



B2SAFE

Johannes Reetz & B2SAFE-Team



What is B2SAFE?

It is a **policy-driven**

data replication service

for **community** and **departmental repositories** to replicate (distribute) their research data **safely** to **highly available** storage service providers.



Why B2SAFE?

Community and departmental repositories need to **replicate their research data** for three compelling reasons:

- to guard against data loss in long-term **archiving** and **preservation**,
- to **optimize access** for users from different regions, and
- to bring **data closer to powerful computers** for compute-intensive analysis.



Who uses B2SAFE & Why?

- **EUDAT (collaboration) partners**
 - EPOS: INGV/CINECA, ORFEUS-KNMI/SARA
 - CLARIN: MPI-PL/SARA + EKUT/RZG + CU/CINES
 - VPH
 - ENES
 - INCF: INCF-SR/KTH
 - DiXa
- **Extensions (B2SAFE-fedora, B2SAFE-Dspace, B2SAFE-ePrints)**
 - Account for long tail of small data organized and stored through Fedora Commons and/or DSpace from departments with little or no IT capabilities



B2SAFE in three steps

- Data ingestion
- Data replication
- PID assignment

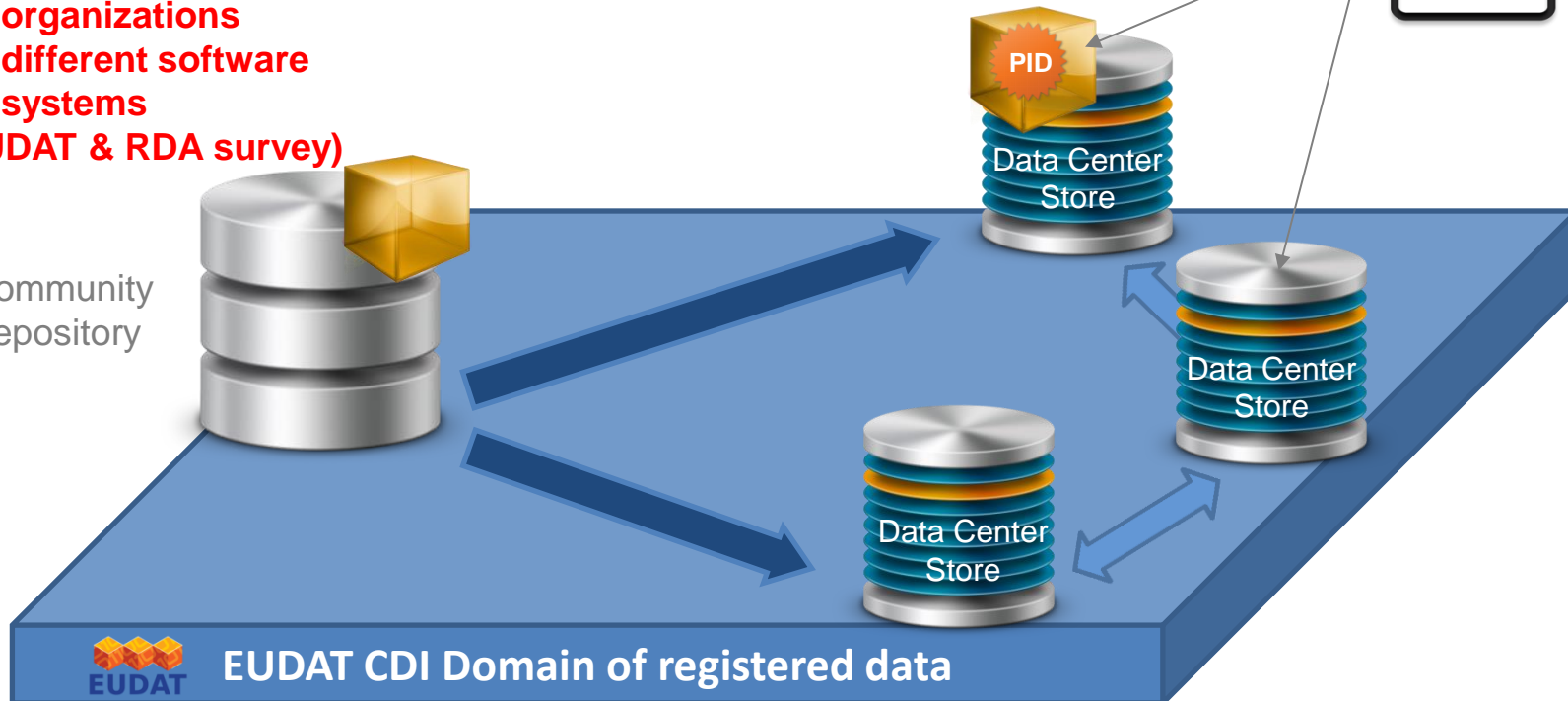


B2SAFE explained

Community repository

- different data organizations
 - different software systems
- (EUDAT & RDA survey)

Community repository



EPIC
service



CLARIN (MPI-PL/TLA)





CLARIN (EKUT)



CLARIN (CUNI/LINDAT)





ENES (DKRZ/CMIP5)

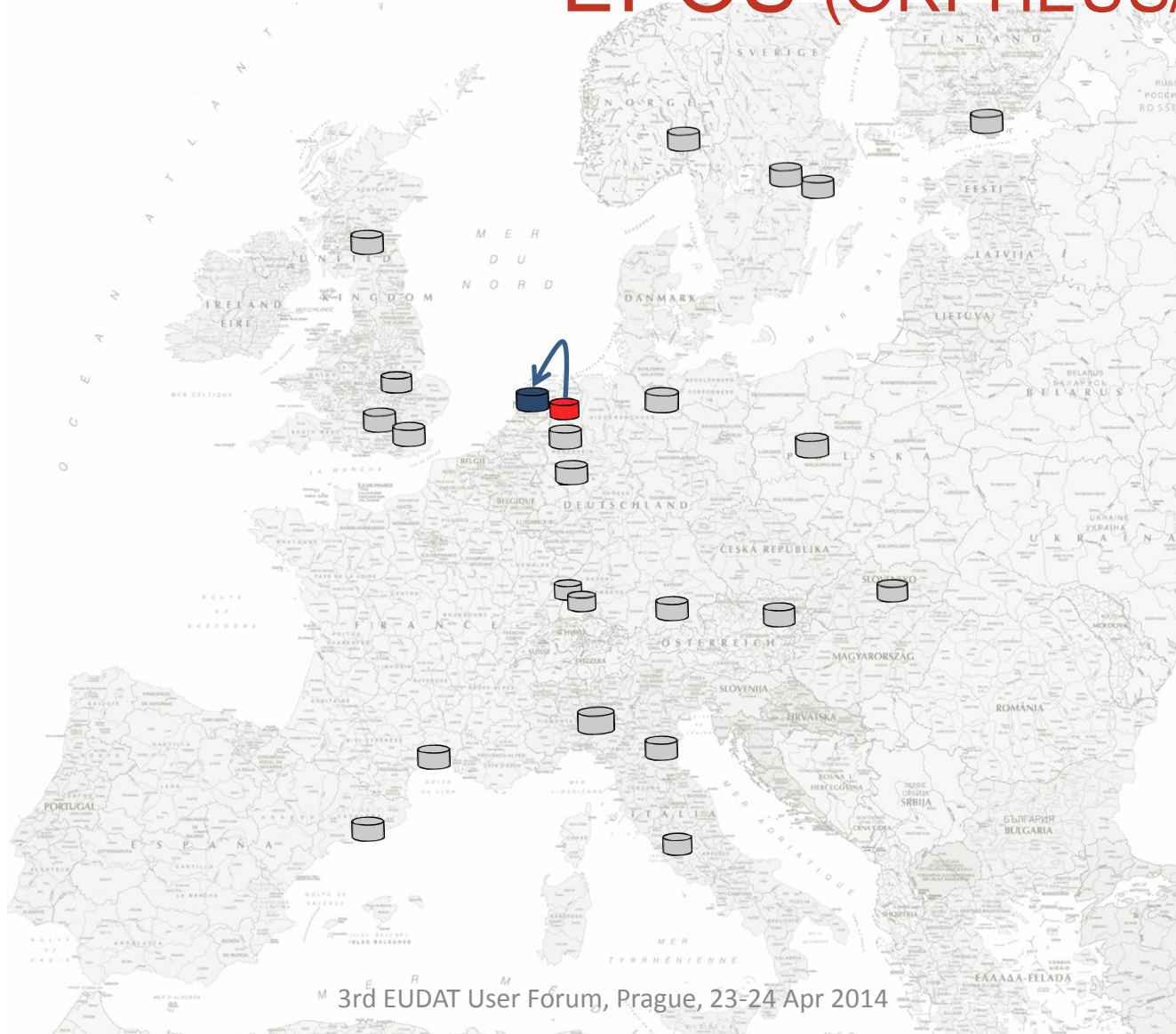


EPOS (INGV)





EPOS (ORPHEUS/ KNMI)





INCF (KTH)



DiXa (EBI)





B2SAFE Future Plans

- Facilitate the use of the service and improve its capabilities
 - Develop **web services** interface (with B2STAGE) to increase the engagement of e.g. skilled communities
 - **Gain experience** from additional community data centers to improve/correct the service, work on data management policies
 - **Easy installable packages** for popular repository software stacks
- Improve **Scalability** (both size and # of files)
- Improve the service with “**dynamic data**” capabilities
- Establish a **Data Policy Management** framework