any questions or feedback.



EUDAT for DCH Simple Storage Service



EUDAT B2Share

https://trng-b2share.eudat.eu/deposit/

2nd eudat conference

maciekb

B2SHARE Search Deposit Help

[Home](#) / [Deposit](#)

1. Upload files (drag and drop or select files)

Deposits are done in 3 steps:

1. you select and upload one or several data resources
2. you then select the metadata set you want to use and describe the resource by metadata
3. you deposit the selected resource including its metadata elements. The deposit takes a moment to be processed. You will get a reference URL at once.

Select files

Start upload



EUDAT for DCH Simple Storage Service



B2SHARE

[Search](#) [Deposit](#) [Help](#)

maciekb

[Home](#) / [Deposit](#)

1. Upload files (drag and drop or select files)

Deposits are done in 3 steps:

1. you select and upload one or several data resources
2. you then select the metadata set you want to use and describe the resource by metadata
3. you deposit the selected resource including its metadata elements. The deposit takes a moment to be processed. You will get a reference URL at once.

Select files

Start upload

Filename	Size	Status	
apples.jpg.JPG	784 KB	<input type="text"/>	
mountains.jpg.jpg	159 KB	<input type="text"/>	



EUDAT for DCH Simple Storage Service



1. Upload files (drag and drop or select files)

Deposits are done in 3 steps:

1. you select and upload one or several data resources
2. you then select the metadata set you want to use and describe the resource by metadata
3. you deposit the selected resource including its metadata elements. The deposit takes a moment to be processed. You will get a reference URL at once.

Select files

Start upload

Stop upload

Filename	Size	Status	
apples.jpg.JPG	784 KB	<div style="width: 100%; height: 15px; background-color: #007bff;"></div>	
mountains.jpg.jpg	159 KB	<div style="width: 80%; height: 15px; background-color: #007bff;"></div>	

2. Select a domain or project

Please select a metadata description set which best fits to your data. The EUDAT Generic set includes a common set of elements which are



EUDAT for DCH Simple Storage Service



Filename	Size	Status	
apples.jpg.JPG	784 KB	<div style="width: 100%; height: 15px; background-color: #007bff;"></div>	
mountains.jpg.jpg	159 KB	<div style="width: 100%; height: 15px; background-color: #007bff;"></div>	









2. Select a domain or project

Please select a metadata description set which best fits to your data. The EUDAT Generic set includes a common set of elements which are inspired by Dublin Core, MARC and DataCite and this set is included in all sets as basis. So if you just want the common set select Generic, if you want in addition a specific community set select that community. Soon there will be more communities. You can always revise your decisions and select another set, however you may lose what you have already entered.

Generic	Ecology
EUDAT	DRIHM

3. Add metadata

Generic

Title	* Maciejs photos	
Description	Photos take in 2013	
Creator	Maciej	
Open Access	<input checked="" type="checkbox"/> ON 	
Licence	Public	
Publisher	Maciej Brzezniak	
Publication Date	2013	
Tags	art, photography, landscape, still nature	

Add more details?

4.

Deposit



EUDAT for DCH Simple Storage Service



< B2SHARE >

Search

Deposit

Help

maciekb

Home / Deposit

Deposit Successful

Your submission will shortly be available at:

<https://trng-b2share.eudat.eu/record/2>

Please note that it may take a few minutes to process your submission.

Deposit another item

Search 61 records for



Search

Add to search

AND

OR

AND NOT



Author

Title

Report number

Affiliation

Collaboration

Keywords

Journal

Any



Vol

Pg

[simple](#)[advanced](#)

Browse

Search

of Research

(Store) service providing
ation of research data

er EUDAT services to ensu
rieved.

ing phase. Users are encour
that is **not ready for produ**

service in its current state is

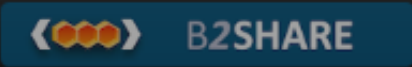
. Also note that all submiss
ricting access to files.

uestions or feedback.

Image courtesy of Frisno Bostr



EUDAT for DCH Simple Storage Service



Search Deposit Help

maciej

Search 11 records for

art

Search

Any Collection

Any Author

Any Year

Showing records 1 to 1 out of 1 results.

Display

Sort

1

2013 Photos by Maciej

Still nature and landscape photographs

by [Maciej Brzezniak](#) | [30 Oct 2013, 11:57](#) | [Similar records](#) | [art](#) [photography](#) [still nature](#) [landscape](#)

[Add tags](#)

Copy Bookmark

Dublin Core

Export

« <

[Home](#) / [2013 Photos by Maciej](#)

2013 Photos by Maciej

[Maciej Brzezniak](#)

Maciej Brzezniak

Abstract: Still nature and landscape photographs**Keyword(s):** [art](#) ; [photography](#) ; [still nature](#) ; [landscape](#)*The record appears in these collections:*[eudat](#)[Add tags](#)[← Back to search](#)

Files ▾

Name	Date	Size	
mountains.jpg.jpg	30 Oct 2013	163.3 kB	Download
apples.jpg.JPG	30 Oct 2013	803.2 kB	Download

Export

[BibTeX](#), [MARC](#), [MARCXML](#), [DC](#), [EndNote](#),
[NLM](#), [RefWorks](#)

Metadata

Checksum: 3f19aeac3c40de9465
6b9f3db978b3383898
70466a9bf223f6aa2e2
dc4bee458**PID:** [ad99da14-4149-11e3-
bc23-14feb57d12b9](#)**Publication:** Maciej Brzezniak**Licence:** Public**Uploaded by:** [maciekb@man.poznan
.pl](#)**Domain:** eudat



EUDAT for DCH Safe Replication Service

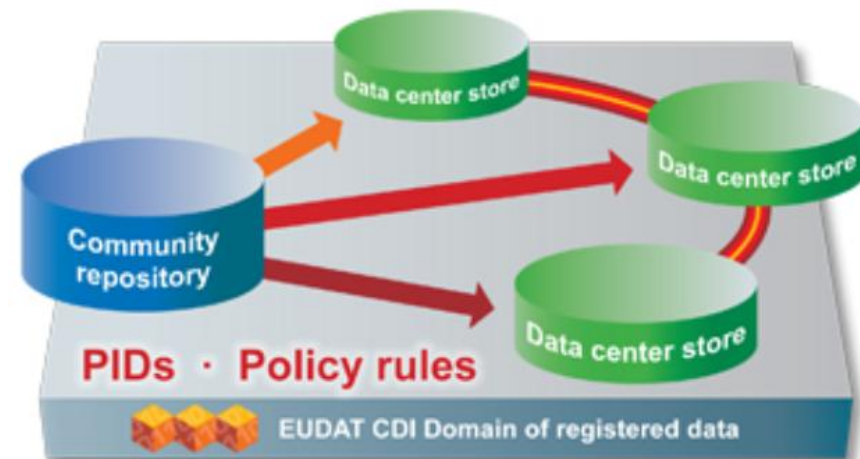


□ Safe Replication Service:

- Allows communities to **replicate data to EUDAT data centers**
- Can be **integrated with portals and community tools**

□ Functionality:

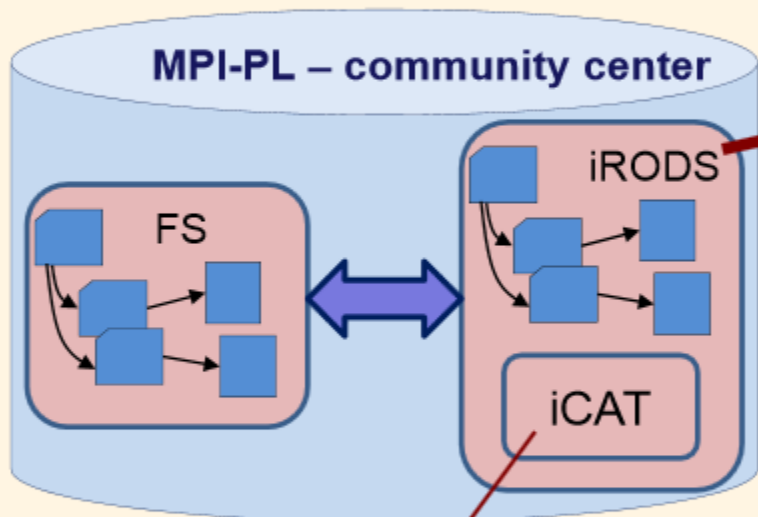
- Ingested file/data are:
 - automatically **replicated** to many data centres
 - **get persistent identifiers registered** (PIDs based on EPIC)
- Various interfaces supported
 - iRODS: icommand, API,
 - WebDAV, GridFTP
- Can replicate data on top of various different data stores:
 - Disks, tapes, HSMs
 - Clouds (e.g. S3)



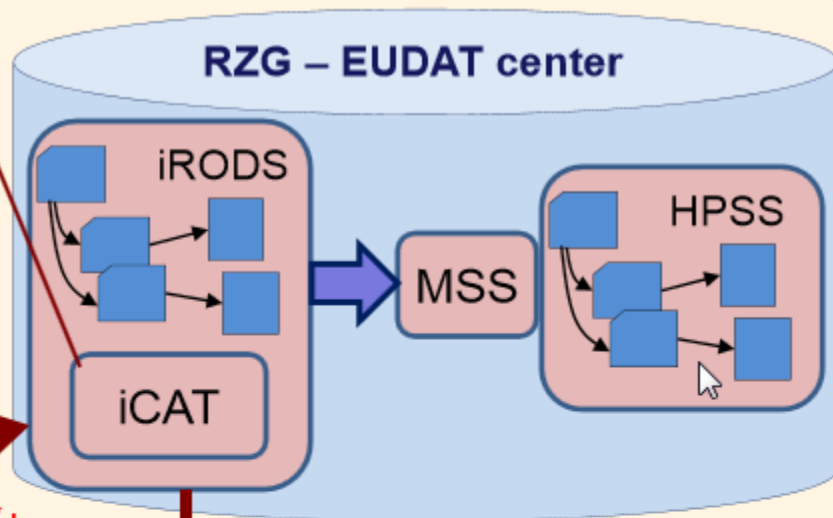
- **More:**
<http://www.eudat.eu/safe-replication>

Zone Federation

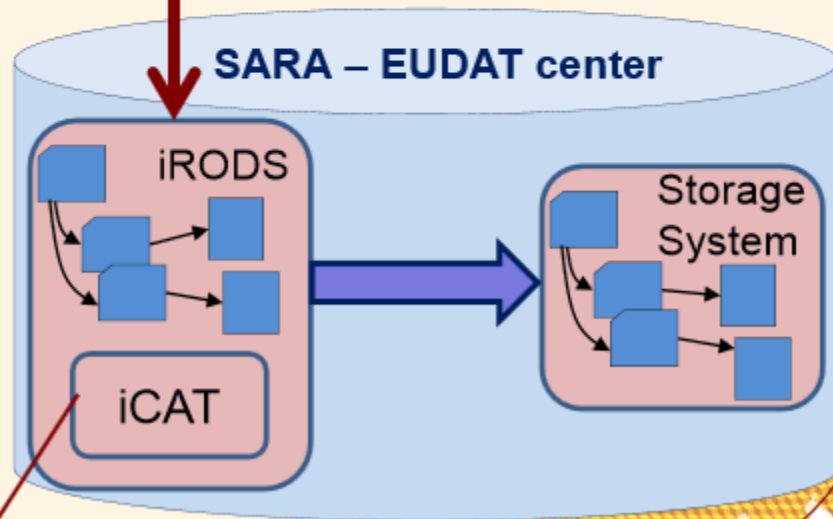
Local zone : /vzRZGE
Remote zones: /vzMPI
/vzSARA



Local zone : /vzMPI
Remote zones: /vzRZGE
/vzSARA



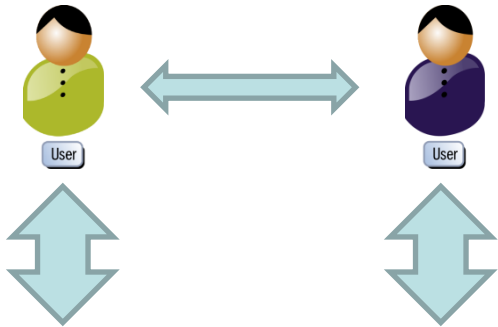
Replication (+
PID creation)



Local zone : /vzSARA
Remote zones: /vzMPI
/vzRZGE



Integration option 1: EUDAT Simple Store + EUDAT Safe Replication



**Support
for sharing**



**Easy deposit
& access to
the data**

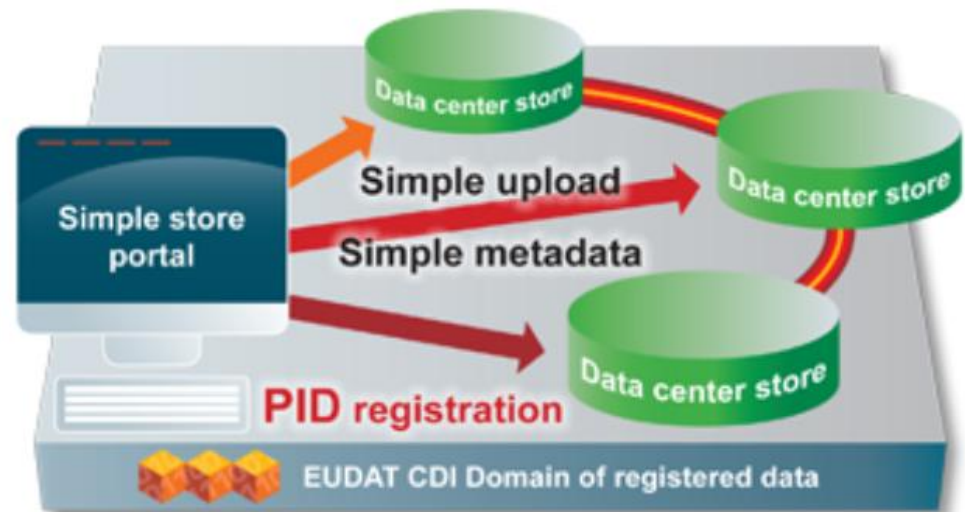


**Transparent
replication of data,
persistent
storage**



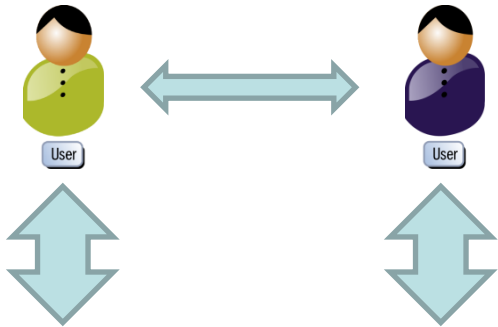
**Data
Store1**

**Data
Store2**





Integration option2: Community portal/solution + EUDAT Safe Replication



Support
for sharing



Storage
& access typical
for community

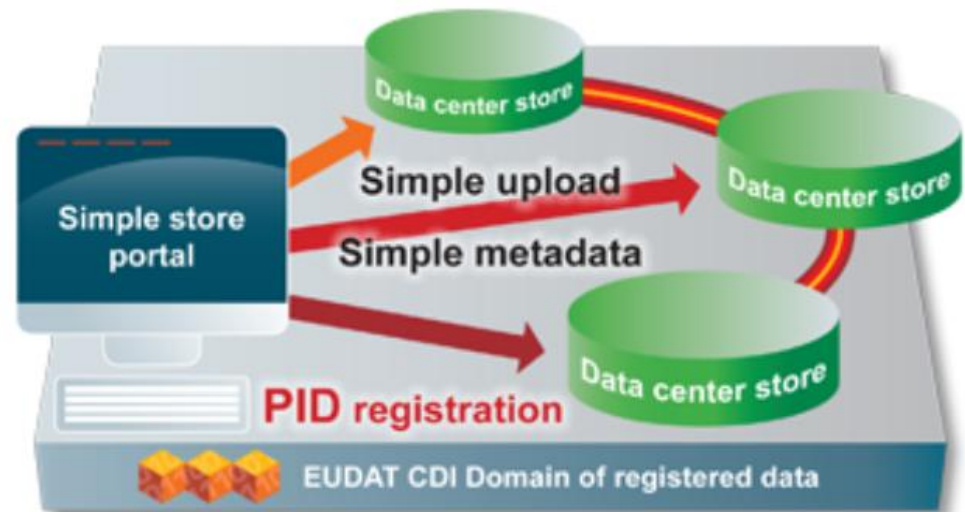


Transparent
replication
of data



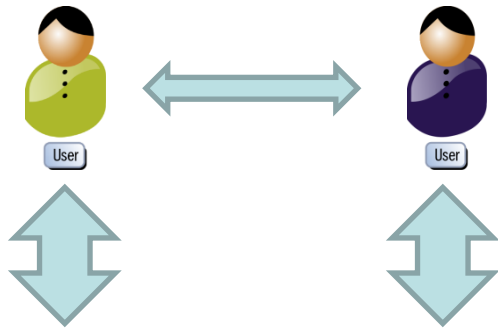
Data
Store1

Data
Store2





Option 1 in practice: EUDAT Simple store + EUDAT Safe Replication *DCH-RP <-> EUDAT PoC 1*



**Support
for sharing**

**Memory
institutions:**



**Simple Store
Service**

**Easy storage
& access to
the data**

**EUDAT
Simple Store**

Provided by PSNC



**Safe
Replication**

**Transparent
replication
of data**

**EUDAT
Safe
replication**

*Provided by
PSNC (Poznan)
& EPCC (Edinburgh)*



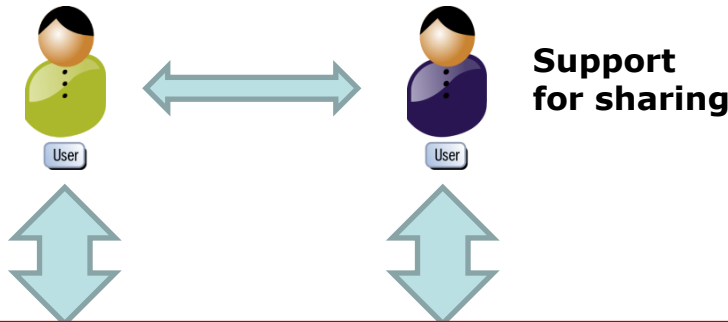
**Data
Store1
@PSNC**



**Data
Store2
@EPCC**



Option 2 in practice: Community CMS solution + EUDAT Safe Replication *DCH-RP <-> EUDAT PoC 2*



Support
for sharing

Memory
institutions:



Community
service

Feature-full
domain-specific
tool with large
user base

**EUDAT
Simple Store**

*Provided by PSNC
or EUDAT partners*



Safe
Replication

Transparent
replication
of data

**EUDAT
Safe
replication**

*Provided by
PSNC (Poznan)
& EPCC (Edinburgh)*



Data
Store1
@PSNC



Data
Store2
@EPCC

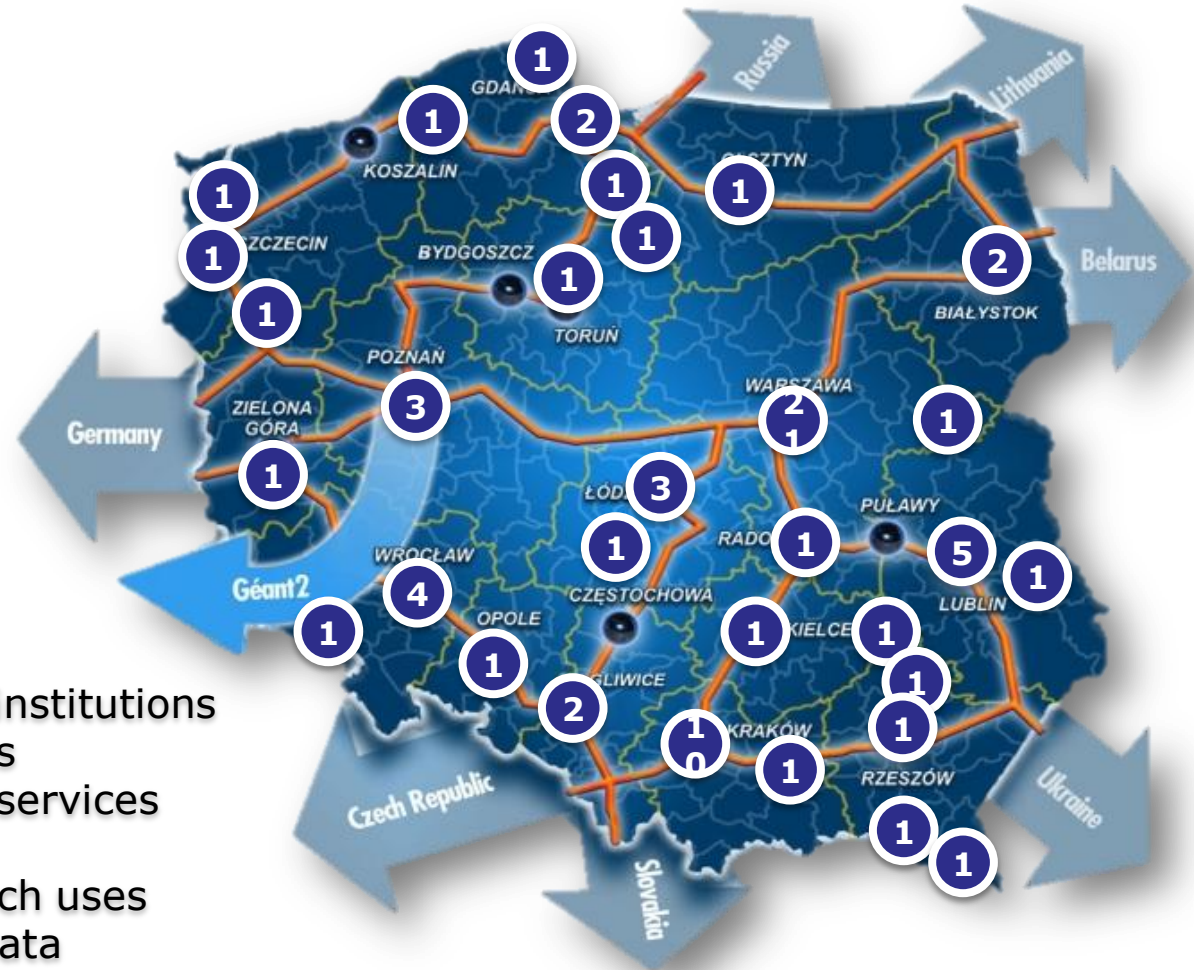


Domain-specific solutions dLibra/dArceo/dLab



DCH-RP <-> EUDAT PoC 2

<http://dlab.psnc.pl/>



Basic statistics:

- ± 100 digital libraries
- several hundreds memory institutions
- over 1,1 M of digital objects
- 98% content delivered via services based on dLibra software (<http://dlibra.psnc.pl/>) which uses Solr for content and metadata indexing and searching



EUDAT <-> DCH-RP PoC summary



- ❑ **We investigate two possible ways to offering data preservation services for DCH**
 - Top-down solution for ,citizen scientists' / ,citizen DCH people'
 - Well-established solution backed by EUDAT Safe Replication

- ❑ **We exploit the layered EUDAT CDI architecture**
 - In theory: It enables integration with existing solutions
 - We try to understand how it works in practice



Planned extensions of simple store Discussion trigger 1



□ Development roadmap

- Premium service:
 - Customisation for layout, metadata – for community needs
 - increased storage capacity
 - increased support
 - increased bandwidth
- Premium vs regular service:
 - Providing premium service requires enrolling with EUDAT
 - Regular services to be offered to ‚citizen scientists/users‘ – no close relationship needed
- AAI integration
 - *On the roadmap*



Possible extensions of simple store

Discussion trigger 2



❑ From yesterdays discussion about Simple Store:

- **Thousand of files?**
- **Upload reliability / robustness?**

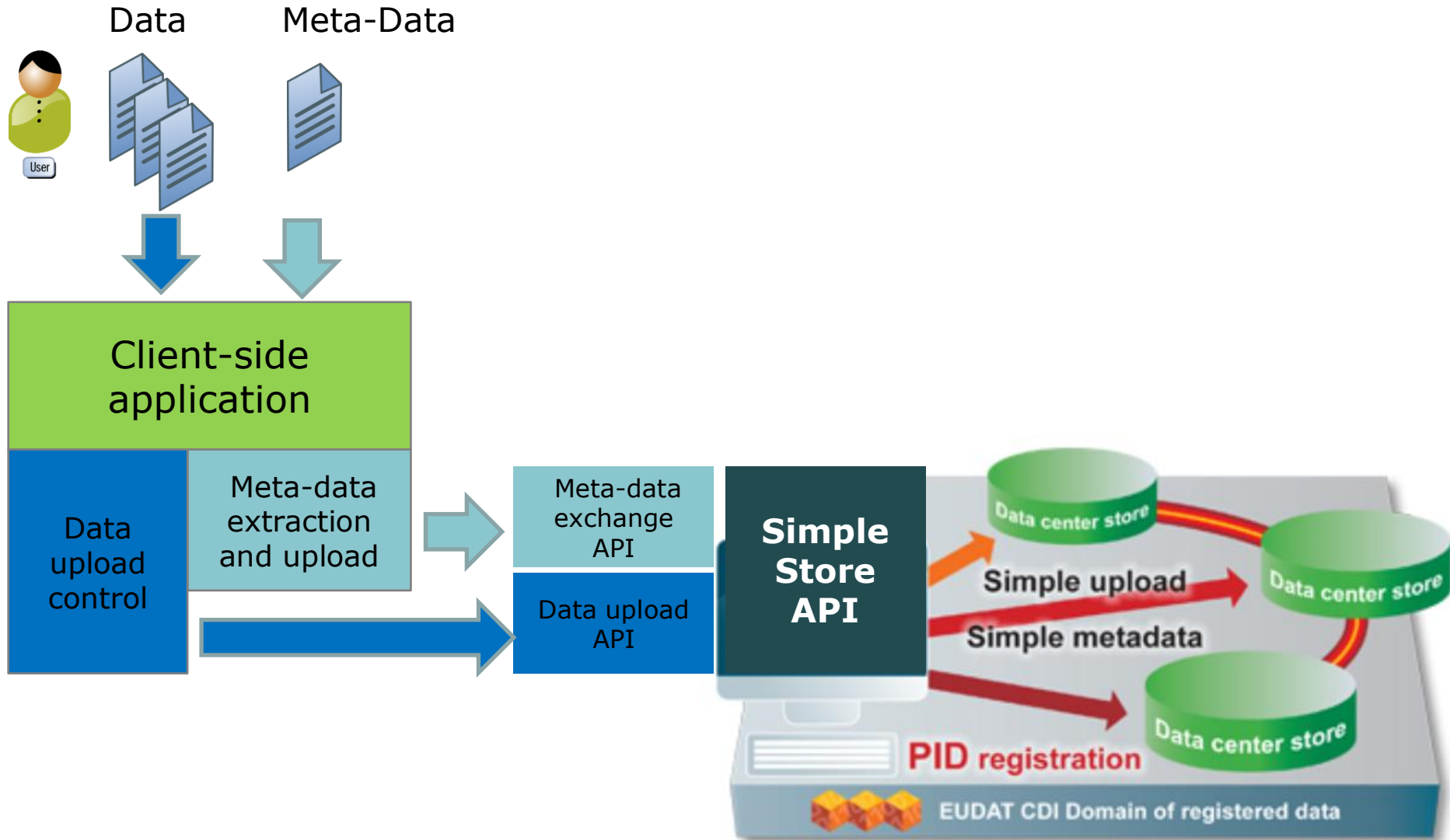
=> Batch upload

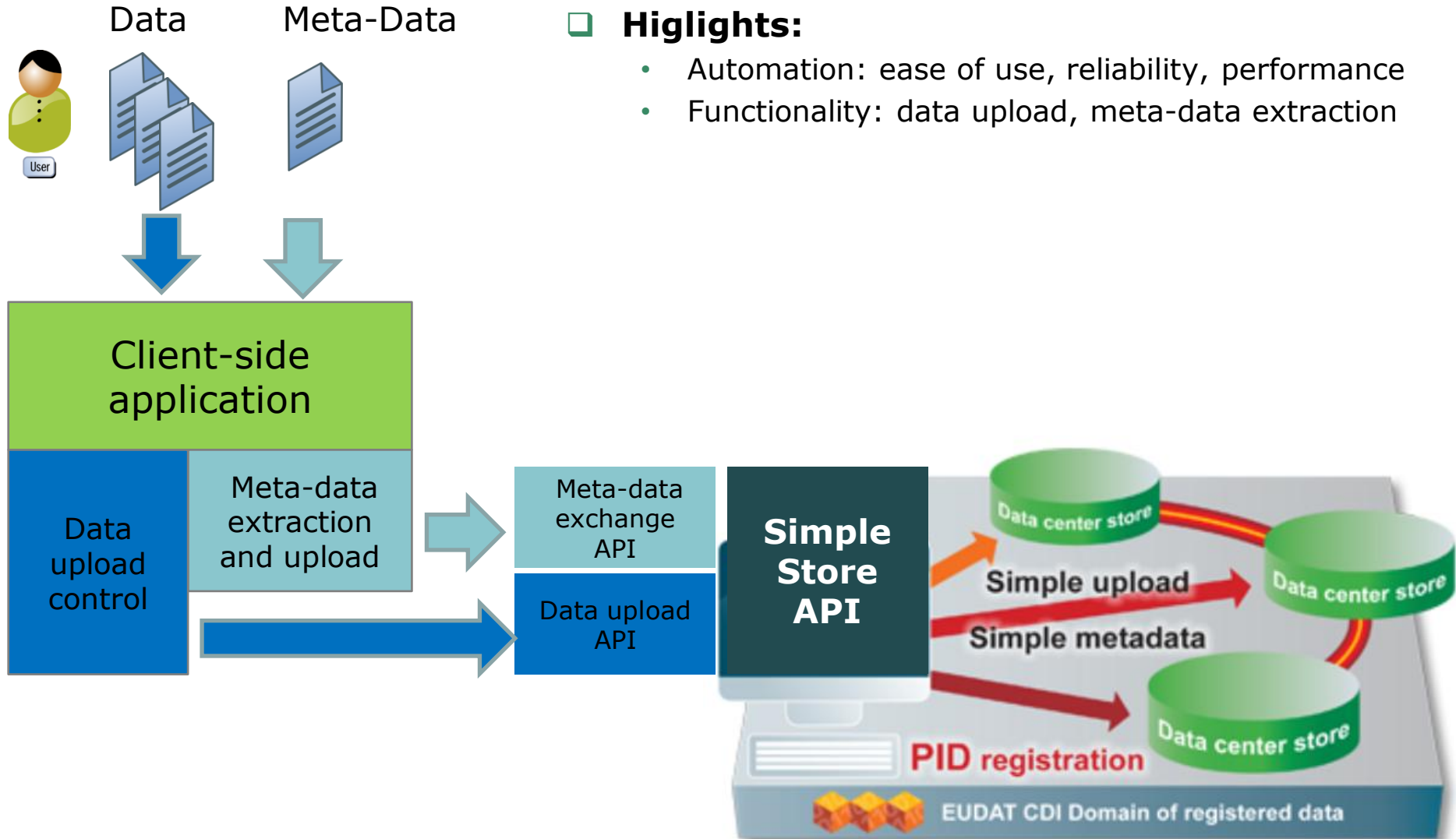
- API to be developed
 - » Enables integration with existing systems
 - » Tools can be offered by EUDAT to support batch upload

- **Collections upload?**

=> Support for meta-data extraction

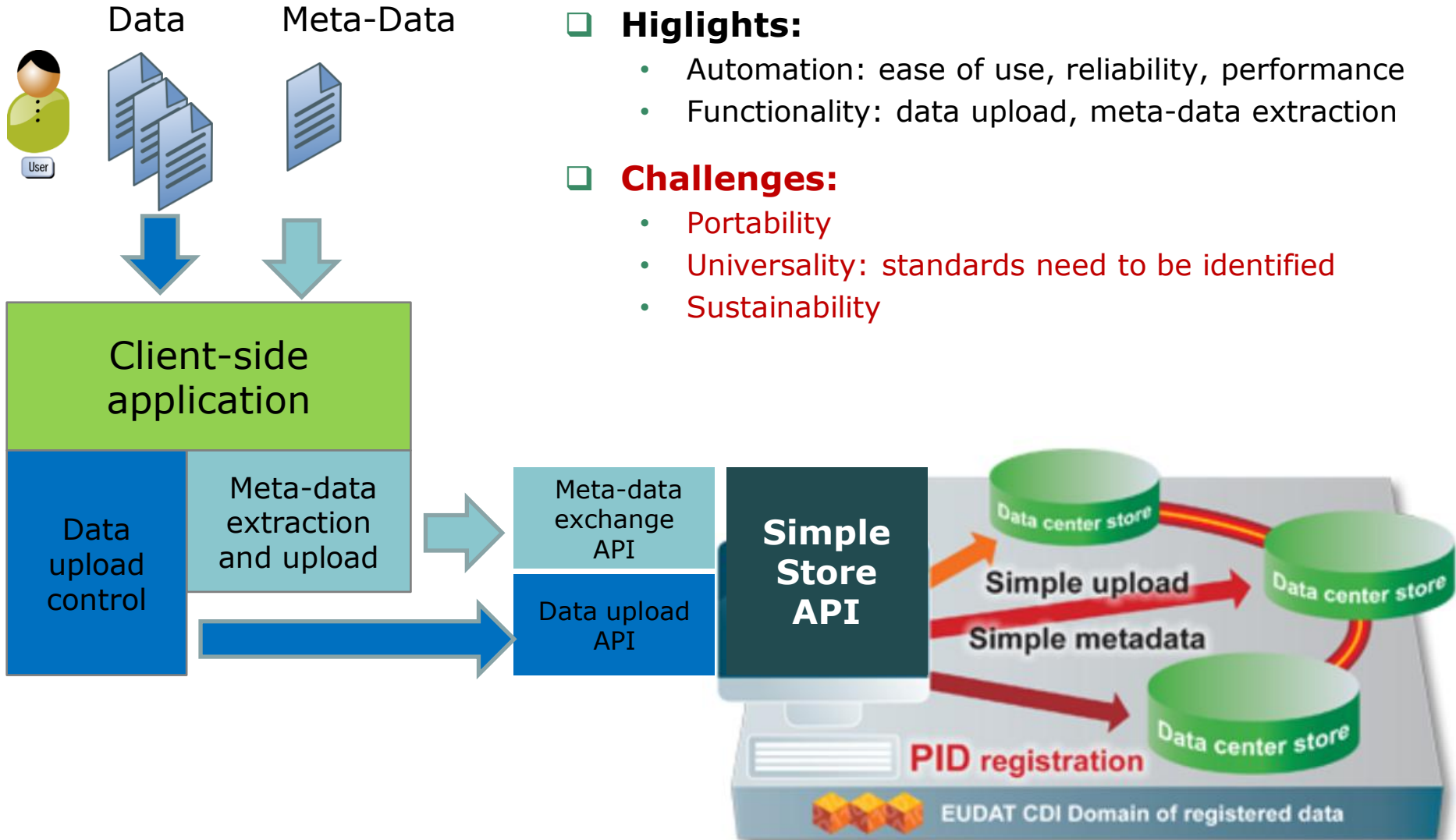
- Implemented client-side?
- E.g. based on pre-prepared collections (e.g. DIPs)





Highlights:

- Automation: ease of use, reliability, performance
- Functionality: data upload, meta-data extraction

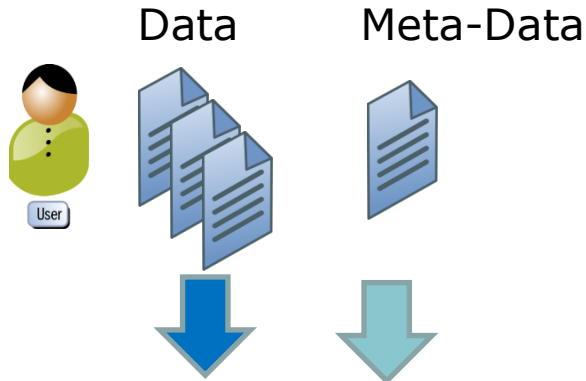


Highlights:

- Automation: ease of use, reliability, performance
- Functionality: data upload, meta-data extraction

Challenges:

- Portability
- Universality: standards need to be identified
- Sustainability



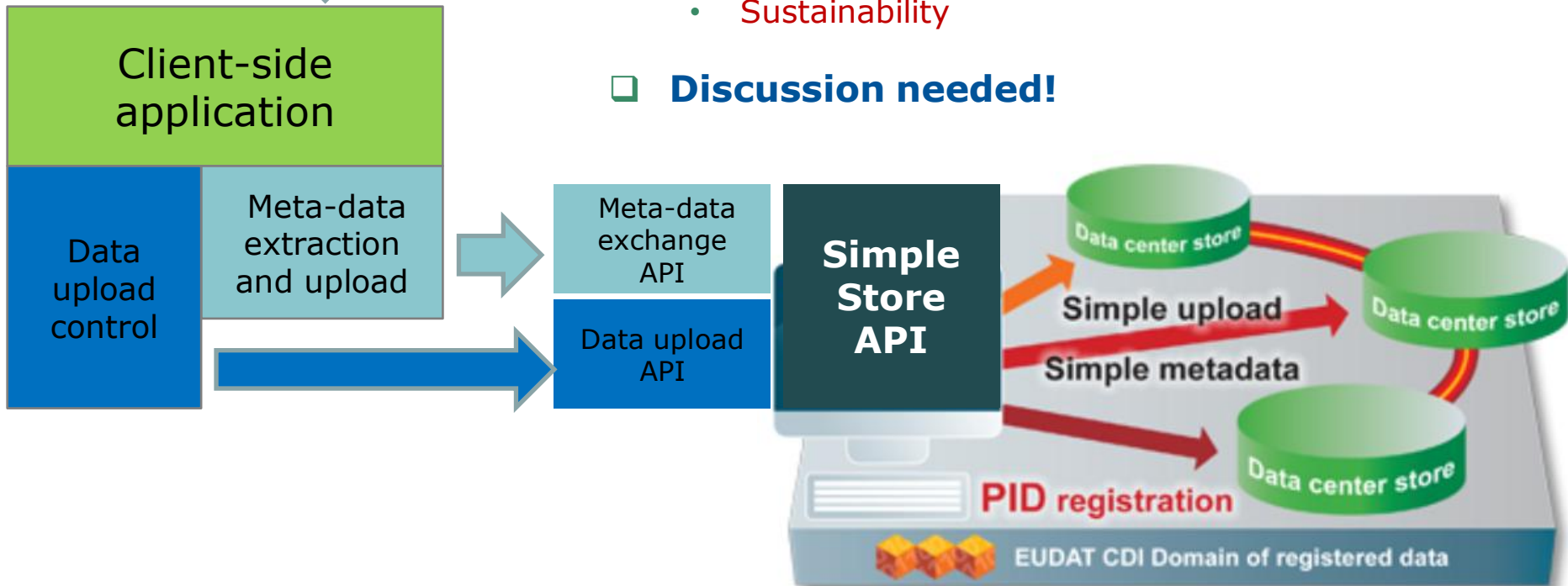
Highlights:

- Automation: ease of use, reliability, performance
- Functionality: data upload, meta-data extraction

Challenges:

- Portability
- Universality: standards need to be identified
- Sustainability

Discussion needed!





Possible extensions of simple store

Implementation: user-side tool?



NDS2GUI - build 0.9.5 - zenon : MANY_VIDEOS

File

Messages
No notices

Show Details

zenon

Name	Size	Type	Status
DCPsi2_800x600x30-coded.avi	2.5 MB	Video file	✓
eucentre.mpg	3.0 MB	Video file	✓
ns1280x960x50-coded.mpg.!kmd2	1.8 MB	Incomplete file (part)	↑ 52% [Pause] [Cancel]
rotate_0000.avi	6.5 MB	Video file	↑ 0% [Pause] [Cancel]
spot_pcass.avi	7.9 MB	Video file	↑ 0% [Pause] [Cancel]
ZZ_clip0003.avi.!kmd2	3.1 MB	Incomplete file (part)	↑ 39% [Pause] [Cancel]
ZZZ_clip0002.avi.!kmd2	6.2 MB	Incomplete file (part)	↑ 0% [Pause] [Cancel]
ZZZ_clip0004.avi.!kmd2	5.2 MB	Incomplete file (part)	↑ 0% [Pause] [Cancel]
ZZZ_clip0005.avi	7.9 MB	Video file	↑ 0% [Pause] [Cancel]

Uploading files to remote server...

Number of uploading tasks: 5
Total Size: 5 files; 33.8 MB

Name	Size	Status	Progress	Action
Task : 1 files	3.0 MB	Completed		
Task : 1 files	2.5 MB	Completed		
Task : 1 files	5.2 MB	Awaiting	0%	[Pause] [Cancel]
Task : 1 files	7.9 MB	Awaiting	0%	[Pause] [Cancel]
Task : 1 files	7.9 MB	Awaiting	0%	[Pause] [Cancel]
Task : 1 files	6.5 MB	Pending	0%	[Pause] [Cancel]
Task : 1 files	6.2 MB	Pending	0%	[Pause] [Cancel]
Task : 1 files	1.8 MB	Completed		
Task : 1 files	3.1 MB	Completed		

Clear completed
Resume all
Pause all
Cancel all
Show on remote
Show on local

Connected [3] ↑ 274 kB/s (7) ↓ 0 B (0) Elements: 9



Extensions of EUDAT services discussion Summary



□ Message:

- EUDAT infrastructure and services are layered, modular
 - This enables integration
- Further extensions possible
 - Users are welcome to influence them
- We want to make sure that we recognised and support necessary standards
- Technical details / organisation / etc. to be discussed



Using EUDAT services to replicate, store, share, and find cultural heritage data

...in PSNC... and beyond...

Maciej Brzeźniak, Norbert Meyer, PSNC
Damien Lecarpentier, CSC

THANK YOU!

