



## JISC Final Report

### Title Page: iREAD e-Infrastructure Project

Project Information			
<b>Project Acronym</b>	iREAD		
<b>Project Title</b>	iRODS Evaluation and Demonstrator		
<b>Start Date</b>	1 <sup>st</sup> March 2008	<b>End Date</b>	31 <sup>st</sup> May 2009
<b>Lead Institution</b>	University of York		
<b>Project Director</b>	Professor Jim Austin		
<b>Project Manager &amp; contact details</b>	Dr Tom Jackson, Department of Computer Science, University of York, York YO10 5DD, <a href="mailto:tom.jackson@cs.york.ac.uk">tom.jackson@cs.york.ac.uk</a> tel 01904 567711		
<b>Partner Institutions</b>	None		
<b>Project Web URL</b>	<a href="http://www.wrg.york.ac.uk/iread">www.wrg.york.ac.uk/iread</a>		
<b>Programme Name (and number)</b>			
<b>Programme Manager</b>	James Farnhill/Chris Brown		

Document Name			
<b>Document Title</b>	iREAD Final Report		
<b>Reporting Period</b>			
<b>Author(s) &amp; project role</b>	Martyn Fletcher, Tom Jackson. Project Manager		
<b>Date</b>	03.07.09	<b>Filename</b>	iREAD.MR.09.001
<b>URL</b>	<a href="http://www.wrg.york.ac.uk/iread">www.wrg.york.ac.uk/iread</a>		
<b>Access</b>	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
1.a	03.07.09	Initial Draft, for JISC evaluation

## Table of Contents

Title Page: iREAD e-Infrastructure Project .....	1
Table of Contents.....	2
Acknowledgements .....	2
Executive Summary .....	3
Background.....	4
Aims and Objectives .....	4
Methodology .....	5
Implementation .....	5
Outputs and Results .....	5
Outcomes.....	6
Conclusions .....	6
Implications .....	6
References.....	6
Appendixes (optional) .....	6

## Acknowledgements

The iREAD project was funded under the JISC E-Infrastructure Programme.

## Executive Summary

The iREAD (iRODS Evaluation and Demonstration) project was an in-depth evaluation and demonstration of the iRODS system<sup>1</sup>, assessing its capabilities and role within a distributed data management scheme and complex virtual organisation driven by real-world requirements from the CARMEN<sup>2</sup> e-Science project.

The project was a 15month, in-depth evaluation to investigate and demonstrate the capabilities of iRODS, within the context of the UK e-Science activities and initiatives.

The project addressed this evaluation through four main work-packages:

1. A demonstration implementation of the iRODS system data management system deployed for open access and evaluation via the White Rose Grid (WRG) and NGS system, which will demonstrate controlled access to distributed data across a collaborative VO environment;
2. An evaluation of the iRODS system for fine-grained role-based access in a real-world Grid based virtual organisation scenario, driven by the use-case requirements from an active national large-scale e-Science project, CARMEN;
3. An Analysis of the capabilities of iRODS for integration with existing e-Science SOA infrastructure and middleware systems, notably the York PMC middleware stack for distributed data-mining;
4. Wide-scale dissemination and technology transfer of the project results via national initiatives including the National Grid Service, White Rose Grid e-Science centre, the OMII and WUN.

iREAD has completed a thorough investigation of iRODS with a view to assessing its maturity and applicability for integration within the UK e-Science programme. Further details of this are included in Appendix A.

It has been shown that iRODS has many additional features and capabilities over and above SRB that make its adoption beneficial within UK e-Science projects. Rules and micro-services have been shown to be both flexible and powerful methods for providing complex management of data across diverse requirements in VO's or VRE's.

The bulk of these findings will be carried forward into the CARMEN project, and will also lead to longer-term deployment across the WRG

---

<sup>1</sup> See [http://irods.sdsc.edu/index.php/Main\\_Page](http://irods.sdsc.edu/index.php/Main_Page)

<sup>2</sup> See <http://www.carmen.org.uk>

## Background

The iREAD (iRODS Evaluation and Demonstration) project was an in-depth evaluation and demonstration of the iRODS system<sup>3</sup>, assessing its capabilities and role within a distributed data management scheme and complex virtual organisation driven by real-world requirements from the CARMEN<sup>4</sup> e-Science project.

Distributed data management issues are still a major challenge in addressing the ubiquity and usability of Grid systems, particularly within the context of complex virtual organisations (VO) where flexible security models and fine-grained role based access are a pre-requisite. The Storage Request Broker (SRB) software from SDSC has established itself as one of the leading Grid middleware applications to support the management of highly distributed large scale datasets for science applications. SRB provides the capability to virtualise distributed datasets, and to provide standardised access to a broad range of underlying storage technologies, spanning flat file systems through to database servers and tape archiving systems. Through the use of SRB, end-users are freed from concerns about the location of data and determining the correct procedures to recall or transfer data to their local or host compute environment. SRB abstracts these challenging aspects of distributed data management away from the end-user, and provides a simplified and uniform way to recall data via indexing systems (metacatalogs) which keep a logical mapping of the underlying distributed data. SRB has been widely adopted within large-scale Grid applications, particularly in the science communities, and provides the data management backbone for the National Grid Service (NGS).

However, SRB does have limitations particularly in regard to integration with complex VO's. These have been recognised by the developers of SRB at SDSC and a new version, iRODS, has been released to address the short-comings within SRB. Many of the limitations within SRB relate to the metadata schemes used to describe and annotate the data collections. In many cases these metadata mechanisms are too restrictive and inflexible to support complex meta-data schemes for complex role-based Grid systems. iRODS is now in general release, and as part of the objectives of the JISC e-infrastructure initiative, this project has assessed the impact that the new features and functions within iRODS will have on the UK e-Science community, and assess its potential for deployment within the NGS.

## Aims and Objectives

The project was a 15month, in-depth evaluation to investigate and demonstrate the capabilities of iRODS, within the context of the UK e-Science activities and initiatives.

The project addressed this evaluation through four main work-packages:

1. A demonstration implementation of the iRODS system data management system deployed for open access and evaluation via the White Rose Grid (WRG) and NGS system, which will demonstrate controlled access to distributed data across a collaborative VO environment;
2. An evaluation of the iRODS system for fine-grained role-based access in a real-world Grid based virtual organisation scenario, driven by the use-case requirements from an active national large-scale e-Science project, CARMEN;
3. An Analysis of the capabilities of iRODS for integration with existing e-Science SOA infrastructure and middleware systems, notably the York PMC middleware stack for distributed data-mining;
4. Wide-scale dissemination and technology transfer of the project results via national initiatives including the National Grid Service, White Rose Grid e-Science centre, the OMII and WUN.

A fifth workpackage, to investigate the interaction of iRODS within security infrastructures, such as Kerberos and Grid-Shib, was scaled down due to work that was being funded on a sister E-Infrastructure project, ASPIS, (see <http://www.kcl.ac.uk/iss/cerch/projects/portfolio/aspis.html>).

The White Rose Grid (WRG) computing infrastructure, operating since 2002 between the Universities of Leeds, Sheffield and York will serve as the basis for the investigations. The WRG will a well

---

<sup>3</sup> See [http://irods.sdsc.edu/index.php/Main\\_Page](http://irods.sdsc.edu/index.php/Main_Page)

<sup>4</sup> See <http://www.carmen.org.uk>

established Grid infrastructure has a large user base and supports a broad range of science and research software applications. In particular, the WRG hosts one node of the CARMEN e-Science infrastructure, which provides a data archiving and collaborative working environment for neuroscience research. The CARMEN system is being actively developed, already deploys SRB and has complex security and data management requirements, hence it provided an ideal case study environment to assess the capabilities of iRODS within a wide user community.

## Methodology

The project has focussed on evaluation of iRODS via the development of two interactive demonstrators of the iRODS system, deployed on the White Rose Grid.

This approach has permitted a detailed assessment of the iRODS technology, requiring a full deployment and development of rules and micro-services. It has also provided a mechanism whereby other interested developers can interact with iRODS and find working examples of both rules and services deployed.

A website has been developed ([www.wrg.york.ac.uk/iread](http://www.wrg.york.ac.uk/iread)) which details the development and deployment methods for the demonstrators. End-users can hence track back from the interactive demo's into the details of the development process.

The study was driven by real-world requirements coming from the CARMEN e-Science project. CARMEN was chosen as it is highly representative of the types of complex interactions that can be carried out in Grid based VO's and Virtual Research Environments (VRE's). CARMEN requirements also provided the opportunity to assess iRODS interoperability against existing web-service standards, and with existing SRB infrastructures.

## Implementation

The iREAD demonstrators were built using a mix of the iRODS browser tool and the extrods extensions with PHP and Java script extensions, hosted within a typical Apache/AXIS host environment.

This provided an opportunity to assess and detail the iRODS installation and maintenance process.

To assess integration with standard web-services, example services from the CARMEN project, which are hosted on the WRG, were used as trial remote services, which were instantiated as remote execution calls via micro-services and rules within iRODS. Full details of this are included in the final technical report, attached as Appendix A.

A number of deployment issues arose within the scope of the evaluation; largely relating to the fact that iRODS is in early release and still under development. A number of bugs or incomplete features were found. These were generally addressed through collaboration with the SDSC development team, via email or iRODS developer's forums. Bug fixes that were found will be carried through into later versions of iRODS. Intermediary solutions for issues (or gaps in documentation) were typically found via the dialogue with the SDSC development team.

## Outputs and Results

The iREAD project has provided a comprehensive evaluation of the new functions and services within the iRODS middleware, from the perspective of integrating it within a typical large scale e-Science project or Grid based VRE.

To make these results accessible to the wider community, interactive demonstrators have been built which provide a detailed means to assess how rules and services can benefit end-users or system's developers. As such, the project has taken a very user or sys-admin focused approach to the evaluation, and has not looked at issues relating to performance and scalability of iRODS (these were considered outside the scope of the evaluation, and to some extent these criteria have already been established through the deployment of SRB systems, the precursor to iRODS).

A by product of the evaluation is that the project has produced a very detailed user guide on installing and deploying iRODS, particularly from the perspective of building rules and services. Although iRODS is still developing, this documentation is likely to be of significant benefit to any users or system-administrators coming new to iRODS.

## Outcomes

iREAD has completed a thorough investigation of iRODS with a view to assessing its maturity and applicability for integration within the UK e-Science programme. Further details of this are included in Appendix A.

It has been shown that iRODS has many additional features and capabilities over and above SRB that make its adoption beneficial within UK e-Science projects. Rules and micro-services have been shown to be both flexible and powerful methods for providing complex management of data across diverse requirements in VO's or VRE's.

The bulk of these findings will be carried forward into the CARMEN project, and will also lead to longer-term deployment across the WRG.

## Conclusions

The iREAD project has completed a detailed evaluation of iRODS from a users and sys-admin perspective, and the assessment has demonstrated that iRODS is a significant enhancement on the already well established functionality offered by SRB. It has largely been shown that iRODS is maturing rapidly, and is currently at a release stage where it could usefully and beneficially be deployed within the UK e-Science framework.

The use of iRODS rules and micro services has been evaluated and demonstrations are provided which perform automated data conversion and executing of external web service examples including pattern searching. The demonstration of rules and micro services have been provided using the iRODS web browser, which has been modified for this purpose, however, the rules and micro services can be executed automatically or as required without the browser. Integration of iRODS into a system such as CARMEN would be beneficial for data conversion and other automatic operations.

iRODS could be used as the underlying system in a distributed architecture because rules and micro services can be run remotely on other servers (on other resources in the same zone and in other zones) using the *remoteExec* micro service. These could also be used to run web service as necessary. In future, parallel execution and broadcast execution modes may also be provided by iRODS, these would greatly assist the use of iRODS for distributed operations.

## Implications

The most relevant implication of the work for a JISC/UK perspective is a review of the iREAD project for the NGS.

## References

The demonstrators and documentation for the iREAD project can all be found at [www.wrg.york.ac.uk/iread](http://www.wrg.york.ac.uk/iread)

## Appendixes (optional)

The final technical report for iREAD is attached as Appendix A