Implementing Trusted Digital Repositories

Reagan W. Moore
Richard Marciano
Arcot Rajasekar
Wayne Schroeder
Mike Wan

{moore, schroede, mwan, sekar, marciano}@sdsc.edu

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

UCSD

SDSC

Topics

- **Representation information for preservation environments**
  - How can preservation policies and procedures be characterized?
- **Rule-based data management systems**
  - How do we make assertions about the trustworthiness of a preservation environment?
- **Theory of digital preservation**
  - What are the components on which a theory could be based?

UCSD

SDSC

Digital Preservation

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

UCSD

SDSC

Assessment Criteria

- **Authenticity**
  - Management of descriptive information about record provenance, record representation information
- **Integrity**
  - Minimization of the risk of data loss
- **Chain of custody**
  - Verification of archivist management policies
- **Respect des fonds**
  - Preservation of the original arrangement of the records
- **Trustworthiness**
  - RLG/NARA assessment criteria - 174 rules

Controlling Remote Operations

iRODS - integrated Rule-Oriented Data System

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

UCSD

SDSC

Representation Information for Preservation Environments

- **Assessment criteria**
  - Mapped to sets of persistent state information
- **Management policies**
  - Mapped to sets of rules
- **Preservation processes**
  - Mapped to sets of micro-services

- **Rules generate persistent state information by controlling the execution of sets of micro-services at remote storage systems**

UCSD

SDSC

Example Rule

- **Rule composed of four parts:**
  - Name | condition | micro-service set | recovery set
- **Rule to automate replication of data for a specific collection**

  acPostProcForPut |

  $objPath like /tempZone/home/rods/nvo/* | msiSysReplDataObj(nvoReplResc,null) |

  nop

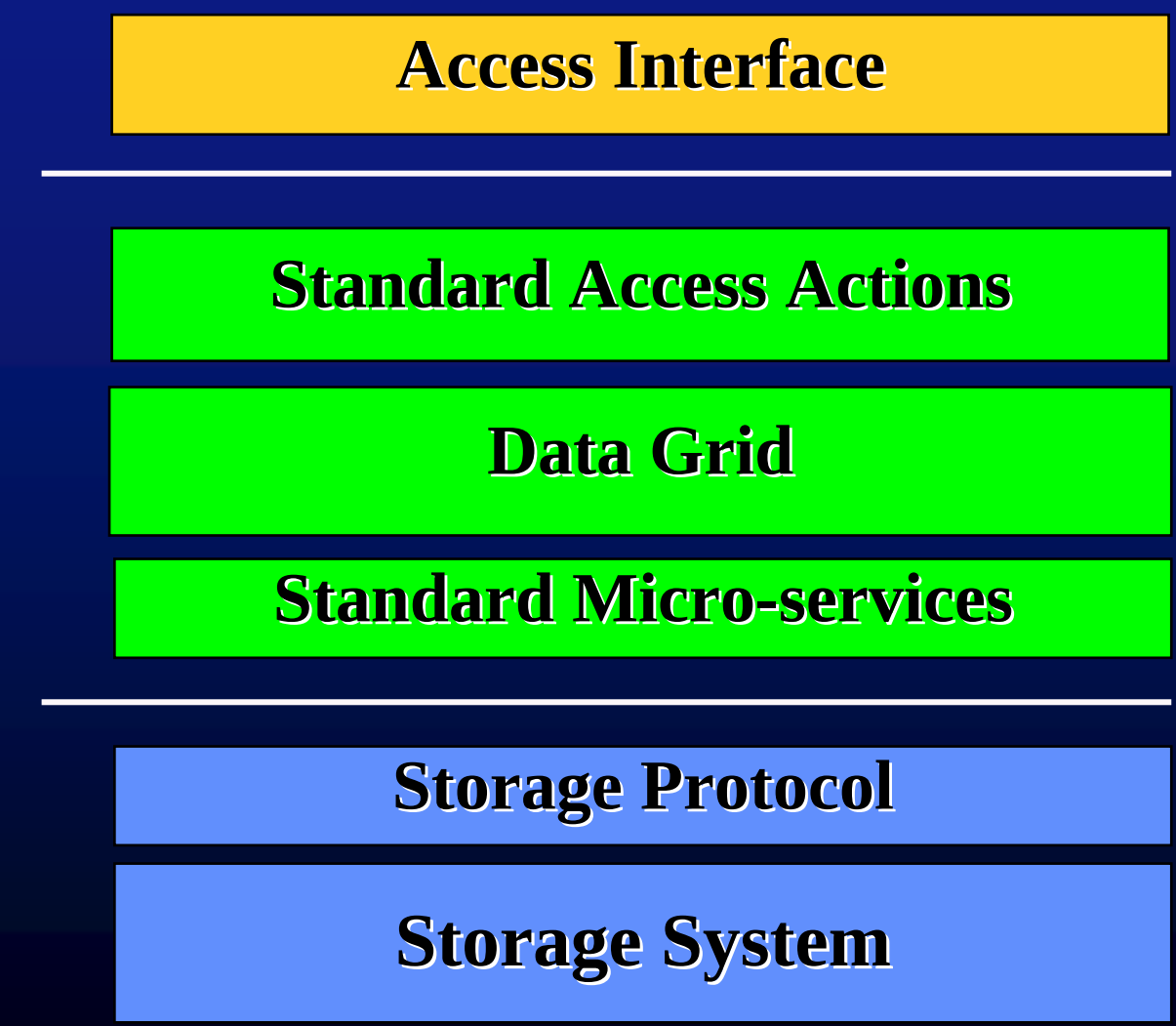
Infrastructure Independence

- **Distributed Data Management**
  - Data virtualization
    - Storage protocol independence
  - Trust virtualization
    - Administrative domain independence
  - Federation
    - Manage interactions between independent data grids
- **Rule-based Data Management**
  - Management virtualization
    - Automating execution of management policies
    - Coupling management policies to assertions about data

UCSD

SDSC

Data Virtualization

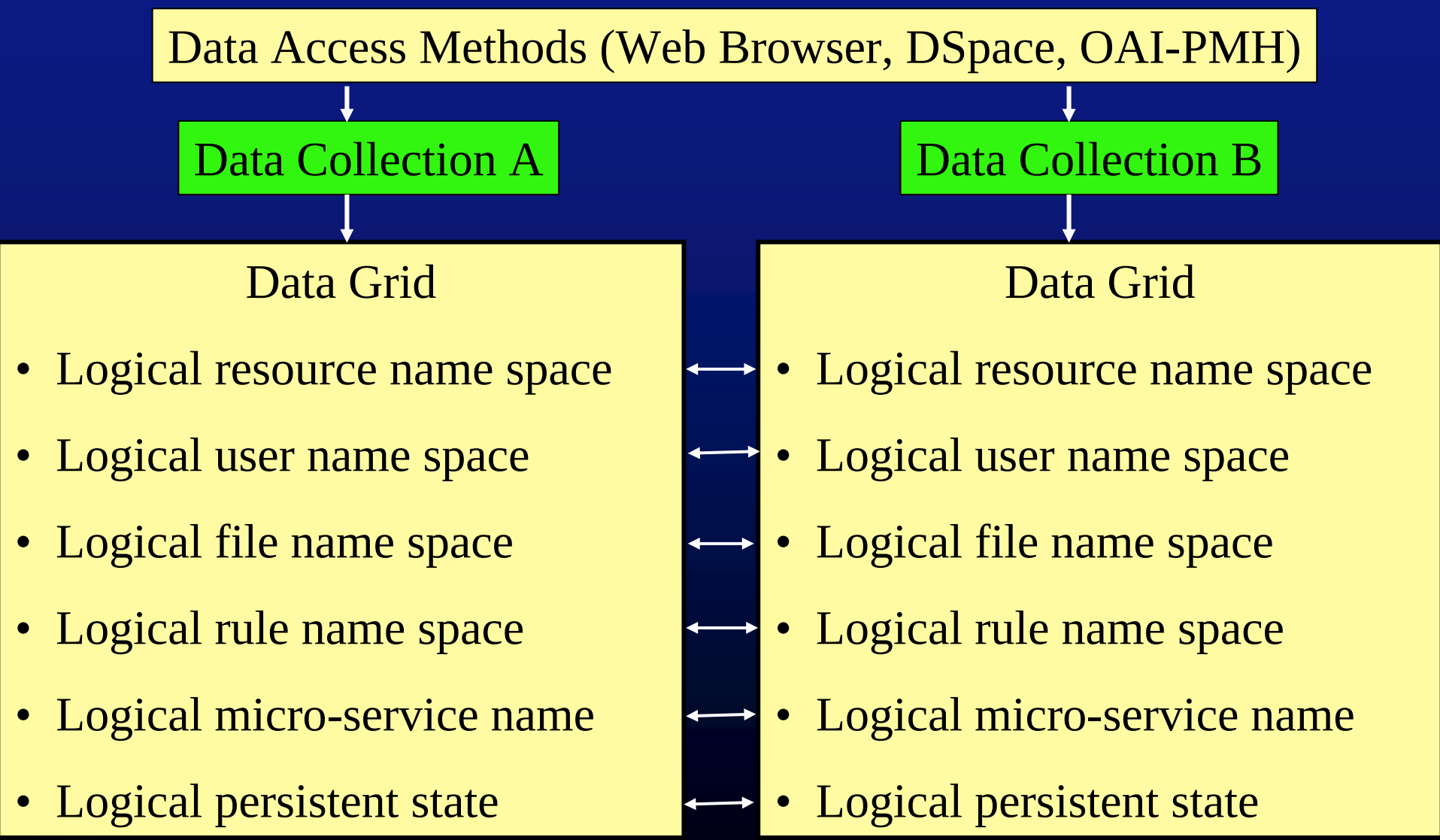


Map from the actions requested by the access method to a standard set of micro-services used to interact with the storage system

Micro-services

- **Examined Electronic Records Archive capabilities list**
  - Identified 174 micro-services for manipulation of data and structured information
  - Identified 212 metadata attributes (persistent state information) across six name spaces
    - Users
    - Files
    - Storage systems
    - Rules
    - Micro-services
    - Persistent state information

Federation Between Data Grids



Theory of Digital Preservation

- Definition of the persistent name spaces
- Definition of the operations that are performed upon the persistent name spaces
- Characterization of the changes to the persistent state information associated with each persistent name space that occur for each operation
- Characterization of the transformations that are made to the records for each operation
- Demonstration that the set of operations is complete, enabling the decomposition of every preservation process onto the operation set.
- Demonstration that the preservation management policies are complete, enabling the validation of all preservation assessment criteria.
- Demonstration that the persistent state information is complete, enabling the validation of assessment criteria.
- The assertion is then: if the operations are reversible, then a future preservation environment can recreate a record in its original form, maintain authenticity and integrity, support access, and display the record.
- A corollary is that such a system would allow records to be migrated between independent implementations of preservation environments, while maintaining authenticity and integrity.

SDSC

For More Information

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

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

UCSD

SDSC