



# Gestion de données avec iRODS

J. Pansanel

Réunion X/Stra – 2 février 2021



```
011011101101000011010010 1100
01100101001000000011000100111010
00100000011001101110011110011
00101110011 11011101011100100
011011111110101110 000101110
011011011100100100101110100
01100101001011 11001101110100
01110010001000100001100001
01101110 00100011 1101101101
01 11001100100110000101101110
01100100011010010111001110100
00101000001100000010111000 0000
00110001001010010010100101001
```

# iRODS

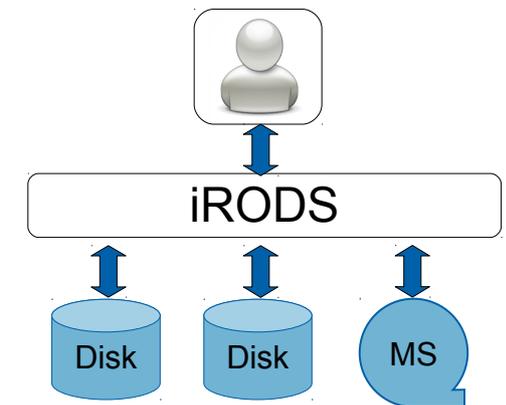
## iRule Oriented Data Systems

- Projet né en 2006, successeur de SRB
- Open Source, distribué sous licence libre BSD
- Développé par le groupe DICE et un consortium (iRODS Consortium)
- Centré sur un moteur de règle et les métadonnées



## 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
- Implémentation forte des politiques de gestion de données
- Documentation complète
- <https://irods.org/>



0111011101101000011010010  
0110010100100000001100010111010  
00100000011001101110011110011  
0010111001110111011001100100  
0110111101110101110000101110  
0111011101110000100100101110100  
011001010010111001101110100  
01110100010000100001100001  
01101110000001110110110101  
0011001100100110000101101110  
0110010001101001011011101100  
00101000001100000010110000000  
00110001001010010010100100101001

# Aux commandes



RESEARCH \ ENGAGEMENT \ INNOVATION



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

0111011101101000011010010 01100  
01100101001000000011000100111010  
001000000110011011100101110011  
00101110011 011011101001100100  
011011111101011100000101110  
01110111011000000000101110100  
01100101001010 011001101110100  
011101000100000000000100001  
01101110 001000011 0110110101  
0110011001001100000101101110  
0110010001101001011011100110100  
00101000001100000010111000 0000  
00110001001010010010100100101001

# Aperçu du consortium



```
0111011101101000011010010
0110010100100000001100010111010
00100000011001101110011110011
00101110011101101101100100
0110111101110101110000101110
0111011101100001001001110100
011001010010111001100110110100
01110010001000010001001100001
01101110001000011101101101
0110010001101000101101101100
00101000001100000010110000
00110001001010010010100101001
```

# Éléments techniques

## Éléments principaux définissant une zone iRODS

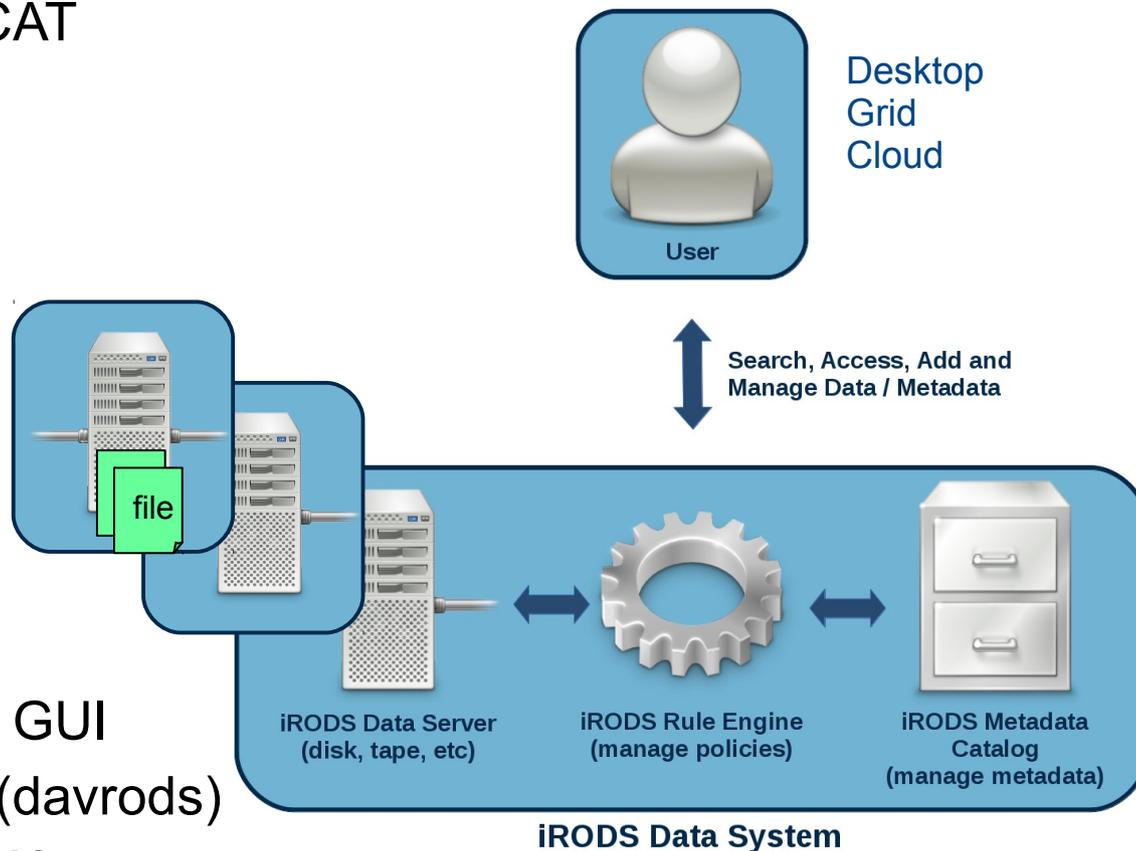
- Base de données + serveur iCAT
- Ensemble de ressources

## 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



```
011011101101000011010010 01100
01100101001000000011000100111010
00100000011001101110011110011
00101110011 010110101100100
0110111011101010110 000101110
01101101110000100100101110100
01100101001010 11001101110100
01100100010000000001100001
01101110 00100011 1101101101
01 11001110010011000010110110
011001000110100010110110110100
001010000011000000111000 0000
00110001001010010010100100101001
```

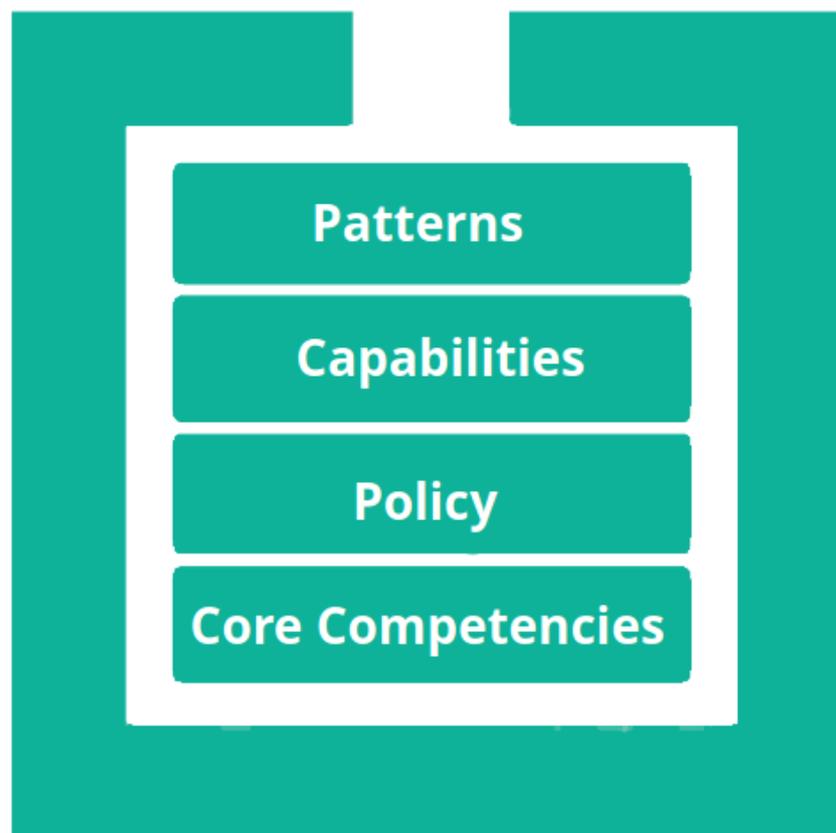
# Interfaces utilisateurs

## Plusieurs modes d'accès

- CLI (unix-like)
- APIs (Python, Java, etc)
- Interface Web : <https://metalnx.github.io/>
- Brocoli : <https://github.com/mesocentre-mcia/brocoli>
- NFS : [https://github.com/irods/irods\\_client\\_nfsrods](https://github.com/irods/irods_client_nfsrods)
- Webdav : <https://github.com/UtrechtUniversity/davrods>
- Fuse : [https://github.com/irods/irods\\_client\\_fuse](https://github.com/irods/irods_client_fuse)

```
0111011101101000011010010 01100
0110010100100000001100010111010
00100000011001011100101110011
00101110011 0101110101100100
0110111101110101110 000101110
011101110110000100101110100
01100101001010 11001101110100
011101000100000000000100001
01101110 00100011 1101101101
00 11001100100110000101101110
011001000110100101101110110100
001010000011000000111000 0000
00110001001010010010100100101001
```

# Modèle iRODS



```
0111011101101000011010010 01100
01100101001000000011000100111010
001000000110011011100101110011
00101110011 011011101001100100
011011110110101110 000101110
011101101110000100101110100
01100101001010 11001101110100
0111001000100001000100001
01101110 00100011 1101101101
00 1100110010011000010110110
011001000110100101101110110100
0010100000110000001011000 0000
00110001001010010010100101001
```

# Core Competencies



```
011011101101000011010010 1100
01100101001000000011000100111010
00100000011001101110011110011
00101110011 1101110101100100
011011111101010110000101110
011011101110000100100101110100
011001010010111100110110100
01110010001000000001100001
01101110 00100011 1101101101
01 110011100100110000101101110
0110010001101001011011100100
0010100000110000001011000 0000
00110001001010010010100100101001
```

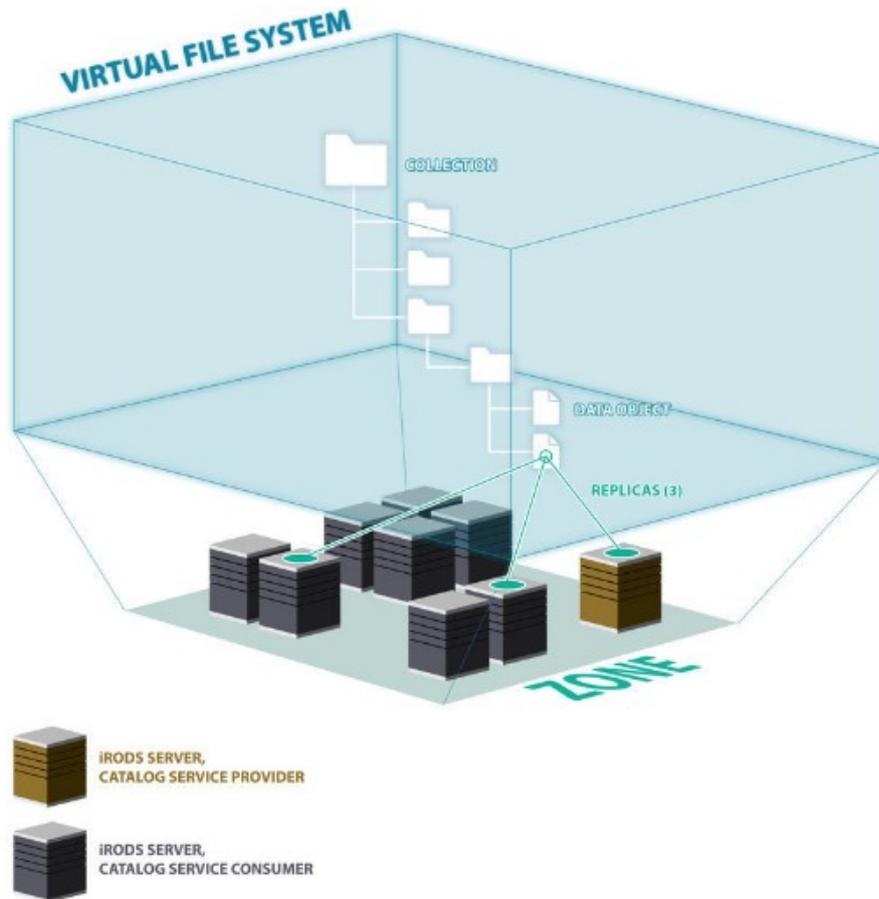
# Virtualisation des ressources

## Virtualisation

- Accès simultané à différentes technologies à travers un seul espace de nom (zone) :
  - Systèmes de fichiers existants (POSIX)
  - Systèmes spécifiques (DDN, etc)
  - Stockage Cloud (S3)
  - Données sur bande (HPSS)
- Vue logique d'une représentation physique qui peut être complexe, géographiquement distribuée et à différentes échelles



# Virtualisation des ressources (projection)



Chemin logique

Chemin(s) physique(s)

```
0111011101101000011010010 1100
0110010100100000001100010111010
00100000011001101110011110011
00101110011 11011101011100100
011011111110101110 000101110
011101110110000100101110100
01100101001011 11001101110100
01110100010001000001100001
01101110 00100011 1101101101
01 11001100100110000101101110
0110010001101001011011100110100
0010100000110000001011000 0000
00110001001010010010100100101001
```

# Data Discovery

## Métadonnées

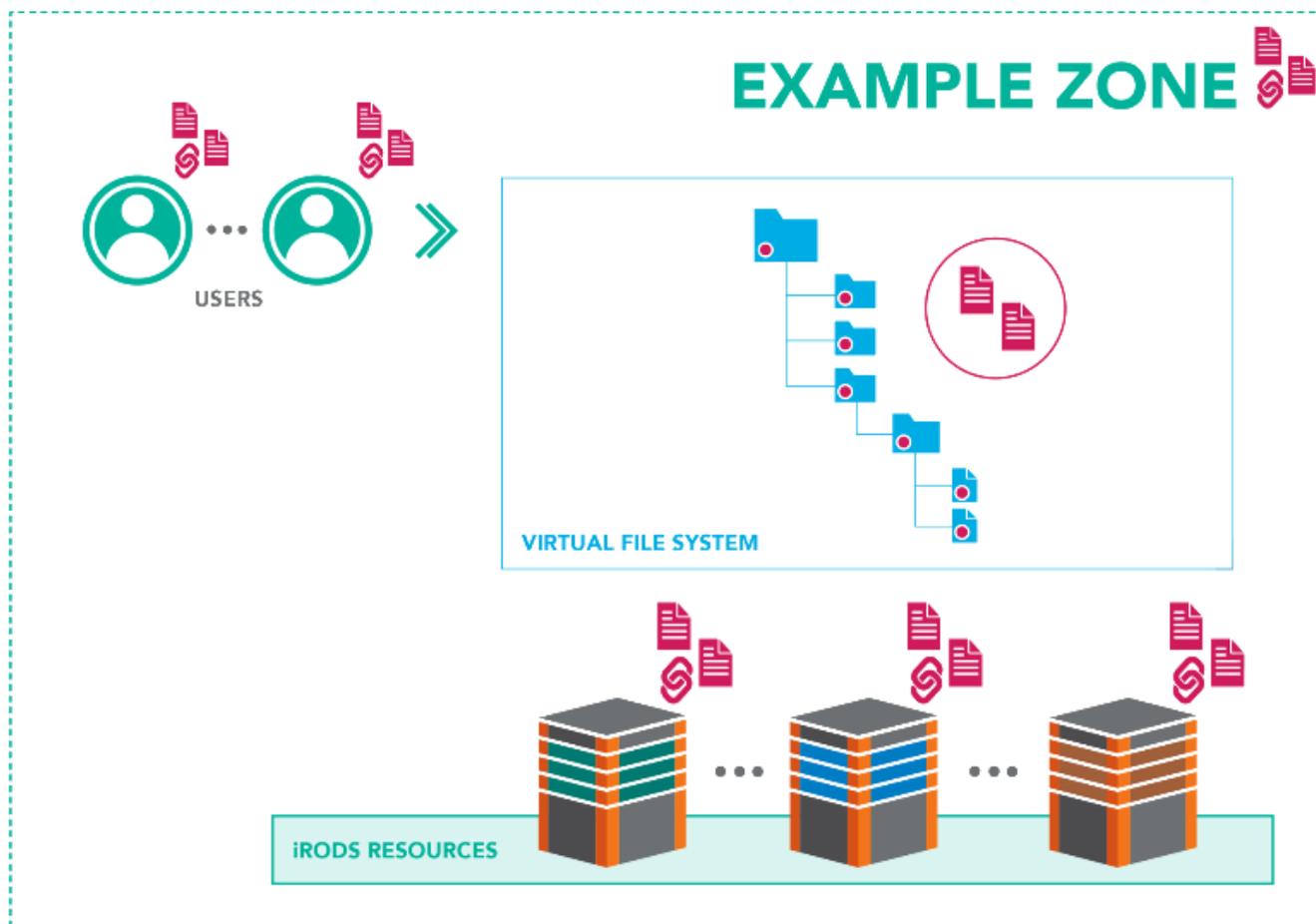
- Possibilité d'attacher des métadonnées à chaque type d'entité dans une zone iRODS :
  - Données (*data objects*)
  - Répertoires (*collections*)
  - Utilisateurs
  - Ressources de stockage
  - Espace de nom
- iRODS fournit un mécanisme de métadonnées permettant à la fois d'automatiser leur attribution, ainsi qu'aux utilisateurs de définir les leurs.
- Une infrastructure de données qui est plus accessible, opérationnelle et valorisable.

DATA  
DISCOVERY



011011101101000011010010 01100  
01100101001000000011000100111010  
00100000011001101110011110011  
00101110011 0101110101100100  
01101111110101110000101110  
01101101110000100100101110100  
01100101001010 11001101110100  
011100100010000000001100001  
01101110 00100011 1101101101  
01 110011100100110000101101110  
011001000110100101101110110100  
0010100000110000001011000 0000  
00110001001010010010100101001

# Des métadonnées partout



# Exemple de manipulation de métadonnées

```
[user ~]$ imeta add -d run0977156_123.raw length 10 cm
[user ~]$ imeta add -d run0977156_123.raw hall east
[user ~]$ imeta ls -d run0977156_123.raw
AVUs defined for /FranceGrillesZone/projet/run0977156_123.raw:
attribute: length
value: 10
units: cm
----
attribute: hall
value: east
units:
[user ~]$ imeta qu -d hall = east
collection: /FranceGrillesZone/projet
dataObj: run0977156_123.raw
```

```
0111011101101000011010010 01100
01100101001000000011000100111010
001000000110011011100101110011
00101110011 01101110101100100
01101111110101110 000101110
011011101110000100101110100
01100101001010 11001101110100
011100100010000000001100001
01101110 00100011 1101101101
011001100100110000110000101110
011001000110100101101100110100
001010000011000000111000 0000
00110001001010010010100100101001
```

# Workflow Automation

## Automatisation du flux de données

- Intégration d'un langage de script qui est appelé à chaque opération :
  - Authentification
  - Accès au stockage
  - Interaction avec la base de données
  - Activité réseau
  - API RPC extensible
- Le moteur de règle iRODS fournit la capacité d'implémenter des politiques réelles de données (== définies par des humains) à travers des traitements activables qui autorisent, refusent ou ajoutent du contexte aux opérations à un système informatique

WORKFLOW  
AUTOMATION

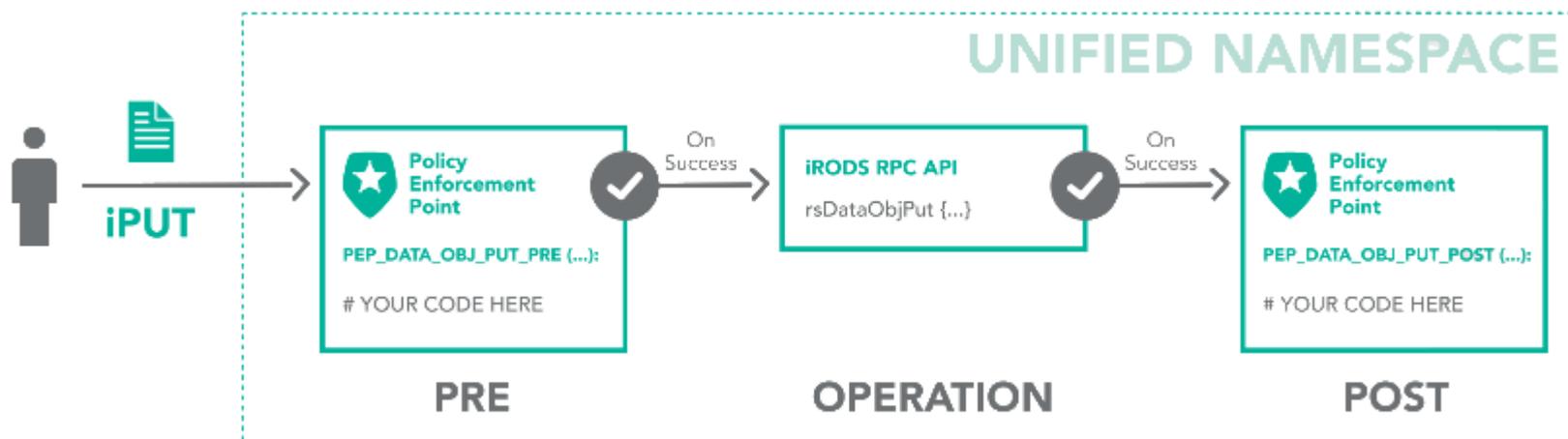




# Policy Enforcement

## Fonctionnement

- Un simple appel API intègre de nombreux plugins en opération
- Chacun d'entre eux invoque l'application de politiques :
  - Restriction d'accès (utilisateurs, chemin, type de données, ...)
  - Enregistrement des informations pour les audits et les rapports
  - Ajout de contextes additionnels (métadonnées)
  - Micro-services (envoi de notification, ...)
  - etc



```
0111011101101000011010010 01100
01100101001000000011000100111010
001000000110011011100101110011
00101110011 0101101001100100
011011111101010110 000101110
011011101110000100100101110100
01100101001010 11001101110100
01110010001000010001001100001
01101110 00100011 1101101101
01 110011100100110000101101110
011001000110100101101110110100
0010100000110000001011000 0000
00110001001010010010100100101001
```

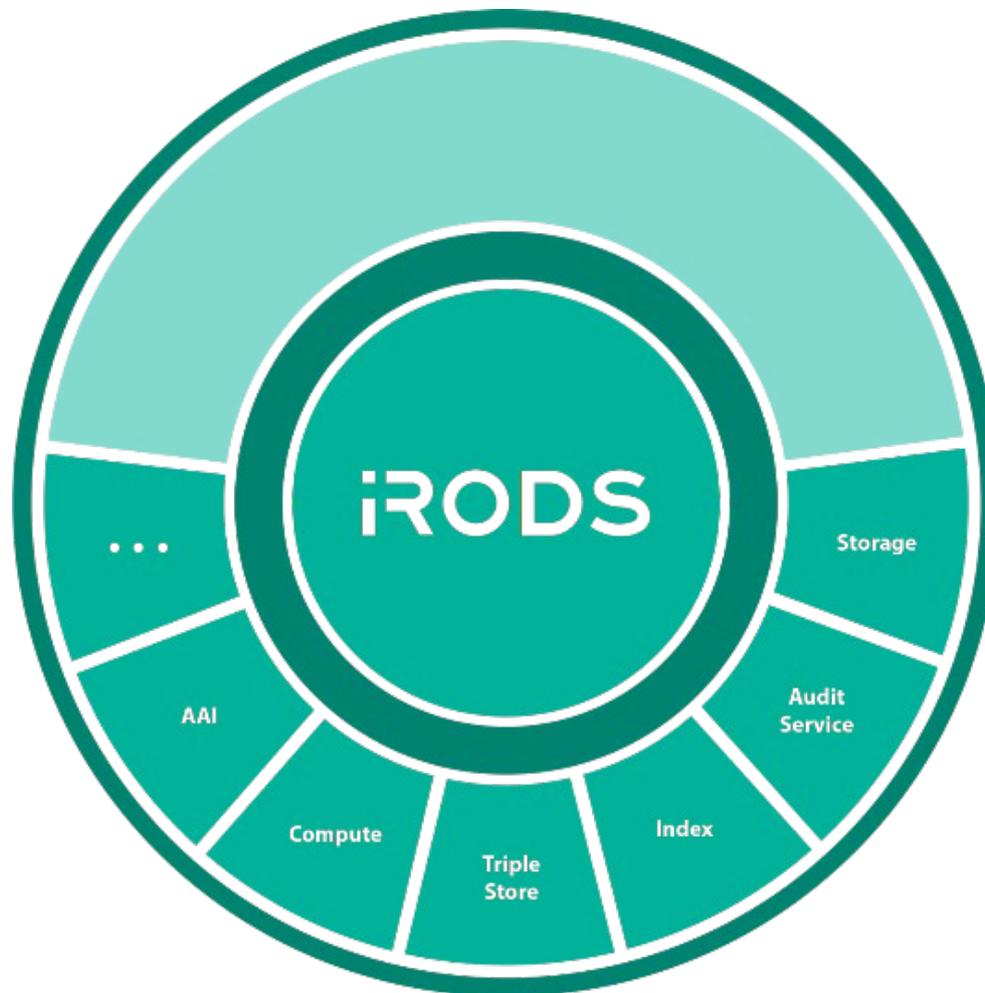
# Secure Collaboration

## Faciliter les collaborations

- Possibilité de mettre en place des collaborations
- 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



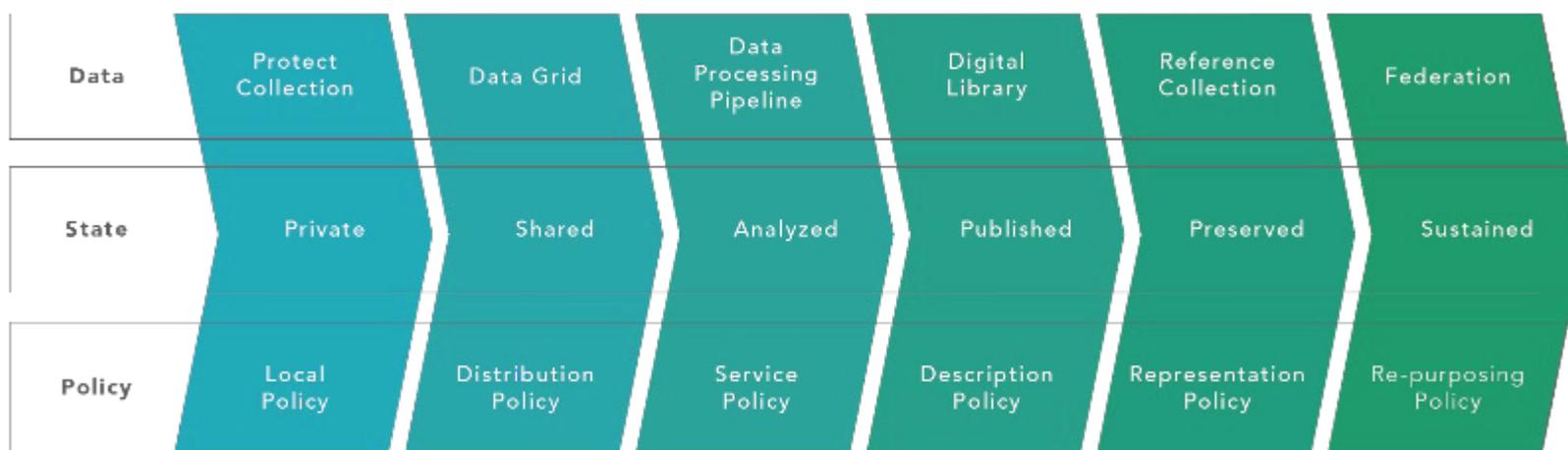
# Une interface unique pour un ensemble de services



# 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

# Composition des politiques et fonctionnalités

## Les politiques à la base des fonctionnalités

- Les politiques qui ont été utilisées dans le cadre d'une fonctionnalité sont nommées selon une convention :
  - irods\_policy\_access\_time
  - irods\_policy\_data\_movement
  - irods\_policy\_data\_replication
  - irods\_policy\_data\_verification
- Chaque politique peut être réutilisée et combinée pour créer de nouvelles fonctionnalités
- Chaque politique peut être outrepassée par un autre moteur de règle, ou modifiée, afin de s'adapter aux nouvelles utilisations et technologies
- Réutilisation et pérennité des fonctionnalités

```
0111011101101000011010010 01100
01100101001000000011000100111010
001000000110011011100101110011
00101110011 011011101011100100
0110111101110101110 000101110
0111011101100001000101110100
01100101001010 11001101110100
0111010001000000000001100001
01101110 00100011 1101101101
00 11001100100110000101101110
0110010001101001011011100110100
001010000011000000111000 0000
001100010010100100101001001001
```

# Fonctionnalités



Automated Ingest



Storage Tiering



Auditing



Provenance



Indexing



Publishing



Data Integrity



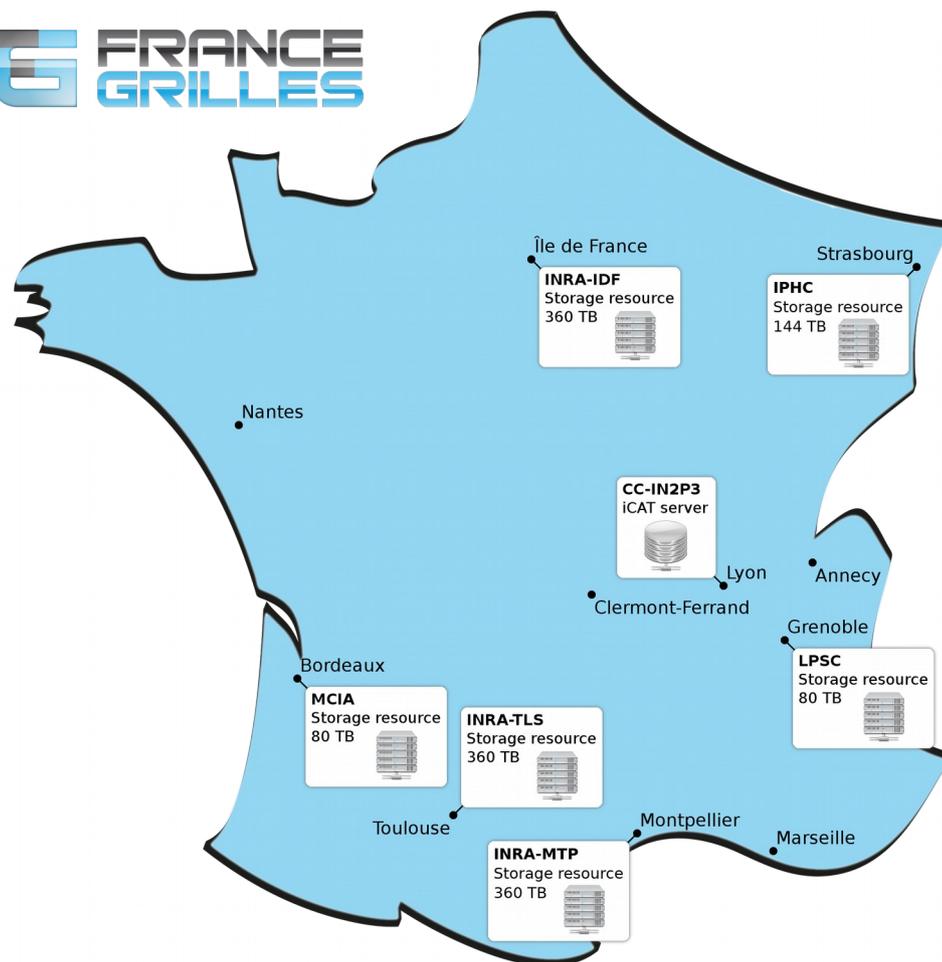
Compliance

```
011011101101000011010010 01100
01100101001000000011000100111010
001000000110010111100101110011
00101110011 01011010101100100
011011101110101110000101110
01101101110000100100101110100
01100101001010111001101110100
0111010001000000000001100001
01101110 001000111 1101101101
0110011000110100101101110110
0110010001101001011011100110100
001010000011000000111000 0000
00110001001010010010100101001
```

# Exemple de fédération

## L'instance France Grilles

- Une instance en production depuis 2014
- Facilité l'accès à un service de niveau production aux petits et moyens projets
- Accompagnement des utilisateurs
- Terrain d'essai
- Ouverte à tous les domaines scientifiques
- Fédération de ressources humaines et matérielles
- Investissement financier minimal et granulaire
- <http://www.france-grilles.fr/catalogue-de-services/fg-irods/>



# À noter

## Formation Administrateur

- Deuxième quinzaine de mars
- En visio
- Choix de la date en cours :  
<https://evento.renater.fr/survey/results/ynqkxaro>
- Contact : [jerome.pansanel@iphc.cnrs.fr](mailto:jerome.pansanel@iphc.cnrs.fr)
- Programme :
  - fonctionnement d'iRODS (catalogue, ressource, moteur de règles, . . . ) ;
  - installation et configuration d'iRODS sur un système CentOS 7 ;
  - développement de règles et gestion des métadonnées ;
  - les types d'authentification ;
  - installation d'outils complémentaires (Metalnx, supervision, . . . ) ;
  - aspects de sécurité.

0111011101101000011010010 01100  
0110010100100000001100010111010  
001000000110011011100101110011  
00101110011 011011101011100100  
011011110110101110 000101110  
01110111011000100100101110100  
01100101001011 11001101110100  
011101000100001000100110001  
01101110 00100011 1101101101  
00 110011100100110000101101110  
0110010001101001011011101110100  
0010100000110000001011000 0000  
00110001001010010010100100101001

# Questions ?