



# Storage virtualization @ CC-IN2P3: iRODS

Pascal Calvat

# Virtualiser le stockage : pourquoi?



- Les scientifiques sont dispersés géographiquement et ont besoin d'accéder des données distantes.
  - Les données peuvent être aussi localisées sur des différents sites
  - Utilisation de technologies de stockage et systèmes d'exploitation hétérogènes
  - Pour les groupes d'utilisateurs:
    - Besoin d'authentification et de droits d'accès uniformisés sur les données.
  - Virtualisation du stockage:
    - Pour être indépendant des technologies et évolution du hardware.
    - Pour être indépendant de l'organisation locale des fichiers
    - Pour répliquer plus facilement les données
- ➔ La vue logique des données est indépendante de la localisation physique des données.

- **iR**ule **O**riented **D**ata **S**ystems est un système de virtualisation de données permettant de gérer de manière simple et cohérente des données distribuées. IRODS supporte :
  - Plusieurs centaines de millions de fichiers
  - Des dizaines de petaBytes de données
  - Plusieurs centaines de ressources physiques
  - Des centaines de micro-service pour composer des workflow côté serveur
- Projet démarré en 2006 par l'équipe DICE (USA)
- Equipe de développement très réactive et à l'écoute des utilisateurs
- **Open source (Licence BSD)**
- Financé par : NSF, NARA (National Archives and Records Administration).
- Partenaires internationaux : CC-IN2P3 (France), e-science (UK), ARCS (Australia)

- Chaque communauté à ses interfaces préférées:
  - Unix shell commands – icommands (Existe aussi pour Windows)
  - Java I/O library (JARGON) et C I/O library
  - Archive numérique – Fedora Common / Dspace
  - Workflows - Kepler / Taverna
  - Web browsers / Windows browser
  - User level file systems - FUSE / WebDAV
  - Grid APIs – JSAGA
  - iDROP – Copie de fichiers par simple drag and drop
- Authentification:
  - User/password, GSI, kerberos, shibboleth
- Micro-services:
  - Côté serveur, ils permettent d'effectuer des actions sur les fichiers

# Production iRODS @ CC-IN2P3



- iRODS: 780 TB avec un taux de croissance de 5 TB/jour  
[http://cctools.in2p3.fr/mrtguser/compta\\_irods.php](http://cctools.in2p3.fr/mrtguser/compta_irods.php)
- Sciences humaines et sociales (Adonis) : 40 TB et diffusion des données archivées via Fedora Commons
- Physique des particules (BABAR) : SLAC/LYON 3TB/jour
- Astroparticule (AMS, Double chooz): 2 TB/jour
- Biologie (Phylogénie, embryogénese, neuroscience, cardiologie) : 30 TB