Rule-Based Distributed Data Management

iRODS 1.0 - Jan 23, 2008
http://irods.sdsc.edu

Reagan W. Moore
Mike Wan
Arcot Rajasekar
Wayne Schroeder

San Diego Supercomputer Center

{moore, mwan, sekar, schroede}@sdsc.edu
http://irods.sdsc.edu
http://www.sdsc.edu/srb/

Data Management Applications

- 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

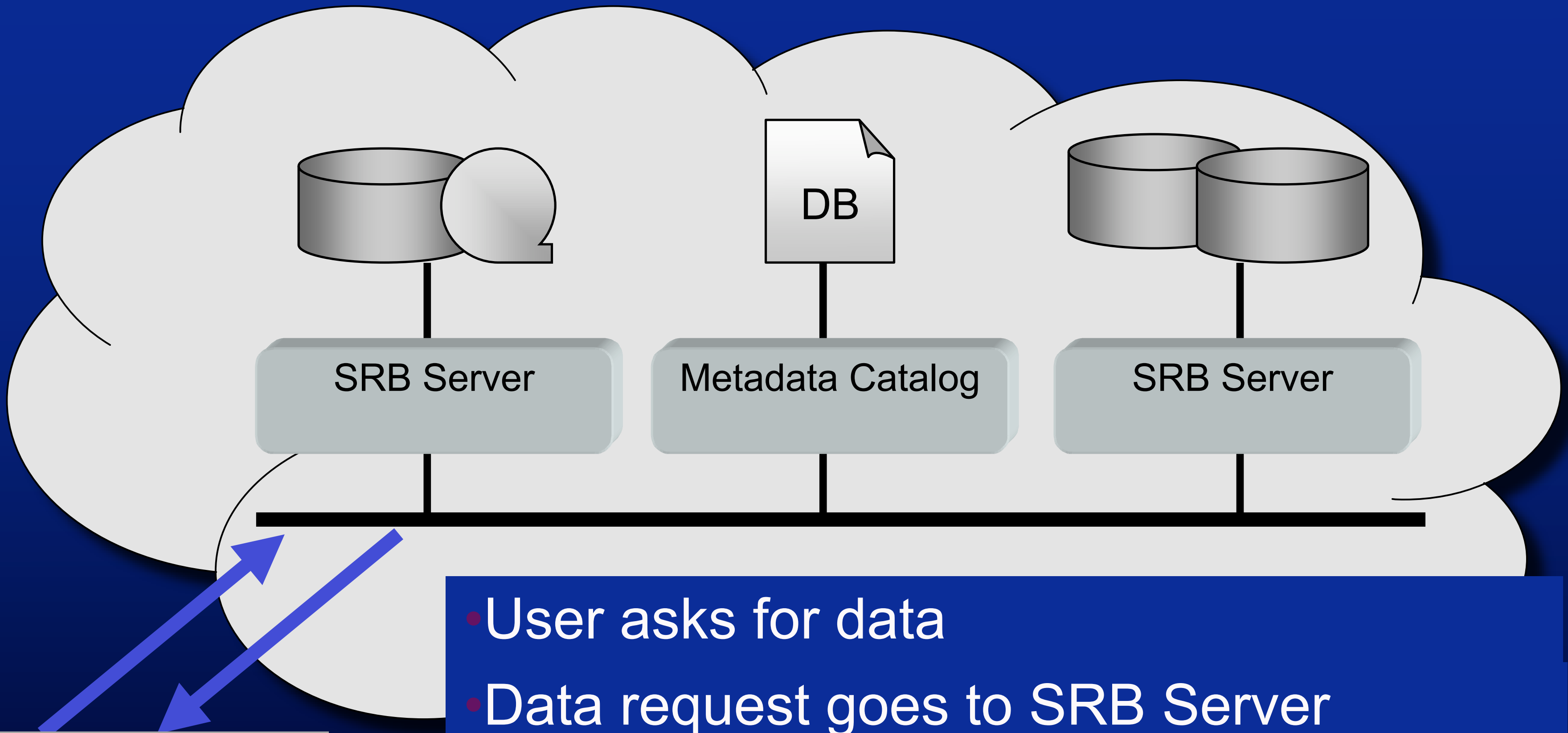
Data Management Goals

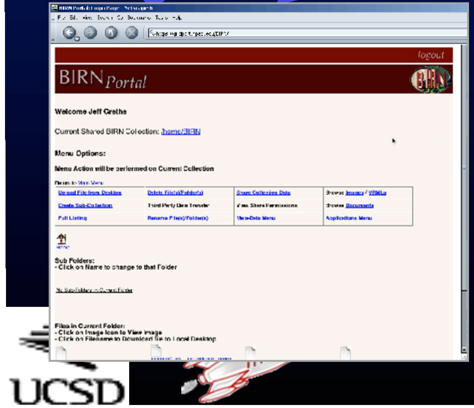
- **Support for data life cycle**
  - Shared collections -> data publication -> reference collections
- **Support for socialization of collections**
  - Process that governs life cycle transitions
  - Consensus building for collection properties
- **Generic infrastructure**
  - Common underlying distributed data management technology
  - iRODS - integrated Rule-Oriented Data System

Why Data Grids (SRB)?

- **Organize distributed data into shared collections**
  - Improve the ability for researchers to collaborate at national and international scale
  - Provide generic distributed data management mechanisms
    - Logical name spaces (files, users, storage systems)
    - Collection metadata
    - Replicas, versions, backups
    - Optimized data transport
    - Authentication and Authorization across domains
    - Support for community specific clients
    - Support for vendor specific storage protocols
    - Support for remote processing on data, aggregation in containers
    - Management of all phases of the data life cycle

Using a SRB Data Grid - *Details*




BIRN Portal: Login Page - Netscape 6
File Edit View Search Go Bookmarks Tools Help
logout
BIRN Portal
BIRN

Welcome Jeff Grethe

Current Shared BIRN Collection: /home/BIRN

Menu Options:

Menu Action will be performed on Current Collection

Return to Main Menu

Upload File from Desktop	Delete File(s)/Folder(s)	Share Collection Data	Browse Images / VRMLs
Create Sub-Collection	Third Party Data Transfer	View Share Permissions	Browse Documents
Full Listing	Rename File(s)/Folder(s)	Meta-Data Menu	Applications Menu

Sub Folders:
- Click on Name to change to that Folder

No Sub-Folders in Current Folder

Files in Current Folder:
- Click on Image Icon to View Image
- Click on Filename to Download file to Local Desktop

- 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 supplies the data

Extremely Successful

- Storage Resource Broker (SRB) manages 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

Date	5/17/02		6/30/04			11/29/07		
Project	GBs of data stored	1000’s of files	GBs of data stored	1000’s of files	Users with ACLs	GBs of data stored	1000’s of files	Users with ACLs
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

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

Why iRODS?

- **Need to verify assertions about the purpose of a collection**
  - Socialization of data collections, map from creator assertions to community expectations
- **Need to manage exponential growth in collection size**
  - Improve support for all phases of data life cycle from shared data within a project, to published data in a digital library, to reference collections within an archive
  - Data life cycle is a way to prune collections, and identify what is valuable
- **Need to minimize labor by automating enforcement of management policies**

Starting Requirements

- **Base capabilities upon features required by scientific research communities**
  - Started with features in SRB data grid, but needed to understand impact of management policies and procedures
- **Incorporate trustworthiness assessment criteria from the preservation community**
  - Other criteria include human subject approval, patient confidentiality, time-dependent access controls
- **Promote international support for iRODS development to enable research collaborations**

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

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

Tension between Common and Unique Components

- **Synergism - common infrastructure**
  - Distributed data
    - Sources, users, performance, reliability, analysis
  - Technology management
    - Incorporate new technology
- **Unique components - extensibility**
  - Information management
    - Semantics, formats, services
  - Management policies
    - Integrity, authenticity, availability, authorization

Data Grid Evolution

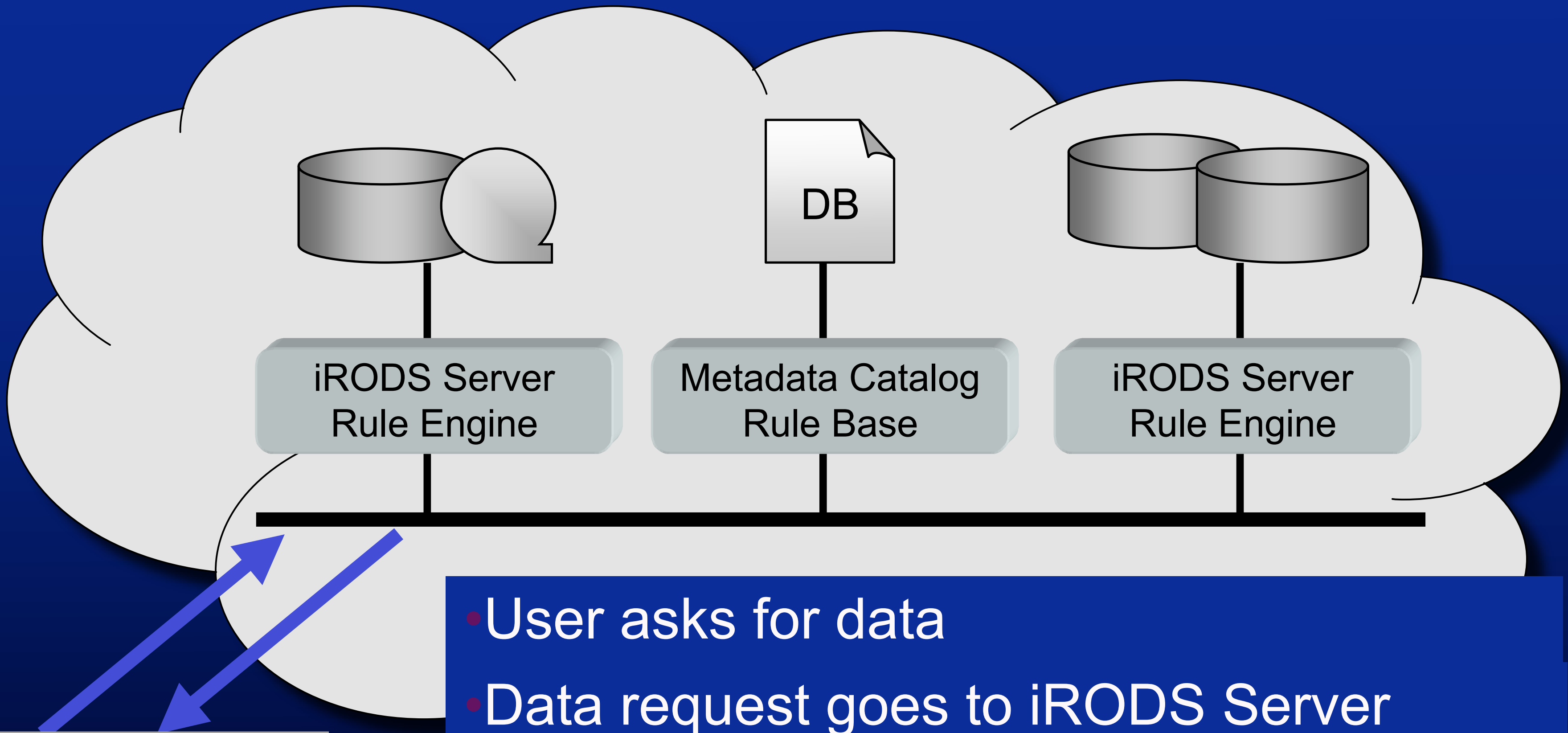
- **Implement essential components needed for synergism**
  - Storage Resource Broker - SRB
  - Infrastructure independence
  - Data and trust virtualization
- **Implement components needed for specific management policies and processes**
  - integrated Rule Oriented Data System - iRODS
  - Policy management virtualization
  - Map processes to standard micro-services
  - Structured information management and transmission

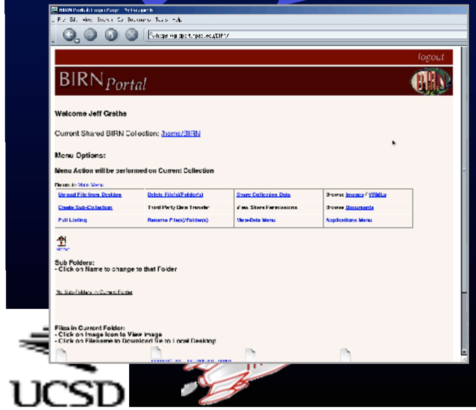
Initial iRODS Design

Next-generation data grid technology

- **Open source software - BSD license**
- **Unique capability - Virtualization of management policies**
  - Map management policies to rules
  - Enforce rules at each remote storage location
- **Highly extensible modular design**
  - Management procedures are mapped to micro-services that encapsulate operations performed at the remote storage location
  - Can add rules, micro-services, and state information
- **Layered architecture**
  - Separation of client protocols from storage protocols

Using an iRODS Data Grid - *Details*





Welcome Jeff Grethe

Current Shared BIRN Collection: /home/BIRN

Menu Options:

Menu Action will be performed on Current Collection

Sub Folders:
- Click on Name to change to that Folder

No Sub-Folders in Current Folder

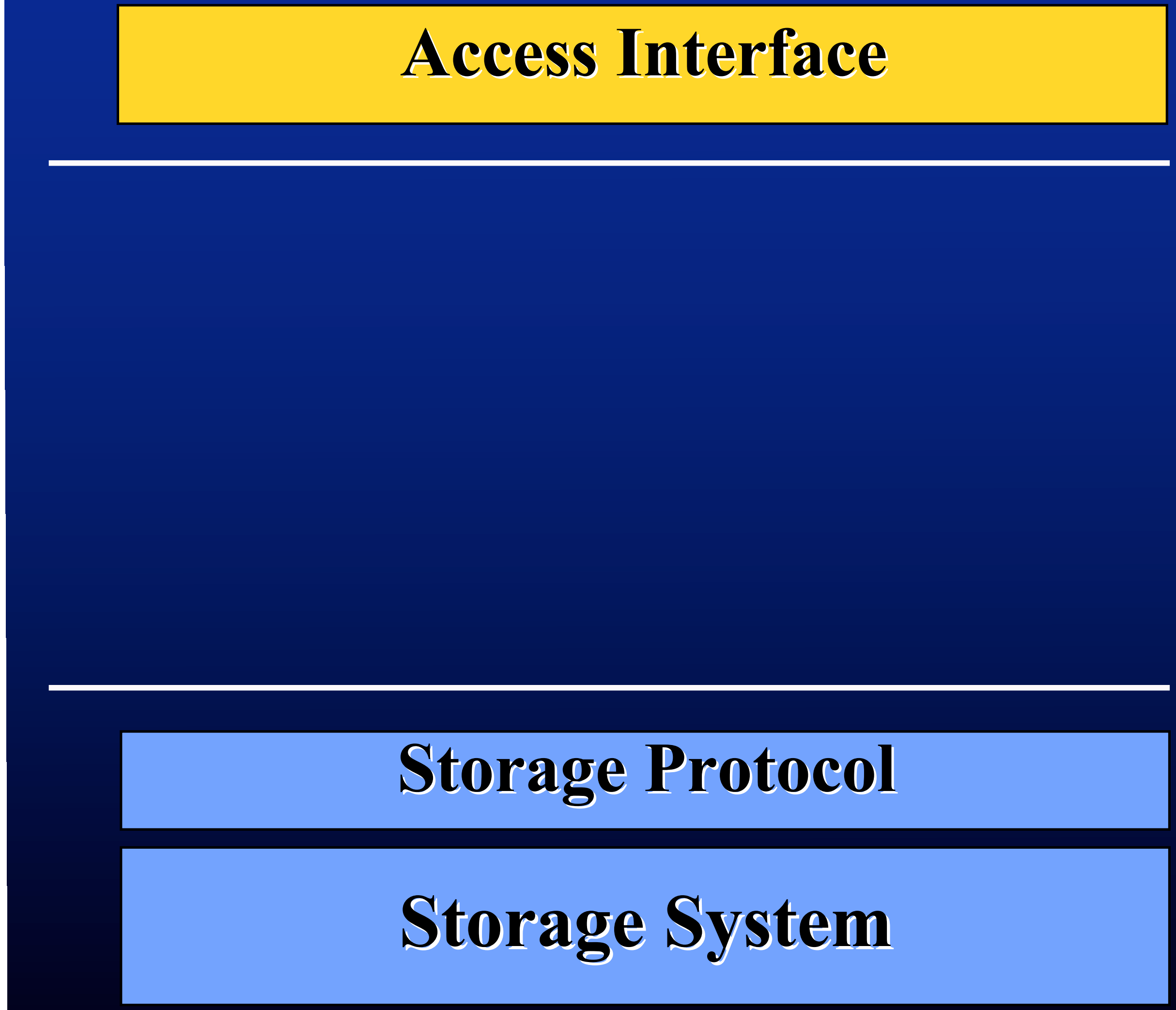
Files in Current Folder:
- Click on Image Icon to View Image
- Click on Filename to Download file to Local Desktop

- 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

Three Usage Models

- **Turnkey data management**
  - User gets / puts data into a shared collection
  - Advanced user adds descriptive metadata
- **Administrative control of data**
  - Administrator modifies rule base to impose management policies
- **Highly modular data management design**
  - Develop creates micro-services to implement server-side workflows to process data

Turnkey Data Management



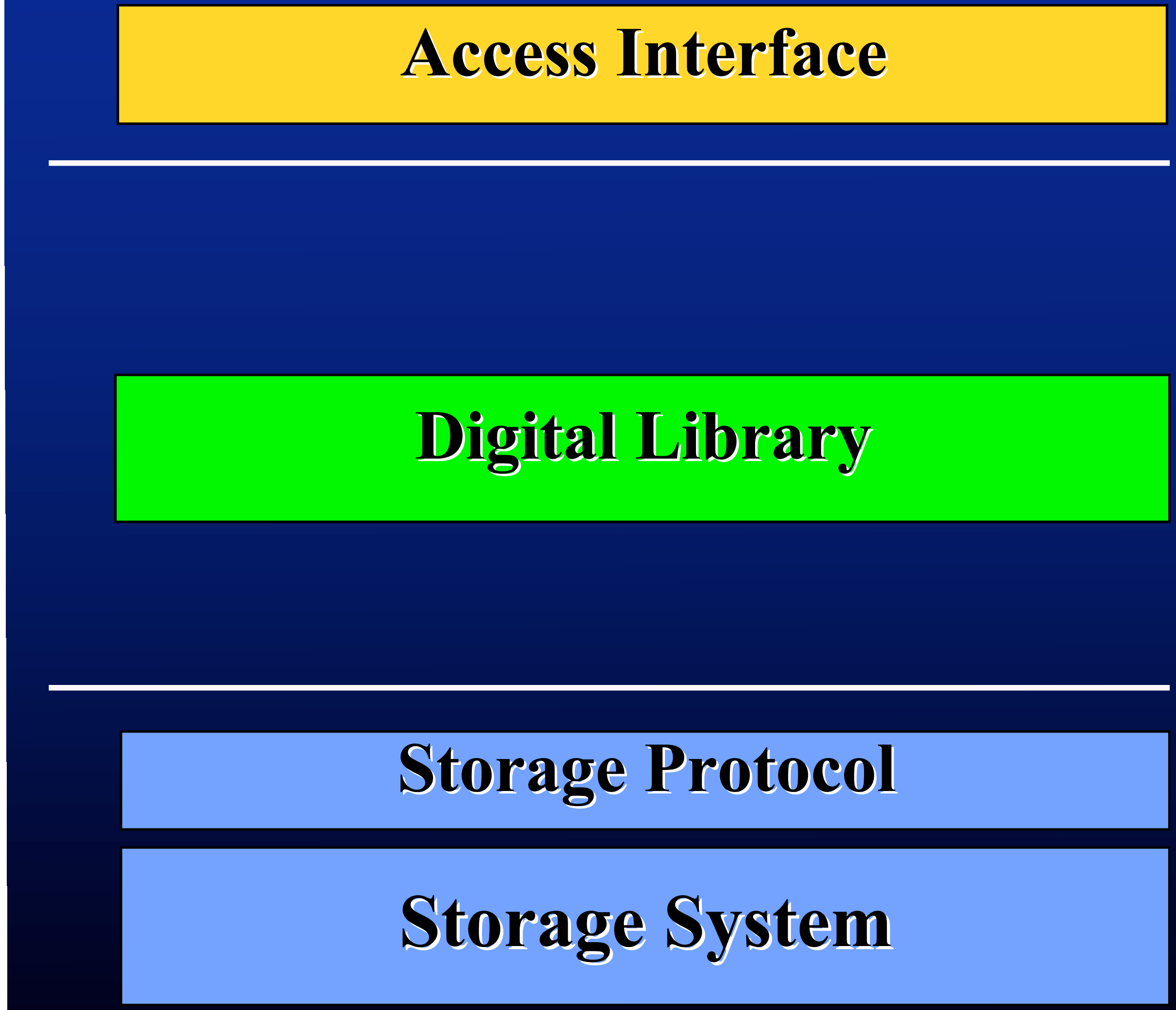
Access Interface

Storage Protocol

Storage System

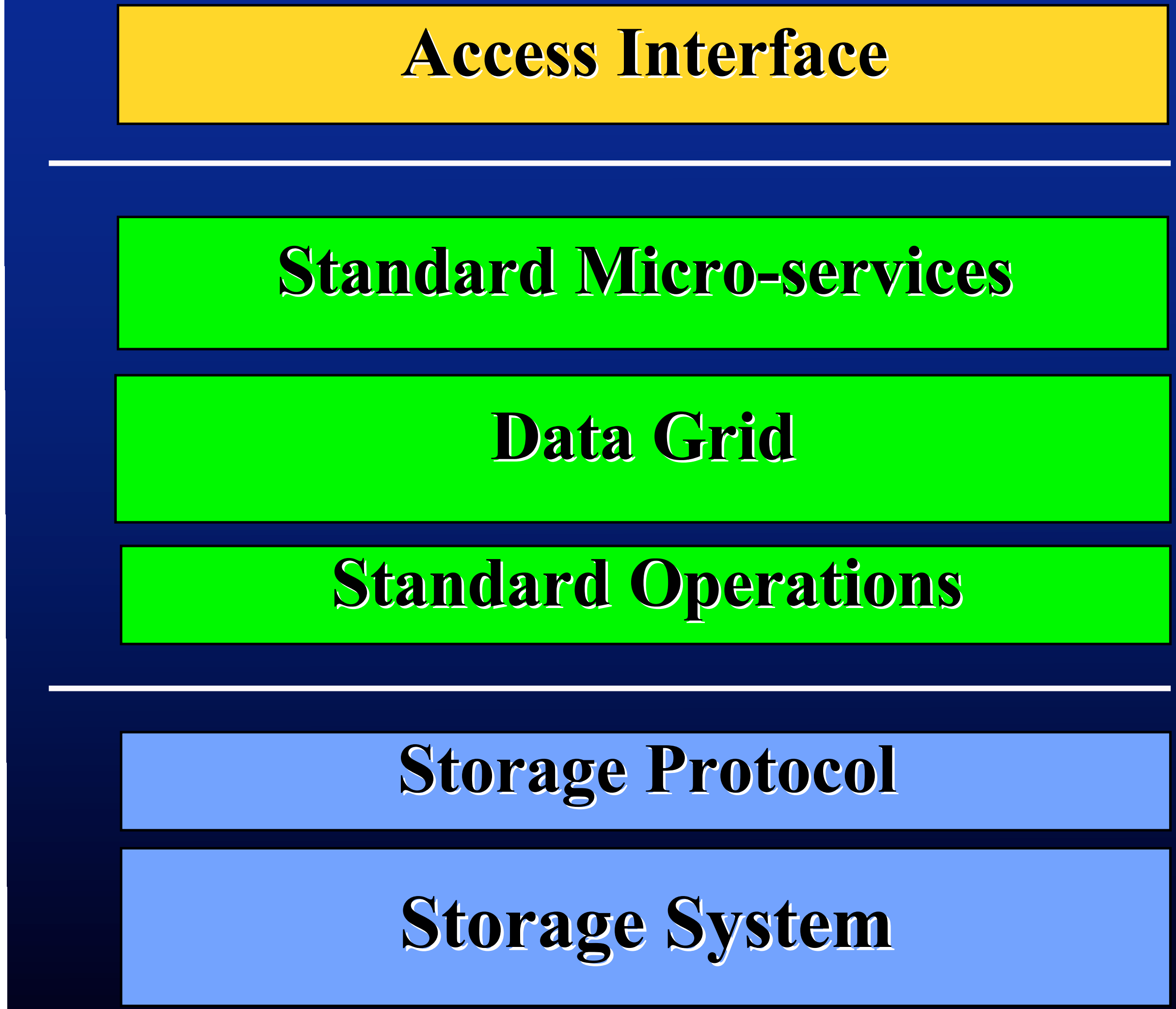
Traditional approach: Client talks directly to storage system using Unix I/O: Microsoft Word

Data Virtualization (Digital Library)



Client talks to the Digital Library which then interacts with the storage system using Unix I/O

Data Virtualization (iRODS)



- Map from the actions requested by the access method to a standard set of micro-services.
- The standard micro-services use standard operations.
- Separate protocol drivers are written for each storage system.

iRODS Release 1.0

- Open source software available at wiki:
  - http://irods.sdsc.edu
- Since January 23, 2008, more than 590 downloads by projects in 18 countries:
  - Australia, Austria, Belgium, Brazil, China, France, Germany, Hungary, India, Italy, Norway, Poland, Portugal, Russia, Spain, Taiwan, UK, and the US
- Release 1.1 scheduled for June 2008

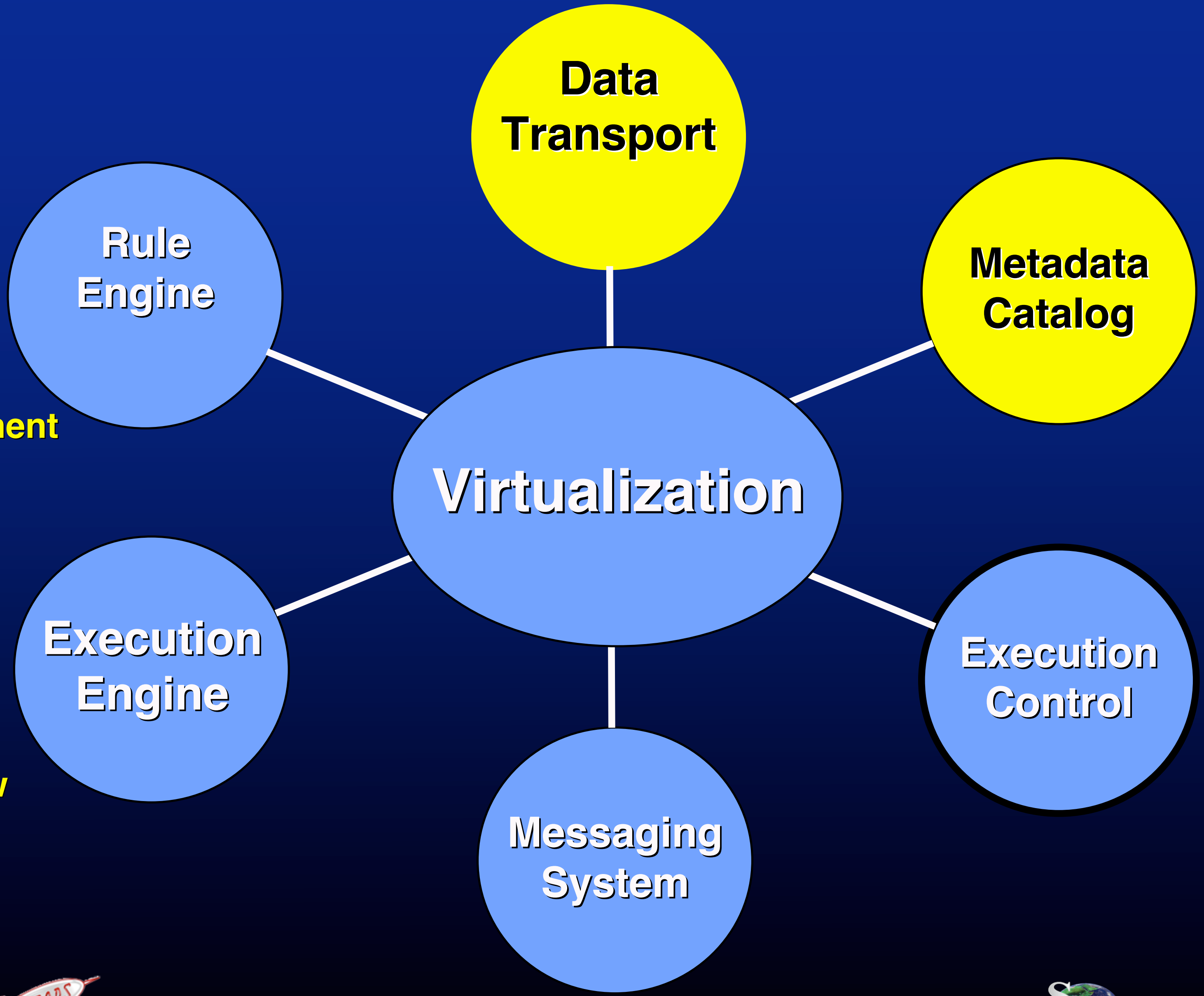
Core Components

- **Framework**
  - Infrastructure that ties together the layered environment
- **Drivers**
  - Infrastructure that interacts with commercial protocols (database, storage, information resource)
- **Clients**
  - Community specific access protocols
- **Rules**
  - Management policies specific to a community
- **Micro-services**
  - Management procedures specific to a community
- **Quality assurance**
  - Testing routines for code validation
- **Maintenance**
  - Bug fixes, help desk, chat, bugzilla, wiki

Rule Specification

- **Rule - Event : Condition : Action set : Recovery Procedure**
  - Event - atomic, deferred, periodic
  - Condition - test on any state information attribute
  - Action set - chained micro-services and rules
  - Recovery procedure - ensure transaction semantics in a distributed world
- **Rule types**
  - System level, administrative level, user level

Distributed Management System



Policy Management

Persistent State information

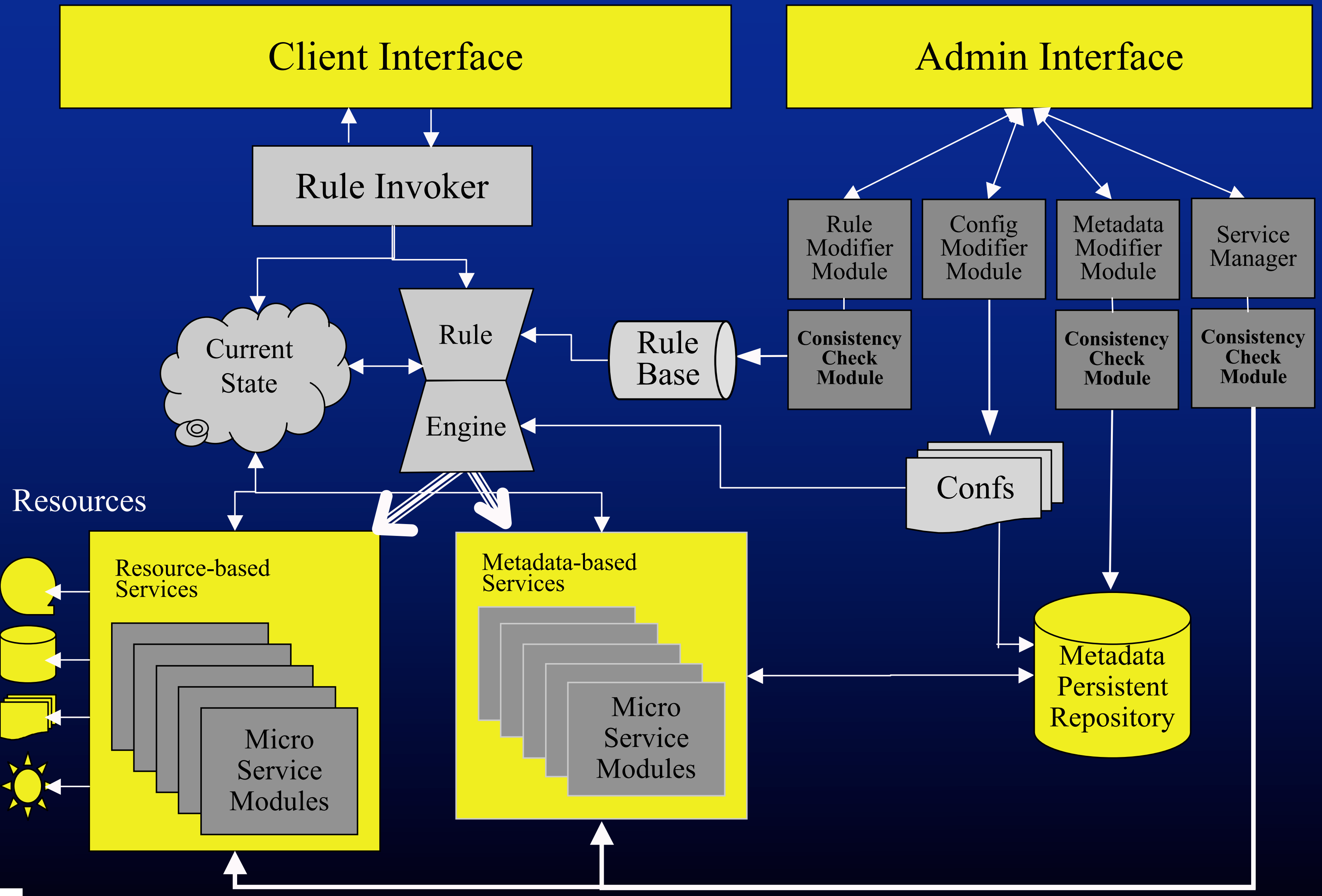
Server Side Workflow

Scheduling

Data Transfer Mechanisms

- **Large files**
  - Parallel I/O using 4-16 I/O streams
  - Observe 5 TBs/day (60 MB/sec from SLAC to IN2P3)
- **Small files**
  - Optimized protocol, send a small file in a single message during a session
- **Message system**
  - High transaction rate, 5000 messages per second

integrated Rule-Oriented Data System



iRODS Data Grid Capabilities

- **Rules**
  - User / administrative / internal
  - Remote web service invocation
  - Rule & micro-service creation
  - Standards / XAM, SNIA
- **Remote procedures**
  - Atomic / deferred / periodic
  - Procedure execution / chaining
  - Structured information
- **Structured information**
  - Metadata catalog interactions / 205 queries
  - Information transmission
  - Template parsing
  - Memory structures
  - Report generation / audit trail parsing

First Major Innovation

1. Management virtualization

- Expression of management policies as rules
- Expression of management procedures as remote micro-services
- Expression of assertions as queries on persistent state information

- Required addition of three more logical name spaces for **rules, micro-services, and state information**

Second Major Innovation

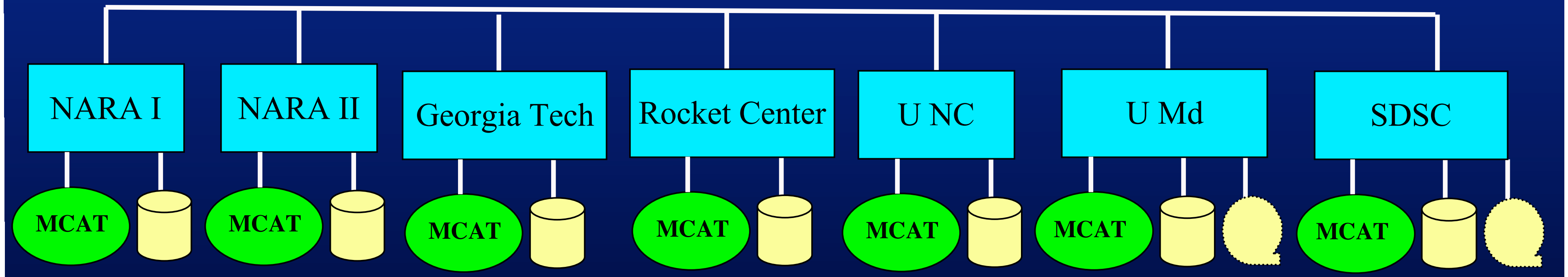
- **Recognition of the need to support structured information**
  - Manage exchange of structured information between micro-services
    - Argument passing
    - Memory white board
  - Manage transmission of structured information between servers and clients
    - C-based protocol for efficiency
    - XML-based protocol to simplify client porting (Java)
    - High performance message system

Third Major Innovation

- **Development of the Mounted Collection interface**
  - Standard set of operations (20) for extracting information from a remote information resource
  - Allows data grid to interact with autonomous resources which manage information independently of iRODS
  - Structured information drivers implement the information exchange protocol used by a particular information repository
- **Examples**
  - Mounted Unix directory
  - Tar file

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.

Project Coordination

- **Define international collaborators**
  - Technology developers for a specific development phase for a specific component.
- **Collaborators span:**
  - Scientific disciplines
  - Communities of practice (digital library, archive, grid)
  - Technology developers
  - Resource providers
  - Institutions and user communities
- **Federations within each community are essential for managing scientific data life cycle**

Scientific Data Life Cycle

- **Shared collection**
  - Used by a project to promote collaboration between distributed researchers
  - Project members agree on semantics, data formats, and manipulation services
- **Data publication**
  - Requires defining context for the data
  - Provenance, conformance to community format standards
- **Reference collections**
  - Community standard against which future research results are compared

Scientific Data Life Cycle

- **Each phase of the life cycle requires consensus by a broader community**
- **Need mechanisms for expressing the new purpose for the data collection**
- **Need mechanisms that verify**
  - Authoritative source
  - Completeness
  - Integrity
  - Authenticity

Why iRODS?

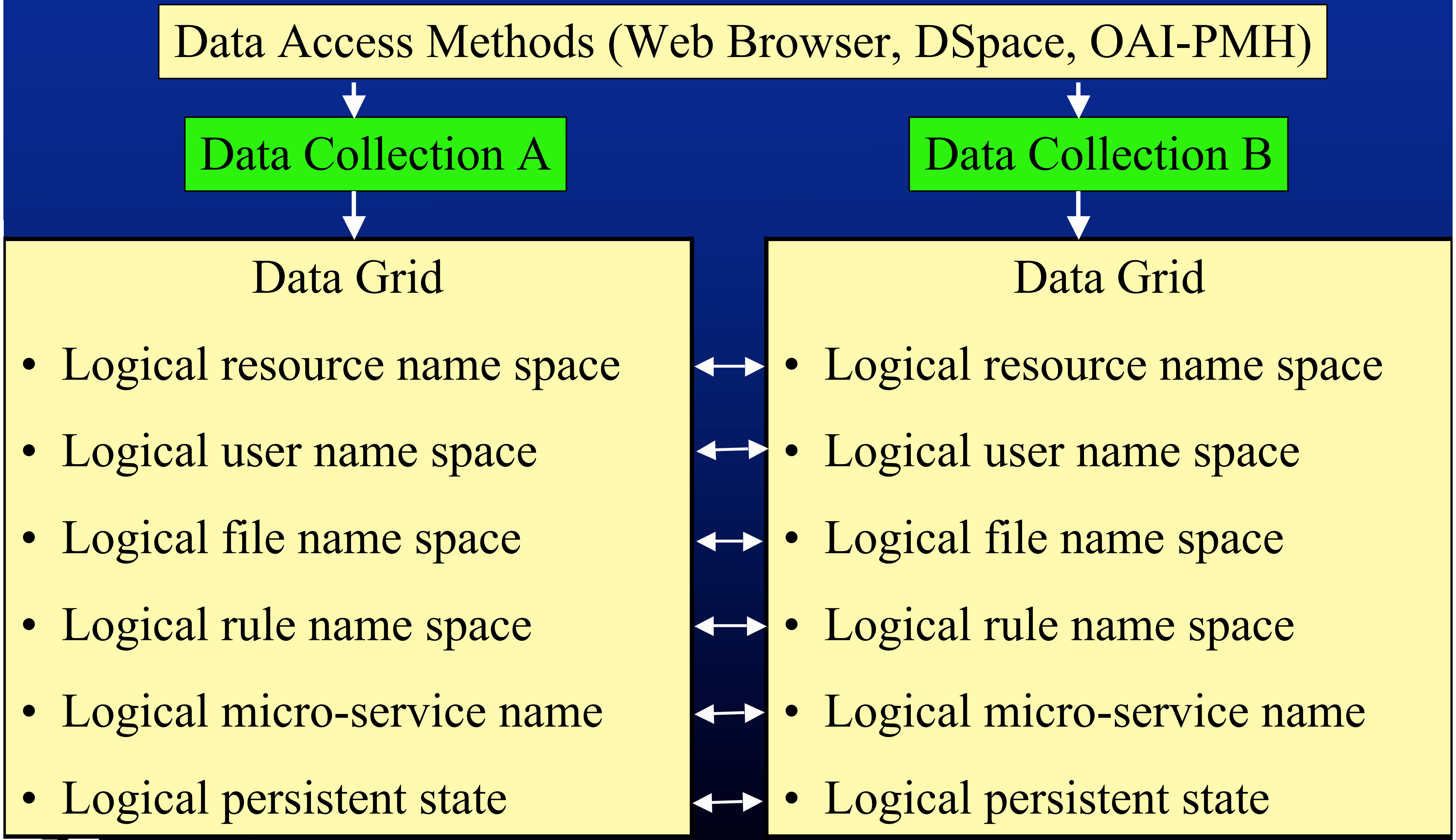
- **Collections are assembled for a purpose**
  - Map purpose to assessment criteria
  - Use management policies to meet assertions
  - Use management procedures to enforce policies
  - Track persistent state information generated by every procedure
  - Validate criteria by queries on state information and on audit trails

Data Management

iRODS - integrated Rule-Oriented Data System

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

Federation Between IRODS Data Grids



Data Access Methods (Web Browser, DSpace, OAI-PMH)

Data Collection A

Data Collection B

Data Grid

- Logical resource name space
- Logical user name space
- Logical file name space
- Logical rule name space
- Logical micro-service name
- Logical persistent state

Data Grid

- Logical resource name space
- Logical user name space
- Logical file name space
- Logical rule name space
- Logical micro-service name
- Logical persistent state

Major iRODS Research Question

- Do we **federate** data grids as was done in the SRB, by explicitly cross-registering information?
- Or do we take advantage of the **Mounted Collection** interface and access each data grid as an autonomous information resource?
- Or do we use a **rule-based database access interface** for interactions between iCAT catalogs?

Mounted Collections

- **Minimizes dependencies between the autonomous systems**
  - Supports retrieval of information from the remote information resource that is needed for interaction
  - Can be controlled by rules that automate interactions
    - Chained data grids
    - Central archive (archive pulls from other data grids)
    - Master-slave data grids (slaves pull from master)

Rule-based Database Access Interface

- **Support interactions by querying the remote iCAT catalog’s database**
  - Expect to support publication of schemata
  - Ontology-based reasoning on semantics
  - Can be used for both deposition and retrieval of information
  - Simplifies exchange of rules and possibly of micro-services

Theory of Data Management

- **Prove compliance of data management system with specified assertions**
  1. Define the purpose for the collection, expressed as assessment criteria, management policies, and management procedures
  2. Analyze completeness and closure of the system
     - For each criteria, persistent state is generated that can be audited
     - Persistent state attributes are generated by specific procedure versions
     - For each procedure version there are specific management policy versions
     - For each criteria, there are governing policies
  3. Audit properties of the system
     - Periodic rules validate assessment criteria

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
- MCAT to ICAT migration tools (2)
- Extensible Metadata including Databases Access Interface (6)
- Zones/Federation (4)
- Auditing - mechanisms to record and track iRODS metadata changes

For More Information

Reagan W. Moore
San Diego Supercomputer Center
moore@sdsc.edu

http://www.sdsc.edu/srb/
http://irods.sdsc.edu/