

Building interoperable systems for SeaDataNet community

Digital infrastructures for Research,
October 9 – 11, 2018, Lisbon

Christopher Ariyo
CSC-IT Center for Science
Finland

SeaDataNet community

- The SeaDataNet project offers a robust and state-of-the-art Pan-European infrastructure to harmonise metadata and data from marine data centers in Europe, and offers the technology to make these data accessible
- The SeaDataNet community partnered with EUDAT consortium of e-infrastructure service providers in SeaDataCloud project to **upgrade their existing infrastructure**



B2ACCESS
Identity & Authorisation



B2DROP
Synch and Share Research Data



B2FIND
Find Research Data



B2SHARE
Store and Publish Research Data



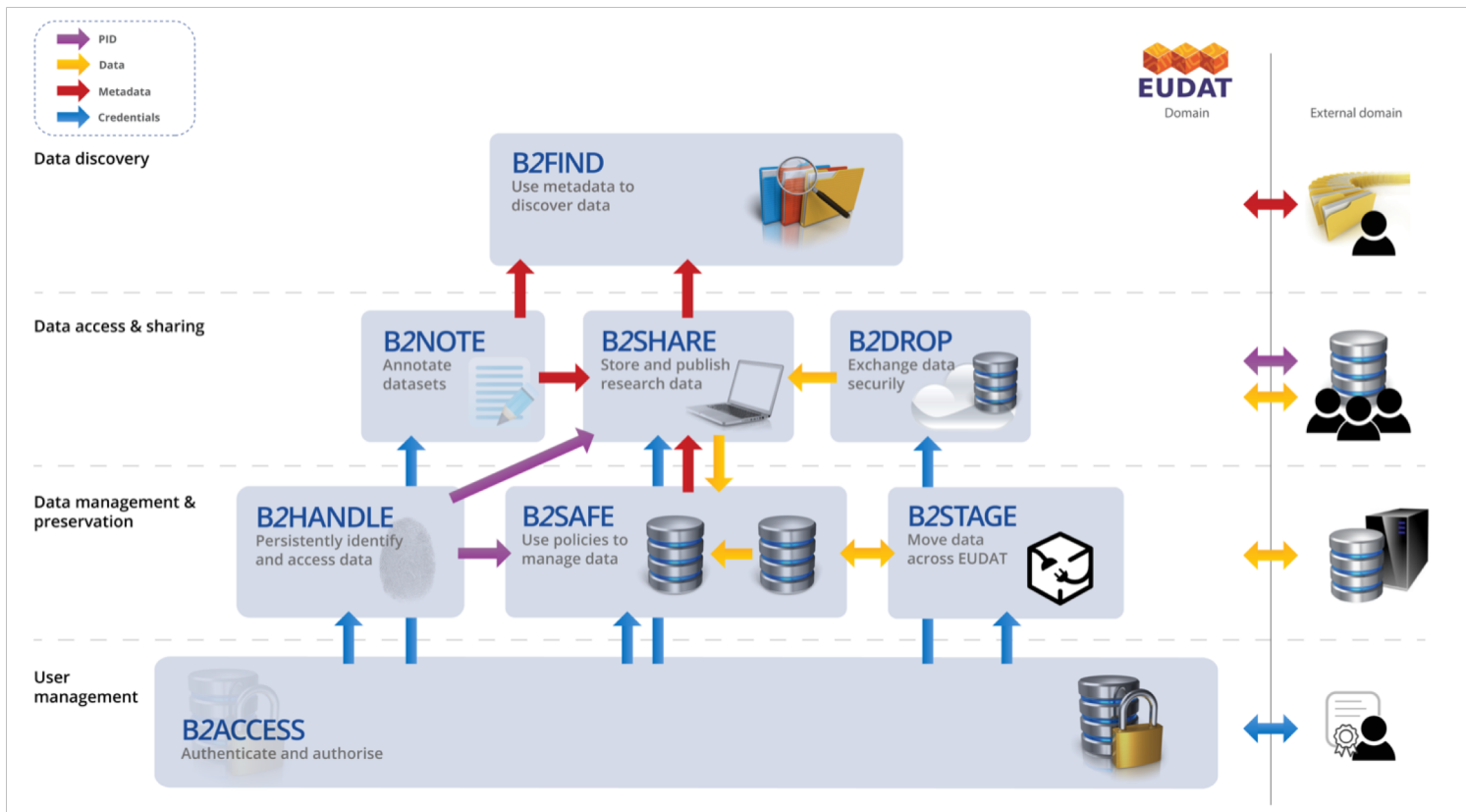
B2SAFE
Replicate Research Data Safely



B2STAGE
Get Data to Computation

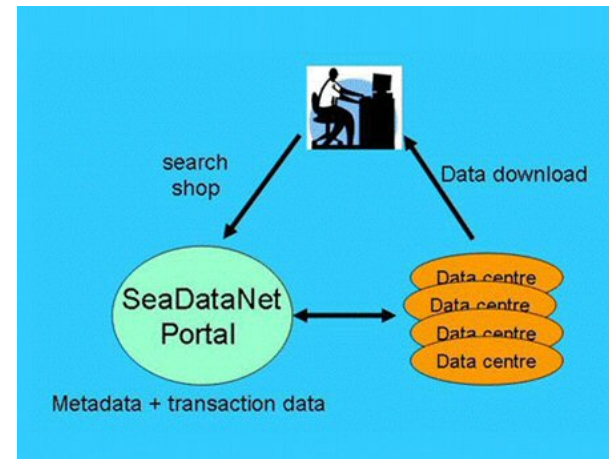
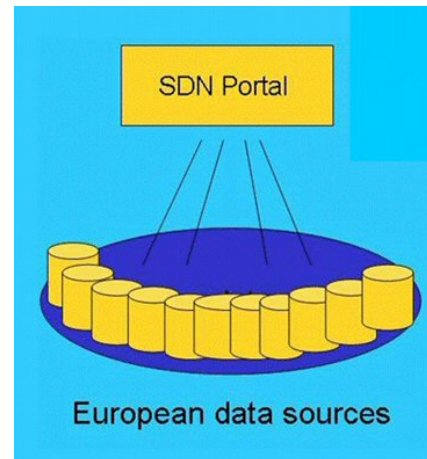


B2NOTE



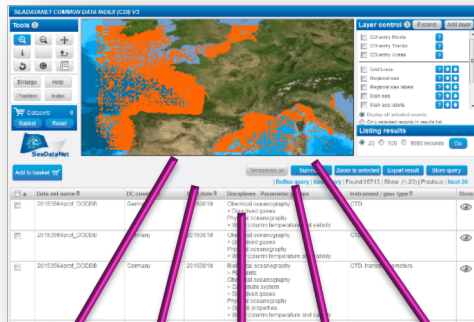
- Offers on-demand computing and data access to researchers across Europe
- B2HOST allows communities to deploy and operate their own applications and data-oriented services on machines next to the data storage location

- One of the core services of the SeaDataNet infrastructure
- Providing a highly detailed insight and unified access to the large volumes of marine and oceanographic data sets managed by the distributed data centres



CDI service for discovery and unified data access

SeaDataNet portal

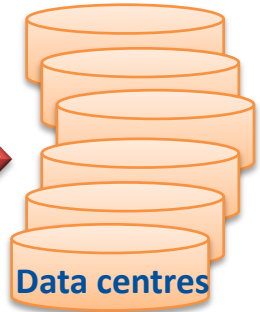
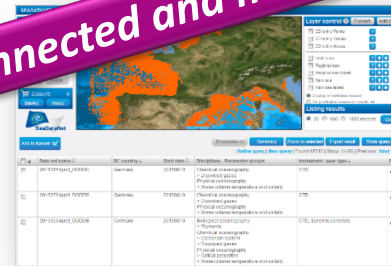


Search and Shop



Data

Already 110 data centres connected and more underway



Metadata

+ transaction data

*European data sources
data centres > 650 originators*

Existing CDI service – components



User interface

CDI

Shopping basket

Request Status Manager

Services for quality management

Central User Register

CDI Import Manager

Download Manager

Download Manager

data

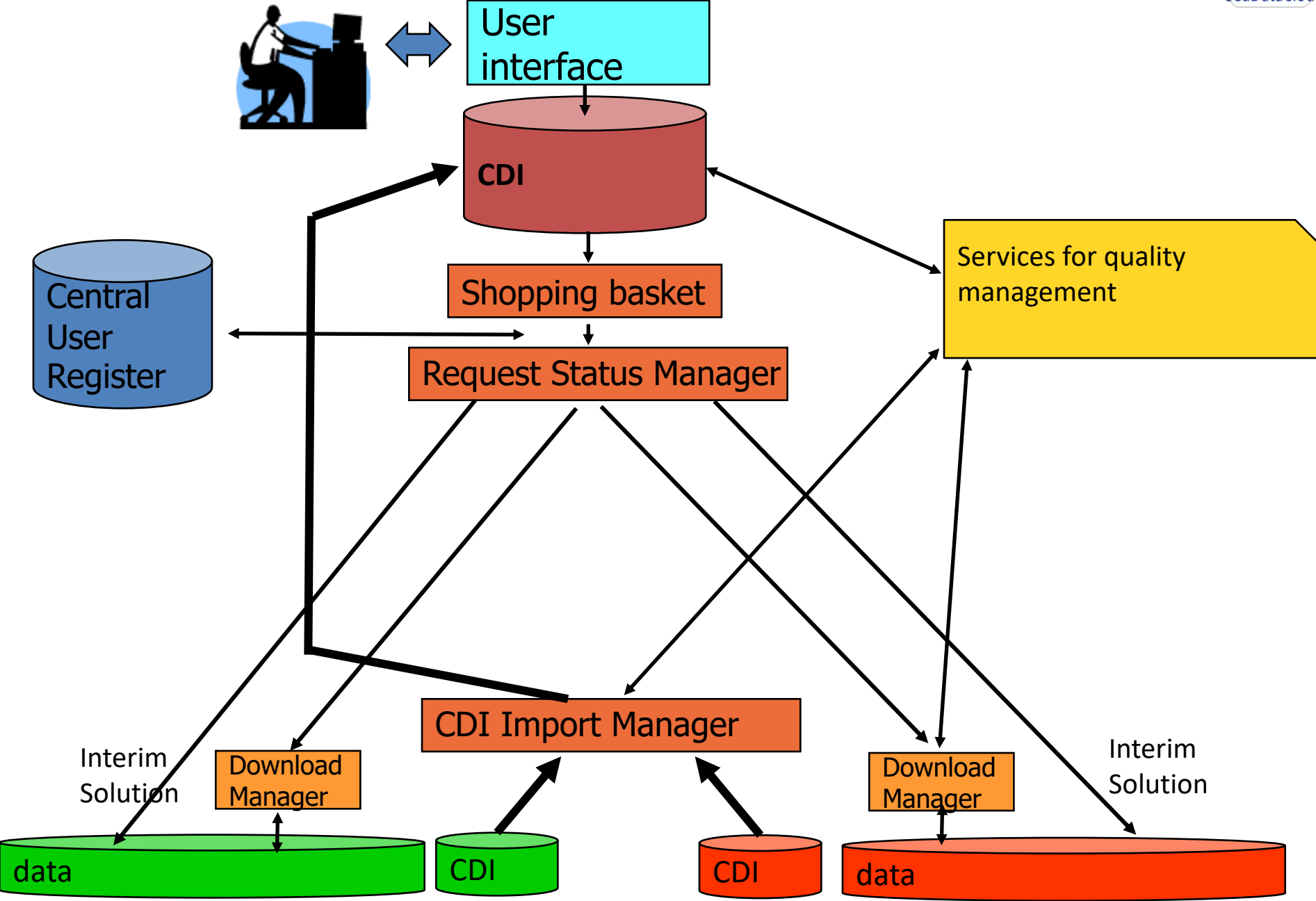
CDI

CDI

data

Interim Solution

Interim Solution



- ◆ **SeaDataNet CDI User Interfaces:** for users to discover relevant data sets by combinations of search criteria or a faceted search or a matrix search and to enter the shopping mechanism
- ◆ **SeaDataNet CDI Catalogue:** contains all active CDI entries, which have been populated by the connected SeaDataNet data centres.
- ◆ **SeaDataNet Central User Register:** details of registered users, their organisations, Marine-Id user and passwords, Roles; based upon Marine-ID
- ◆ **SeaDataNet Shopping Basket:** used for preparing access requests of multiple data sets, max 10.000 records per transaction, and routing requests to the Request Status Manager; multiple transactions allowed
- ◆ **SeaDataNet Request Status Manager (RSM):** processing and administration of all requests and data deliveries (downloads) for users in communication Download Managers at data centres for restricted data and data centres directly in case of interim solution.
- ◆ **CDI Import Manager (IM):** online Content Management System (dashboard) for managing and steering the process for import and validation of CDI metadata
- ◆ **Download Manager:** intermediates between Request Status Manager (Master) and local data management systems; Takes care that requested files are made ready for downloading by User (if cleared!) in the agreed formats via a local agreed website address.

Upgrading CDI service – with EUDAT cloud



User interface

CDI

Shopping basket

Request Status Manager

Services for quality management, versioning, transformation

Central User Register

Eudat cloud

CDI Import Manager

Interim Solution

Interim Solution

Download Manager

Replication Manager

Replication Manager

Download Manager

data

CDI

CDI

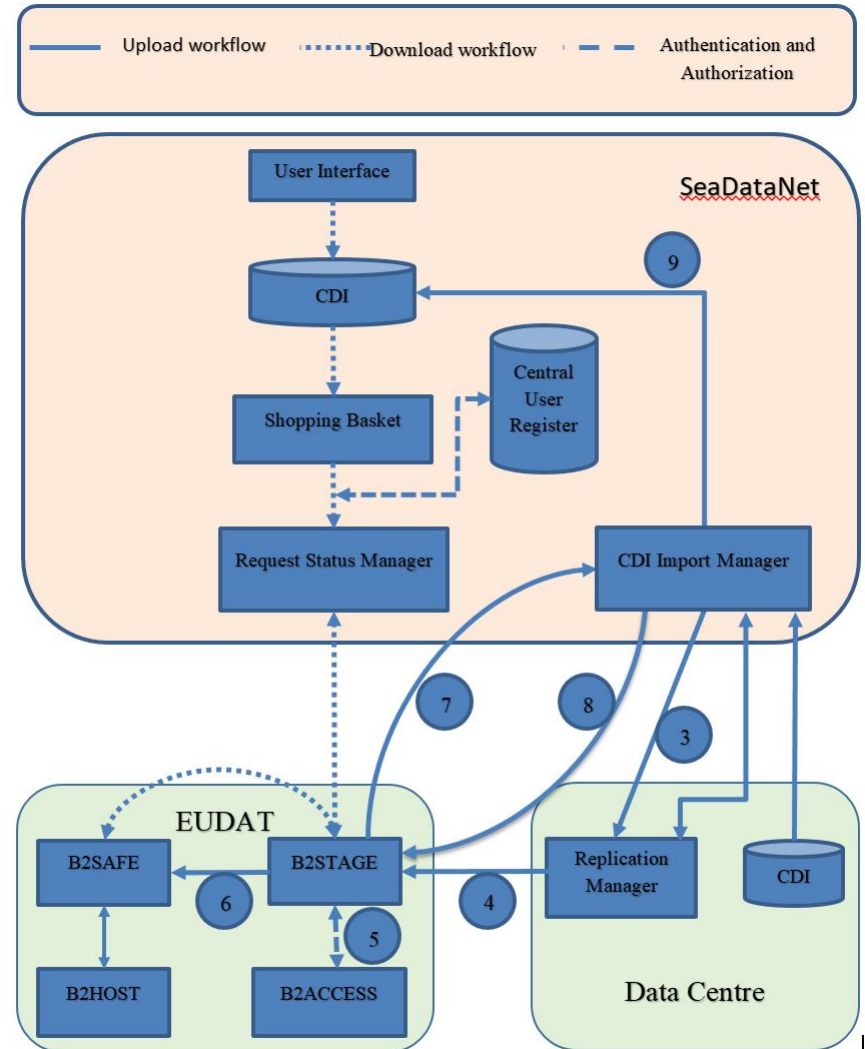
data

Components of the upgraded CDI service for data access

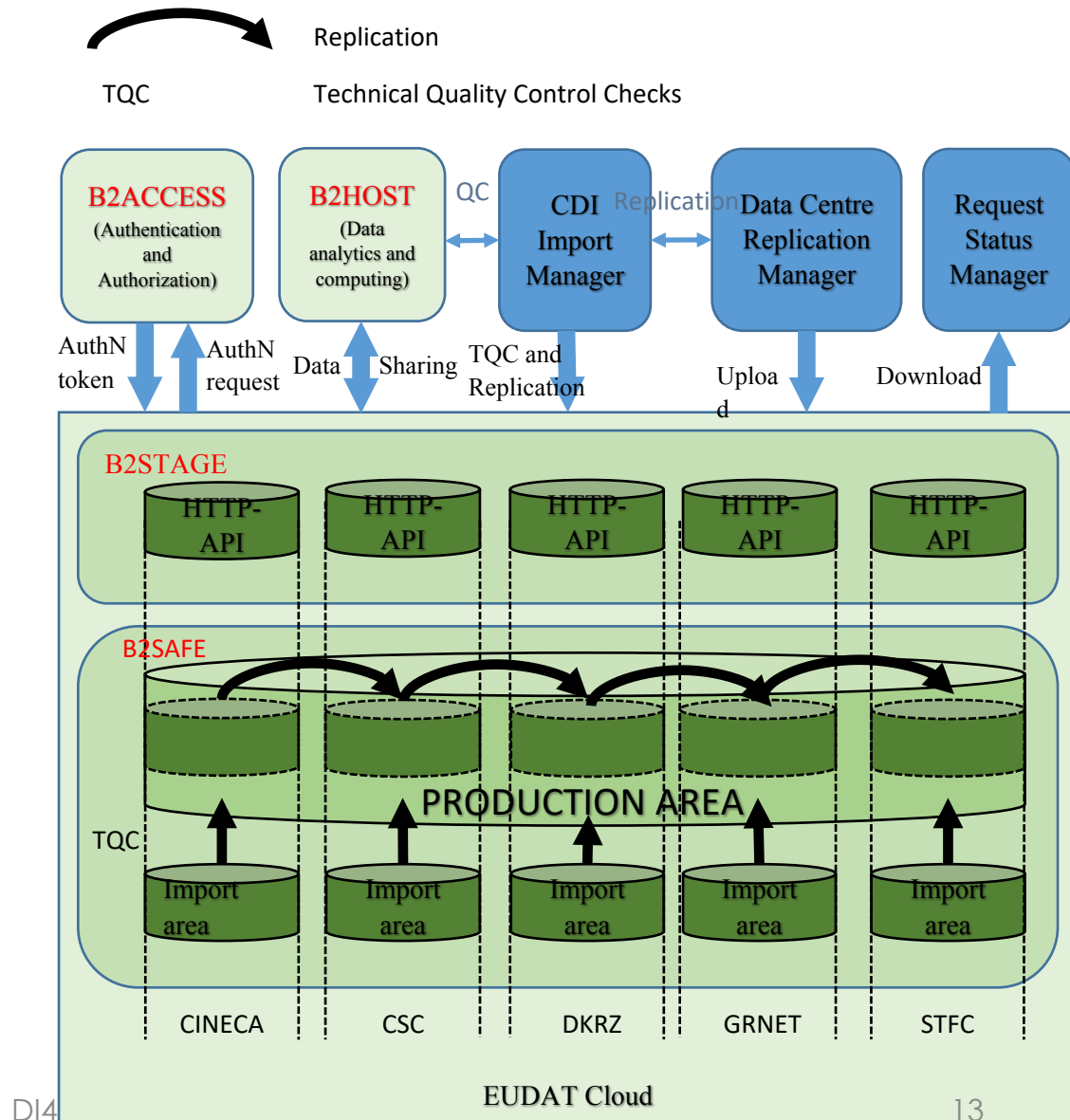
- **Eudat Cloud:** central data cache for all unrestricted data sets. Is driven by the CDI Import Manager. Interacts with the CDI Import Manager and the Replication Managers.
- **Replication Manager:** intermediates between CDI Import Manager, local data management systems and Data Cloud to replicate unrestricted data files to the Eudat cloud

Designing EUDAT cloud

- **EUDAT cloud** is formed by EUDAT CDI services such as B2ACCESS, B2STAGE, B2SAFE and B2HOST
- Local software tools at data centres to prepare ingestions
- **Replication Manager (RM)** at data centres for exchanging to Import Manager and EUDAT cloud
- Upgraded **CDI User Interface**, ordering and downloading facility



- Importing data from data centers is distributed across the five EUDAT partners
- Production area (i.e., quality passed data sets) replicated across the five EUAT partner



Integrating CDI system with EUDAT cloud

- To develop the new CDI system, different systems such as Replication Manager, Import Manager, Request Status Manager and EUDAT cloud have to be interoperable
- Challenges:
 - Systems are developed at different geographical locations
 - The work cultures and backgrounds of involved organizations were quite different

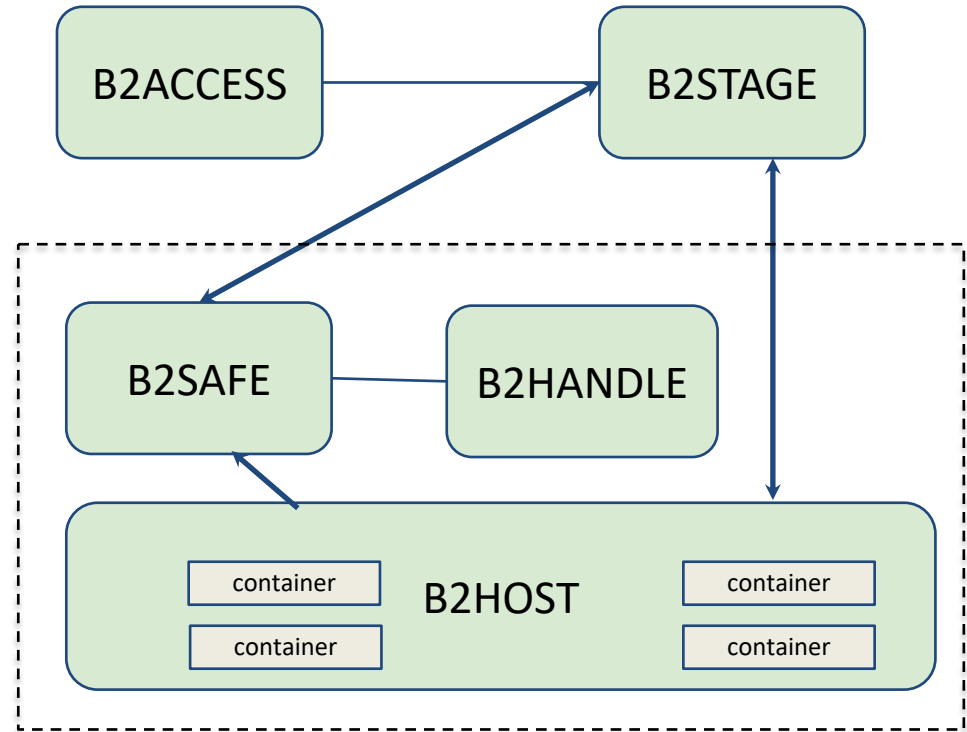
Upgrading CDI system: chosen solutions

- We have chosen REST APIs to enable interaction between EUDAT services and other components
- Defining the data format to exchange information between the systems – chosen data format is json
- Defining REST interfaces facilitated the independent development of systems and understanding the interfaces of different systems; helped in realizing the seamless communication between different systems

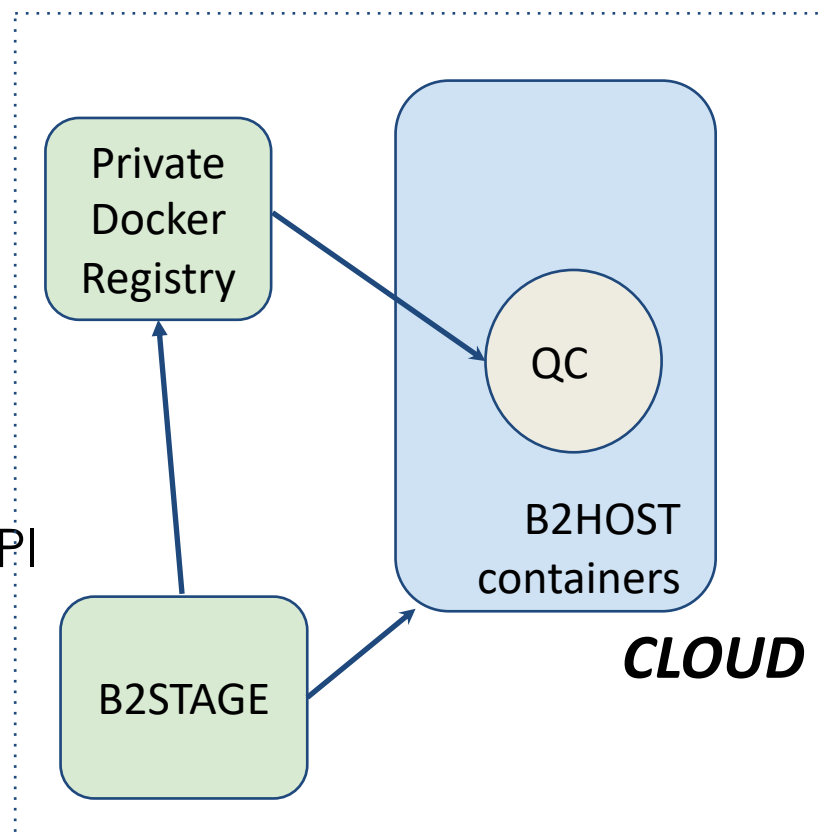
Upgrading CDI system: adapting EUDAT services

- The EUDAT service suite has been adapted in order to seamlessly integrate with existing CDI components

- **B2ACCESS**: User Authentication
- **B2SAFE**: Replicated Data Store
- **B2HANDLE**: PID Registration
- **B2STAGE**: API Server
- **B2HOST**: Containers Cluster
- **B2DROP**: User personal storage



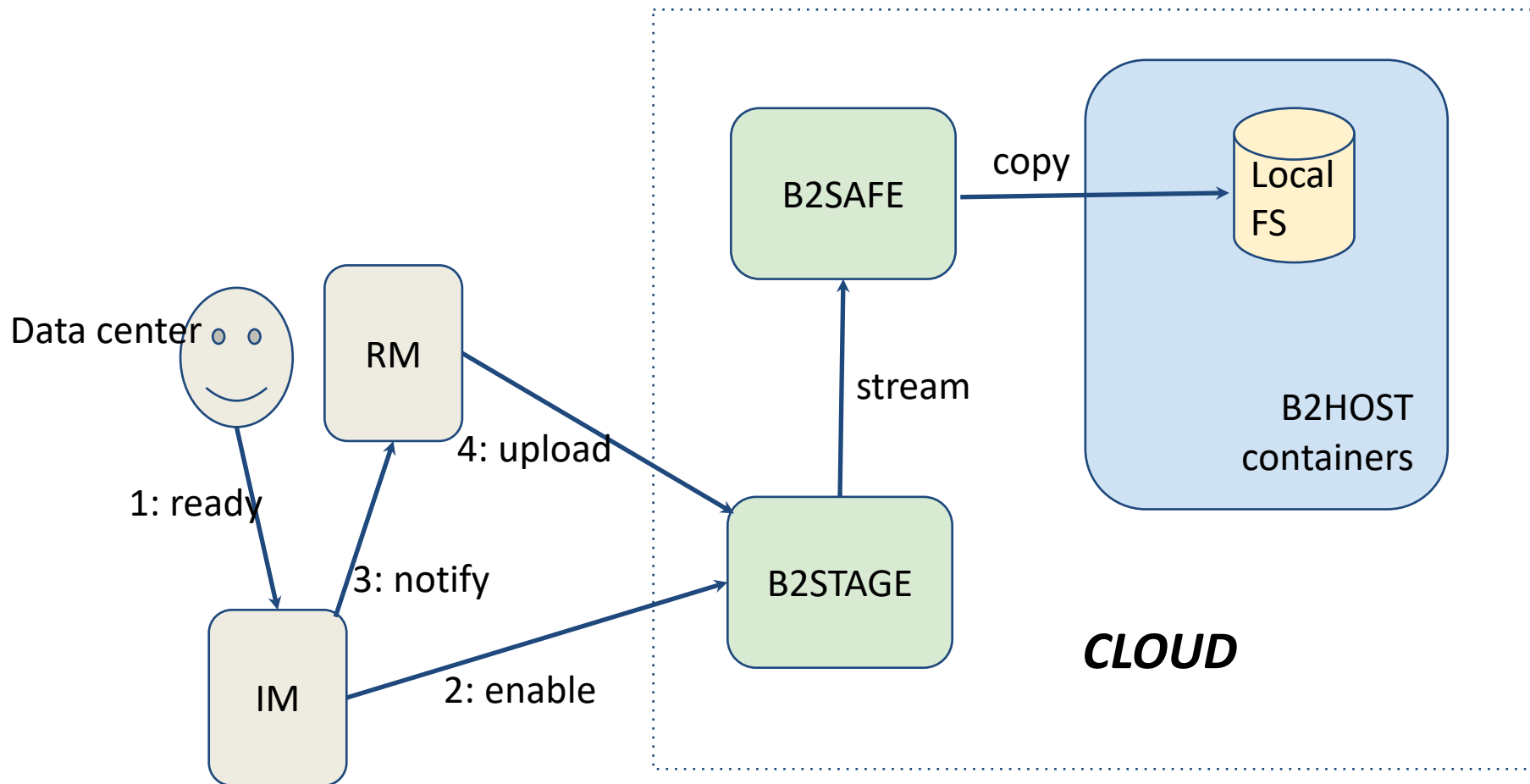
- ◆ B2HOST environment
 - ◆ Containers cluster
 - ◆ micro-services oriented
 - ◆ Flexible to add/remove services
 - ◆ Host and containers monitoring
 - ◆ Private HUB (container images)
 - ◆ One container per quality check
 - ◆ Managed through the B2STAGE API
 - ◆ Also other services are containers
 - ◆ Logging
 - ◆ RabbitMQ
 - ◆ ELK stack
 - ◆ Containers monitoring



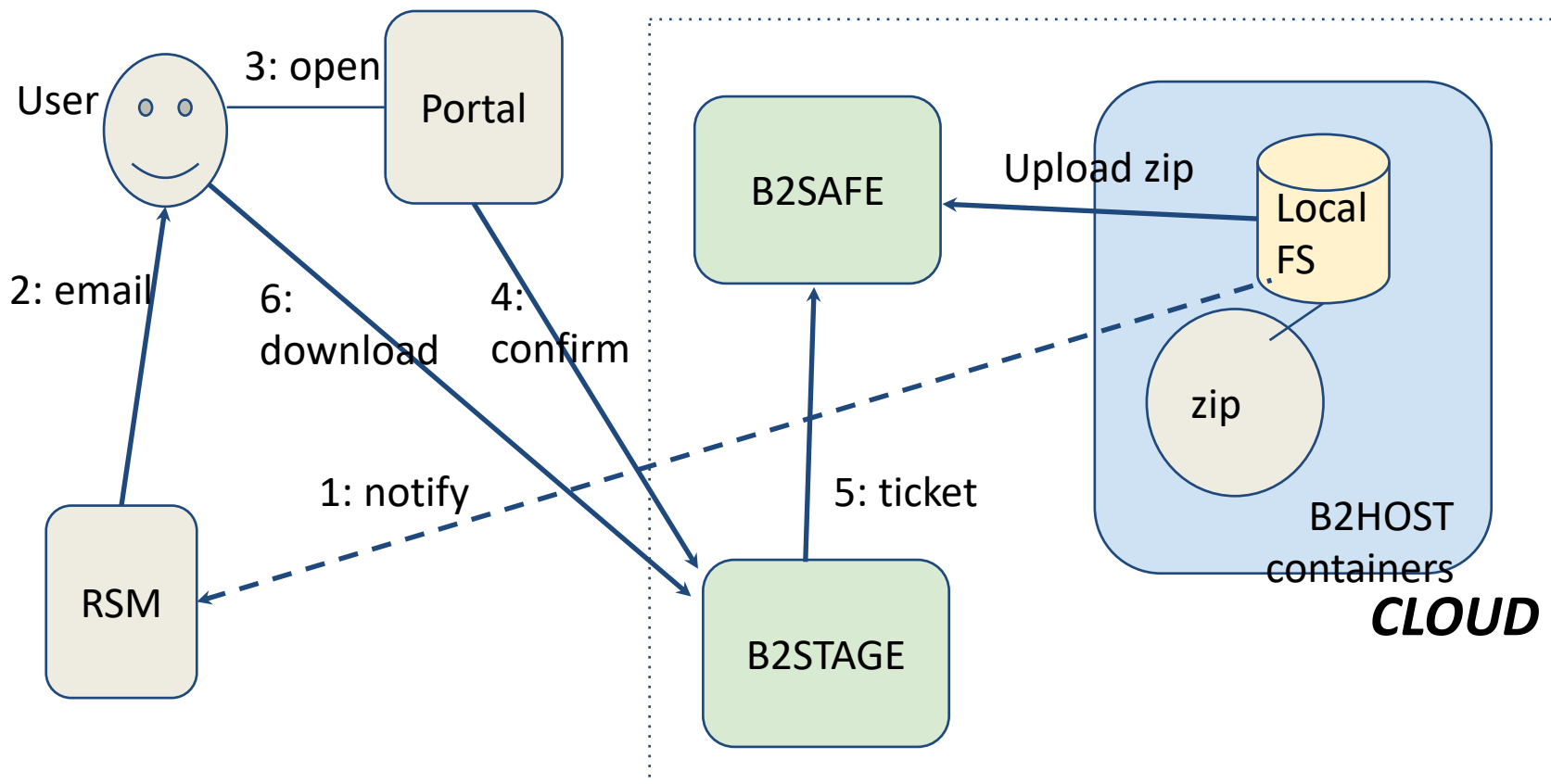
Upgrading CDI system: adapting B2STAGE

- B2STAGE REST APIs is used for all the interactions between the systems
 - for uploading data from data centers to EUDAT cloud storage
 - downloading data requested by users from EUDAT cloud storage
 - performing data management on the data in the cloud, etc.

Uploading data from data centers to EUDAT cloud storage



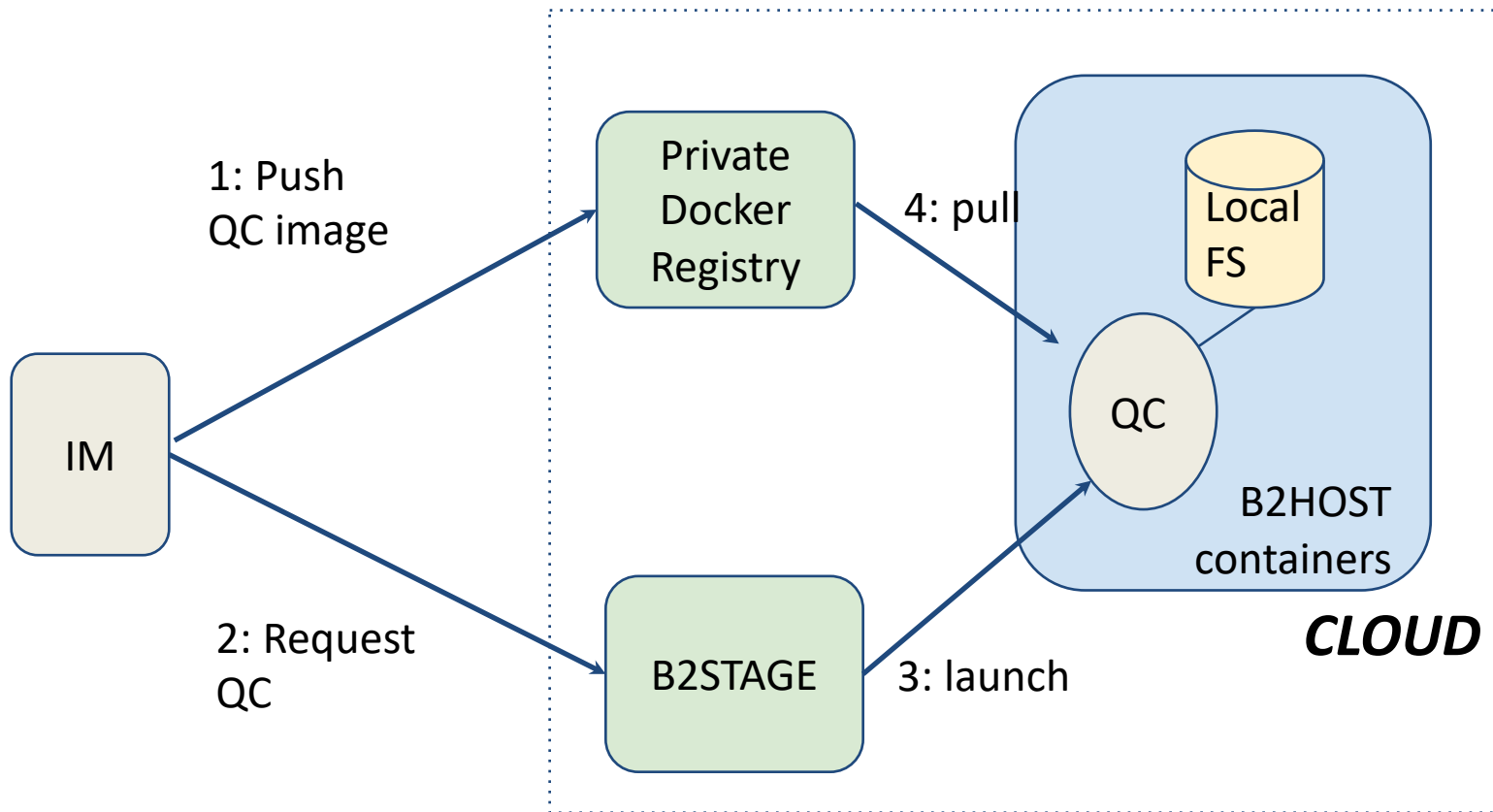
IM – import manager
 RM – replication manager



portal – CDI user portal

RSM – request status manager

Performing quality checks on the data deposited in cloud



IM – import manager

- The performance will be speeded up, discovery and data requests improved, and downloading made more easy
- Overall quality and coherence (data – metadata) will improve
- Possibility to evaluate the data deposited in the cloud, etc.

- It is possible to perform extra actions on the data in the cloud buffer:
 - checking overall quality of metadata and data, as extra check on top of local QA-QC by data centres
 - checking integrity of data files and metadata relations
 - results of checks to be reported back to data centres for amendments of their submissions and/or local configurations for mapping data and metadata
 - create transformation services for converting data sets to relevant INSPIRE data models
 - introduce versioning of metadata and data as part of provenance

Conclusions

- ◆ Following best modern practices
- ◆ Infrastructure is much more flexible
- ◆ Automated data management in the cloud could be applied to other communities that deposit data in the cloud
- ◆ Opens new possibilities to automatically apply different artificial intelligence models on the data
- ◆ Allows communities to monitor and evaluate the quality of ingested data before delivering data to users



Contact information:

Chris Ariyo Chris.Ariyo@csc.fi
Service Manager, Research Data Services, CSC – IT Center for Science, Finland

Dick M.A. Schaap (MARIS - Netherlands)
Technical Coordinator SeaDataCloud Project

Damien Lecarpentier damien.lecarpentier@csc.fi
Project Director, Research Infrastructures, CSC – IT Center for Science, Finland
EUDAT Project Director

[https://b2\(service\).eudat.eu/](https://b2(service).eudat.eu/)
<http://www.eudat.eu/support-request>