Rule-Oriented Data Systems
A Grid-Based Cyberinfrastructure for Large Scale & Long-Term Data Management
Arcot (Raja) Rajasekar, Frank Vernon & Reagan Moore, UCSD

What is iRODS?
Data grid system—data virtualization
- A distributed file system, based on a client-server architecture
- Allows users to access files seamlessly across a distributed environment, based on their attributes rather than just the names or physical locations
- Replicates, syncs and archives data, connecting heterogeneous resources in a logical and abstracted manner

Distributed workflow system—policy virtualization
- Policy can be coded as functions (micro-services)
- Remote micro-services can be chained
- The chains (workflows) are interpreted at run-time
- Micro-services communicate through parameters, shared contexts, and out-of-band message queues

Data Virtualization with iRODS
Logical name space
Location independent identifier; Persistent identifier
Collection owned data
Access controls; Audit trails
Checksums; Descriptive metadata
Inter-realm authentication
Single-sign-on mechanism

Policy Virtualization with iRODS
MICRO-SERVICES: Functions with well-defined semantics; Transactional - recovery; Context of application; Message Queues
RULES: Triggers by events; Conditional execution of alternative rule declarations; System constructs: loops, recursion, branching
WORKFLOWS: Distributed Execution; Immediate; Deferred; Periodic

Automation of Management Tasks
Integrated Rule-Oriented Data System—iRODS
Express management policies as rules that control the execution of micro-services
- Micro-service is a standard operation performed on a remote storage system
Manage persistent state information that describes outcome of the micro-service
- Persistent Metadata catalog stores state information

Virtualize the management policies
- Logical name space for rules
- Logical name space for micro-services
- Logical name space for state information

iRODS System
Applications include
- Data grids for sharing data
- Digital libraries for publishing data
- Persistent archives for preserving data
- Real-time sensor data collections
- Large scale data analysis