



# iRODS et la gestion de données

Jérôme Pansanel et Emmanuel Medernach

17 mai 2021



## Crédits

Cette présentation est basée sur la présentation cadre d'iRODS réalisée par Jason Coposky (directeur exécutif, consortium iRODS) :

- <https://slides.com/jasoncoposky>

## Le logiciel iRODS

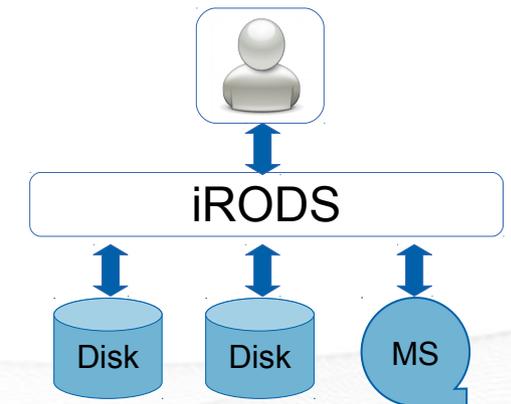
### iRule Oriented Data Systems

- Gestion de données
- Projet Open Source (BSD) succédant en 2006 à SRB
- Développé par le groupe DICE et un consortium (iRODS Consortium)
- Fonctionnement centré sur un moteur de règles

# iRODS

### Système de gestion de données

- Gestion de collection de données géographiquement distribuées (sites lointains)
- Infrastructures hétérogènes
- Utilisation intensive des métadonnées (modèle AVU)
- Organisation logique des fichiers indépendantes de leur organisation physique
- Consistance et homogénéité des formats peuvent être contraintes
- Documentation complète



## Éléments techniques

### Éléments principaux définissant une zone iRODS

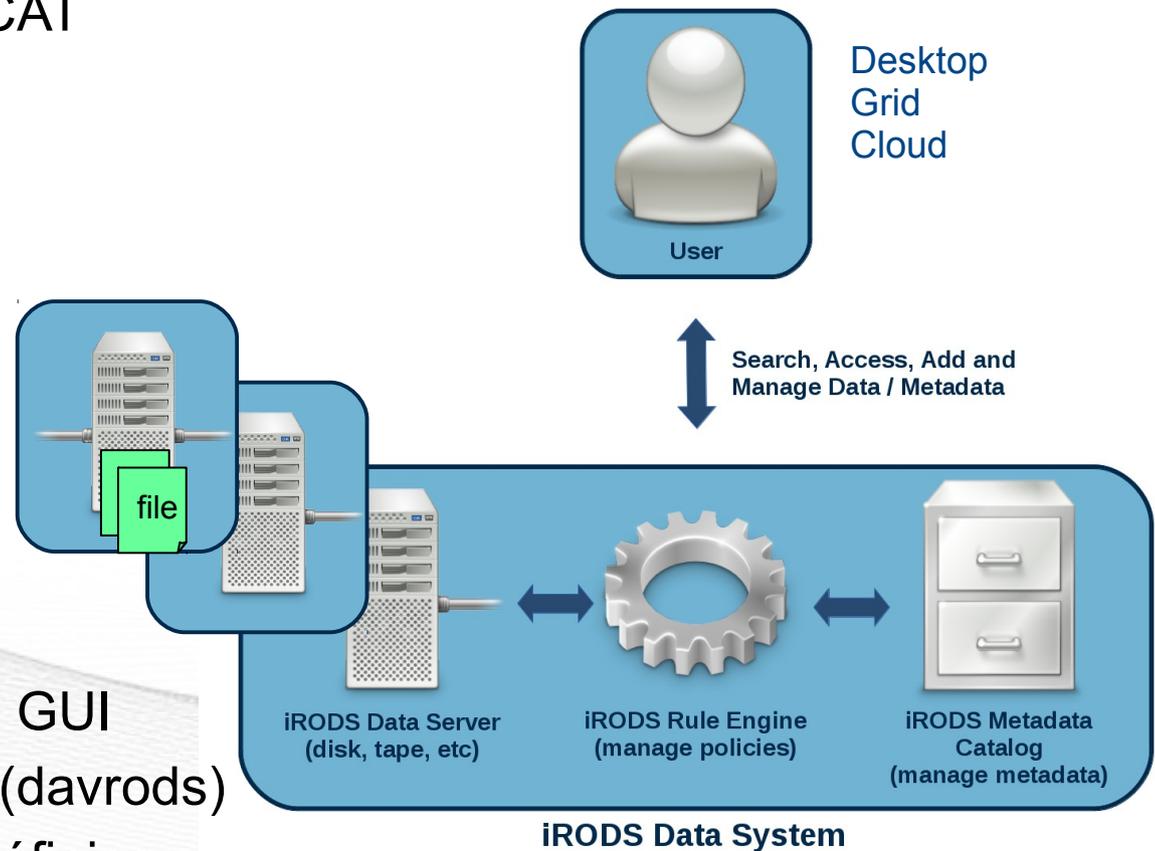
- Base de données + serveur iCAT
- Ensemble de ressources

### Authentification

- Mot de passe (base interne)
- PAM / LDAP, Kerberos
- Certificat, ...

### Gestion des données

- Interface utilisateur : CLI, API, GUI
- (metalnx, brocoli) et WebDav (davrods)
- Règles (flot de données) prédéfinies (ou pas) → transparence





## iRODS

# iRODS

— CONSORTIUM —

renci

RESEARCH \ ENGAGEMENT \ INNOVATION



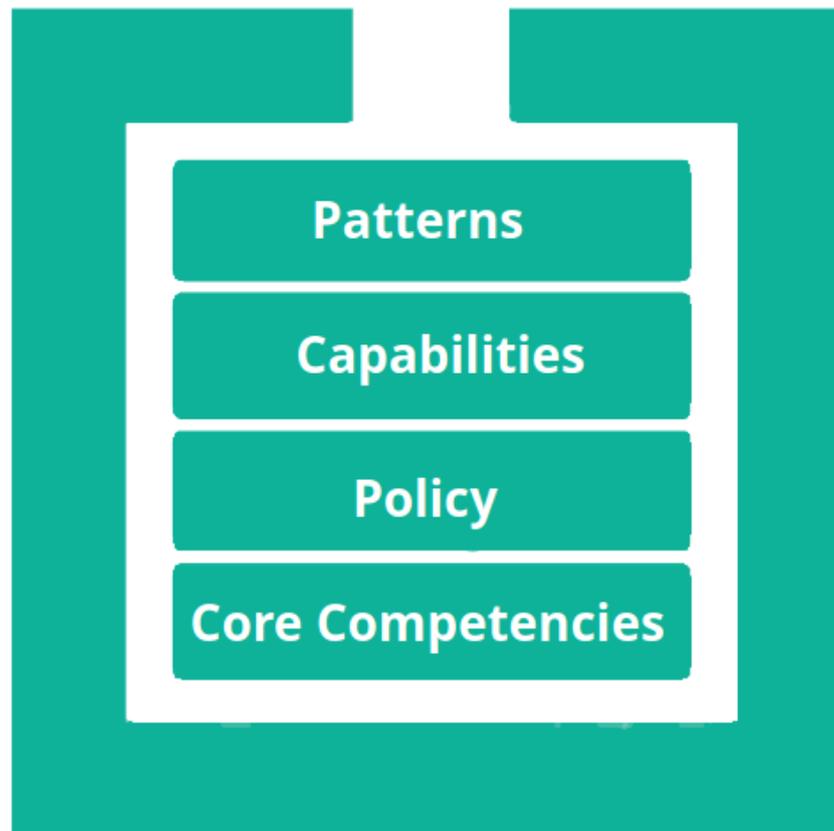
THE UNIVERSITY  
*of* NORTH CAROLINA  
*at* CHAPEL HILL

## Le consortium iRODS



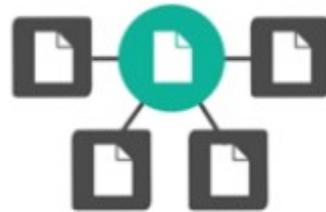


## Implémentation dans iRODS



## Core Competencies

**DATA  
VIRTUALIZATION**



**DATA  
DISCOVERY**



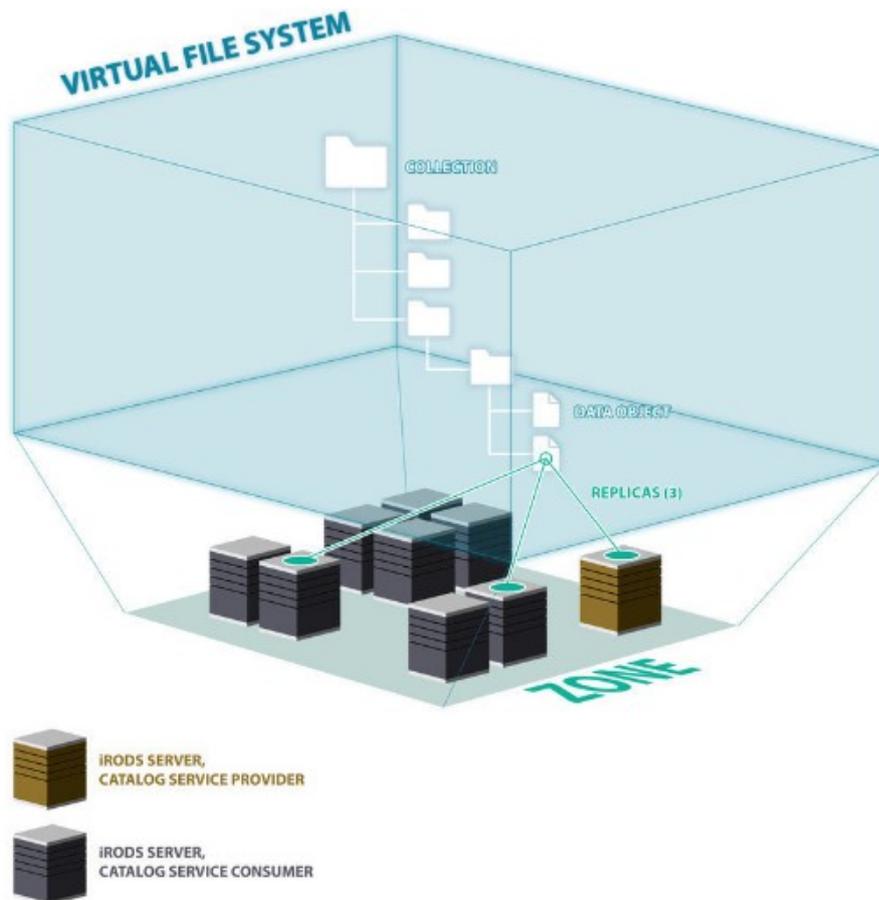
**WORKFLOW  
AUTOMATION**



**SECURE  
COLLABORATION**



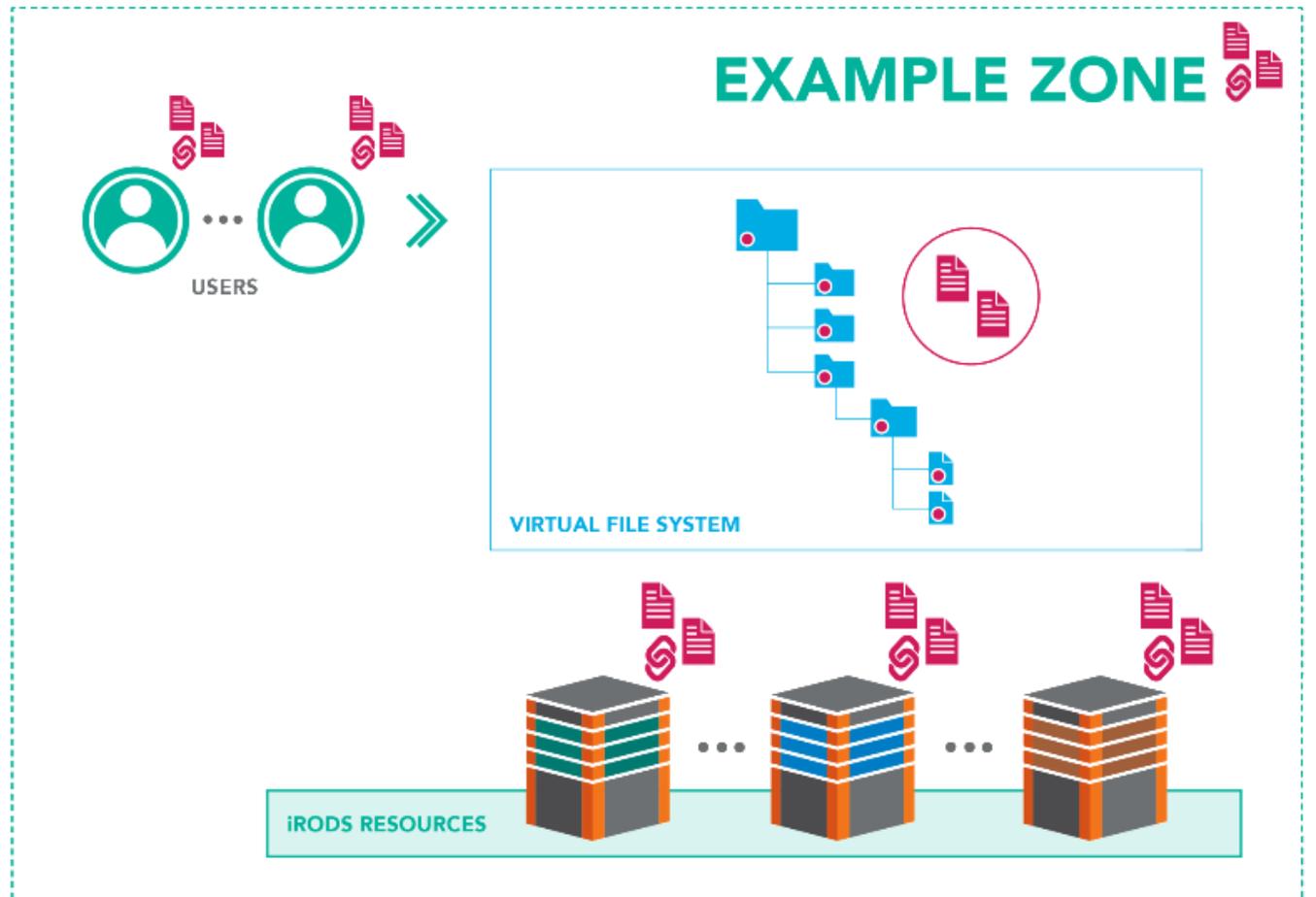
# Virtualisation des données



Chemin logique

Chemin(s) physique(s)

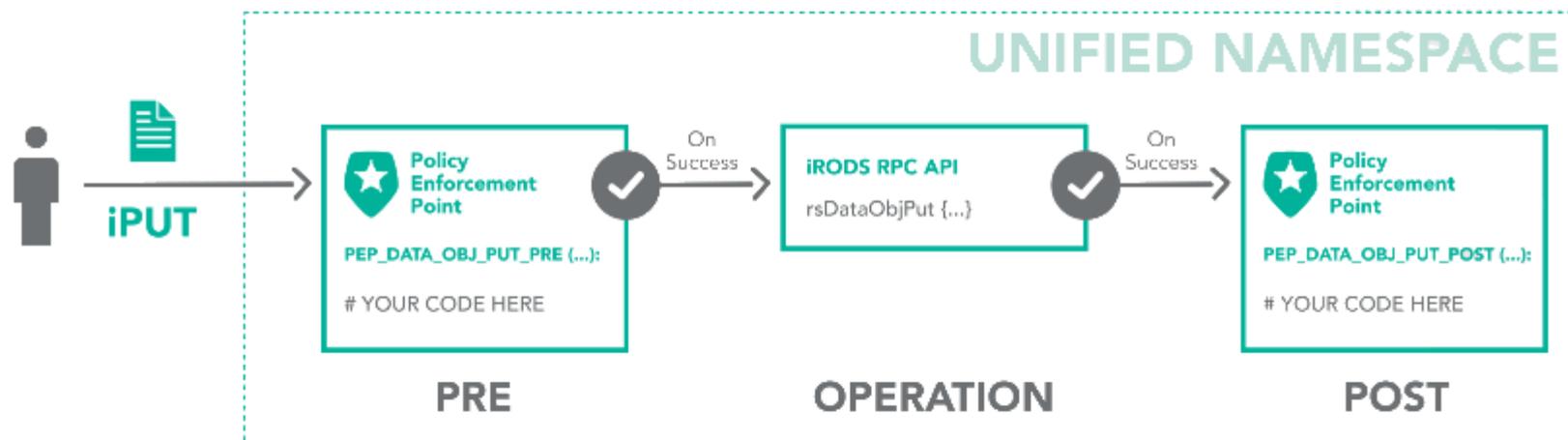
## Des données trouvables et exploitables



## Automatisation des flux de données

### Basée sur l'utilisation des règles

- Définie au niveau du système ou exécutée par l'utilisateur
- Exécution d'actions ou de micro-services
- Application immédiate ou différée



## Des collaborations sécurisées

### Fédérer les instances iRODS

- Possibilité de mettre en place des collaborations
- Chiffrement de bout en bout
- Fédération de zone
- À n'importe quel moment du cycle de vie de l'infrastructure
- Infrastructures restent indépendantes
- Stratégie d'évolution et de financement différentes entre les zones
- Collaborations temporaires

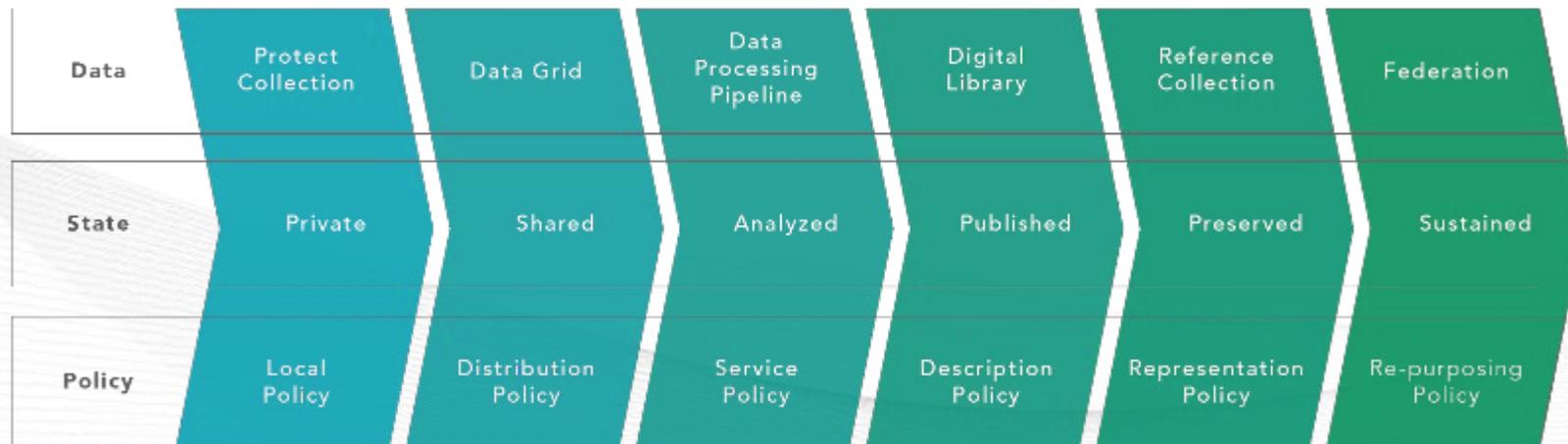




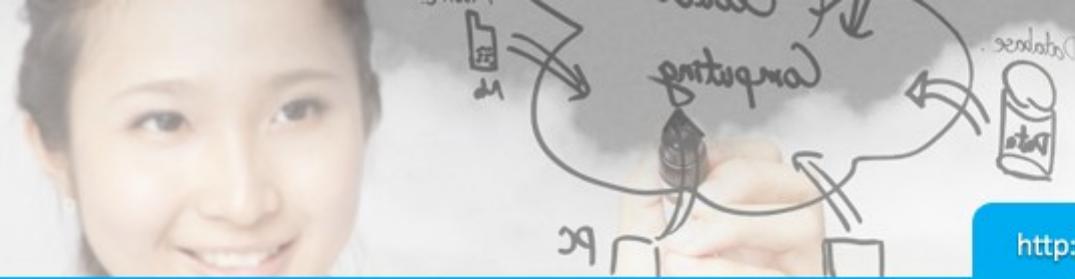
## De l'ingestion au dépôt institutionnel

À chaque fois que les données évoluent et qu'elles atteignent une communauté plus large, la politique de gestion des données doit évoluer pour répondre aux nouvelles exigences.

### DATA LIFECYCLE



iRODS virtualizes the stages of the data lifecycle through policy evolution



## Fonctionnalités



Automated Ingest



Storage Tiering



Auditing



Provenance



Indexing



Publishing



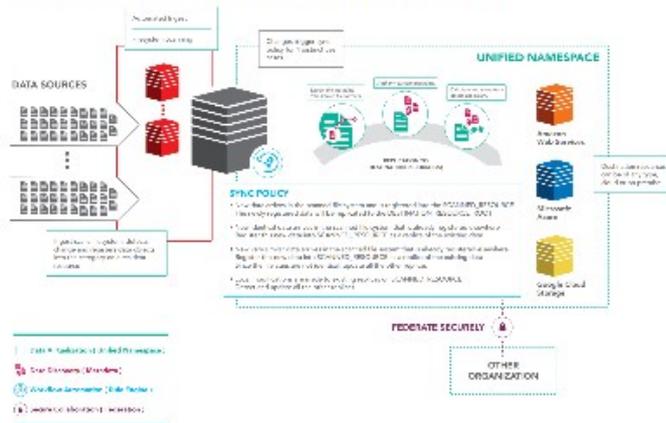
Data Integrity



Compliance

# Déploiements types (patterns)

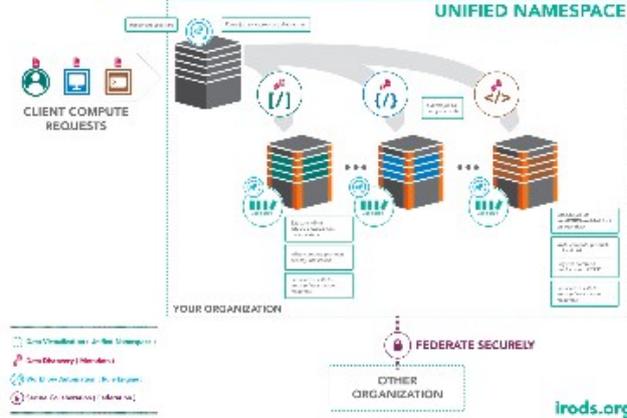
## Filesystem Synchronization



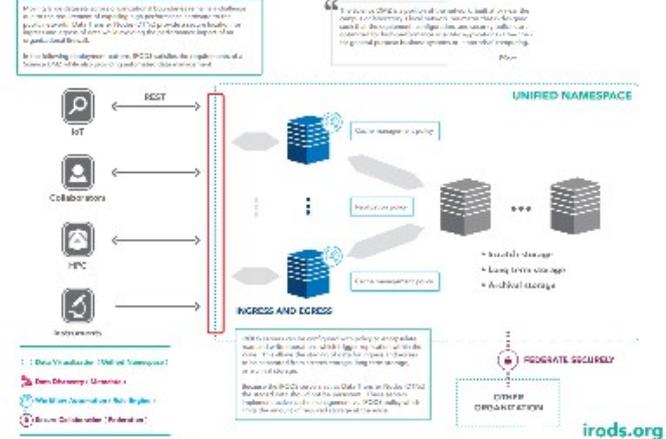
## Data to Compute



## Compute to Data



## Data Transfer Nodes





# Questions ?