

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

