

Policy-Driven Data Preservation

Reagan W. Moore

University of North Carolina at Chapel Hill

rwmoore@renci.org

<http://irods.diceresearch.org>

NSF OCI-0848296 “NARA Transcontinental Persistent Archives Prototype” (2008-2012)
NSF SDCI 0721400 “Data Grids for Community Driven Applications” (2007-2010)



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Preservation is an Integral Part of the Data Life Cycle

- Organize project data into a shared collection
- Publish data in a digital library for use by other researchers
- Preserve reference collection for use by future research initiatives
- Compare new reference collection against prior state-of-the-art data
- Enable data-driven analyses that dynamically optimize research



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Policy-based Data Management

iRODS - integrated Rule Oriented Data System

- Turn management policies into computer actionable rules
 - Support dynamic rule base updates
- Turn management processes into remotely executable computer procedures
 - Apply procedural workflow at the storage system to filter, subset, manipulate data
 - Minimize the amount of data pulled over the network
 - Automate administrative tasks
- Validate assessment criteria
 - Automate validation of collection properties
 - ISO MOIMS-rac



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Generic Data Management Systems

iRODS - integrated Rule-Oriented Data System

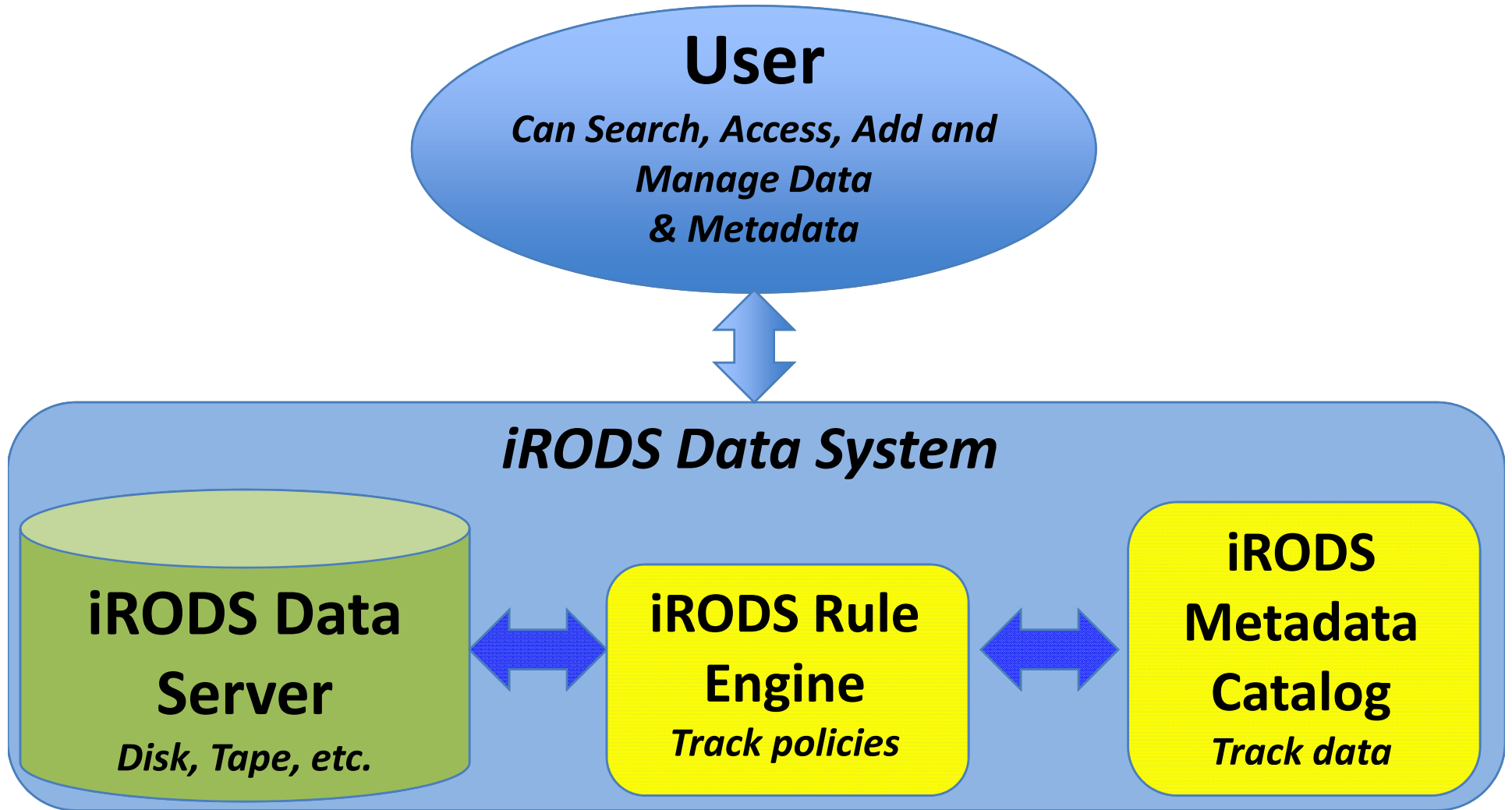
Data Management Environment	Conserved Properties	Control Mechanisms	Remote Operations
Management Functions	Assessment Criteria	Management Policies	Management Procedures
Data Management virtualization			
Data Management Infrastructure	State Information	Rules	Micro-services
Data and trust virtualization			
Physical Infrastructure	Database	Rule Engine	Storage System



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Overview of iRODS Data System



*Access data with Web-based Browser or iRODS GUI or Command Line clients.

Policies

- Administrative
 - Retention, disposition, distribution, replication, deletion, registration, synchronization, integrity checks, IRB approval flags, addition of users, addition of resources
- Ingestion / Access
 - Metadata extraction, logical organization, derived data product generation, redaction, time-dependent access controls
- Validation
 - Authenticity verification, chain of custody, repository trustworthiness, audit trails



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

