
Project name :
Towards an
e-infrastructure for plant
phenotyping
Update : 16/07/20

PI: Vincent Nègre, INRAE France

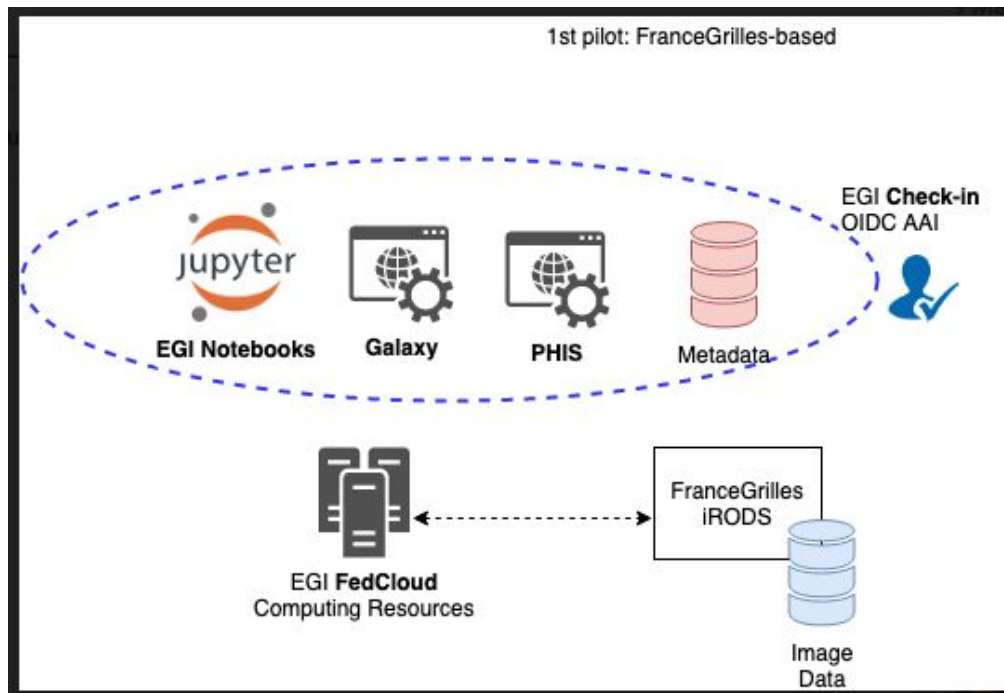
Shepherd : Nicolas Cazenave, CINES France

Context

Services we are expecting to use from **three pilots** are :

- **cloud services** : virtual machines and containers
 - **storage services** : B2SAFE service (iRODS) and EGI DataHub (One data)
 - **authentication** : EGI check In service
 - **computing service**: EGI notebook
-

First pilot



- PHIS + Galaxy environment will be deployed on **EGI virtual machines**
- The storage layer is based on the existing **FranceGrilles iRODS** infrastructure
- An authentication layer based on the **EGI check-in service**
- A computing layer provided with the **EGI Notebooks service** will be added.

First pilot

- **RQ1 Deploy virtual machines** : CESNET-MCC (or other sites if more performant), vo.emphasisproject.eu: 1VM ; 4CPUs with 32GB RAM; 80GB of storage for the system + 100GB of additional storage (Mongodb) - EGI cloud compute
- **RQ2 Install our information system (PHIS)** on the VMs - PHIS team
- **RQ3 Connect iRODS data with PHIS IS** - PHIS team
- **RQ4 Deploy Jupyter Notebook** : community-deployment for notebooks for 4 concurrent users (2 vCPUs cores, 4GB of RAM and 40GB of storage per notebook) - EGI
- **RQ5 Data available in Notebooks** - PHIS team
- **RQ6 Deploy Galaxy environnements** - Galaxy.eu
- **RQ7 Provide persistent identifier to the data** - B2HANDLE EUDAT/GRNET
- **RQ8 Federated authentication** should be integrated within PHIS IS. - check-in EGI/GRNET

done

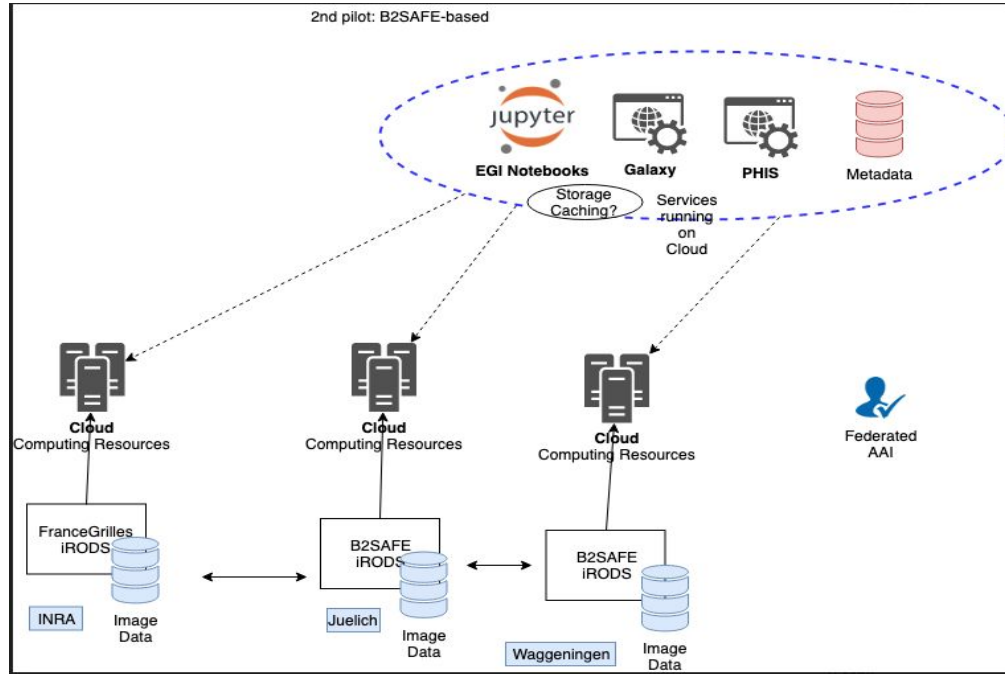
in progress

not started

RQ8 - Federated authentication

- **EGI check-In service.**
 - [GGUS ticket](#) is open.
 - OpenID protocol.
-

Second pilot



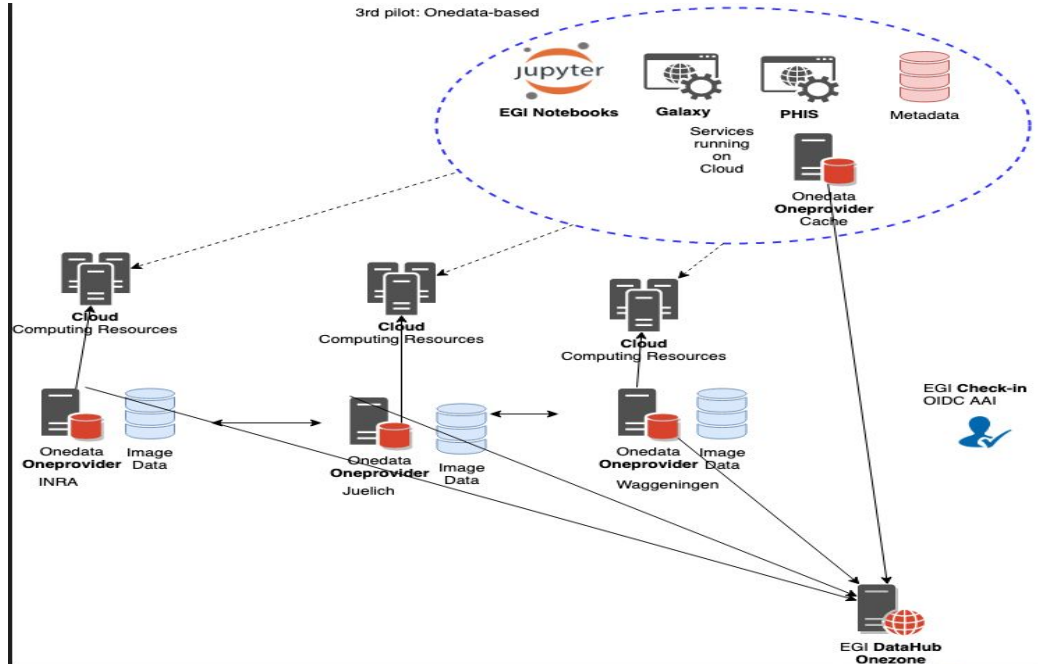
Compared to the previous pilot the **storage layer is based on the B2SAFE service** supported by the EGI infrastructure.

Second pilot

- Current status : not yet deployed



Third pilot



Compared to the previous pilot the **storage layer is based on the Data Hub service** supported by the EGI infrastructure.

Third pilot - Current status

- **RQ17 Provide virtual machines** : IN2P3-IRES vo.emphasisproject.eu - 1VM for PHIS IS; 4CPUs with 32GB RAM; 80GB of storage for the system + 100GB of additional storage (Mongodb) - Oneprovider VM with 8 vCPU, 32GB RAM with SSD - EGI cloud compute
- **RQ18 Install PHIS information system** on the VMs - PHIS team
- **RQ19 Provide 10TB of storage in EGI DataHub** - EGI
- **RQ20 Support to Connect PHIS IS EGI DataHub** - EGI
- **RQ21 Provide persistent identifier to the data** - B2HANDLE EUDAT/GRNET
- **RQ22 Deploy Jupyter Notebook** : community-deployment for notebooks for 4 concurrent users (2 vCPUs cores, 4GB of RAM and 40GB of storage per notebook) - EGI
- **RQ23 Data available in Notebooks** - PHIS team
- **RQ24 Deploy Galaxy environnements** - EGI

done

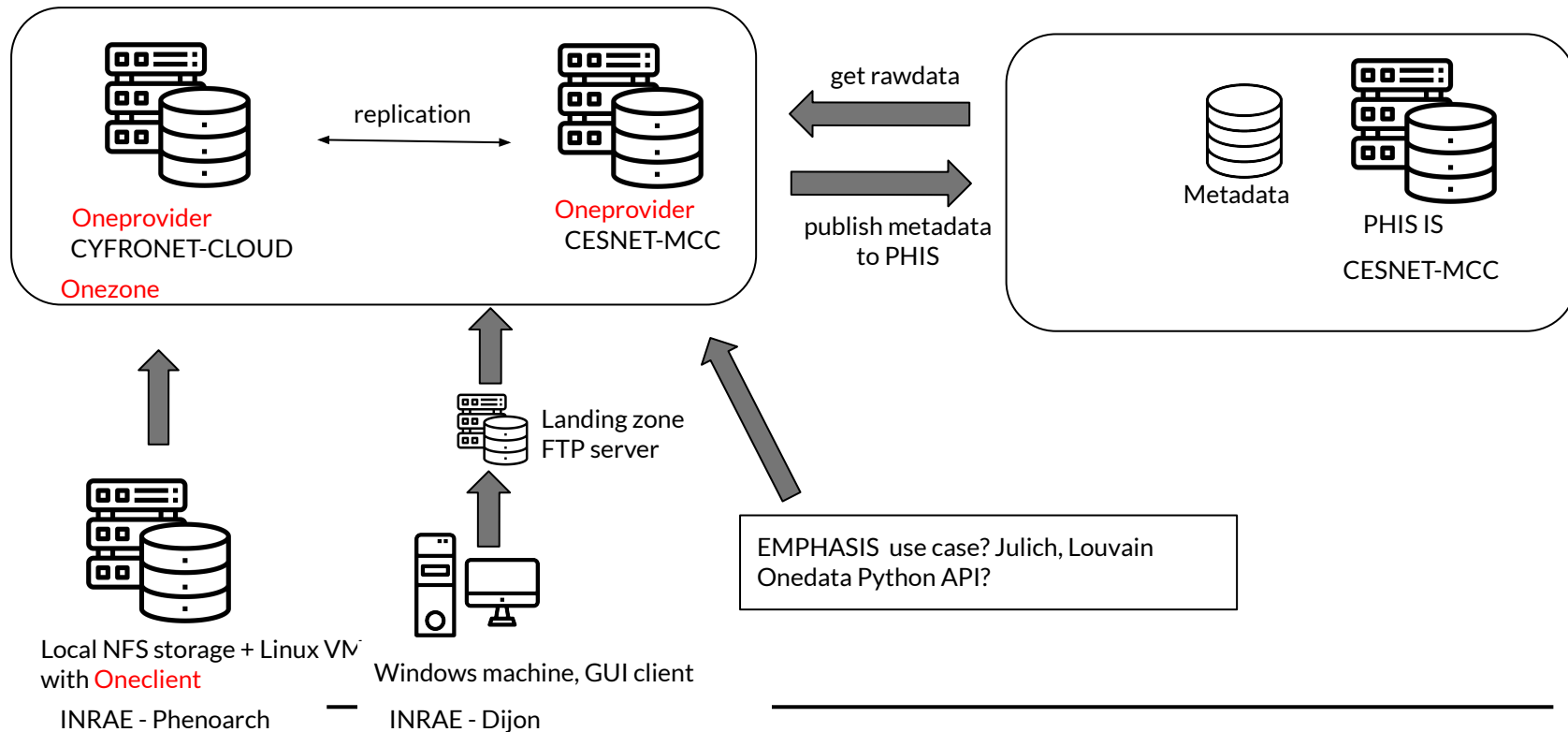
in progress

not started

Third pilot - RQ 20 Data replication

- Onedata team have been contacted.
 - Architecture have been proposed.
 - OneData space is available (10Tb replicated on 2 sites).
 - Data replication is under testing (performance issue has been solved, work in progress for automatic replication).
 - Data ingestion scenarios have to be tested.
 - Meta data registration have to be tested.
-

Third pilot - architecture



Third pilot - Data ingestion

- **scenario 1** : set up one VM (4 CPUS, 12Gb RAM, 20Gb stockage) connected (readonly) to the local storage server (NFS).
 - **scenario 2** : set up a FTP setup where users could upload data.
 - **scenario 3** : use oneclient (running on Linux, in testing on Windows with docker integration)
 - **scenario 4** (preferred scenario by OneData team). API Onedata (API http). Python package.
-

Questions?

