**iRODS: integrated Rule-based Data System**

http://www.irods.org

The integrated Rule-Oriented Data System (iRODS) organizes distributed data into shared collections, while automating the application of management policies. Each policy is expressed as a set of rules that control the execution of a set of micro-services. Persistent state information is maintained to track the results of all operations. The result is a data grid that enforces community policies across administrative domains, while enabling the sharing of files through a common name space across multiple types of storage systems. The persistent state information can be queried to verify properties of the shared collection and validate assertions about trustworthiness, time dependent access controls, data distribution, data replication, data redaction, patient confidentiality constraints, retention and disposition, and any other desired management policy.

The iRODS system builds upon concepts from multiple communities; data grids (data and trust virtualization, and well defined data operations implemented as remote micro-services) workflows (the composition of micro-services executed as server-side workflows), database triggers (deferred and periodic rule execution and transactional operations), and logic programming (recursive rule expression and forward chaining).

Figure 1 defines the software layers in an iRODS data management server. The top layer provides standard access interfaces to clients and supports a robust protocol for communication between clients and remote servers and between peer-level servers. The bottom layer provides translation drivers for connecting to and communicating with storage systems such as file and tape archives, databases and object-based sensor streams. The middle layer is the intelligent software system that not only provides physical transparency, hiding the idiosyncrasies of the client and driver levels, but also provides the data management functionality for performing operations needed by data grids and long-term preservation systems. In the Storage Resource Broker data grid (SRB), the middle layer is a single software package providing hard-coded functions. In iRODS, this layer is split as shown, defining a standard set of actions that are composed from micro-services (and possibly other actions and rules). By differentiating the software system into distinct layers we provide flexibility in defining operational semantics. Each action is applied under the control of rules that can be adapted to meet the needs of the community. iRODS provides an adaptive framework that can be modified to meet the needs of each user community.

Rules are expressed as event:condition:action-sets:recovery-sets. The condition is implemented as an operation on any of the persistent state information attributes. Examples are rules that are specific to a user group, collection, data type, or storage system, or that depend upon a prior event, or that are triggered by a time stamp. The action-sets contain both micro-services and rules. A recursive rule hierarchy can be defined, as long as an exit condition guarantees closure on the recursion.

Standard micro-services include support for queries on the metadata persistent state repository shown in Figure 2 and for looping over results sets. Micro-services can be chained, with output parameters from one micro-service used as input for a subsequent micro-service. The rules are managed in a rule base and executed in the distributed rule engine. Since a rule may invoke multiple micro-services, the current state is maintained in a white board. On successful completion, the state information is registered into the persistent state repository. Transaction semantics are maintained through the invocation of recovery procedures on the failure of a micro-service.

![Figure 1. iRODS Software Layers](image)

![Figure 2. iRODS Client-Server Architecture](image)

The iRODS system is being developed as an international collaboration. Collaborators include institutions in Europe, the UK, the US, the Far East, and Australia. The first production release is scheduled for fall quarter 2007, and will be available at the wiki [http://irods.sdsc.edu](http://irods.sdsc.edu) as an open source software release distributed under a BSD license. Applications of the iRODS technology range from Ocean Observatory Infrastructure to Science of Learning Centers data grids.

This project is supported by the National Archives and Records Administration under NSF cooperative agreement 0523307 through a supplement to SCI 0438741, “Cyberinfrastructure; From Vision to Reality” and by the National Science Foundation grants ITR 0427196, “Constraint-based Knowledge Systems for Grids, Digital Libraries, and Persistent Archives” and SDCI 0721400, “SDCI Data Improvement: Data Grids for Community Driven Applications”.