

**INRAE**

**➤ Dataverse : a key component of  
INRAE's institutional data  
infrastructure**

European Dataverse Workshop 2020 - 22/01/20

Michel Bamoumi

Esther Dzalé (@edzale)

Dimitri Szabo (@Dimitri\_Szabo)

# INRAE : a new dawn



+



# INRAE

INRAE

Data INRAE is a key component of INRAE's institutional infrastructure  
2019-01-23 / European Dataverse Workshop 2020 / Esther Dzale & Dimitri Stabio

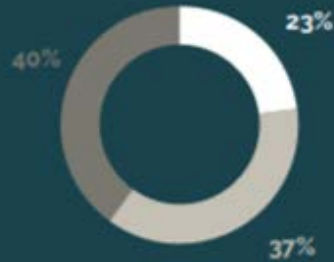


## Key Figures

Combined data from 2018 for INRA and IRSTEA

### Personnel

(employees at INRA and IRSTEA in 2018)



- > 1,965 researchers
- > 3,115 engineers and assistant engineers
- > 3,333 technicians

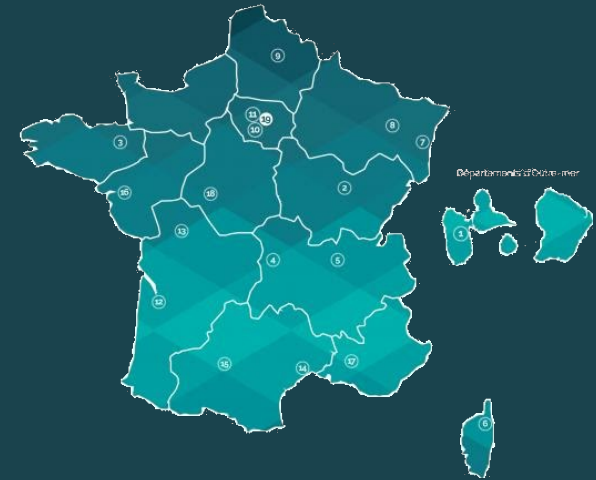


### Structure

**18**  
centres

**14**  
scientific  
divisions

**268**  
research units,  
experimental research  
units, and support units



## 6 major thematics

- Climate change and threats
- Agroecology
- Biodiversity
- Food, global safety
- Bioeconomy
- Society and *territoires*

**INRAE**

Dataverse : a key component of INRAE's institutional infrastructure  
2019-01-23 / European Dataverse Workshop 2020 / Esther Dzalé & Dimitri Szabo

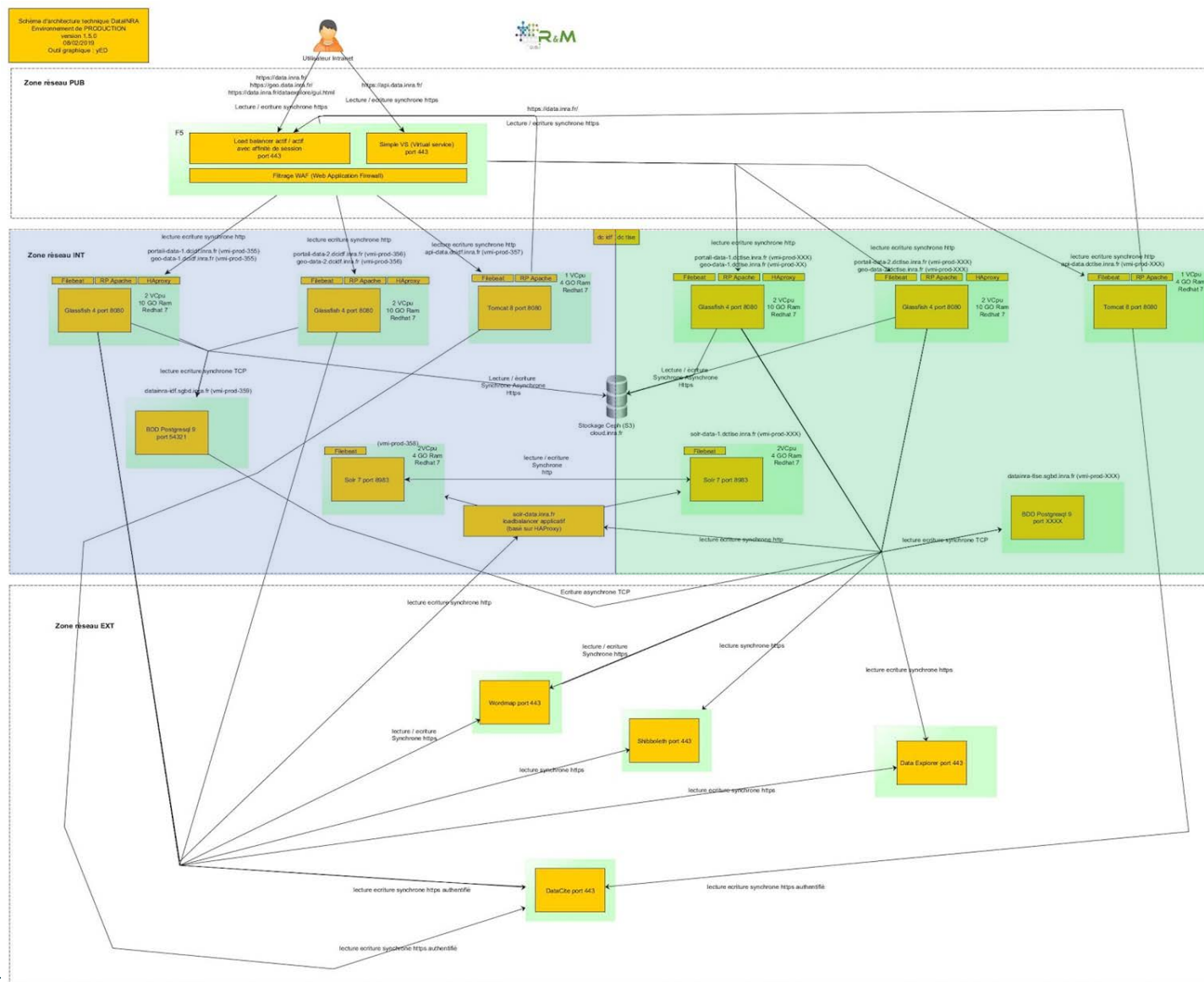
# Our Dataverse journey until now

- Preliminary study (6 months)
- A prototype (6 months)
- Go / no go decision
- Building and testing the production architecture (3 months)
- Deploying in production (march 2018)
- Upgrades



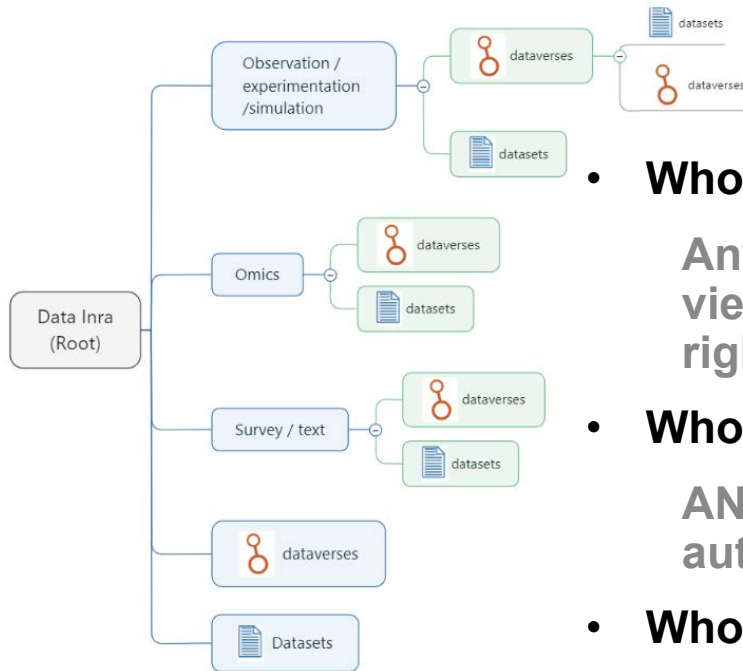
# How we deployed Dataverse

- Deployment on 2 Data centers
- 2x5 VMs in production
- CEPH S3 Object Storage
- Solr instances synchronize by Cross Data Center Replication (CDCR)
- 6 weeks data retention & data loss detection tool



Dataverse : a key cor

# Data management in Data INRAE



- **Who can view and access the data?**

Anyone can view the published metadata; the data are viewable and accessible depending on their access-rights

- **Who can deposit data in the repository**

Anyone with our institutional account + external authorized partners

- **Who can create a dataverse?**

Anyone with our institutional account who has been authorized and who accepted our administrators charter

- **Who is responsible of the quality of the metadata and the data?**

Each dataverse is managed by its owners and we provide training and a helpdesk to all users

**Data files associated with datasets are deposited in Data INRAE or in an external data system**



**INRAE**

Dataverse : a key component of INRAE's Institutional Infrastructure

2019-01-23 / European Dataverse Workshop 2020 / Esther Dzalé & Dimitri Szabo

# Challenges

- Unexpected data loss
- Training and raising awareness
- Engaging depositors to provide rich metadata and data in sustainable file formats
- Foster data reuse



# Perspectives

In addition to the regular updates

- Implement data governance principles
- Implement tools to facilitate the flow from active storage to Data INRAE then to long term archiving as smoothly as possible for the end-users
- Enable computing from Data INRAE
- Integrate Data INRAE with vocabulary repositories (metadata)



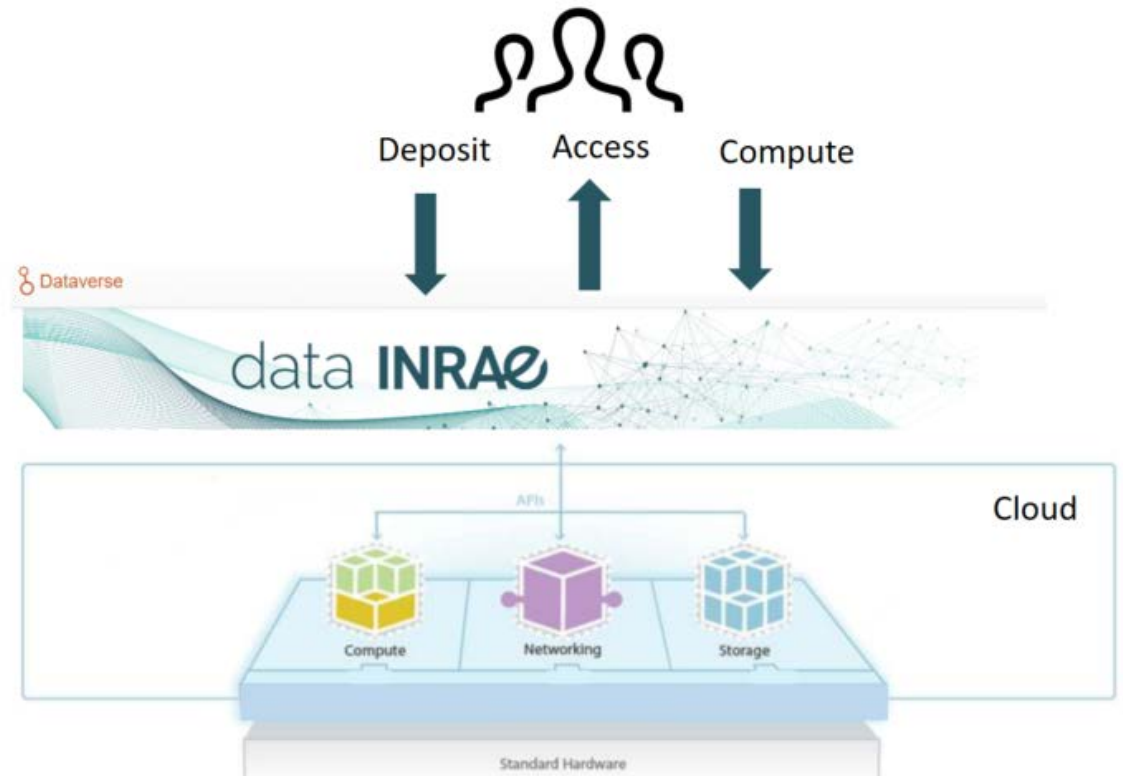


# Perspectives

Link with national and european projects

- BRIDGE a national funded project towards a common roadmap to improve institutional Dataverse repositories' fairness
- EOSC-Pillar

## ➤ Data INRAE in the cloud



**INRAE**

Dataverse : a key component of INRAE's

2019-01-23 / European Dataverse Workshop 2020 / Estelle Dzale & Dimitri Szabo

# Towards a national data repository?



**INRAE**

Dataverse : a key component of INRAE's institutional infrastructure  
2019-01-23 / European Dataverse Workshop 2020 / Esther Dzalé & Dimitri Szabo

# Thanks for listening



**INRAE**

Dataverse : a key component of INRAE's institutional infrastructure  
2019-01-23 / European Dataverse Workshop 2020 / Esther Dzalé & Dimitri Szabo