



# SeaDataCloud

Deployment of EUDAT components for upgraded CDI  
service and securing operations



**ICT Solutions for Brilliant Minds**

Chris Ariyo (CSC – Finland)

SeaDataCloud Week, Brest, Plouzané, October 14-18<sup>th</sup>, 2019

## EUDAT CDI

- EUDAT CDI is a network of nodes that provide a range of services for data upload, retrieval, identification, replication. The nodes are essentially data centers
- CSC is a EUDAT node involved in many activities, but in particular the provision of production and test services

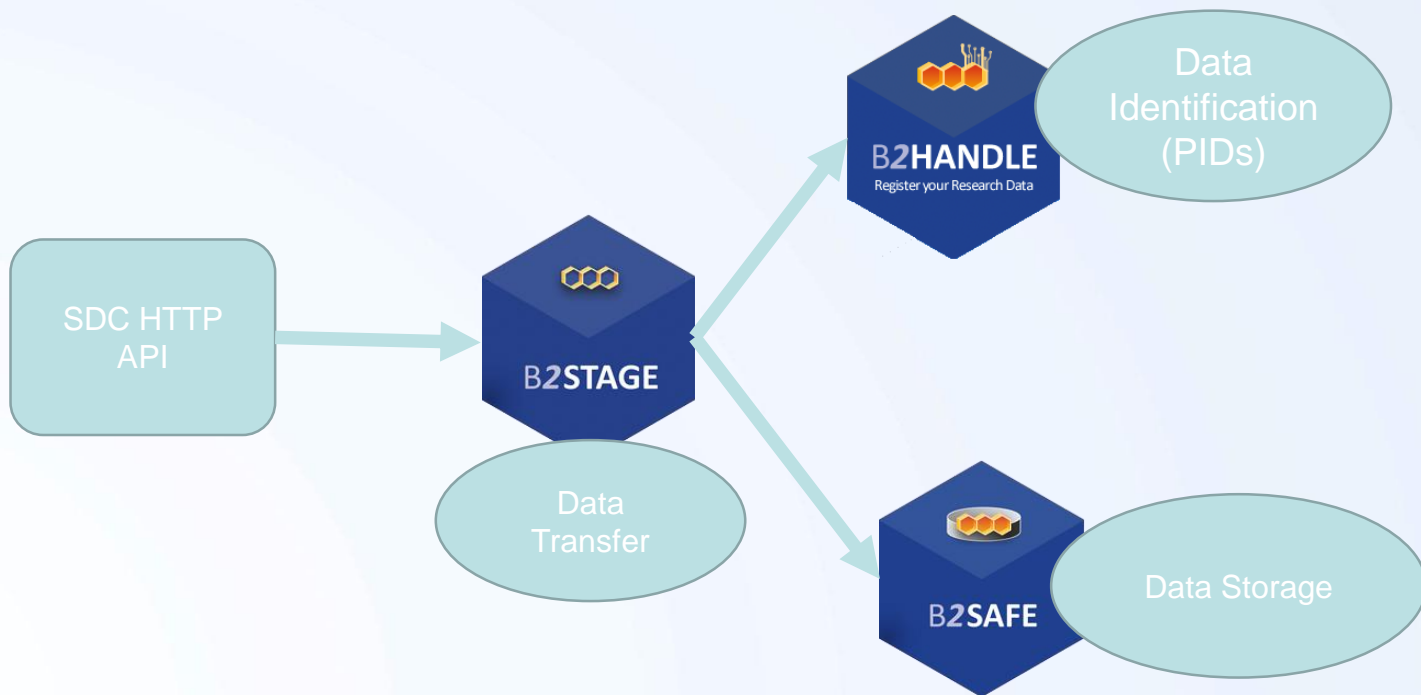
## The EUDAT Team



## B2STAGE

- SeaDataCloud HTTP APIs are part of the EUDAT CDI and it is build on top of the B2STAGE APIs
- One of the **core services** of the EUDAT Data Infrastructure
- Offers functionalities for **data transfer** between EUDAT resources and external facilities
- Built to extend **iRODS**, the core data technology of the EUDAT B2SAFE Service
- Implements an extensible **RESTful HTTP** interface

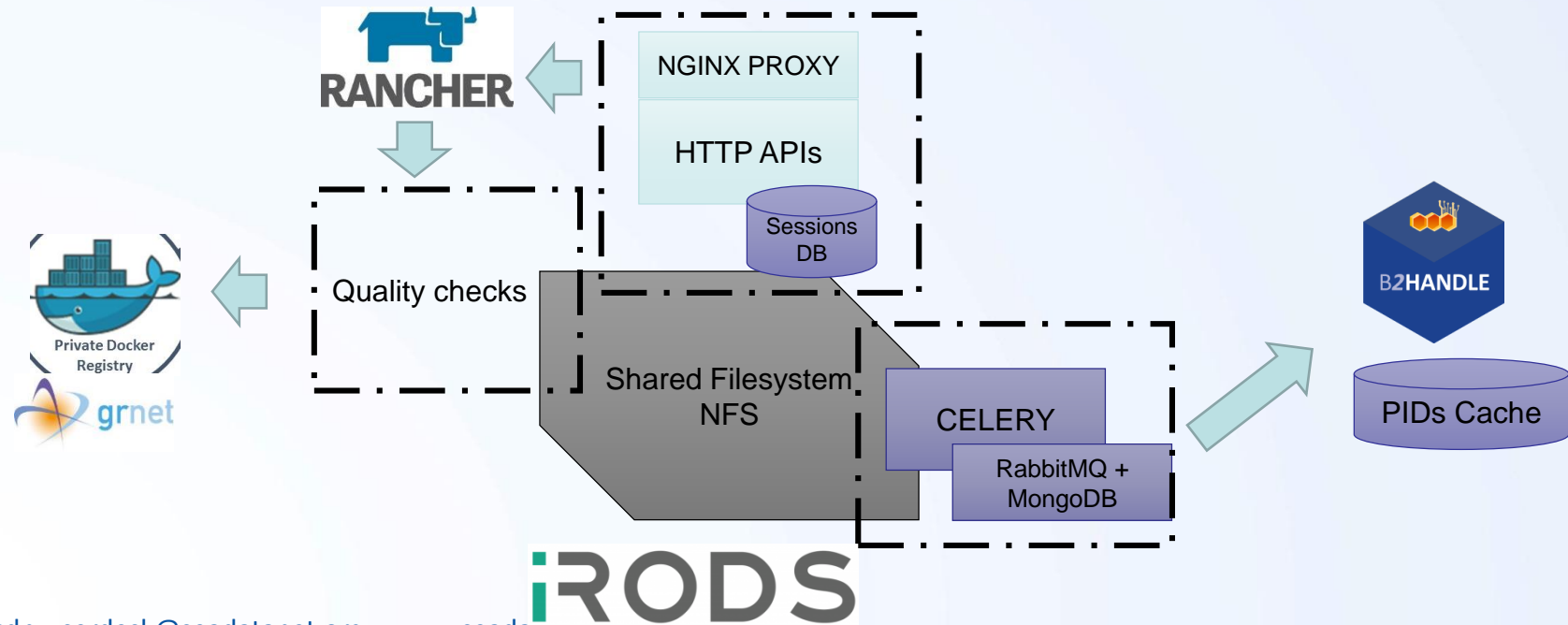
## EUDAT Services



## Service infrastructure & components

- HTTP APIs (B2STAGE + custom SDC functionalities)
  - Data transfer functionalities and interface of all other services
- iRODS (B2SAFE)
  - Data storage (replicated among all involved centers)
- Persistent Identifiers (PIDs, B2HANDLE)
  - Data identification
- Celery
  - Asynchronous operations (e.g. data ingestion and data harvesting)
- Rancher & Docker
  - Execution of quality checks
- RabbitMQ
  - Centralized logging archive
- Monitoring

## SDC HTTP API Architecture



## Three different environments

- Development (DEV) environment
- Test environment
- Production (PROD) environment

## DEVELOPMENT environment

- Deployment of latest **development version**
- Potentially untested/untrusted functionalities
- Only accessible to other **developers** (EUDAT, MARIS, IFREMER)
- **Temporary data**
- Currently installed at **CINECA**

## TEST environment

- Deployment of latest **stable version**
- Working functionalities
- Accessible to **all SDC partners**
- **Temporary data**

## PRODUCTION environment

- Deployment of latest **stable version**
- Fully working functionalities
- Accessible to **all SDC partners**
- **Production data**
- **Currently installed at CSC**

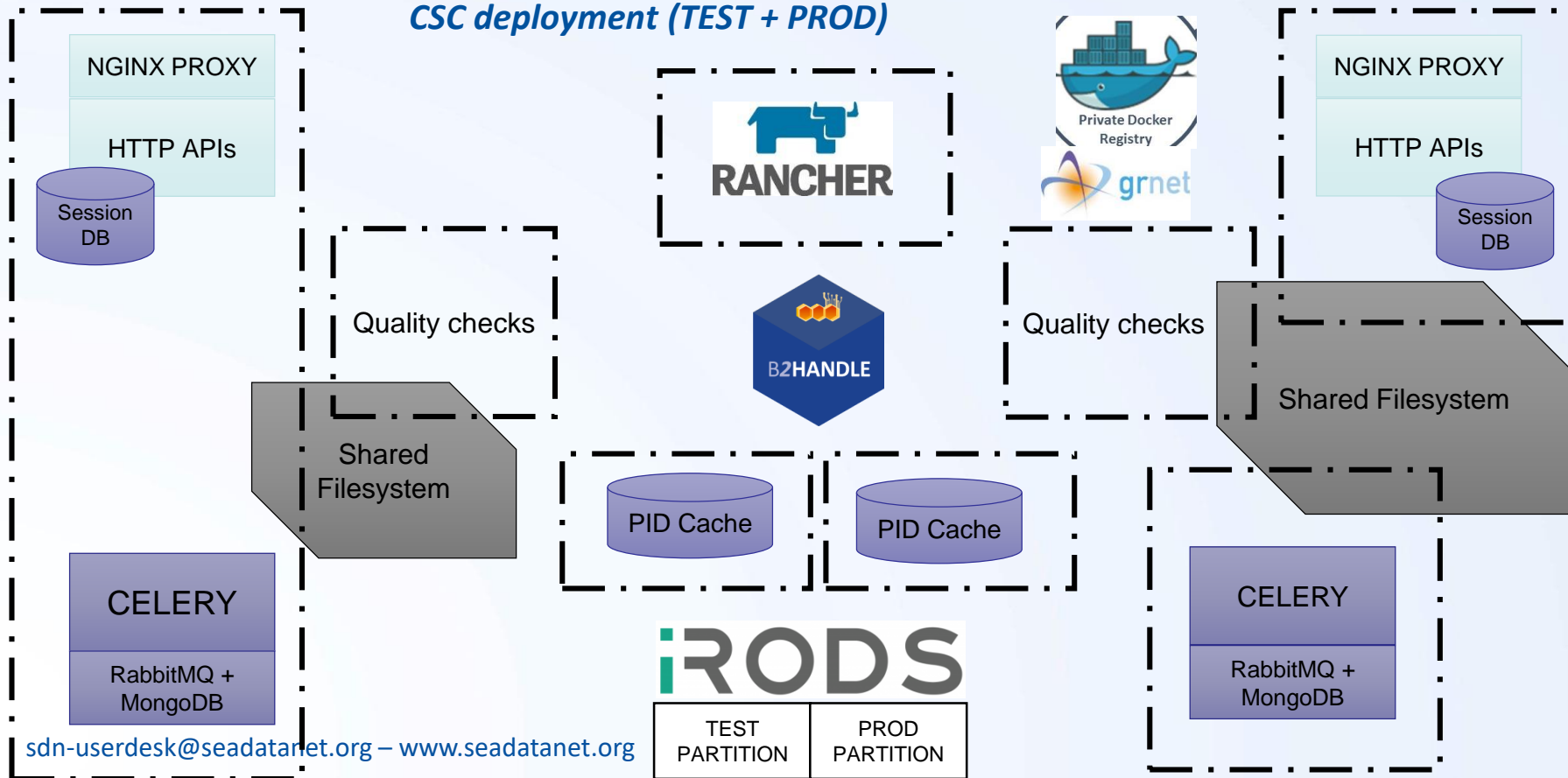
## CINECA deployment (DEV)

- 6 VMs, 42 cores, 126gb RAM
  - B2STAGE / HTTP API (2 cores, 6gb RAM)
  - iRODS (100gb) (8 cores, 24gb RAM)
  - Celery (8 workers) (8 cores, 24gb RAM)
  - Redis + Celery Backend + NFS (150GB) (8 cores, 24gb RAM)
  - Rancher (8 cores, 24gb RAM)
  - Quality checks (8 cores, 24gb RAM)
  - + remote DockerHub (GrNet)
  - + remote RabbitMQ (DKRZ)

## CSC deployment (TEST + PROD)

- 8 VMs, 44 cores, 120gb RAM
  - B2STAGE / HTTP API + NFS (400gb) (6cores, 16gb RAM)
  - iRODS (6+3 cores, 16+4gb RAM)
  - Celery (8 workers) + Redis + Celery Backend (8 cores, 32gb RAM)
  - Rancher (3 cores, 4gb RAM)
  - Quality checks (8 cores, 24gb RAM)
  - B2STAGE TEST + Celery TEST + NFS (350gb) (6 cores, 16gb RAM)
  - Quality Checks TEST (4 cores, 8gb RAM)
  - + remote DockerHub (GrNet)
  - + remote RabbitMQ (DKRZ)

*CSC deployment (TEST + PROD)*



Next step – deployment on all 5 centers



PROD  
TEST

PROD  
TEST

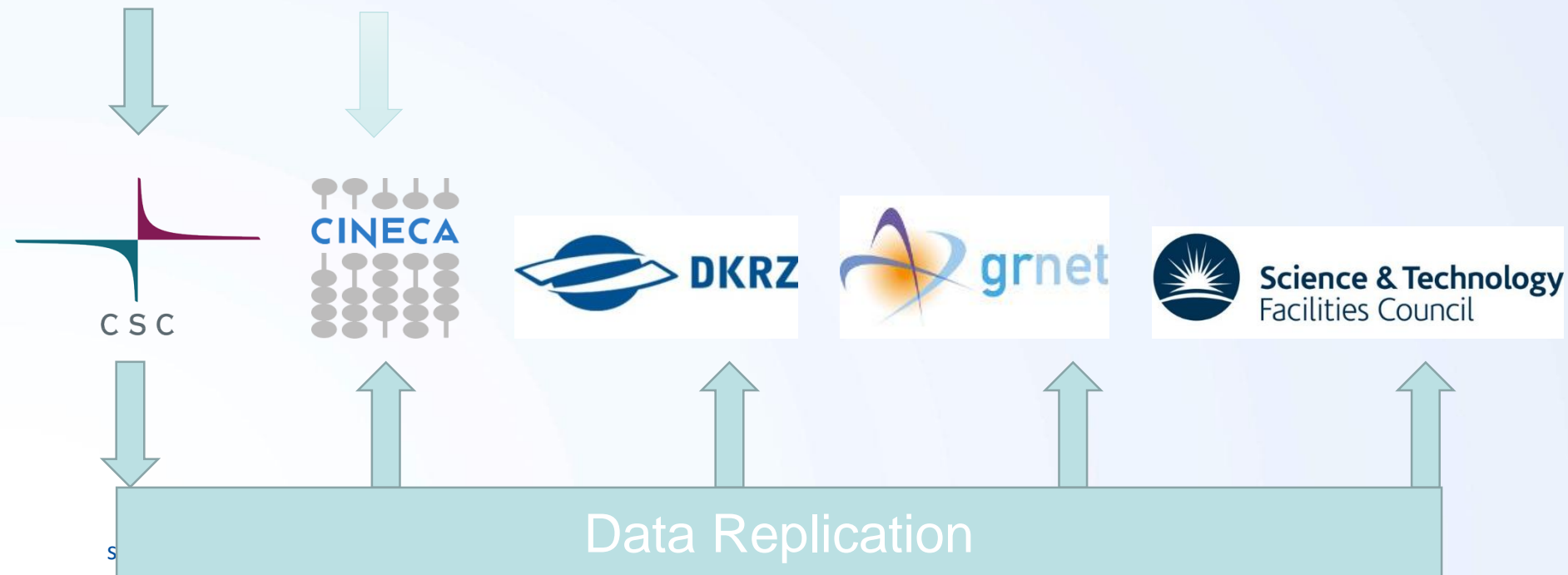
DEV

PROD  
TEST

PROD  
TEST

PROD  
TEST

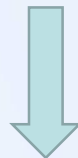
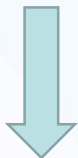
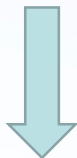
## Extended data ingestion workflow: data replication



## Extended data harvesting workflow



Proxy



Thank you for your attention



**ICT Solutions for Brilliant Minds**



*Chris Ariyo (Chris.Ariyo@csc.fi)*

sdn-userdesk@seadatanet.org – [www.seadatanet.org](http://www.seadatanet.org)