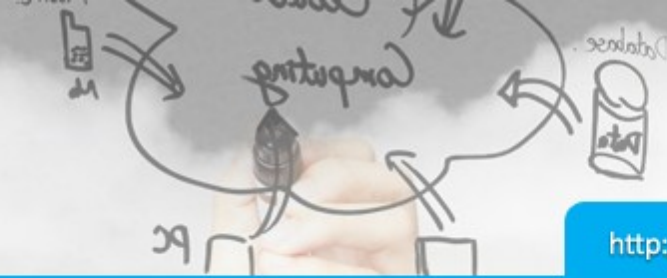




# iRODS et la gestion de données

Jérôme Pansanel et Emmanuel Medernach

15 mars 2021



## Crédits

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

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



## iRODS

# iRODS

— CONSORTIUM —

# renci

RESEARCH \ ENGAGEMENT \ INNOVATION



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

# Le consortium iRODS

The image displays a collection of logos for various organizations and institutions associated with the iRODS consortium. The logos are arranged in a grid-like fashion. Some logos include the text "Consortium Member".

- Bayer**
- DDN STORAGE**
- Maastricht University**
- SUSE** (We adapt. You succeed.)
- Western Digital**
- Consortium Member** (with people icon)
- Universiteit Utrecht**
- UCL**
- welcome sanger institute**
- renci**
- university of groningen**
- Research Computing UNIVERSITY OF COLORADO BOULDER**
- Quantum**
- Consortium Member** (with people icon)
- CLOUDIAN**
- BIH Berlin Institute of Health Charité & MDC**
- OpenIO**
- AGRICULTURE VICTORIA**
- TACC TEXAS ADVANCED COMPUTING CENTER**
- KU LEUVEN**
- SURF**
- SNIC**
- MSC medical science & computing**
- NetApp**
- NIH National Institute of Environmental Health Sciences**

## iRODS et la gestion des données

### iRODS

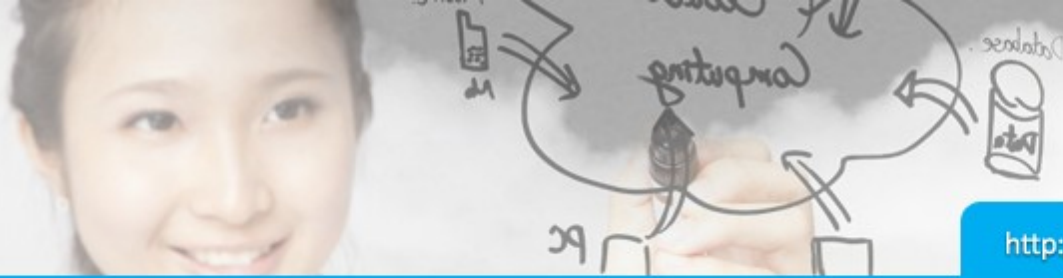
- Une solution pérenne pour la gestion des données et de l'infrastructure qui les entoure
- « Le développement, l'exécution et la supervision de plan de gestion, politiques, programmes et pratiques qui contrôlent, protègent, mettent à disposition et valorisent les données et les informations associées. »

A woman with dark hair is smiling and looking towards the camera. She is holding a black marker and drawing a diagram on a whiteboard. The diagram includes arrows, a central figure, and some handwritten text. A blue banner at the bottom right of the image contains the URL 'http://www.france-grilles.fr'.

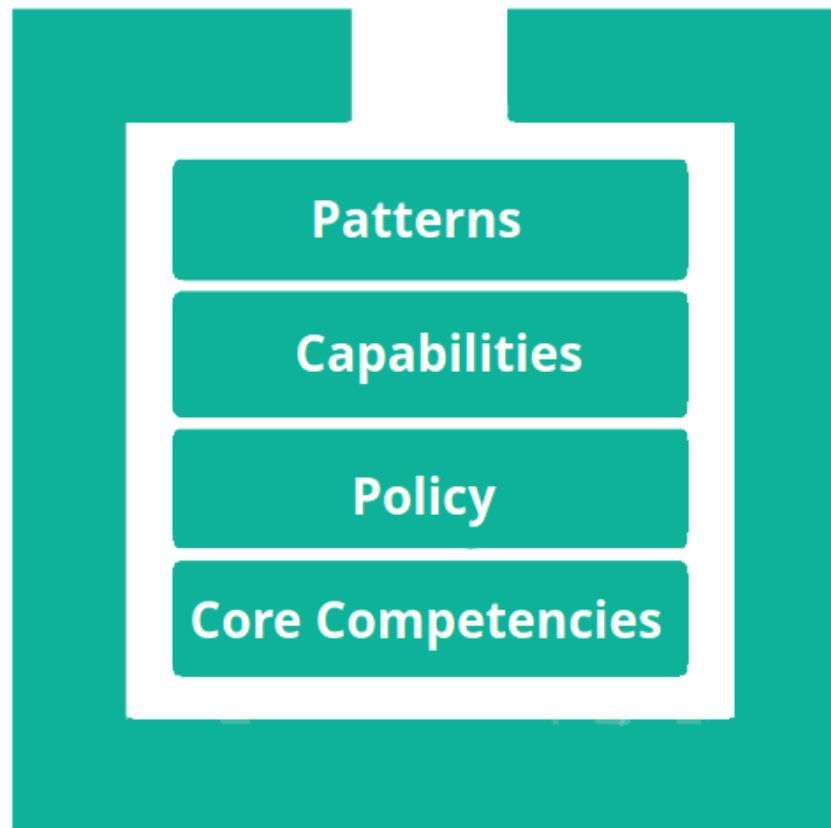
## Politique des données

### Politique des données ?

- « Un ensemble d'idées ou de plans de gestion décrivant quelles sont les actions à réaliser dans une situation particulière et qui ont été officiellement validés par un groupe de personnes. »



## Implémentation dans iRODS



## Core Competencies

**DATA  
VIRTUALIZATION**



**DATA  
DISCOVERY**



**WORKFLOW  
AUTOMATION**



**SECURE  
COLLABORATION**





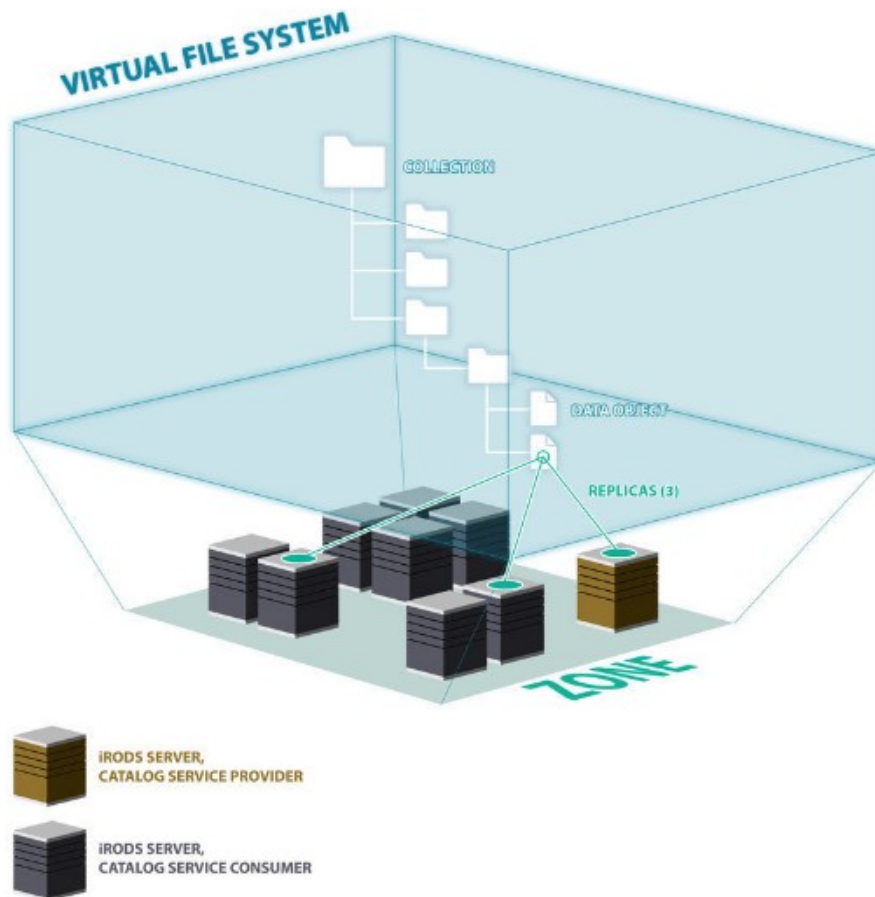
# Data Virtualization

## Virtualisation

- Accès simultané à différentes technologies à travers un seul espace de nom (zone) :
  - Systèmes de fichiers existants
  - 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

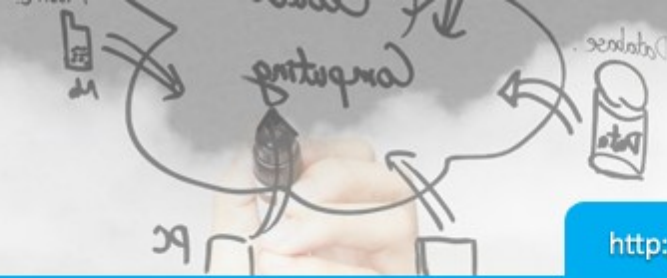


# Projection de l'infrastructure physique vers la virtuelle



Chemin logique

Chemin(s) physique(s)



## Data Discovery

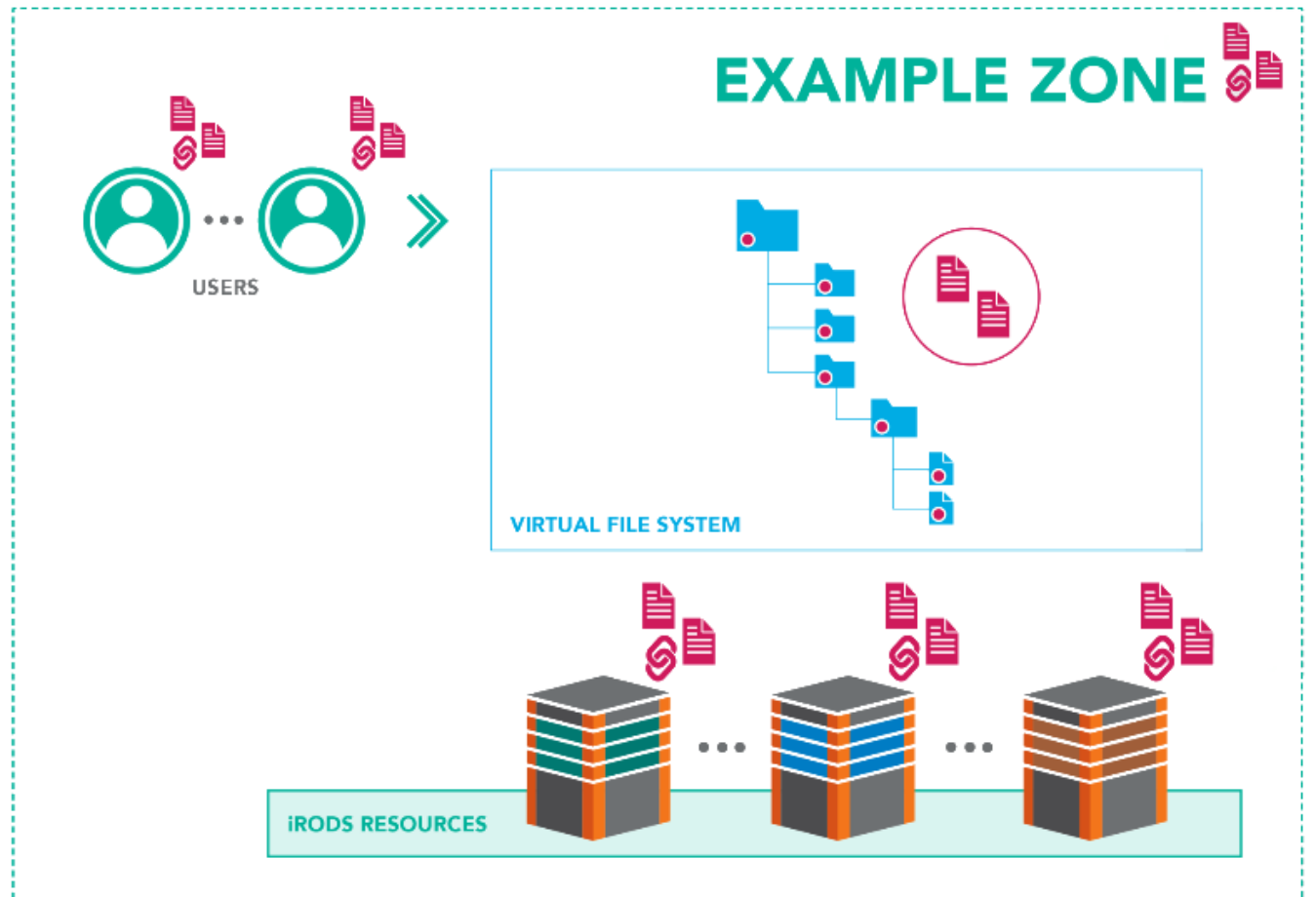
### Métadonnées

- 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



## Des métadonnées partout



## 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

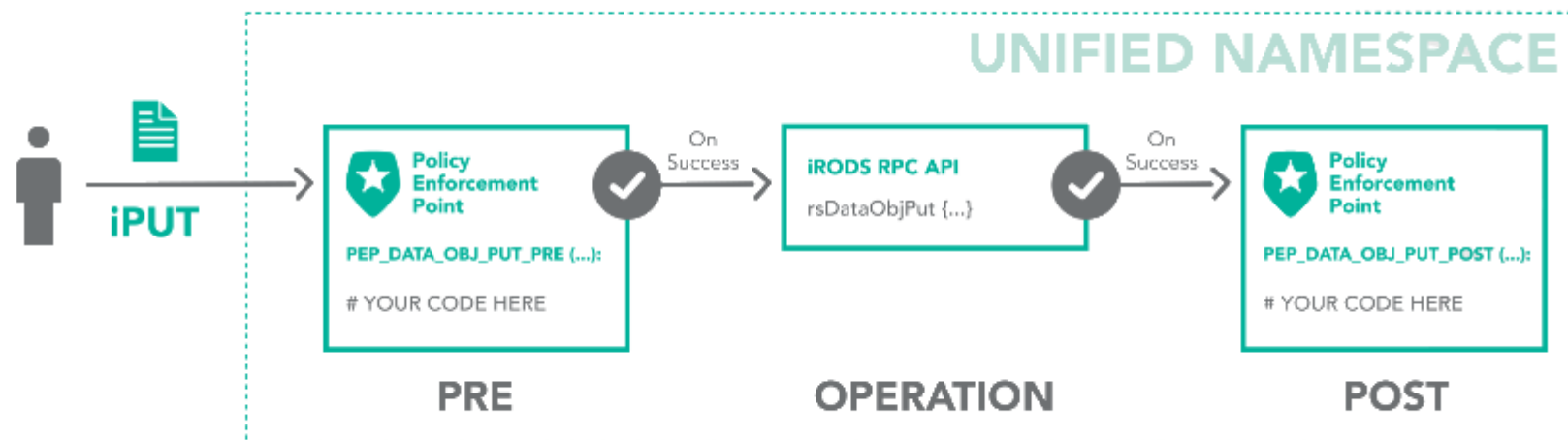
**WORKFLOW  
AUTOMATION**



# Dynamic Policy Enforcement

## Capacités d'une règle

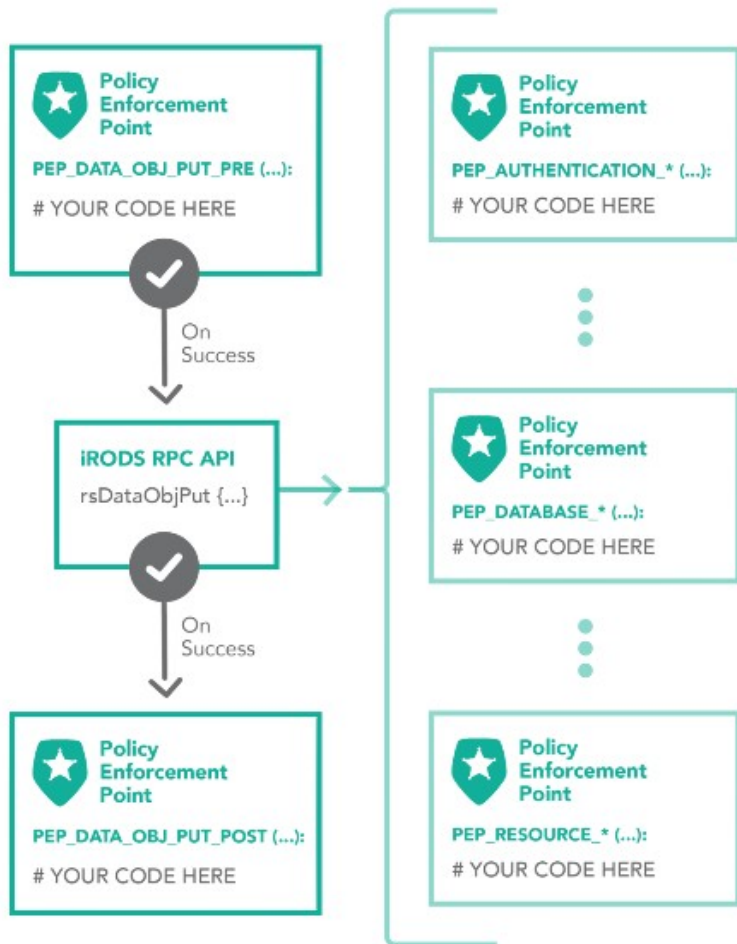
- Restriction d'accès
- Enregistrement des informations pour les audits et les rapports
- Ajout de contexte additionnel
- Envoi de notifications



# Dynamic Policy Enforcement

## Fonctionnement

- Un simple appel API intègre de nombreux plugins en opération
- Chacun d'entre eux invoque l'application de politiques
- Plugins :
  - Authentification
  - Base de données
  - Stockage
  - Réseau
  - Moteur de règle
  - Micro-service
  - API RPC



## Secure Collaboration

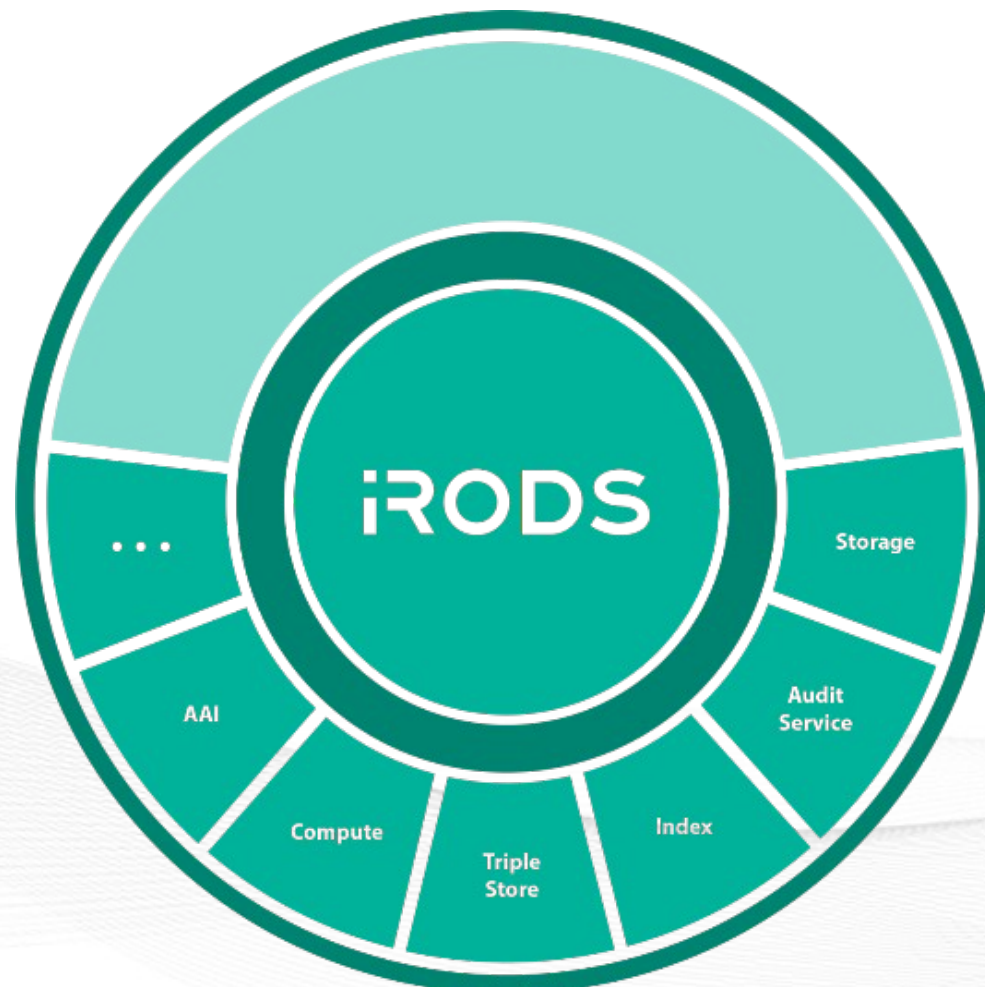
### Sécuriser les collaborations

- 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

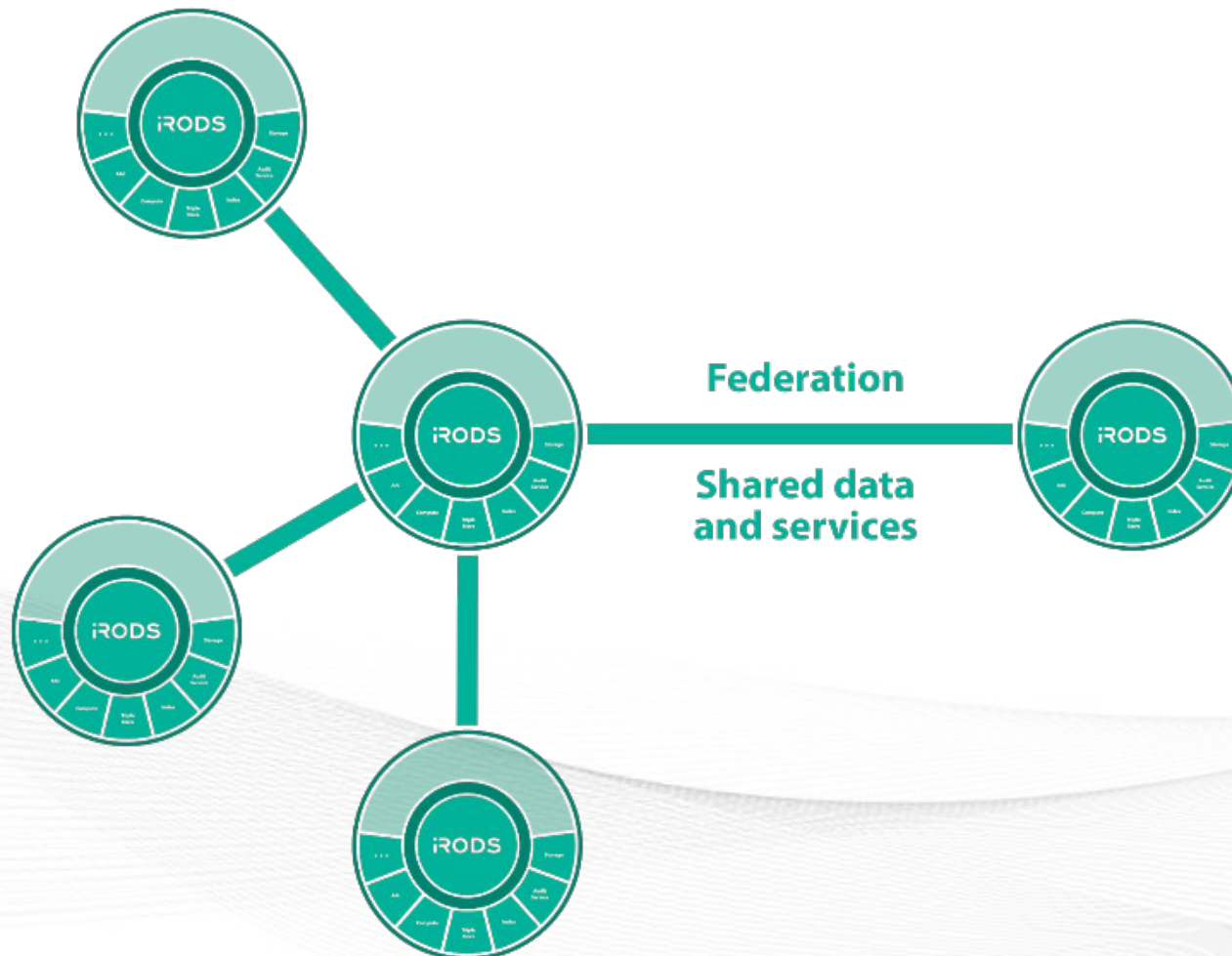


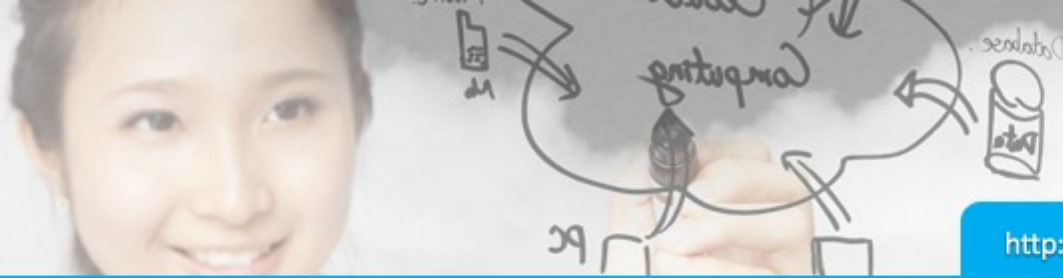


## Une interface pour les services



## Fédération : partage de données et de services

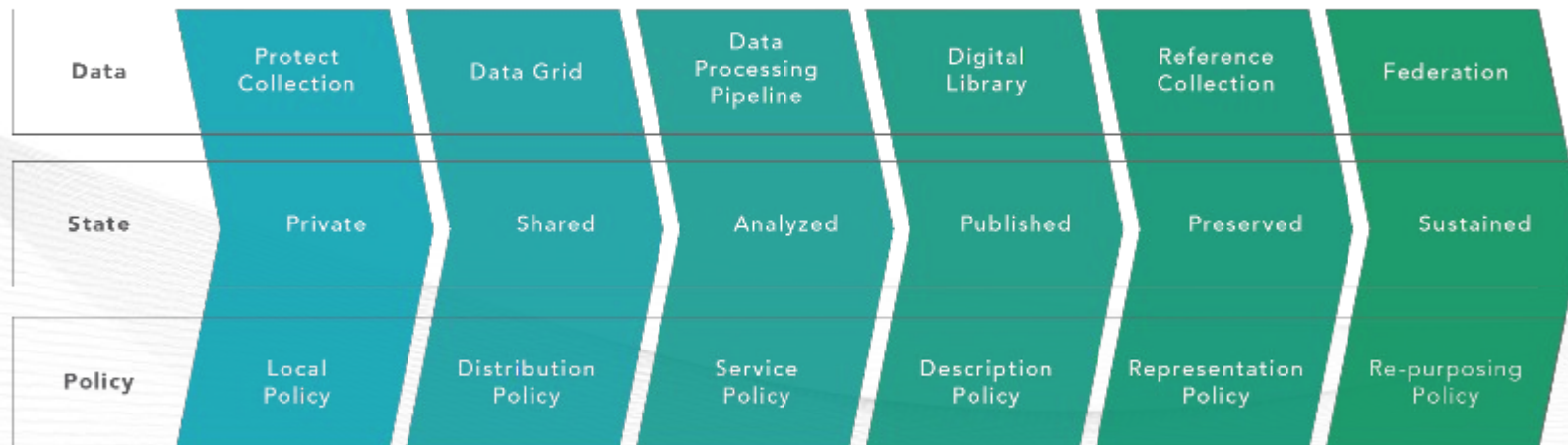




## De l'ingestion au dépôt institutionnel

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

## Policy

### Les politiques disponibles

- Déplacement de données
- Vérification de données
- Rétention des données
- Réplication des données
- Choix du placement des données
- Calcul de *checksum*
- Extraction de métadonnées
- Application de métadonnées
- Conformité des métadonnées

## Composition avec les règles de base

### Les règles de base

- Par exemple : `pep_data_obj_put_post(...)`
- Extraction et application de métadonnées
- Réplication asynchrone
- Démarrage de l'indexation
- Application de métadonnées avec l'horodatage des accès
- Calcul asynchrone de *checksum*
- Séparer les implémentations en éléments individuels de base et permettre le passage de la règle à travers eux
- Simplification de la maintenance

## Policy Composition and Capabilities

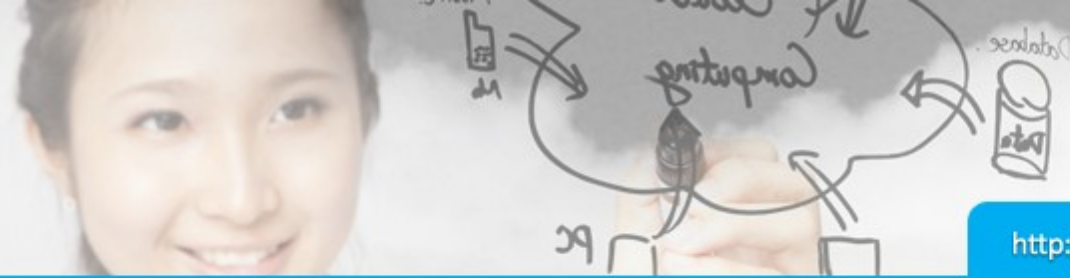
### Exemple du stockage hiérarchique

- Date d'accès à la donnée
- Identification des objets violant une contrainte par rapport à cette date
- Réplication de la donnée sur un autre stockage (par ex. bande)
- Vérification de la donnée
- Suppression de la première réplique
- Cette fonctionnalité est implémentée comme une composition qui délègue chaque étape à l'application d'une politique particulière

## Policy Composition and Capabilities

### Réutilisation des politiques

- 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



## Fonctionnalités



Automated Ingest



Storage Tiering



Auditing



Provenance



Indexing



Publishing



Data Integrity

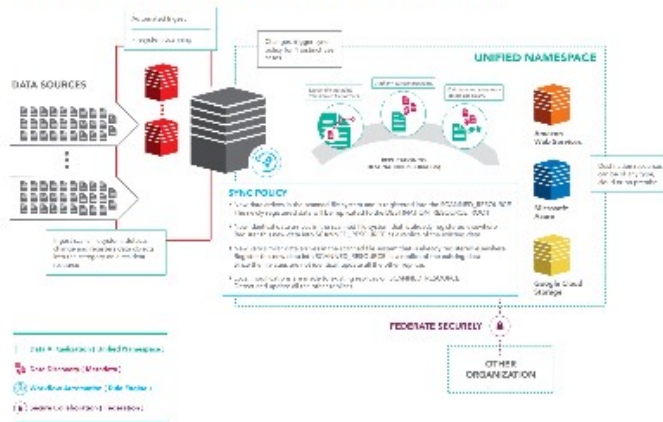


Compliance



# Deployment Patterns

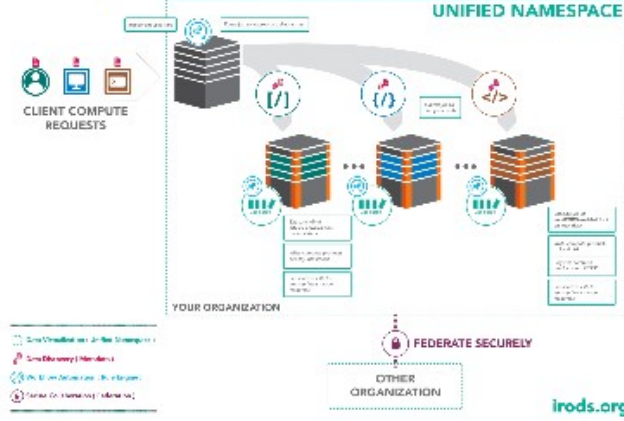
## Filesystem Synchronization



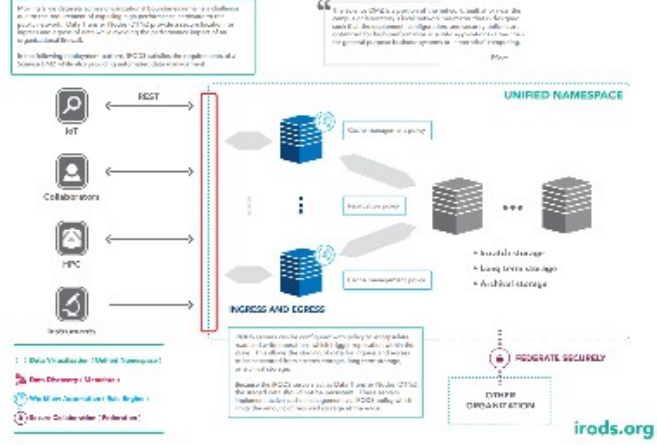
## Data to Compute

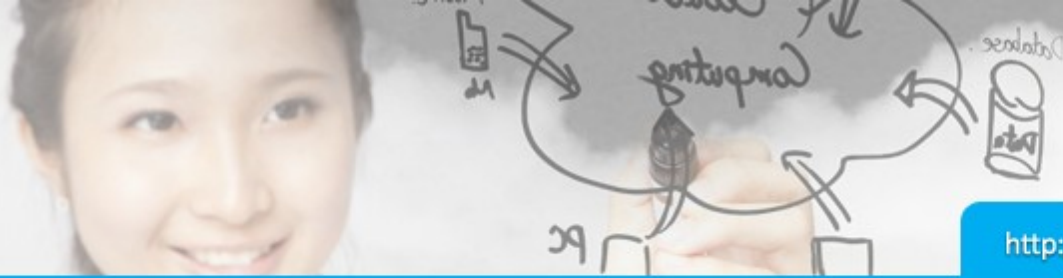


## Compute to Data



## Data Transfer Nodes





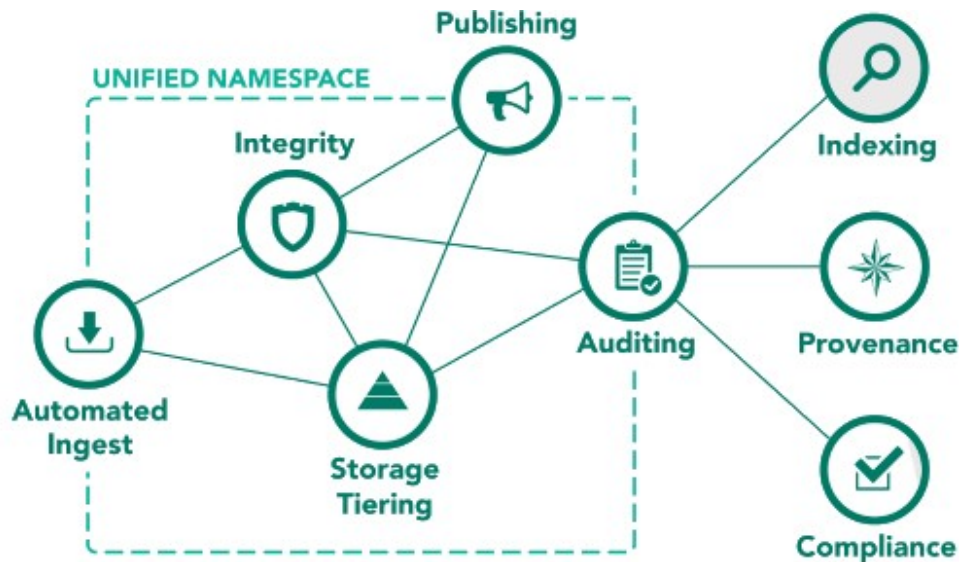
# The data management model

iRODS provides eight packaged capabilities, each of which can be selectively deployed and configured.

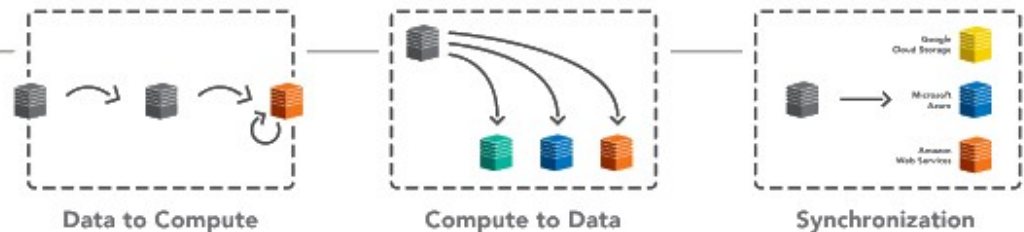
These capabilities represent the most common use cases as identified by community participation and reporting.

The flexibility provided by this model allows an organization to address its immediate use cases.

Additional capabilities may be deployed as any new requirements arise.



A pattern represents a combination of iRODS capabilities and data management policy consistent across multiple organizations. Three common patterns of iRODS deployment have been observed within the community:





# Questions ?