

Minutes
MAGIC Meeting
April 1, 2009, 2:00-4:00
NSF, Room 1160

Attendance:

Doug Baggett	NSF	dbaggett@nsf.gov
Bob Bohn	NCO	bohn@nitrd.gov
Bob Chadduck	NARA	robert.chadduck@nara.gov
Frederica Darema	NSF	fdarema@nsf.gov
Ken Klingenstein	Internet2	kjk@internet2.edu
Ravi Madduri	ANL	Madduri@mcs.anl.gov
David Martin	IBM	martinde@us.ibm.com
Don Middleton	NCAR	don@ucar.edu
Grant Miller	NCO	miller@nitrd.gov
Reagan Moore	Un North Carolina	rwmoore@renci.org
Mike Nelson	GU	mnelson@pobox.com
Don Petravick	DOE	Don.Petravick@science.doe.gov
Alan Stone	DOE	
Susan Turnbull	DOE/SC	susan.turnbull@ascr.doe.gov
Liming Yang	NIH/NCRR	lyang@mail.nih.gov

Action Items

Proceedings

This meeting of MAGIC was chaired by Susan Turnbull of DOE

iRODS: Reagan Moore

Reagan Moore gave a presented on Transcontinental Persistent Archive Prototype. Data Grids organize distributed data into shared collection to provide virtualization of retention, disposition distribution, replication, integrity, authenticity, chain of custody, access controls, provenance, representation information, description information and logical arrangement. The Grid provides uniform interfaces to multiple heterogeneous storage systems and provides scalable mechanisms for data transport. The data grids provide:

- Extraction of records from the environment where they were created
- Import of records into a persistent archive
- Management of the properties of the collection independent of the choice of technology and archival form
- For evolution/migration to new archival technology
- Preservation of policies and procedures under which data was archived in the past
- Periodic evaluation/validation of the trustworthiness of the data.

Challenges for archives include:

- Massive amounts of data
- Multiple, remote data sources
- Wide range of data types
- Collections contain up to millions of files
- Metadata is required for data discovery
- Long-term retention requires migration to a reference collection

Then data management requires

- Logical arrangement
- Metadata
- Policies for managing the data
- Processes for manipulating the data
- Access mechanisms
- Workflows for data analysis and processing

With the iRODS data system the user can search, access, add and manage data and metadata. iRODS has a data server, a rules engine, and a metadata catalog. The iRODS policy-based data management turns management policies into computer actionable rules. Management processes are turned into remotely executable computer procedures. It provides procedural workflows at the storage system to filter, subset, and manipulate data. This minimizes the amount of data sent over the network.

Policies administered provide for retention, disposition, distribution, replication, deletion, registration and other data characteristics. Policies for ingestion and access to data implement metadata extraction, logical organization, derived data product generation and other characteristics. Validation policies provide for authentication, chain of custody, repository trustworthiness, and audit trails.

iRODS provides a unified view of the distributed data archival system across sites and databases. When a user asks for data the request goes to iRODS server 1 which looks up information in the catalog which identifies the iRODS server holding the data. The iRODS server 1 asks for the data from iRODS server 2 and the iRODS server 2 applies the rules.

The integrated Rules-Oriented Data System (iRODS) provides a client interface and an administrative interface to implement the Grid data archive services. It supports up to hundreds of millions of files, petabytes of data, hundreds of metadata attributes, hundreds of collaborators, hundreds of policies, federations of tens of data grids, and tens of thousands of users.

Applications using iRODS include:

- Carolina digital repository at UNC
- Duke medical archive
- RENCi data grid in North Carolina
- NARA Transcontinental Persistent Archive Prototype
- NSF Temporal Dynamics of Learning Center data grid
- NASA Center for Computational Sciences archive
- JPL planetary Data Systems data grid
- International data grids

Challenges are:

- Building a consensus on management policies for a shared collection
- Translating service level agreements into computer actionable rules
- Defining federation policies for sharing data between data grids and institutions.

For further information on iRODS, please see:

<http://irods.diceresearch.org>

Meetings of Interest

April 6-8 at NIST, Interoperability conference focused on metadata.

April 14-16 NIST conference on federated ID sponsored by OASIS, NIST, and Internet2

April 14-16 ESnet workshop to identify user requirements, particularly for the OSCARS environment

April 30, 9:00 AM iRODS Developments briefing to NSF, CISE, at the NSF, Room 110

May 25: OGF Chapel Hill, NC on Open Grid Environments

Next MAGIC Meetings

May 6, 2:00-3:30, NSF, Room 1150

June 3, 1, 2:00-3:30, NSF, Room 1150