Policy Based Data Management

integrated Rule-Oriented Data System
iRODS

Reagan W. Moore
Arcot Rajasekar
Hao Xu
Mike Conway

University of North Carolina at Chapel Hill
Projects using the iRODS data grid
(integrated Rule Oriented Data System)

1. Astrophysics  Auger supernova search
2. Atmospheric science  NASA Langley Atmospheric Sciences Center
3. Biology  Phylogenetics at CC IN2P3
4. Climate  NOAA National Climatic Data Center
5. Cognitive Science  Temporal Dynamics of Learning Center
6. Computer Science  GENI experimental network
7. Cosmic Ray  AMS experiment on the International Space Station
8. Dark Matter Physics  Edelweiss II
9. Earth Science  NASA Center for Climate Simulations
10. Ecology  CEED Caveat Emptor Ecological Data
11. Engineering  CIBER-U
12. High Energy Physics  BaBar / Stanford Linear Accelerator
13. Hydrology  Institute for the Environment, UNC-CH; Hydroshare
14. Genomics  Broad Institute, Wellcome Trust Sanger Institute, NGS
15. Medicine  Sick Kids Hospital
16. Neuroscience  International Neuroinformatics Coordinating Facility
17. Neutrino Physics  T2K and dChooz neutrino experiments
18. Oceanography  Ocean Observatories Initiative
19. Optical Astronomy  National Optical Astronomy Observatory
20. Particle Physics  Indra multi-detector collaboration at IN2P3
21. Plant genetics  the iPlant Collaborative
22. Quantum Chromodynamics  IN2P3
23. Radio Astronomy  Cyber Square Kilometer Array, TREND, BAOradio
24. Seismology  Southern California Earthquake Center
25. Social Science  Odum, TerraPop
26. National library  Bibliotheque Nationale de France
27. Regional repository  Texas Digital Library
28. Institutional repository  Carolina Digital Repository
Generalizing Library Services

• Provide framework for making new services
  – Chain together basic operations to create service
  – Manage information needed to implement the service
  – Automate application of the service

• Provide framework for managing distributed data
  – Implement collaboration environment
  – Manage properties of a shared collection
  – Enforce policies
California Digital Library Services

- **Curation (collection building)**
  - Ingest: Upload data
  - Index: Arrange data into collections
  - Annotation: Add descriptive metadata
  - Access: Set access controls
  - Transformation: Convert format
  - Search: Query metadata

- **Preservation (administrative management)**
  - Characterization: Representation information
  - Inventory: Usage report
  - Replication: Replicate files
  - Fixity: Manage checksums
  - Storage: Manage storage locations
  - Identity: Persistent identifier
Policy-based Data Environments

- **Purpose** - reason a collection is assembled (which stage)
- **Properties** - attributes needed to ensure the purpose
- **Policies** - control for ensuring maintenance of properties
  - Computer actionable rules
- **Procedures** - functions that implement the policies
  - Computer executable workflows composed from micro-services
- **Persistent state information** - results of applying the procedures
  - System, descriptive, and provenance metadata
- **Property assessment** - validation that state information maintains the desired properties
  - Periodic policies, queries on audit trails
- **Policy-based Federation** - controlled sharing of logical names for collection re-use
  - Policies to control interaction between data grids

These are the necessary elements to build any data management application
iRODS Distributed Data Management

- HDF Viewer For iRODS
- Visualization Of HDF5 File
- iRODS Rich Web Client
- WebDAV On iPod
- Windows Browser
- iCommands Command Line

---

- iSEC
- iXMS
- iCAT
- iRES
- iRES
- iRES

---

- Schedule & Compute Queue
- Message Queue
- Metadata Database
- Storage & Compute Resources
  - File Systems, Archives, Databases, Sensor Systems, Clusters, ...