

**Online data storage using
*iRODS*TM
(Integrated Rule-Oriented Data
System).**

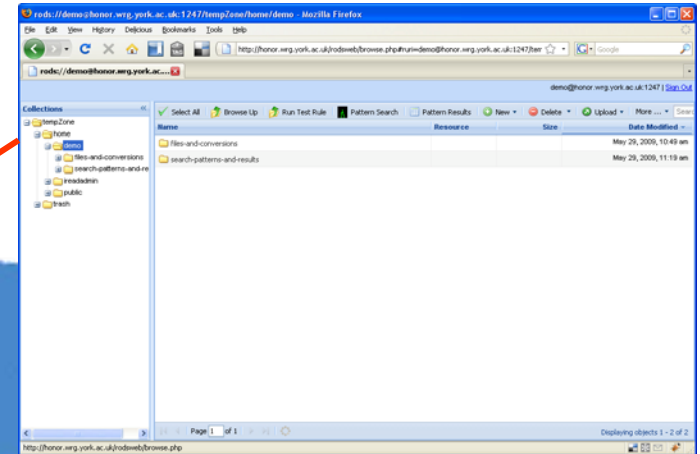
- What is iRODS?
- What is iRODS System?
- What does iRODS provide?
- iREAD: a project to assess iRODS.
- iREAD conclusions / experience.
- Potential future work.
- The iREAD project documentation and demonstrations.

■ iRODS™ (Integrated Rule-Oriented Data System)

<https://www.irods.org/> :

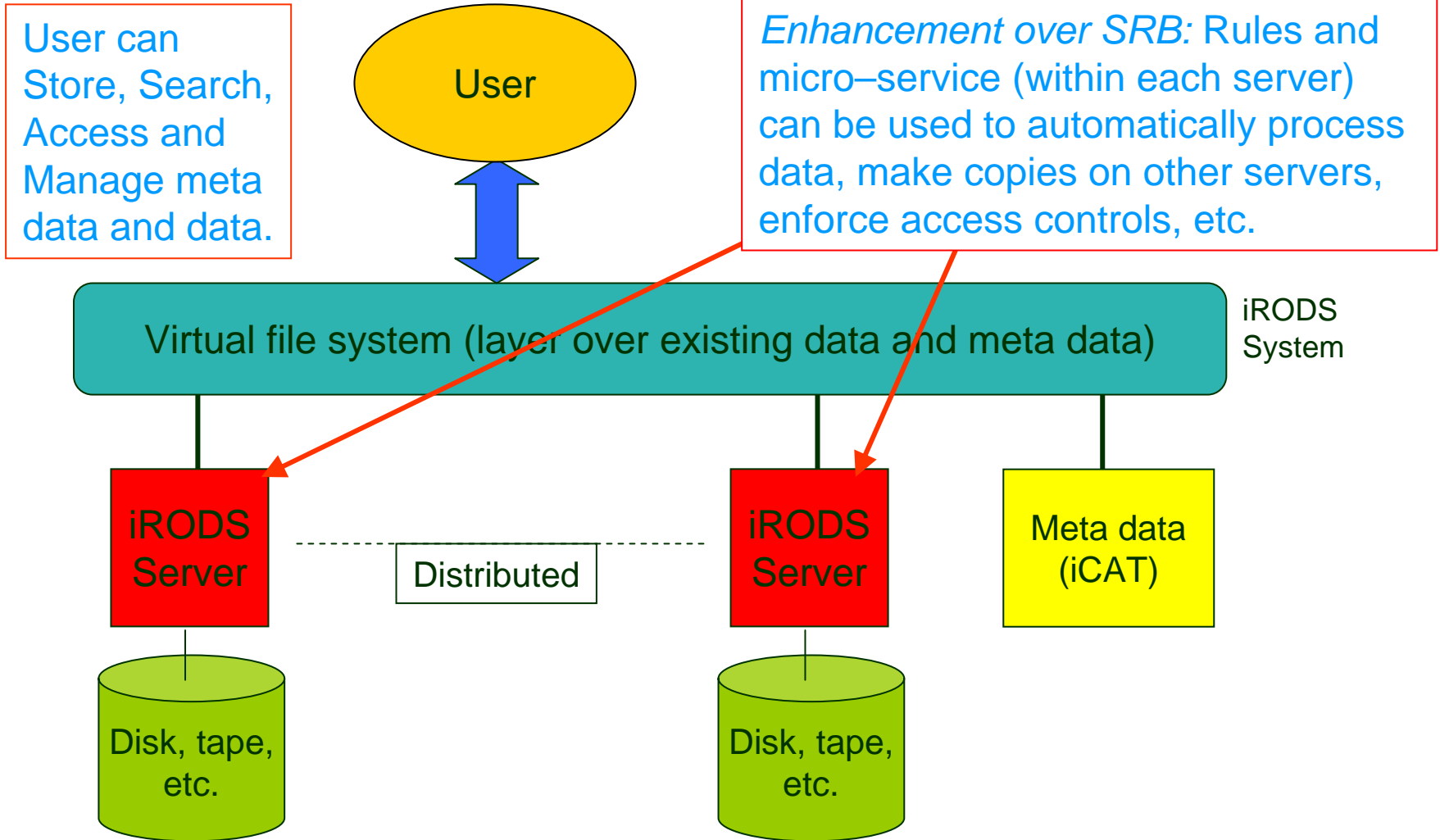
- Development of Storage Request Broker (SRB) by the Data Intensive Cyber Environments (DICE)
- Virtual file system.
- Supports Data Grids, Digital Libraries, Persistent Archives, etc.
- Main addition (cf SRB):
 - Rules Engine executes rules: to decide how the system is to respond to various requests and conditions, apply management policies, etc.
 - Rule execute other rules and micro services.
- Eventually iRODS is expected to replace SRB.
- Open source under a BSD license.

System of distributed iRODS nodes (servers).



Users can interact with all repositories through a single interface: command line, desktop application or browser.

- One or more interconnected resources:
 - Host machine each running an iRODS server.
 - Providing storage - in a “Vault” area within it’s file system.
 - Provides for the execution of operations.
- Manage data files.
- Manages data about the files (meta data) stored in a catalog known as an iCAT (iRODS CATalog):
 - Managed by a Postgres server:
 - Contains all metadata for everything that is managed locally including Domains, Resources and Metadata of Objects, list of administrators within the zone.



- iRODS (version 2.0.1.) can use a Unix file system or a Windows file system as a resource.
- A user:
 - has access to all resources and can usually direct files to be stored on any resource.
 - can operate on remote or local data on different types of resources through a common interface.
- Resources and clients distributed across the Internet.
- So far only one system considered - can interconnect iRODS separate systems (zones) i.e. federate iRODS.

- Storage and sharing of a file system of distributed repositories.
- Users can upload and share data with other users.
- Users can view all data through a single interface.
- Authentication (encrypted password, GSI, Kerberos) and Authorisation (ichmod, groups).
- Operations can be executed on data at the data location:
 - Automatically.
 - On request.

- Operations are carried out by:
 - rules:
 - contain other rules and micro services.
 - executed under defined conditions (automatically on events, time etc.) or on command from the user through the programmatic interface or command line.
 - micro services (inbuilt and user created):
 - Actually perform the operations.
 - Only executed as part of a rule.
- Rules and micro services are flexible and may be updated in for domain-specific needs. For example, can carry out:
 - selection of a target repository.
 - identification of replica repositories.
 - check user for fine grained authorisation.
 - post-processing for derived products.
 - Etc.

- User clients available:
 - Command line (Linux and Windows).
 - iRODS Explorer (Windows only - similar to Windows Explorer)
 - iRODS Web Browser (any platform).
- Programmatic clients:
 - JARGON – a Java API.
 - PRODS - a PHP API - used by the browser client. Use in iREAD to run the search rules from the browser.
 - Recent additions:
 - Python.
 - Perl.

- The iREAD (iRODS Evaluation and Demonstration) project at York carried out an evaluation of iRODS:
 - Produced a demonstration of the iRODS on the White Rose Grid – publicly available.
 - Evaluated iRODS system mainly in respect of iRODS rules and micro services.
 - Reviewed iRODS authentication and authorisation.
- Work funded by JISC:
<http://www.jisc.ac.uk/whatwedo/programmes/einfrastructure/iread.aspx>

- During the course of the iREAD project three demonstrations were produced:
 - Automated data conversion – potentially for the CARMEN (Code Analysis, Repository and Modelling for e-Neuroscience www.carmen.org.uk) project.
 - Use of web external services (operations are web services).
 - Use of signal search service.

- iRODS provides a virtual file system for sharing (selective sharing) of data, etc. and is being used.
- iRODS can be used to perform:
 - automated / manual operations on data e.g. iREAD demonstration of data conversion.
 - execution of external web services e.g. iREAD demonstration of pattern searching.
- Disadvantages:
 - Micro services (and rules) need to be installed on every server where they need to be executed.
 - Rules need to be managed carefully otherwise they can become very long and unwieldy.
 - Currently, changes / additions to micro service (not rules) require a recompile and restart of the iRODS system.

- Provide automatic:
 - Generation of the search databases (as new data arrives).
 - Search for known conditions in bulk data.
- Wide range of potential applications:
 - Many engineering disciplines.
 - Health care (particularly if image recognition micro services were also utilized).
- Utilise iRODS to perform automatic data conversion within the CARMEN project.
- Perform testing on the scalability of larger iRODS system and federation.

- All documentation, results and demonstrations are available on the public website see:
<http://www.wrg.york.ac.uk/iread>
- A paper has been submitted to the All Hands Conference 2009.
- Demonstrations and questions?