Rule-Based Distributed Data Management

Reagan W. Moore
Richard Marciano

{moore,marciano}@diceresearch.org
http://irods.diceresearch.org
http://srb.diceresearch.org

UCSD

UNC

Preservation Concepts

- **Infrastructure independence**
  - Manage technology evolution
  - Manage authenticity, integrity, chain of custody
  - Scalability
- **Management virtualization**
  - Characterize preservation policies
  - Characterize preservation procedures
  - Federation, interoperate with other archives

Data Management Applications
(What do they have in common?)

- **Data grids**
  - **Share data** - organize distributed data as a collection
- **Digital libraries**
  - **Publish data** - support browsing and discovery
- **Persistent archives**
  - **Preserve data** - manage technology evolution
- **Real-time sensor systems**
  - **Federate sensor data** - integrate across sensor streams
- **Workflow systems**
  - **Analyze data** - integrate client- & server-side workflows

Assertion

- **Data management applications apply many of the same procedures**
  - Data format parsing
  - Metadata manipulation
  - Data administration tasks
- **Each application applies different management policies**
  - Migrate records between data life cycle stages by changing the management policies
  - Can create generic infrastructure that can be used to implement collections, digital libraries, persistent archives

DICE Group Objectives

- Build technology to organize distributed data into a shared collection
- Manage properties of the shared collection independently of the local administrative domains
- Automate application of management policies
- Automate validation of assessment criteria

Digital Preservation

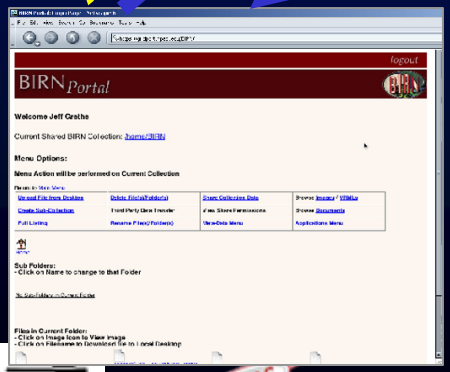
- **Preservation is communication with the future**
  - How do we migrate records onto new technology (information syntax, encoding format, storage infrastructure, access protocols)?
  - SRB - Storage Resource Broker data grid provides the interoperability mechanisms needed to manage multiple versions of technology
- **Preservation manages communication from the past**
  - What information do we need from the past to make assertions about preservation assessment criteria (authenticity, integrity, chain of custody)?
  - iRODS - integrated Rule-Oriented Data System

Virtualization of Data Collections

Data Grid

Ask for data

Data delivered



- User asks for data from the data grid
- The data is found and returned
  - Where & how details are hidden

UCSD iRODS SRB UNC

Data Virtualization

Data Access Methods (C library, Unix, Web Browser)

↓

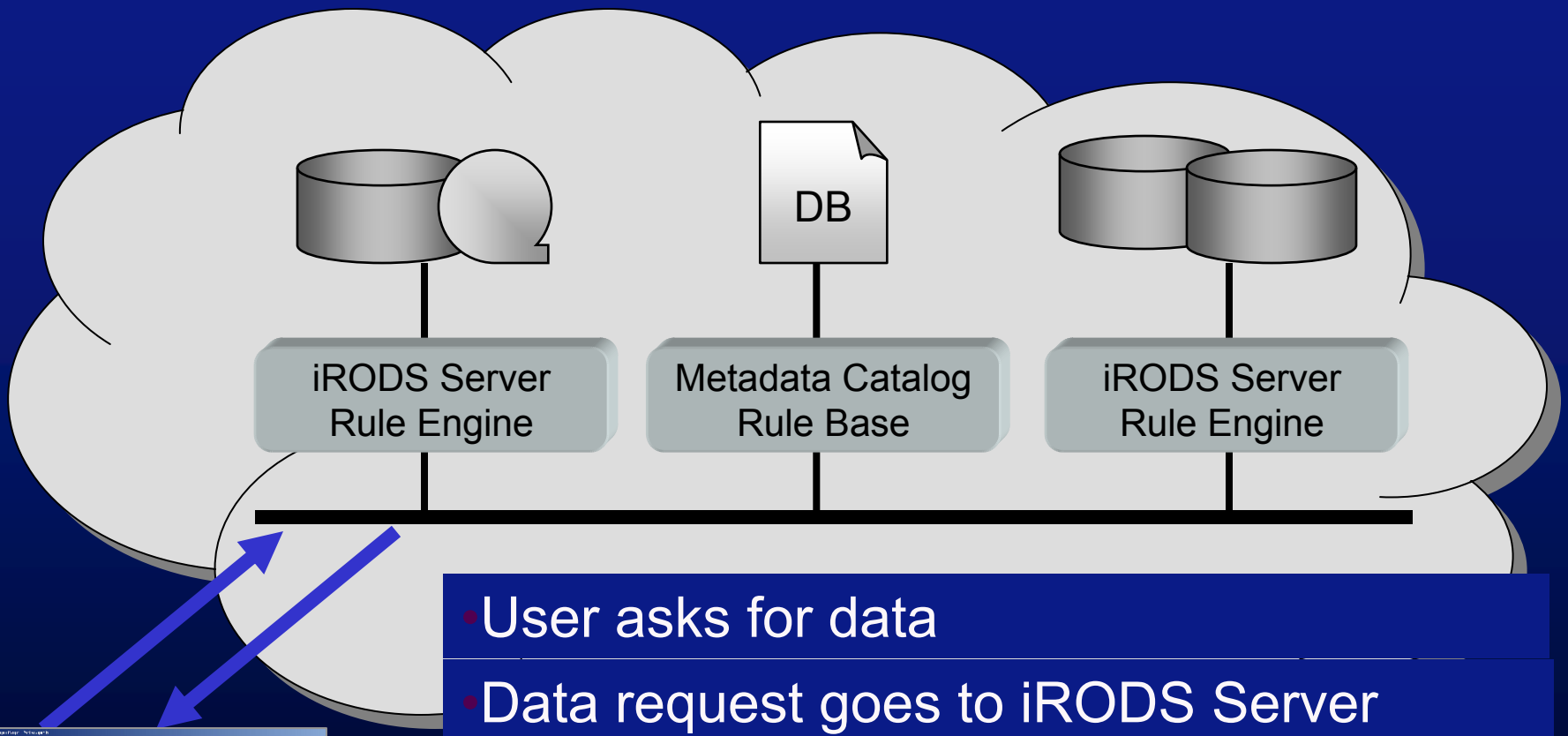
Data Collection

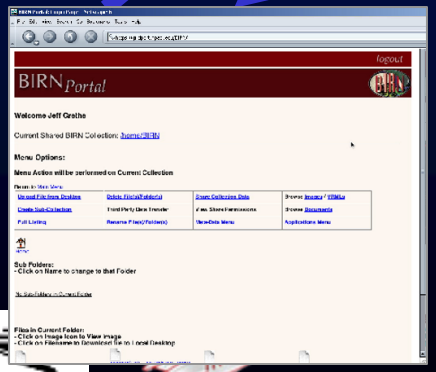
↓

Storage Repository	Data Grid
• Storage location	• Logical resource name space
• User name	• Logical user name space
• File name	• Logical file name space
• File context (creation date,…)	• Logical context (metadata)
• Access controls	• Access constraints

Data is organized as a shared collection

Using a Data Grid - *Details*





- User asks for data
- Data request goes to iRODS Server
- Server looks up information in catalog
- Catalog tells which iRODS server has data
- 1st server asks 2nd for data
- The 2nd iRODS server applies rules

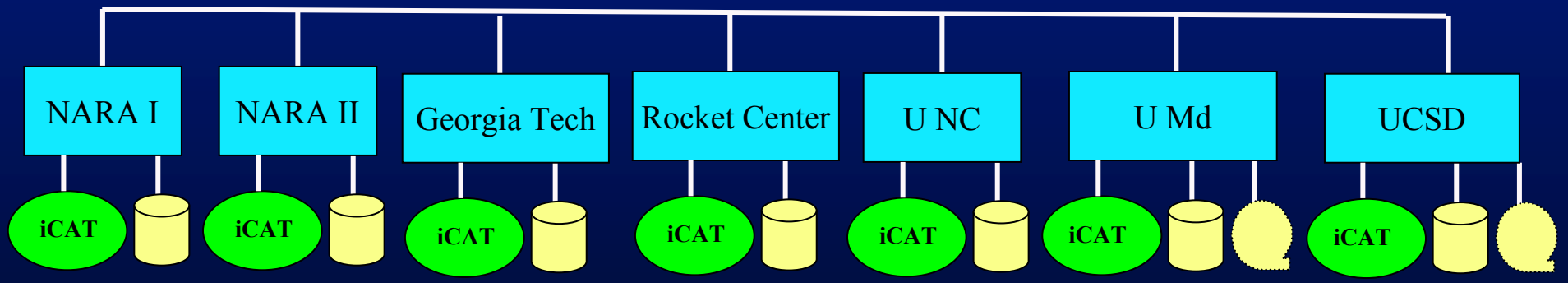
UCSD

UNC

Date	5/17/02		6/30/04			11/29/07		
Project	GBs of data stored	1000Õs of files	GBs of data stored	1000Õs of files	# Curators	GBs of data stored	1000Õs of files	# Curators
Data Grid								
NSF / NVO	17,800	5,139	51,380	8,690	80	88,216	14,550	100
NSF / NPACI	1,972	1,083	17,578	4,694	380	39,697	7,590	380
Hayden	6,800	41	7,201	113	178	8,013	161	227
Pzone	438	31	812	47	49	28,799	17,640	68
NSF / LDAS-SALK	239	1	4,562	16	66	207,018	169	67
NSF / SLAC-JCSG	514	77	4,317	563	47	23,854	2,493	55
NSF / TeraGrid			80,354	685	2,962	282,536	7,257	3,267
NIH / BIRN			5,416	3,366	148	20,400	40,747	445
NCAR						70,334	325	2
LCA						3,787	77	2
Digital Library								
NSF / LTER	158	3	233	6	35	260	42	36
NSF / Portal	33	5	1,745	48	384	2,620	53	460
NIH / AfCS	27	4	462	49	21	733	94	21
NSF / SIO Explorer	19	1	1,734	601	27	2,750	1,202	27
NSF / SCEC			15,246	1,737	52	168,931	3,545	73
LLNL						18,934	2,338	5
CHRON						12,863	6,443	5
Persistent Archive								
NARA	7	2	63	81	58	5,023	6,430	58
NSF / NSDL			2,785	20,054	119	7,499	84,984	136
UCSD Libraries			127	202	29	5,205	1,328	29
NHPRC / PAT						2,576	966	28
RoadNet						3,557	1,569	30
UCTV						7,140	2	5
LOC						6,644	192	8
Earth Sci						6,136	652	5
TOTAL	28 TB	6 mil	194 TB	40 mil	4,635	1,023 TB	200 mil	5,539

National Archives and Records Administration Transcontinental Persistent Archive Prototype

Federation of Seven
Independent Data Grids



Extensible Environment, can federate with additional research and education sites. Each data grid uses different vendor products.

Data Management Systems

iRODS - integrated Rule-Oriented Data System

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

Rules

- **Event**
  - Atomic, deferred, periodic
- **Condition**
  - Test on state information attributes
    - Storage system, collection, user group, file type
- **Action set**
  - Chained micro-services and rules
  - Server-side workflow
- **Recovery set**
  - Maintain transaction semantics in distributed system

Preservation Rules

- **Authenticity**
  - Rules that quantify required descriptive metadata
  - Rules that verify descriptive metadata is linked to records
  - Rules that govern creation of AIPs
- **Integrity**
  - Rules that verify records have not been corrupted
  - Rules that manage replicas
  - Rules that recover from corruption instances
  - Rules that manage data distribution
- **Chain of custody**
  - Persistent identifiers for archivists, records, storage
  - Rules to verify application of access controls
  - Rules to track storage location of records

RLG/NARA Assessment

- **Developed 105 rules that implement the TRAC assessment criteria**

90	*Verify descriptive metadata and source against SIP template and set SIP compliance flag*
91	*Verify descriptive metadata against semantic term list*
92	*Verify status of metadata catalog backup (create a snapshot of metadata catalog)*
93	*Verify consistency of preservation metadata after hardware change or error*

Theory of Data Management

- Characterization
  - Persistent name spaces
  - Operations that are performed upon the persistent name spaces
  - Changes to the persistent state information associated with each persistent name space that occur for each operation
  - Transformations that are made to the records on each operation
- Completeness
  - Set of operations is complete, enabling the decomposition of every management process onto the operation set.
  - Management policies are complete, enabling the validation of all assessment criteria.
  - Persistent state information is complete, enabling the validation of authenticity and integrity.
- Assertion
  - If the operations are reversible, then a future management environment can recreate a record in its original form, maintain authenticity and integrity, support access, and display the record.
  - Such a system would allow records to be migrated between independent implementations of management environments, while maintaining authenticity and integrity.

For More Information

Reagan W. Moore
University of North Carolina, Chapel Hill
rwmoore@renci.org

http://irods.diceresearch.org

Representation Information

- **Defined in OAIS - Open Archival Information System**
  - Constitutes the information required to parse, understand, and manipulate a record
- **Structural information**
  - Data format
- **Semantic information**
  - Descriptive metadata
  - Provenance metadata
- **Knowledge**
  - OAIS punted and provided a link to a knowledge community
- **CASPAR project in Europe is reanalyzing**

What is Information? (Scientific Data)

- **Data**
  - Comprised of bits
- **Information**
  - Labels applied to data
- **Knowledge**
  - Relationships between labels
- **Challenge**
  - The application of a label requires the assertion that specified relationships hold

Scientific data

- **Relationships**
  - Semantic - logical
    - Ontologies that define “is a” and “has a”
  - Structural - spatial
    - Graphical information systems
  - Procedural - temporal
    - Workflows
  - Algebraic - functional
    - Mathematical or physical
  - Systemic - epistemological
    - Properties of the collection

Scientific Data Representation for a Sequence of Bits

- **Impose structure on the bits**
  - Bytes : words : arrays
- **Label the structures**
  - Time : pressure : density : temperature
- **Map the structures to a coordinate system**
  - 1 Dimension : 2D : 3D : 4D : 11D
- **Map the coordinate systems to a geometry**
  - Rectilinear : cylindrical : spherical : toroidal
- **Map derived variables to functions**
  - Algorithms for pressure = density * temperature
- **Map variables to epistemology**
  - Mathematical theory behind the simulation
  - Units - MKS : CGS : EMU

Building a Digital Library for Scientific Data

- **Scale**
  - Size of collections
  - Size of records
- **Distributed over wide area networks**
  - Data sources
  - Researchers
- **Diversity of data types**
  - Semantics
  - Formats
  - Embedded relationships
- **Management policies**
  - Differ for each phase of the data life cycle

Motivation

- Observe that communities need to share data across institutions, administrative domains
- There typically is a driving purpose behind the data sharing
- There typically are specific management policies associated with the shared collection

Evolution of Data Grid Technology

- **Shared collections**
  - Enable researchers at multiple institutions to collaborate on research by sharing data
  - Focus was on performance, scalability
- **Digital libraries**
  - Support provenance information and discovery
  - Integrated with digital library front end services
- **Preservation environments**
  - Support preservation policies
  - Build rule-based data management system
- **Differ in choice of management policies**

Communities Driving Requirements

- DARPA Massive Data Analysis System (1996)
- DARPA/USPTO Distributed Object Computation Testbed
- NSF National Virtual Observatory - survey replication
- DOE Accelerated Strategic Computing Initiative - data aggregation
- NASA Information Power Grid - data grid
- NIH Biomedical Informatics Research Network - federation
- Interuniversity Consortium for Political and Social Research - bulk I/O
- DOE BaBar high energy physics - data grid
- National Optical Astronomy Observatories - data grid
- National Science Digital Library - archive
- UK e-Science data grid - parallel I/O
- NARA Transcontinental Persistent Archive Prototype - archive
- NHPRC Persistent Archive Testbed
- NSF Temporal Dynamics of Learning - data grid (2008)

SDSC Data Grids

- **Storage Resource Broker (SRB)**
  - Initially funded by DARPA in 1996
  - Current version is 3.5.0, released Dec 3, 2007
  - Production system used internationally
- **Integrated Rule-Oriented Data System (iRODS)**
  - Funded by NSF SDCI and NARA
  - Current version is 1.1, released June 2008

Software Development

- **iRODS release controlled by DICE group**
  - Daily build and test procedures
  - CVS maintained by DICE group
  - Multiple releases per year
  - Wiki for distributing information
  - irods-chat e-mail discussion list
  - Bugzilla error report tracking
  - Feature list of proposed developments

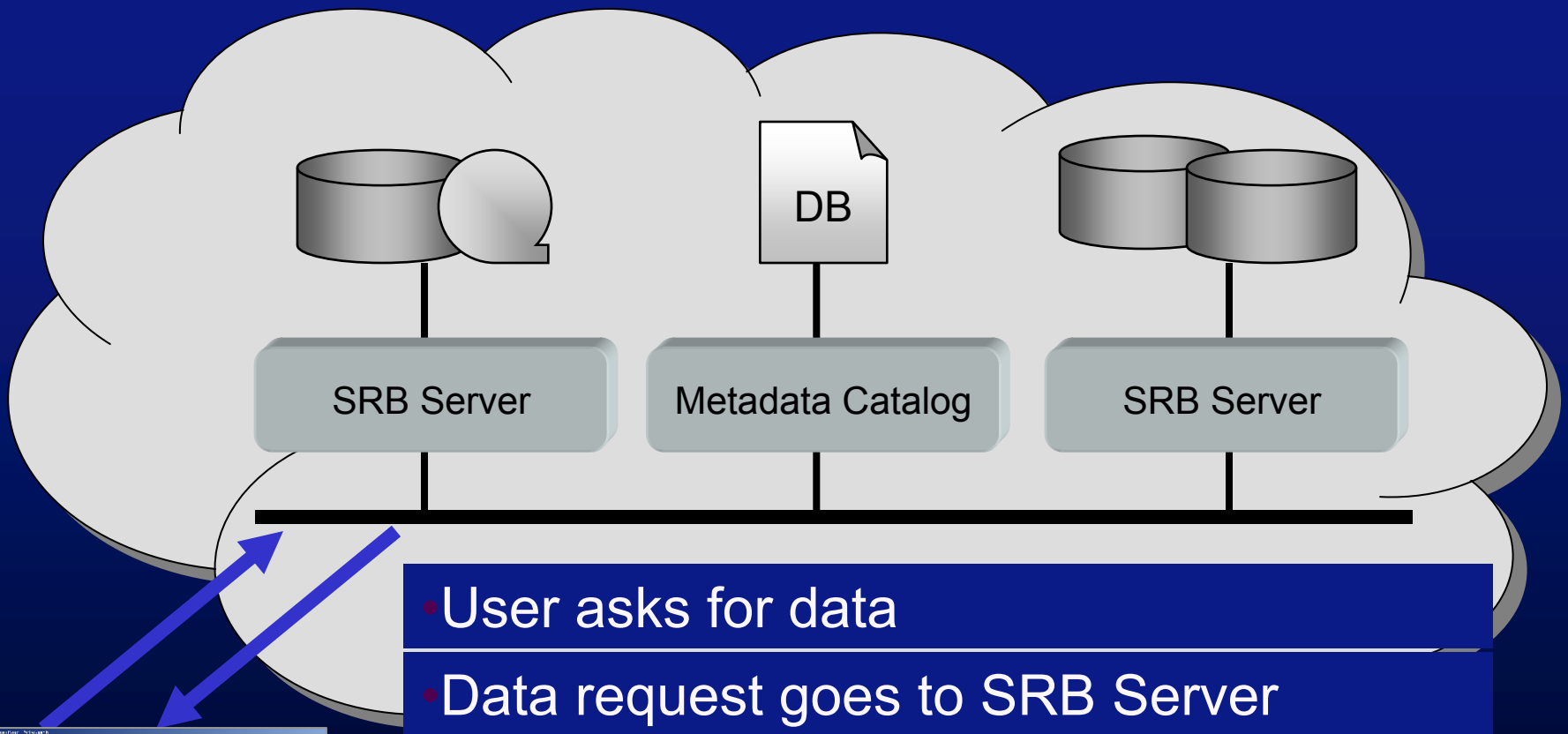
Design Implications

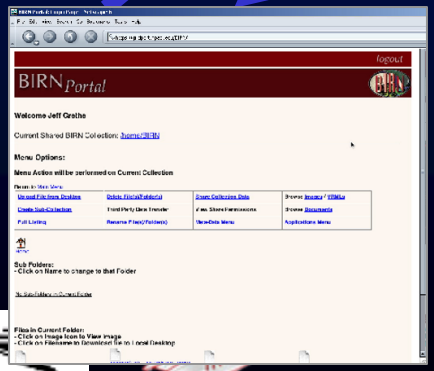
- **Heterogeneous storage systems**
  - Data stored in file systems, archives, databases
- **Global name spaces**
  - Files
  - Users
  - Resources
- **Persistent access controls**
  - Constraints between name spaces
- **Consistent state information**
  - Properties of files, collections, resources, users

Data Grids

- **Data virtualization**
  - Provide the persistent, global identifiers needed to manage distributed data
  - Provide standard operations for interacting with heterogeneous storage system
  - Provide standard actions for interacting with clients
- **Trust virtualization**
  - Manage authentication and authorization
  - Enable access controls on data, metadata, storage
- **Federation**
  - Controlled sharing of name spaces, files, and metadata between independent data grids
  - Data grid chaining / Central archives / Master-slave data grids / Peer-to-Peer data grids

Using a Data Grid - *SRB*





- User asks for data
- Data request goes to SRB Server
- Server looks up information in catalog
- Catalog tells which SRB server has data
- 1st server asks 2nd for data
- The 2nd SRB server returns the data

UCSD

UNC

Extremely Successful

- **Storage Resource Broker (SRB) manages more than 2 PBs of data in internationally shared collections**
- **Data collections for NSF, NARA, NASA, DOE, DOD, NIH, LC, NHPRC, IMLS: APAC, UK e-Science, IN2P3, WUNgrid**

Astronomy	Data grid
Bio-informatics	Digital library
Earth Sciences	Data grid
Ecology	Collection
Education	Persistent archive
Engineering	Digital library
Environmental science	Data grid
High energy physics	Data grid
Humanities	Data Grid
Medical community	Digital library
Oceanography	Real time sensor data, persistent archive
Seismology	Digital library, real-time sensor data

- **Goal has been generic infrastructure for distributed data**

Approach

- **To meet the diverse requirements, the architecture must:**
  - Be highly modular
  - Be highly extensible
  - Provide infrastructure independence
  - Enforce management policies
  - Provide scalability mechanisms
  - Manipulate structured information
  - Enable community standards

Generic Infrastructure

- **Data grids manage data distributed across multiple types of storage systems**
  - File systems, tape archives, object ring buffers
- **Data grids manage collection attributes**
  - Provenance, descriptive, system metadata
- **Data grids manage technology evolution**
  - At the point in time when new technology is available, both the old and new systems can be integrated

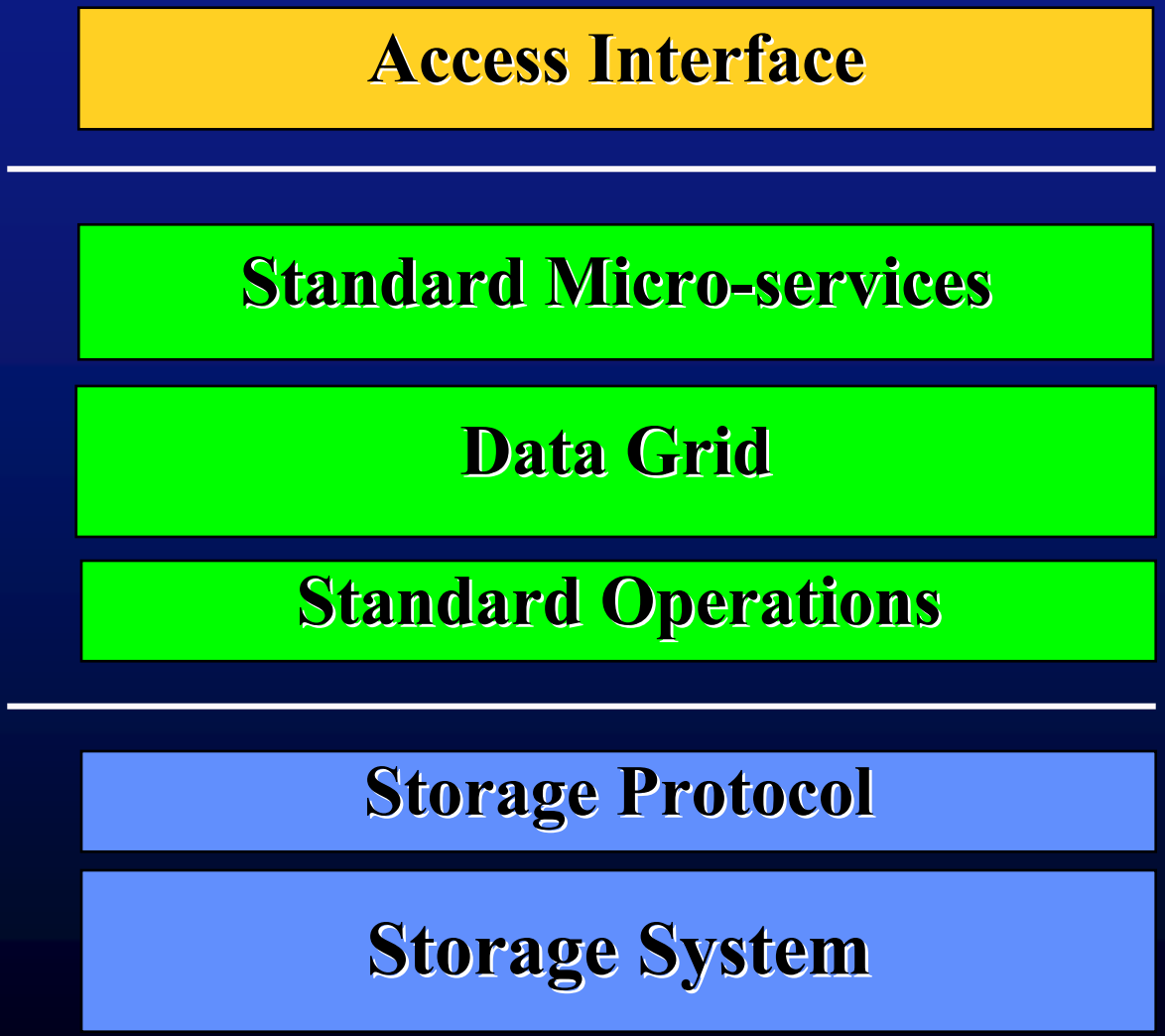
Observations of Production Data Grids

- **Each community implements different management polices**
  - Community specific preservation objectives
  - Community specific assertions about properties of the shared collection
  - Community specific management policies
- **Need a mechanism to support the socialization of shared collections**
  - Map from assertions made by collection creators to expectations of the users

Data Management

- **Observe that efficient systems require**
  - Aggregation of data
    - Bulk operations
    - Bulk storage
    - Bulk I/O
  - Aggregation of operations on data
    - Management policies
    - Remote Procedures
    - Structured information
    - Posix I/O
- **iRODS maps from policies to procedures to operations on structured information to I/O operations on storage systems**

Data Virtualization



Map from the actions requested by the access method to a standard set of micro-services. The standard micro-services are mapped to the operations supported by the storage system

iRODS Clients

- **C library calls**
- **Unix shell commands**
  - iRODS/clients/icommands/bin
- **Jargon Java I/O class library**
  - iRODS/java/jargon
- **FUSE user level file system**
  - iRODS/clients/fuse/bin/irodsFs fmount
- **Parrot**
  - http://www.cse.nd.edu/~ccl/software/parrot
- **Python-based web browser**
  - http://irods.sdsc.edu
- **iRODS rich web client**
  - https://rt.sdsc.edu:8443/irods/index.php
- **DSpace and Fedora in test**

iCommands

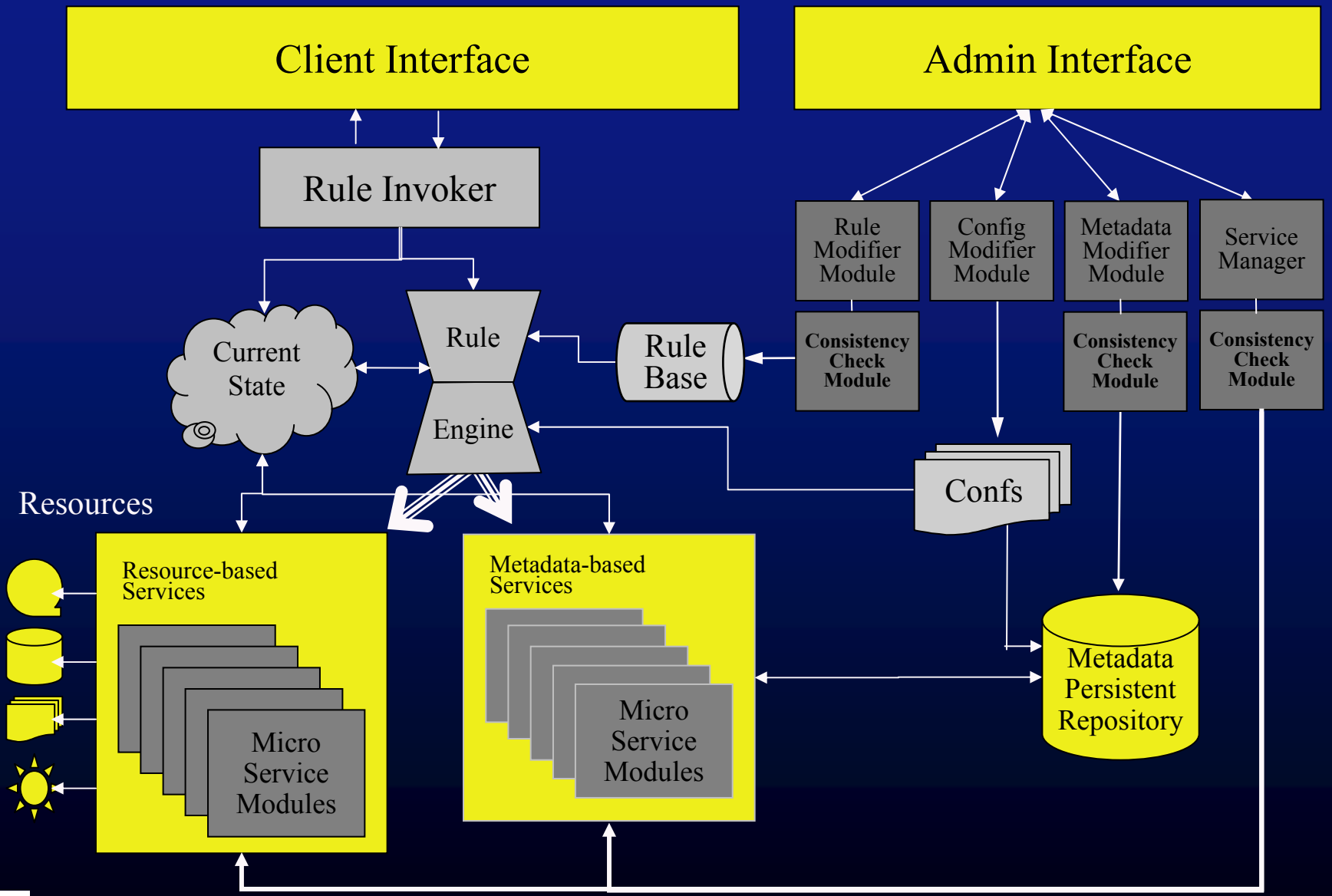
~/irods/clients/icommands/bin

- icd
- ichmod
- icp
- ils
- imkdir
- imv
- ipwd
- irm
- ienv
- ierror
- iget
- iput
- ireg
- irepl
- itrim
- irsync
- ilsresc
- iphymv
- irmtrash
- ichksum
- iinit
- iexit
- iqdel
- iqmod
- iqstat
- iexecmd
- irule
- iuserinfo
- isysmeta
- imeta
- iquest
- imiscsvrinfo
- iadmin

Standard Operations

- **The capabilities needed to interact with storage systems**
  - Posix I/O
  - File manipulation
  - Metadata manipulation
  - Bulk operations
  - Parallel I/O
  - Remote procedures
  - Registration

integrated Rule-Oriented Data System



UCSD

UNC

iRODS Components

- Clients
- Persistent state information catalog - iCAT
- Server middleware at each storage system
- Rule engine at each storage system

- Implements server-side workflows composed from micro-services
- Rules control execution of micro-services

irods Extensibility

- **Rules**
  - Use default rules for data grid capabilities
  - Administrator modification of pre-packaged rules (turn capabilities on and off)
  - Creation of new rules using existing micro-services
  - Write new micro-services and the rules controlling their execution

iROD Exensibility

- **State information**
  - Audit trails on system state information
    - Parse audit trails for compliance with policy
  - Add user-defined metadata
    - Descriptive context
  - Create schema versions
    - Map persistent state name to a different column in the database
  - Add new system metadata

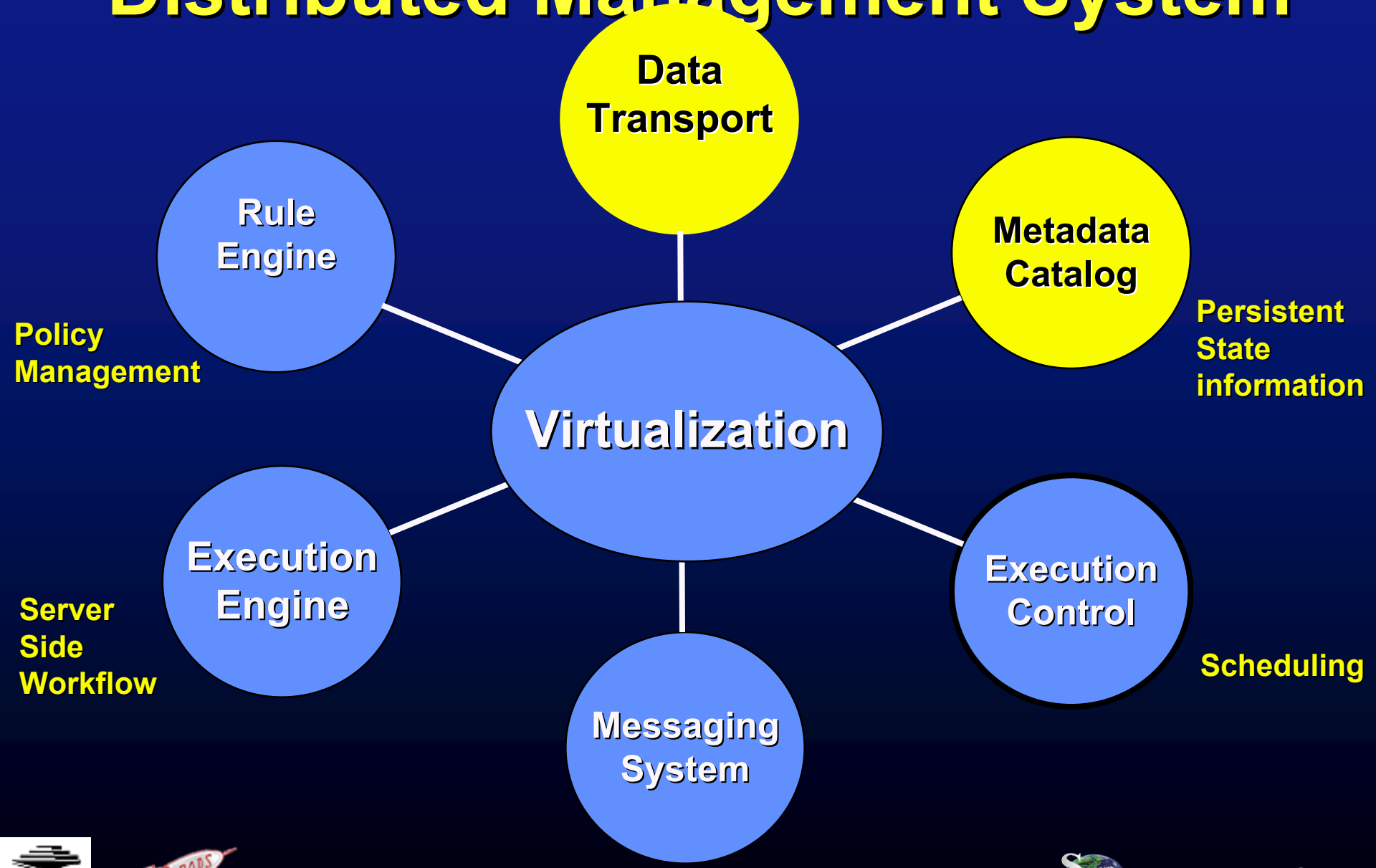
NARA Transcontinental Persistent Archive Prototype

- **Distributed Data Management Concepts**
  - Data virtualization
    - Storage system independence
  - Trust virtualization
    - Administration independence
  - Management virtualization
    - Policy independence
- **Risk mitigation**
  - Federation of multiple independent data grids
    - Operation independence

Data Management Challenges

- **Authenticity**
  - Manage descriptive metadata for each file
  - Manage access controls
  - Manage consistent updates to administrative metadata
- **Integrity**
  - Manage checksums
  - Replicate files
  - Synchronize replicas
  - Federate data grids
- **Infrastructure independence**
  - Manage collection properties
  - Manage interactions with storage systems
  - Manage distributed data

Distributed Management System



iRODS Development

- **NSF - SDCI grant "Adaptive Middleware for Community Shared Collections"**
  - iRODS development, SRB maintenance
- **NARA - Transcontinental Persistent Archive Prototype**
  - Trusted repository assessment criteria
- **NSF - Ocean Research Interactive Observatory Network (ORION)**
  - Real-time sensor data stream management
- **NSF - Temporal Dynamics of Learning Center data grid**
  - Management of IRB approval

Planned Development

- GSI support (1)
- Time-limited sessions via a one-way hash authentication
- Python Client library
- GUI Browser (AJAX in development)
- Driver for HPSS (in development)
- Driver for SAM-QFS
- Porting to additional versions of Unix/Linux
- Porting to Windows
- Support for MySQL as the metadata catalog
- API support packages based on existing mounted collection driver
- iCAT to ICAT migration tools (2)
- Extensible Metadata including Databases Access Interface (6)
- Zones/Federation (4)
- Auditing - mechanisms to record and track iRODS metadata changes

iRODS Development Collaborations

- Shibboleth - UK ASPIS
- GSI - BMBF Germany
- Data format parsing - EU SHAMAN
- Parrot interface - U Notre Dame
- Fedora port - Cornell / NSDL
- DSpace port - MIT
- Container - UK e-Science
- Perl/Python/PHP - Australia
- LStore integration - Vanderbilt

iRODS Tutorials - 2008

- January 31, SDSC
- April 8 - ISGC, Taipei
- May 13 - China, National Academy of Science
- May 27-30 - UK eScience, Edinburgh
- June 5 - OGF23, Barcelona
- July 7-11 - SAA, SDSC
- August 4-8 - SAA, SDSC
- August 25 - SAA, San Francisco

Approach - Data Grid

- **Organize the data into a shared collection**
  - Logical name space for organizing files
  - Access controls for allowed use
- **Manage the properties of the shared collection**
  - Description
  - Policies
  - Procedures
- **Automate**
  - Administrative functions
  - Validation of assessment criteria