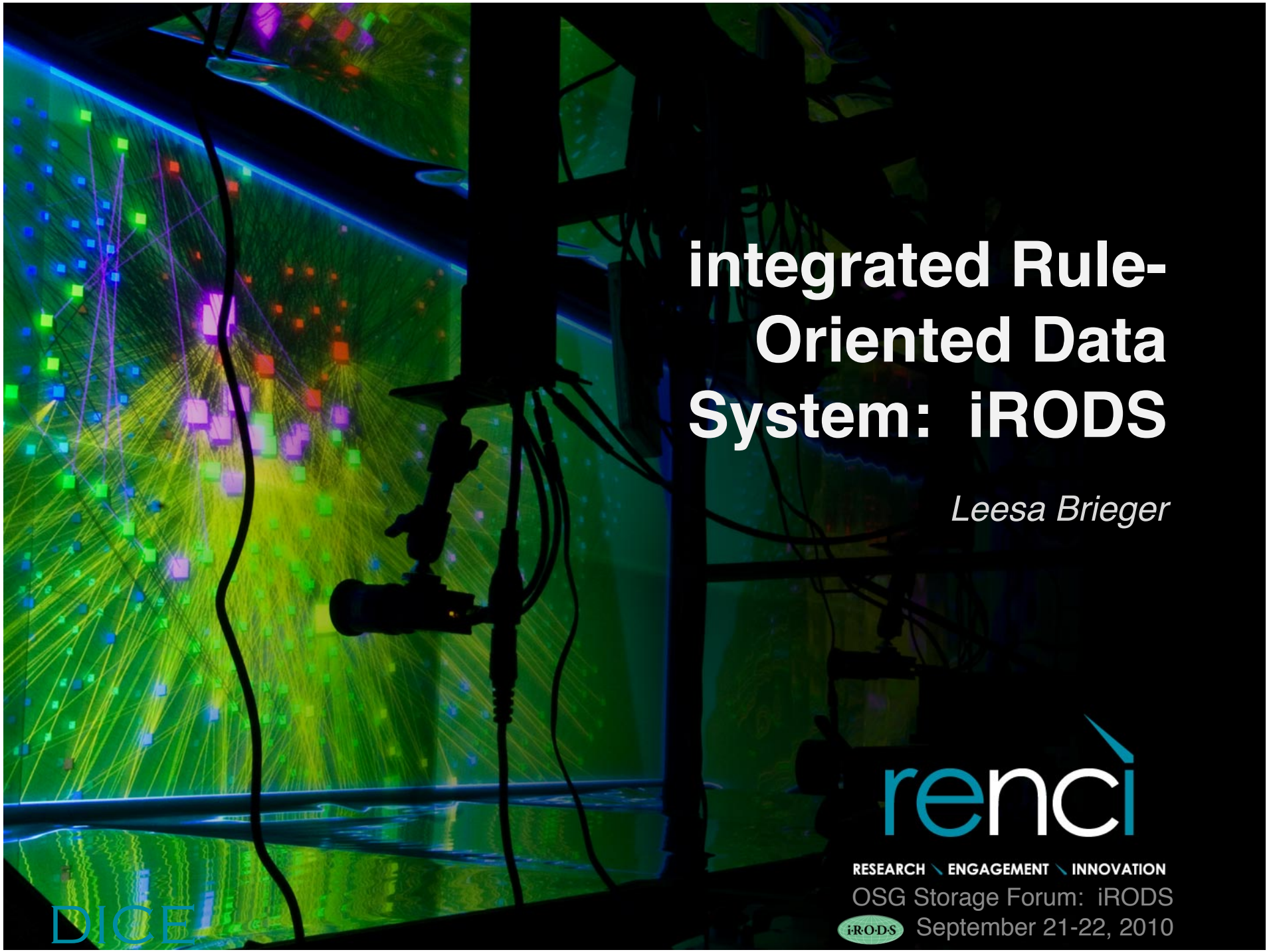


integrated Rule-Oriented Data System: iRODS

Leesa Brieger

renci

RESEARCH \ ENGAGEMENT \ INNOVATION

OSG Storage Forum: iRODS

September 21-22, 2010

DICE

iRODS

- Developed by the *Data Intensive Cyber Environments* (DICE) group at UCSD and UNC
- Based on decade-long experience of the *Storage Resource Broker* (SRB) development
- Community-driven
- Open source (BSD license)
- Supported by RENCI at UNC: iRODS@RENCI

renci
DICE

The Issues

Data…

- collection (physical or virtual)
- sharing/publishing
- security/integrity
- auditing/accounting
- metadata management
- curation (of remote and local data)
- preservation (remotely and locally)

In a nutshell: data *management,* i.e. the application of data *policy* across the data life cycle

renci DICE

The Issues - Examples

- Genome centers:
  - petabytes of data
  - researchers sharing data
  - derived data products from workflows
  - requirements for traceability and reproducibility (provenance and metadata management)
- NOAA's National Climate Data Center (NCDC)
  - repository management
  - publishing public data
  - delivery of services with the data (to the public and to researchers)
  - support for climate modeling (at ORNL, …)

The Issues – More Examples

- Streamline (automate) data movement (OSG)
  - distributed jobs
  - collect data results to a central location for sharing/post-processing
  - archive data output
- Institutional repositories
  - collect a variety of data from a multitude of disparate sources
  - manage collections independently:
    - some public collections
    - some collections shared between select user or institutional groups (across administrative boundaries)
    - varying life spans
    - some privacy-protected data (legal issues)
    - different integrity requirements
    - break-the-glass scenarios (emergency management) – access permissions change as a function of state information

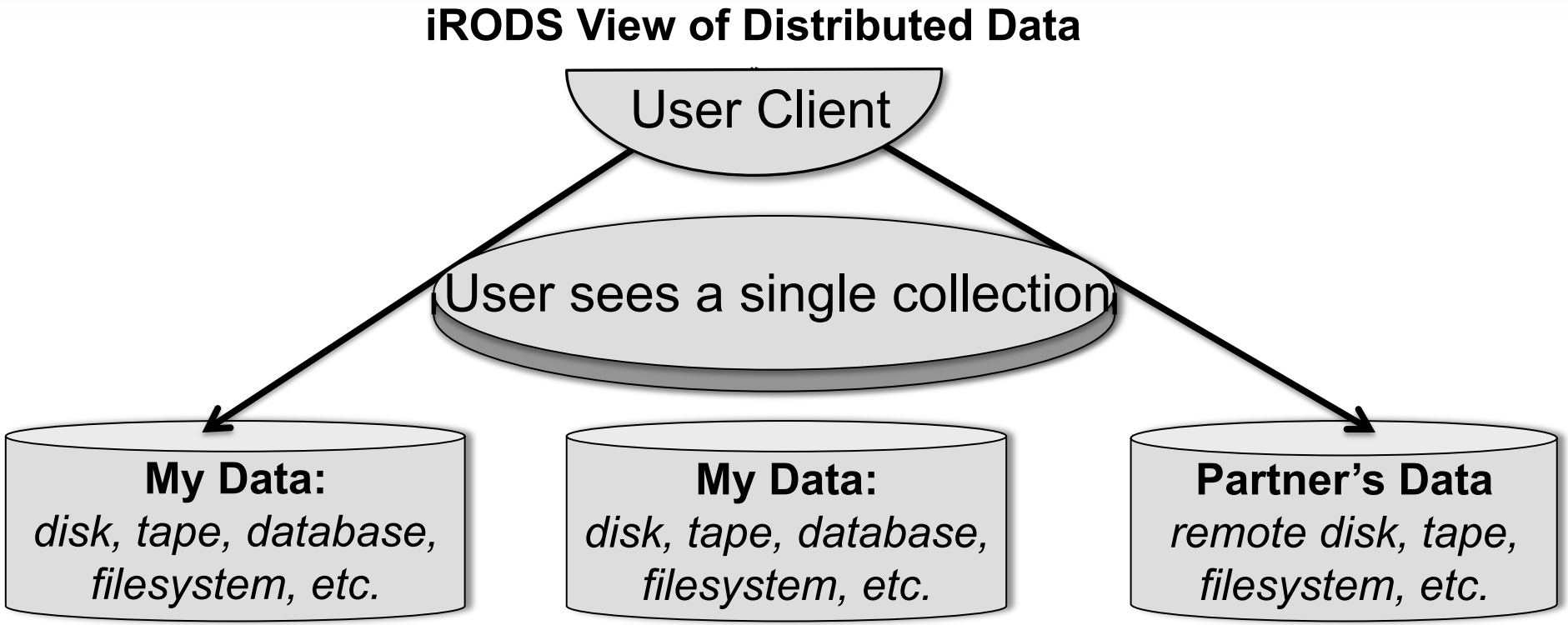
renci DICE

iRODS as a Data Grid

- Sharing data across:
  - geographic and institutional boundaries
  - heterogeneous resources (hardware/software)
- Virtual collections of distributed data
- Global name spaces
  - data/files
  - users
  - storage
- Metadata catalogue (iCAT) manages mappings between logical and physical name spaces
- Beyond a single-site repository model

renci DICE

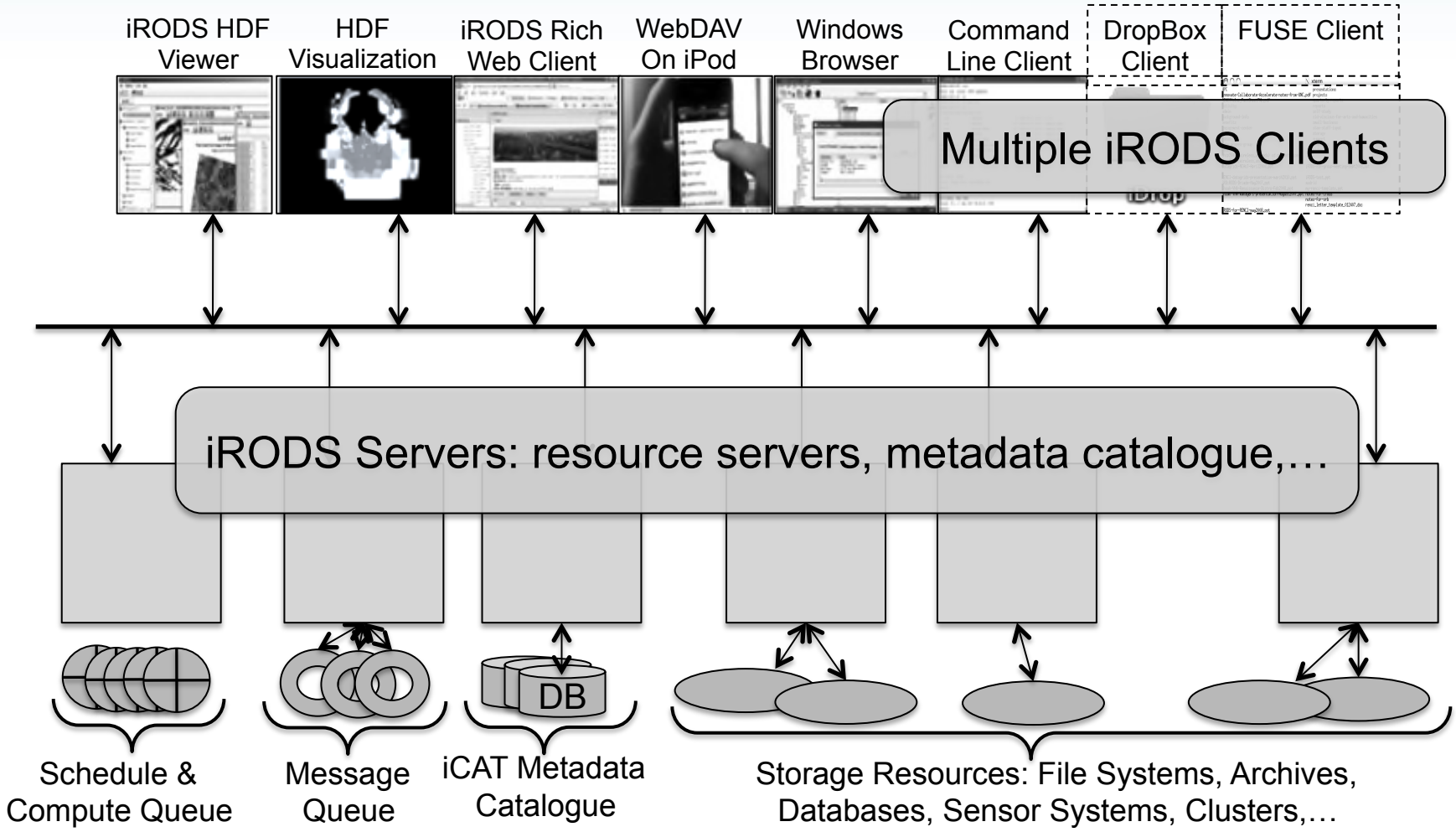
iRODS Virtual Collection



iRODS installs over heterogeneous data resources; users view and manage distributed data as a single collection.

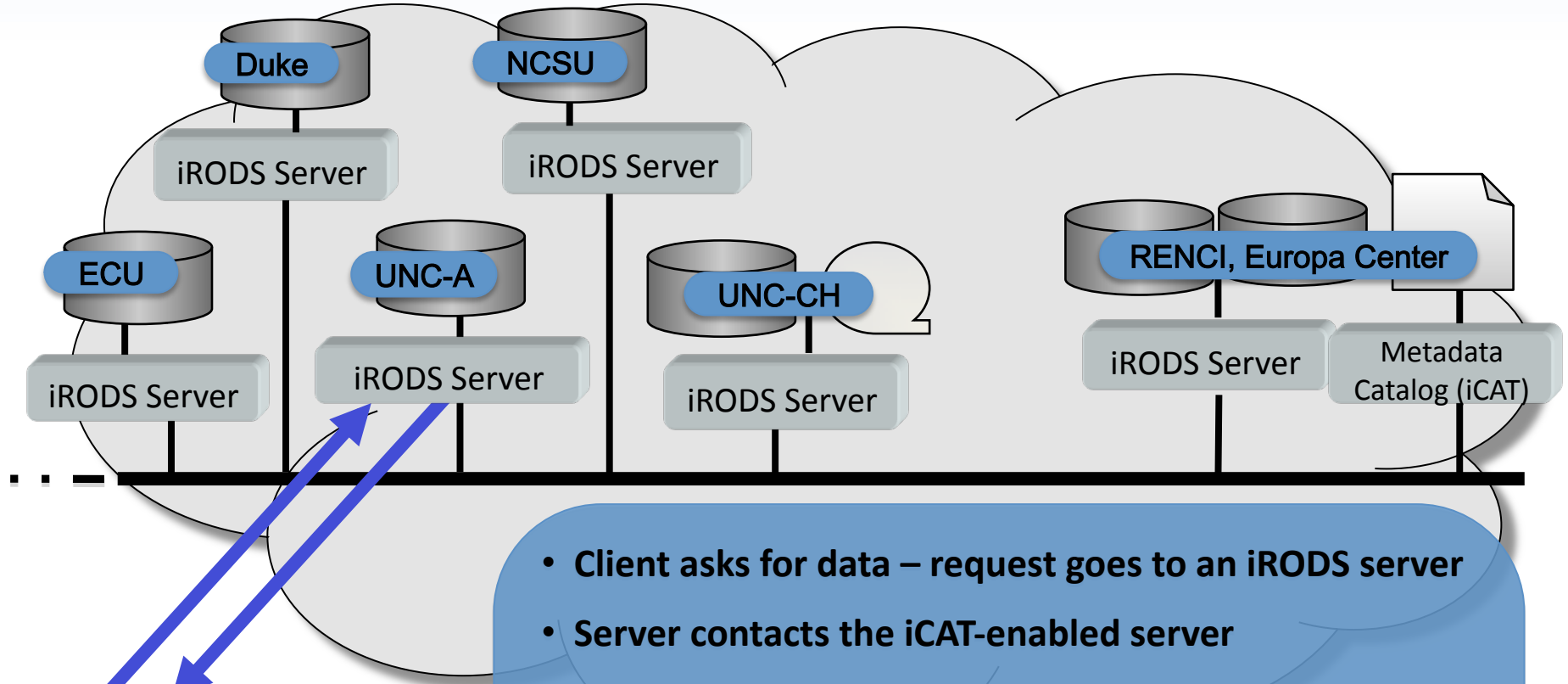
renci DICE

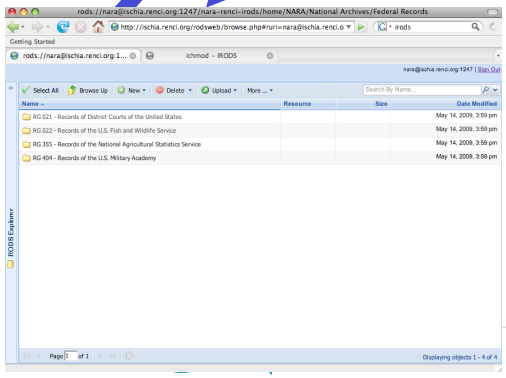
An iRODS Overview



renci DICE

RENCI VO Data Grid





- **Client asks for data – request goes to an iRODS server**
- **Server contacts the iCAT-enabled server**
- **Information (location, access rights, etc) is retrieved from the iCAT**
- **Server containing data is signaled to send data to authorized client**

iRODS Command Line Client: icommands

- ils
- icp
- irepl
- irsync
- iput
- iget
- imeta – add, modify, read metadata
- iquest – query the iCAT database

Data Grid Security

- Authenticate each user access
  - PKI, Kerberos, challenge-response, Shibboleth
  - Use internal or external identity management system
- Access controls as constraints imposed between name spaces
  - Controls on: Files / Storage resources / Metadata
  - Access controls remain invariant as files move within the data grid
- Authorization of operations
  - ACLs (Access Control Lists) on users and groups
  - Conditional rule execution
  - ACLs on services not yet implemented, but coming

renci DICE

iRODS Rules and Microservices

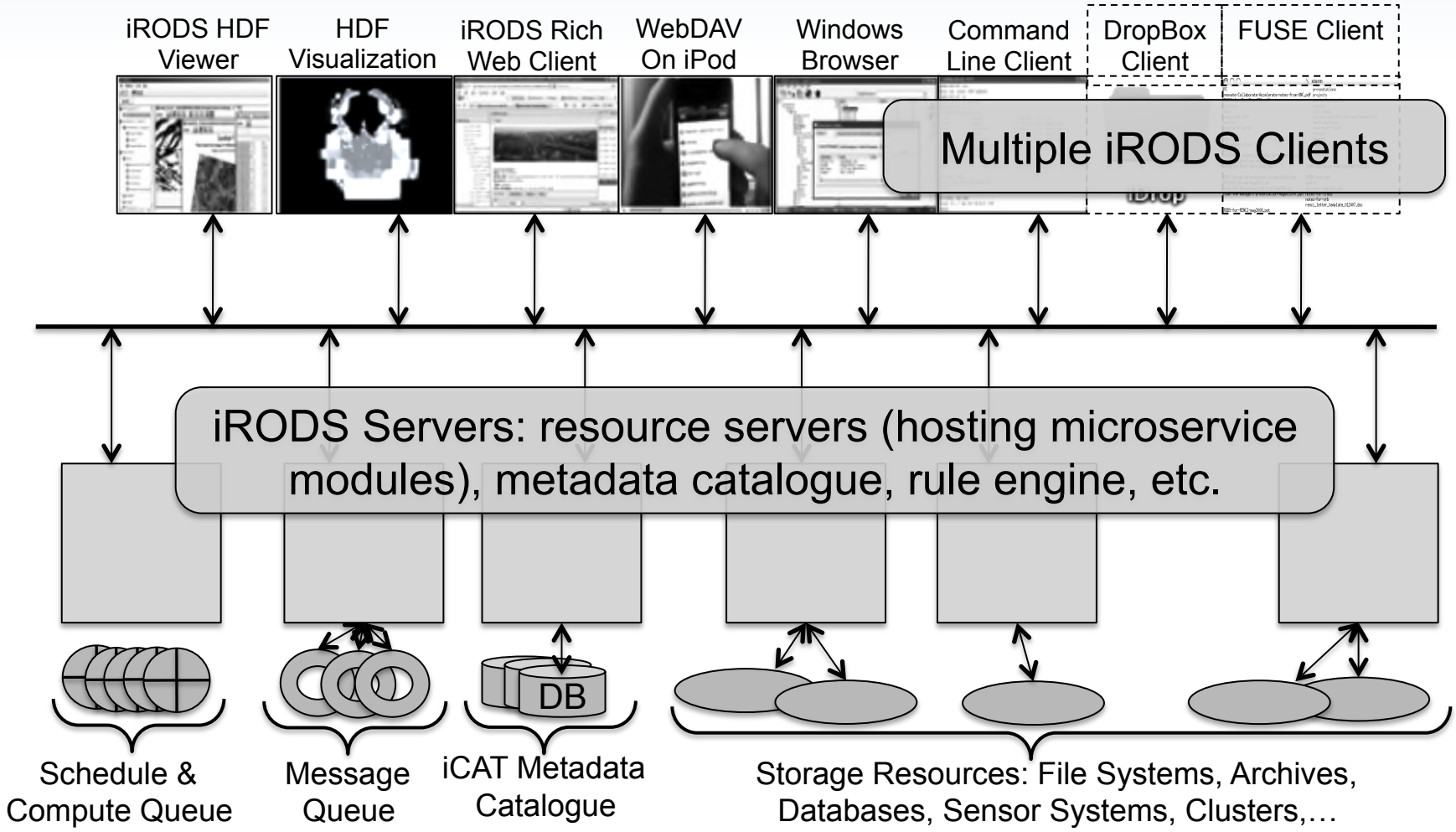
- Functional unit: *microservices* (C programs)
- 185 microservices provided out-of-the-box
- Workflows of microservices: *rules*
- iRODS Rule Base set by iRODS administrator
  (a file that contains the data grid rules: core.irb)
- Remote execution - run where the data reside
- Delayed execution (eg daily synchronization)

renci DICE

iRODS Rules and Microservices

- User communities customize a data grid by writing their own microservices and rules
- Modules of new microservices shared with the larger user community
- Modularity increases sense of community architecture
- User-defined rules can be composed of provided microservices and run on command line (*irule*)

renci DICE


iRODS HDF Viewer
HDF Visualization
iRODS Rich Web Client
WebDAV On iPod
Windows Browser
Command Line Client
DropBox Client
FUSE Client
Multiple iRODS Clients
iRODS Servers: resource servers (hosting microservice modules), metadata catalogue, rule engine, etc.
DB
Schedule & Compute Queue
Message Queue
iCAT Metadata Catalogue
Storage Resources: File Systems, Archives, Databases, Sensor Systems, Clusters,…

Microservice Examples

- msiGetCollectionACL
  get AC list for a collection
- msiGetDataObjAVUs
  retrieve metadata AVU triplets for a data object and return as an XML file
- msiGetAuditTrailInfoByKeywords
- msiCopyAVUMetadata
  copy triplets from one data object to another
- msiRecursiveCollCopy

Core microservices + modules of custom microservices

renci DICE

Rules

- Syntax:
  rule_name | condition | workflow-chain | recovery-chain
- Contained in core.irb file
- getATInfoByObjPath.ir

```
Get Audit Info By Object Path||writeLine(stdout,'<?xml version="1.0"
   encoding="ISO-8859-1"?>')
##writeLine(stdout,"<audit_trail>")
##msiIsData(*objPath,*objID,*foo)
##msiGetAuditTrailInfoByObjectID(*objID,*BUF,*Status)
##writeBytesBuf(stdout,*BUF)
##writeLine(stdout,"</audit_trail>")|nop
*objPath=/dcape-dev/home/leesa/foo.txt
ruleExecOut
```

- RuleGen parser for easier syntax

Policy-driven Data Management

Actions: event-triggered rules

- acCreateUser – services to run when creating a new user
- acPreProcForDataObjOpen
- acPreProcForCollCreate
- acPostProcForPut
- acPostProcForCopy
- acPostProcForCreate
- acSetPublicUserPolicy - set the list of operations that are allowable for the user "public"

Data administrator determines policy by defining actions and by providing rules and services to users.

renci DICE

iRODS Data Management

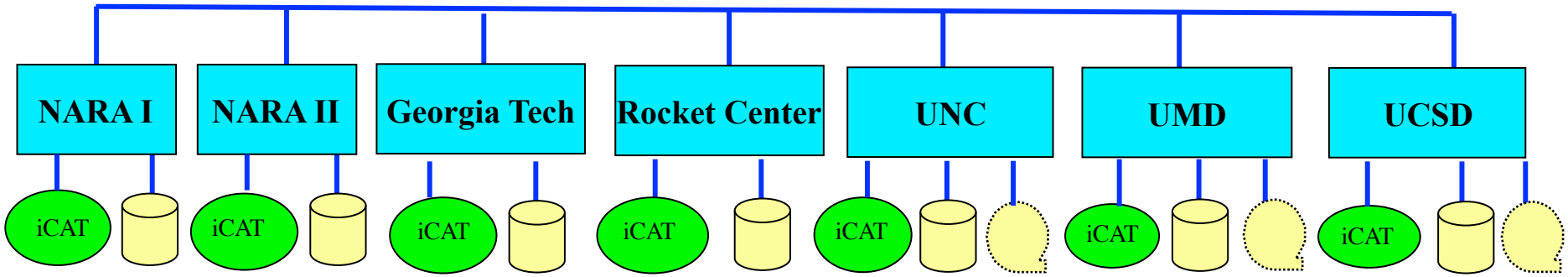
- Sharing & publishing data
- Virtual collections
- Security controls for data access
- Metadata
- Provide services with the data
- Remote services - that run where the data reside
- Policy that follows the data anywhere in the data grid - can even use this to implement your policy on cloud storage

renci DICE

Federated Data Grids

National Archives and Records Administration
Transcontinental Persistent Archive Prototype (TPAP)

Federation of Seven Independent Data Grids



- Extensible environment: can federate with additional research & education sites
- Each data grid uses different vendor products.

renci DICE

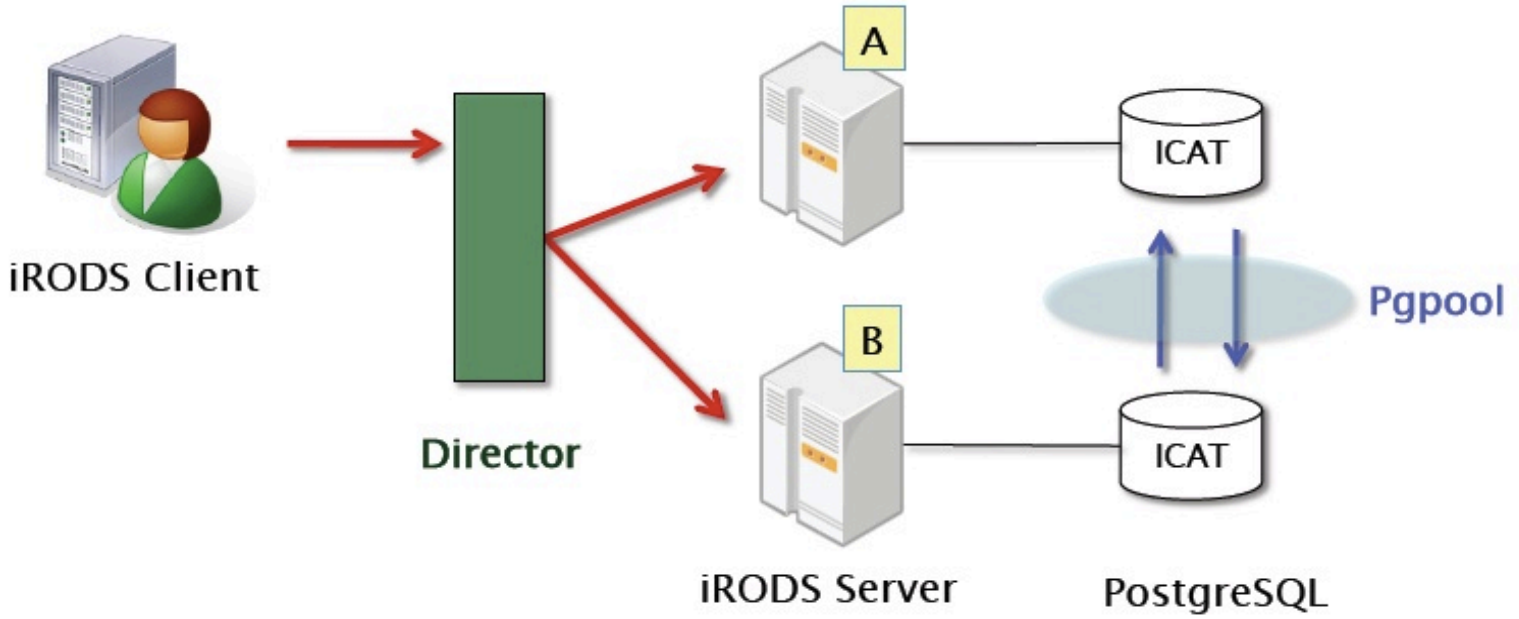
Current iRODS Applications

High Availability iRODS System

KEK: High Energy Accelerator Research Organization, Japan

Redundancy and load balancing with Pgpool and Director

- Place a Director between Client and Server
  - Monitor the iRODS server statuses



iRODS at Centre de Calcul de l'Institut de Physique Nucléaire et de Physique des Particules (CC-IN2P3), France

iRODS setup

- 9 servers:
  - 3 iCAT servers (metacatalog): Linux SL4, Linux SL5
  - 6 data servers (200 TB): Sun Thor x4540, Solaris 10
- Metacatalog on a dedicated Oracle 11g cluster
- HPSS interface
- Use of fuse-iRODS:
  - For Fedora-Commons
  - For legacy web applications
- TSM: backup of some stored data
- Monitoring and restart of the services fully automated
- Automatic weekly re-indexing of the iCAT databases
- Accounting: daily report on web site

CC-IN2P3

- Communities: High Energy Physics, astrophysics, biology, biomedical, Arts and Humanities
- TIDRA: Rhône-Alpes area data grid
  - biology, biomedical applications
    - animal imagery, human data
    - automatic bulk metadata registration in iRODS based on DICOM files content
  - Coming soon: synchrotron data (ESRF – Grenoble)
  - 3 million files registered
  - up to 60,000 connections per day on iRODS
  - authentication: using password or grid certificate
- Beginning:
  - Neuroscience: ~60 TB
  - IMXGAM: ~ 15 TB (X and gamma ray imagery)
  - dChooz (neutrino experiment): ~ 15 TB / year
- Coming soon LSST (astro): For the electronic test-bed: ~ 10 TB

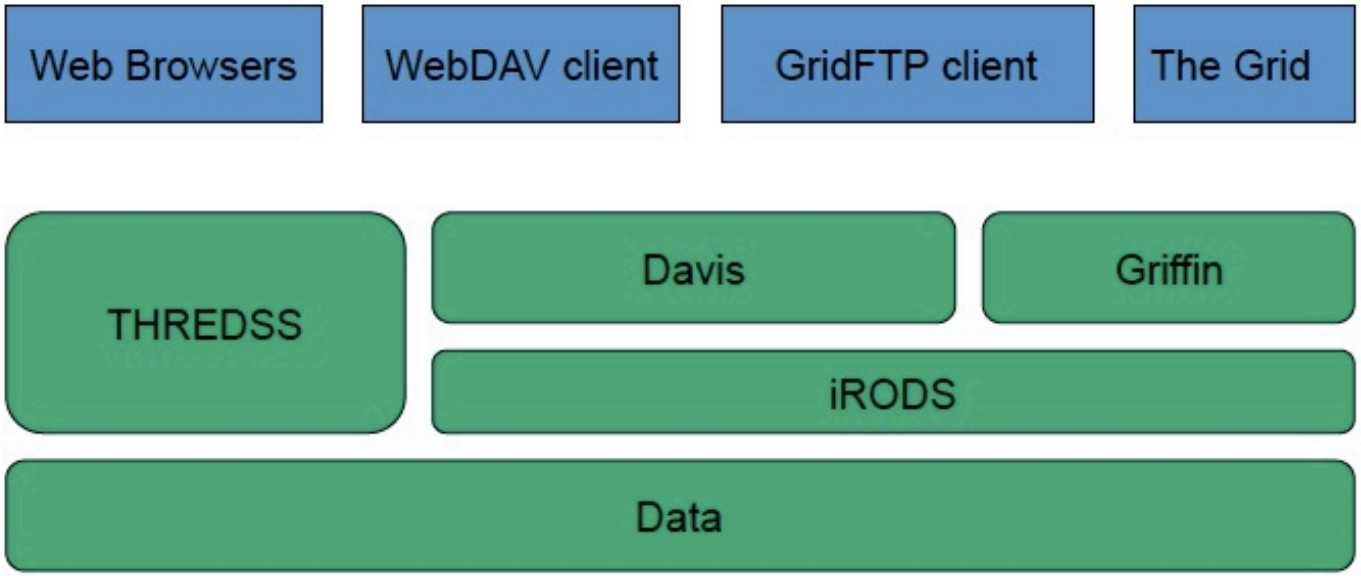
PetaShare and the Louisiana Optical Network Initiative (LONI)

- Petashare: a distributed data archival, analysis and visualization cyberinfrastructure for data-intensive collaborative research, connected through LONI; NSF-funded
- iRODS for distributed data management and storage infrastructure
- Novel approach: Treat data storage resources and the tasks related to data access as first class entities just like computational resources and compute tasks
- Key technologies being developed: Data-aware storage systems, data-aware schedulers (i.e. Stork), and cross-domain meta-data scheme

renci DICE

Australian Research Collaboration Services (ARCS)

- Services: data, compute, collaboration (web, video), authorization
- Architecture



ARCS

Web Interface (Davis)

- Basic file operations
- Metadata editing
- Permissions
- Dynamic objects
  - Configurable interface to run rules or for displaying information

WebDAV (Davis)

- Turns data fabric into a local folder
- Works on Windows, Mac & Linux
- Easy to use: drag and drop folders/multiple files

Griffin: the GridFTP to iRODS interface; developed by ARCS

renci DICE

Further and Future Possible Applications

- NASA JPL (evaluation), Goddard
- OOI (Ocean Observatories Initiative)
- NOAA (National Climate Data Center)
- Genomics: Broad Institute, Sanger Institute (UK)
- NARA
- Bibliothèque Nationale de France
- OSG
- …

renci DICE

Some Current iRODS Development

- Access to external databases
- Netcdf integration
- iDrop client (drag and drop)
- Integration with storage vendors (plug-in data grids)
- Metadata capture systems (bioinformatics)
- Workflow integration

RENCI now begins significant support to the DICE group in these and similar efforts.

renci DICE