



# Partnership for Preserving the Records of the eLegislature

Reagan W. Moore

Richard Marciano

Chien-Yi Hou

{moore, marciano, chienyi}@sdsc.edu

Elizabeth Lighthipe

Shawn Rounds

Robert Horton

{shawn.rounds, beth.lighthipe}@mnhs.org

<http://www.sdsc.edu/srb>

<http://irods.sdsc.edu/>



# 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

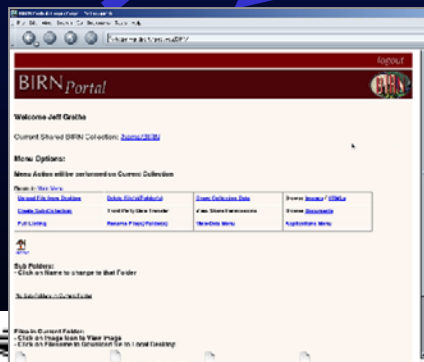
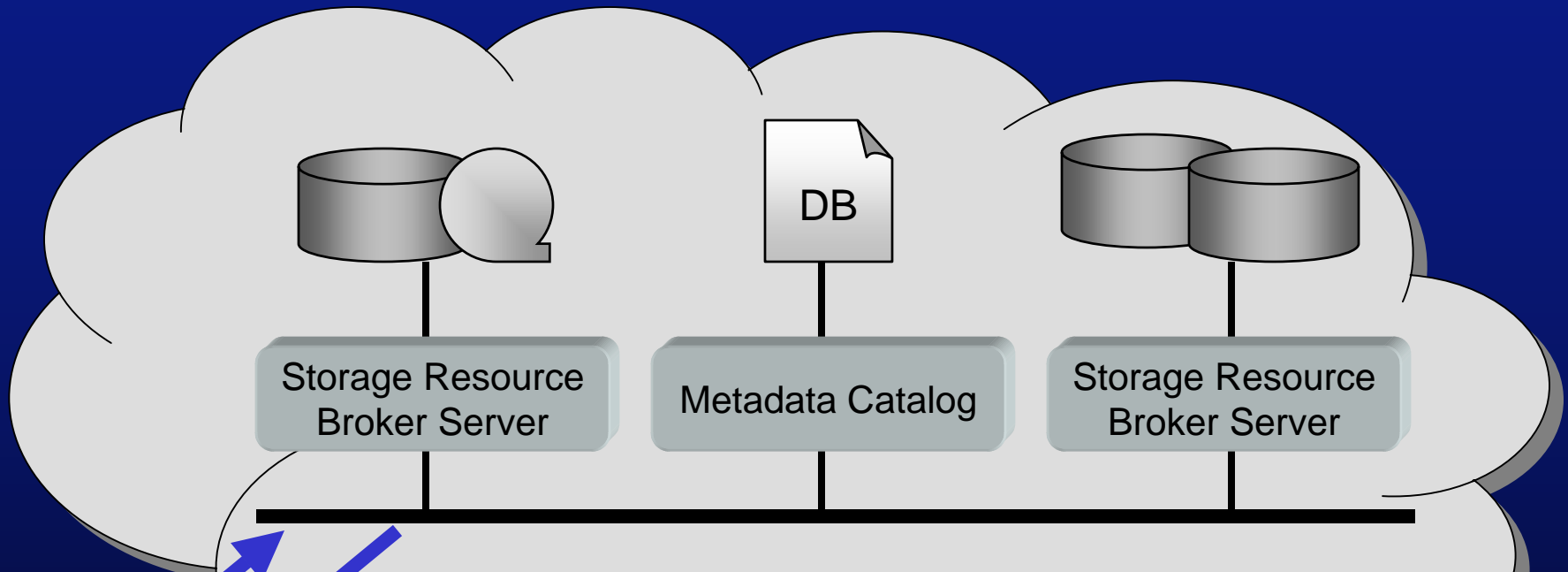


# Site Dependencies

- **Each preservation environment site is unique**
  - Local network configuration (**firewalls**, network infrastructure, storage systems, operating systems)
  - Local management policies (**arrangement**, description, preservation, access)
  - Preferred access mechanisms (web browser, portal, DSpace digital library, **Windows browser**)
- **Challenge is to provide unifying infrastructure that enables a preservation environment to span multiple sites**



# Using a Data Grid - *Details*



- User asks for data
- Data request goes to SRB Server
- Server looks up information in catalog
- Catalog tells which SRB server has data
- 1<sup>st</sup> server asks 2<sup>nd</sup> for data
- The data is found and returned



# Standard Approach

- **Install data grid testbed**
  - Verify network connectivity
  - Integrate with local firewalls
- **Verify access to original records**
  - Identify integrity mechanisms
  - Identify locations of descriptive metadata
  - Identify preferred arrangement
- **Select desired interface**
  - Windows browser, web browser, Unix shell commands



# eLegislature Testbed



- **Installed SRB server on a Windows platform**
  - Supported firewall configuration
  - Registered existing files into the SRB data grid
  - Identified very slow network (2 Mbits/sec) implying a 3-month data transfer time
- **Replicated files to an archive at SDSC**
  - Copied files onto disk, and shipped the disk
  - Registered the files into the SRB collection at SDSC as replicas of the original files
  - Used checksums to identify corrupted files
- **Created workflow process to verify integrity**
  - Checksummed files before writing to disk
  - Checksummed files after reception at SDSC
  - Checksummed files after storage in archive



# eLegislature Challenges



- Validated testbed
  - Identified problems with Windows platform for bulk operations
  - Provided upgrades to the SRB data grid (typically 2 releases per year)
  - Upgraded documentation (wiki site <http://www.sdsc.edu/srb>)
- Now need to automate preservation policies
  - **Next generation data grid technology -iRODS**



# Automating Policies

- **Build disaster recovery data grid**
  - Identify point in business process at which records are created
  - Automate registration of the records into a disaster recovery data grid for the Revisor's Office
  - Replicate directly from the Revisor's Office onto the archive at SDSC
- **Build independent archive data grid at SDSC**
  - Access material from the disaster recovery data grid
  - Create archival information package
  - Register record into the archive data grid
  - Perform operations on records in higher bandwidth network environment at SDSC





# Data Grids

- **SRB - Storage Resource Broker**
  - Persistent naming of distributed data
  - Management of data stored in multiple types of storage systems
  - Organization of data as a shared collection with descriptive metadata, access controls, audit trails
- **iRODS - integrated Rule-Oriented Data System**
  - Rules control execution of remote micro-services
  - Manage persistent state information
  - Validate assertions about collection
  - Automate execution of management policies



# Automate Management Policies



- **Policies may differ between disaster recovery data grid and archive data grid**
  - Periodic integrity / authenticity checks
  - Version control on changes by Revisor's Office
  - Periodic harvesting of records from Revisor's Office
  - Organization of material by year
  - Retention period for Revisor's Office material
  - De-accession approval flags
  - Format migration
  - Access controls



# Preservation Management

iRODS - integrated Rule-Oriented Data System

<i>Data Management Environment</i>	<b>Conserved Properties</b>	<b>Control Mechanisms</b>	<b>Remote Operations</b>
<b>Management Functions</b>	<b>Assessment Criteria</b>	<b>Preservation Policies</b>	<b>Preservation Procedures</b>
	<b>Data grid Š Management virtualization</b>		
<b>Data Management Infrastructure</b>	<b>Persistent State</b>	<b>Rules</b>	<b>Micro-services</b>
	<b>Data grid Š Data and trust virtualization</b>		
<b>Physical Infrastructure</b>	<b>Database</b>	<b>Rule Engine</b>	<b>Storage System</b>



# For More Information

Reagan W. Moore  
San Diego Supercomputer Center  
[moore@sdsc.edu](mailto:moore@sdsc.edu)

<http://www.sdsc.edu/srb/>

<http://irods.sdsc.edu/>