ASPiS - PROJECT PLAN

Project

Project Acronym	ASPiS	**Project ID**	
Project Title	Architecture for a Shibboleth-Protected iRODS System		
Start Date	01/03/2008[1]	**End Date**	31/05/2009
Lead Institution	Centre for e-Research, King's College London		
Project Director	Dr. Mark Hedges		
Project Manager & contact details	Dr. Mark Hedges Deputy Director Centre for e-Research King's College London 26 - 29 Drury Lane LONDON, WC2B 5RL Tel: 0207 848 1970 Fax: 0207 848 1989 Email: mark.hedges at kcl.ac.uk		
Partner Institutions	Science and Technology Facilities Council		
Project Web URL	TBD		
Programme Name (and number)	e-Infrastructure (Federated Tools and Services)		
Programme Manager	James Farnhill		

Document

Document Title	Project Plan		
Reporting Period	n/a		
Author(s) & project role	Dr. Mark Hedges, Project Manager		
Date	21/02/2008	**Filename**	Project Plan – ASPiS – 1b.doc
URL	n/a		
Access	☐ ✓ Project and JISC internal	☐ General dissemination	

Document History

Version	Date	Comments
1a	11/02/2008	Initial draft
1b	21/02/2008	Updated with comments from JISC

[1] The project was due to start on 1st Jan. 2008 and run for 15 months; however, the funding decision for the call was delayed by JISC for approximately 2 months. Consequently, we have assumed a 2 month delay in the project start.

Project Acronym: ASPiS
Version: 1b
Contact: Mark Hedges
Date: 21 February 2008

JISC Project Plan

Overview of Project

1. Background

Research across the academic disciplines is increasingly both driven by and a generator of data on a large scale, the so-called data deluge. In the “big” sciences, the primary emphasis has been the management of very large data sets, but the issue frequently arises in other disciplines as well, as it is increasingly necessary to retain both the original data and processed versions of the data. These requirements arise partly from the need to prove that analysis was carried out correctly, partly because computations can be expensive and time consuming so cannot easily be repeated, and partly to enable future analysis of intermediate datasets. These complex workflows are often supported by data grid middleware, for example the widely used Storage Resource Broker[2] (SRB). STFC support numerous production SRB systems on behalf of researchers in a variety of disciplines.

iRODS[3] is an open source project developed by SDSC as the successor to SRB, with significantly enhanced functionality. A particular feature of iRODS is the ability to represent data management policies in terms of rules. The system incorporates a Rule Engine, which interprets these rules and allows pre-defined sequences of actions (represented as "micro-services") to be executed in response to certain triggers, e.g. when an data object is stored, read, or updated. These rules have great potential for implementing data management strategies that are to take place "under the hood".

With its Rule Engine, iRODS addresses a particular limitation of the SRB, by providing a simple, flexible and integrated means of implementing specific application requirements that are not supported by the core software, and in particular providing a mechanism for accessing external systems. These could include, for example, access management systems, or systems that support more complex metadata than is provided natively by SRB or iRODS. This enables integrated iRODS-based systems that incorporate the best of both (or several) worlds, combining the virtualised storage of the data grid with external services that encapsulating specialised processing and knowledge.

This is of crucial importance when we consider that in many disciplines the “data deluge” concerns not only the size of the data, but also its complexity and diversity, which is reflected in the complexity of the metadata required to manage both the data and access to the data. For example, a humanities researcher may work with digital resources that combine textual objects (enriched with various degrees of XML mark-up), databases and multi-media objects; a medical researcher may use large images with complex, detailed annotations and links to other resources.

When making digital resources more widely available, access management issues are of key importance, particularly in a highly disaggregated grid environment where access may depend on dynamic and ad hoc arrangements between users, institutions and virtual organisations. Much of the research community is simply not going to engage with grids and make resources available in this way, if security issues are not addressed to their satisfaction.

SRB addresses user authentication by means of X.509 certificates or centrally managed user lists and passwords, in combination with access control lists and groups to handle authorisation of access to individual grid resources. Natively, iRODS follows a similar approach. However, identity-based authorisation does not scale well, and does not easily support role-based access in dynamic environments involving virtual organisations working on complex data. In addition, the technical complexities of setting up and using certificates will not encourage a wide uptake of grid use among users who are less likely to be technically knowledgeable, e.g. in arts and humanities disciplines.

2. Aims and Objectives

We propose to address two complementary aspects of access management for virtualised data:

[2] http://www.sdsc.edu/srb
[3] http://irods.sdsc.edu

- access control that is based on user roles (including, e.g., VO membership) allowing access rights to be defined for individual files, modes of access (e.g. get, put, update, copy), and user roles (not just the users' identities).
- generation of audit and provenance information that tracks access to resources.

To address the first of these we plan to integrate Shibboleth with iRODS, enabling the authentication of a user to be devolved onto the user's home institution. Middleware will be implemented to capture a user's Shibboleth attributes (which can include a unique user identifier and multiple roles) and make them available within iRODS, allowing finely grained access control to data resources. We will investigate how these attributes/roles can be used to make authorisation decisions within the iRODS context. A major line of investigation will be the use of iRODS' rule architecture, whereby an attempt to access a resource results in the execution of a "micro-service", which could be used to determine access rights. We will also investigate the use of external systems such as PERMIS, an authorisation system implementing "Role Based Access Control" authorisation.

As well as using the Shibboleth attributes to control access to grid-based data, we intend to investigate the complementary use of these attributes in the generation and recording of provenance metadata. The concept of provenance is of key importance when assessing the quality and accuracy of information and services available in a grid. Typically, in a grid-based research environment many people can create data, and apply services and workflows to analyse and process this data to produce new results. The assessment of the validity or reliability of information will be dependent on the entire sequence of actions (and actors) that culminated in its current state[4]. Broadly, provenance metadata represents the series of steps by which a particular piece of data was derived, and can include such information as the ancestral data from which it was derived, workflows and services that were executed to derive it, the inputs and outputs of those workflows and services, as well as information about users who initiated them. Such information represents added value for e-Science users who subsequently make use of the data, whether by publishing, citation or further processing.

An important part of the provenance of a data object relates to the agent that produced the data. Some data may result from entirely automated workflows, but often a user is the ultimate cause. A user's Shibboleth attributes may contain several items of relevance for recording provenance: in terms of the eduPerson attributes recognised by JISC as providing the necessary core functionality for the UK academic community, these include *eduPersonPrincipalName* (a persistent identifier for the user[5]) and *eduPersonEntitlement* (indicating, e.g., membership of a number of research groups)[6]. We will implement a service (or services) to generate provenance information, incorporating Shibboleth attributes among other metadata, for data objects held in iRODS data grids, when they are created, read or modified. We realise of course that the UK Access Management Federation encourages IdPs to not publish personal data, so we will encourage our user communities not to require this in their provenance data, in the spirit of the federation policies. However, since some communities may genuinely require personal data to be stored in the provenance metadata, we will provide means to associate personal identifiers with the *eduPersonTargetedID*, with the users' consent. This process of association may also involve formal authorisation by the user's project PI for additional security and role assignment.

3. Overall Approach

The overall project strategy may be addressed in terms of three broad phases:

1) definition of user requirements;
2) software architecture and development;
3) evaluation.

1) User requirements

[4] http://www.cs.indiana.edu/~ysimmhan/l/pubs/simmhan-record-2005.pdf
[5] However, there are issues to be addressed with *eduPersonPrincipalName*: an institution will not in general publish this attribute, and also may choose to recycle the value after a user leaves.
[6] http://www.jisc.ac.uk/uploaded_documents/JISC_Fed_doc_full.doc

We will ensure that the implementation is grounded in authentic user requirements by producing a set of use cases in liaison with selected stakeholders in a range of disciplines. These use cases will address:

- Managing access to data, and in particular fine-grained, role-based authorisation in virtualised data grid environments;
- Generation, capture and use of audit/provenance metadata.

We emphasise the *use* of provenance metadata here, as surveys of data provenance in e-Science have indicated that the nature and format of the metadata recorded has been strongly dependent on the requirements for use[7], and diverse requirements have led to an equal diversity of representations. In order to integrate provenance metadata creation with iRODS in a way (or ways) useful to the community, we must determine how the metadata will be used, which may include browsing, citation, automated analysis and reasoning. We aim to develop a machine-interpretable representation that is sufficiently flexible to accommodate a diverse subset of the requirements and applications that may be expected in an iRODS data grid environment. It will also be extensible, to allow additional information to be built in as new requirements are encountered.

2) Architecture and development

See Section 9.

3) Evaluation

To ensure that the software meets the needs of the community as a whole, we will follow a user-driven, evolutionary approach, involving incremental cycles of implementation and assessment in collaboration with potential users and other stakeholders, and in particular the NGS, of which STFC are significant partners. This iterative approach implies that the development and testing/evaluation phases will overlap to a significant degree.

We aim to produce software that is usable in other iRODS systems, at least as a beta version, and consequently rigorous testing is of key importance. All formal testing will be fully documented: test specifications will be written; the tests will be executed and the results logged, together with any issues (such as test failures) that arise. Software will be updated and rebuilt to resolve these issues where possible. Given the incremental approach to development, testing may be repetitive, so we will create sets of automated test scenarios that can be executed against the software without human interaction, to validate the software after any significant changes have been made.

Testing will take place in several environments:

a) In localised test environments at STFC and CeRch.
b) In a test-bed iRODS data grid at CeRch and STFC/RAL (internal to the project).
c) In an prototype iRODS data grid, based at STFC/RAL and CeRch, which is accessible to identified external users for evaluation. The grid will be populated with initial test data, although users will be able to add their own (subject to space limits).

The aim is that a given “release” of the components will pass though these three test environments in succession, becoming increasingly robust as it does so. There may be several “releases” of the software being tested at any one time in different environments, so close attention will be paid to issue tracking and version control: all software changes will be tracked using the Subversion, and bugs/issues will be tracked using an appropriate tool, such as TRAC[8].

The prototype in (c) will be protected by a Shibboleth SP located at STFC/RAL and registered by STFC as an SP with the UK Access Management Federation for Education and Research, of which both STFC and KCL are members and have IdPs. The prototype will be made available as a test bed to NGS users that are affiliated to an institution within the UK federation. Currently, this covers approximately 75% of the UK-based NGS users, and this percentage will increase as more institutions join the federation, and more members set up IdPs.

To enable non-federation users to access the service, depending on the feasibility and the needs of the user communities, we will investigate setting up a non-federation IdP that is not coupled to an

[7] http://www.cs.indiana.edu/pub/techreports/TR618.pdf
[8] http://subversion.tigris.org/; http://trac.edgewall.org/

institution's authentication service, which we can use to manage the users. Naturally, such an IdP cannot join the UK federation. We will thus need to set up an independent WAYF that can redirect users both to the federation IdPs and to the test IdP. Alternatively (or additionally) the prototype iRODS service will also support the standard iRODS authentication and authorisation mechanisms for non-Shibboleth users, thus enabling federation and non-federation authentication to co-exist. We will use the Internet2 implementations of Shibboleth, and test both with Version 1.2, which is the version currently used by the UK Federation, and Version 1.3, which is the most recent group of releases.

4. Project Outputs

The main deliverables will be:

a) A set of use cases and requirements.
b) Modular software components and rules for iRODS (and possibly for Shibboleth), implementing the following functionality: capture of Shibboleth attributes, use of Shibboleth attributes for determining access to iRODS data resources, capture of audit/provenance metadata.
c) Documentation, including technical specifications, test documents and user guides.
d) A prototype Shibboleth-enabled iRODS data grid, based at STFC/RAL with additional storage at CeRch, and using the software from (b), available as a test bed for NGS and other users.
e) Final report, incorporating case study.

5. Project Outcomes

The project outputs will demonstrate to the research and UK higher education communities the utility of iRODS for data management, will simplify access management for iRODS data grids, and will enable data grids to be more easily integrated within the access management framework for UK HE being developed by JISC (via the UK Access Management Federation). As a consequence of this, we expect to see a greater understanding within UK HE of both iRODS and data grids in general, and an increased take up of data grid technologies among researchers.

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
Research projects and staff who generate, manage and consume research data in a data grid environment. In particular, current users of the SRB services provided by STFC, both within and outside the NGS.	Have interest in making the results of their research available, while both controlling access to these results (whether for reasons of sensitivity, privacy, commercial value) and tracking or monitoring access to the results and the uses that are subsequently made of them (e.g. in publications or further research).	HIGH
Research projects and staff who would do the above, if their requirements were better supported.	ditto	HIGH
Funding bodies (including research councils)	ditto, for the results of the research that they funded.	HIGH
JISC, especially the e-Infrastructure team and projects.	The project outputs may form part of a future e-infrastructure.	MEDIUM
The NGS (including the UK e-Science Engineering Task Force) and users of NGS services.	The project may result in a new service for the NGS.	MEDIUM
The iRODS development team at SDSC.	The project will provide an enhancement to the core software.	HIGH
Other grid and e-science bodies: NeSC, OMII-UK, AHeSSC, the Grid Technology Group at the STFC Daresbury Laboratory, the	All have an interest in advancing the use of grid technologies.	MEDIUM

Grid Tools Liaison Group.		
Preservation projects.	iRODS has significant potential for digital preservation	MEDIUM
PERMIS team	The project may make use of the PERMIS software.	MEDIUM
iREAD project (Contact: Tom Jackson)	iREAD is investigating the iRODS software.	MEDIUM

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing: inability to attract and retain staff with appropriate skills and experience	1	4	4	Most of the nominated staff are already employed at CeRch & STFC. There is a broad knowledge of the technologies involved within the partner organisations. Spread expertise throughout the project, document work so that knowledge is not lost.
Dependencies on other projects: the iRODS development project.	1	2	2	A production-quality release (1.0) of iRODS is now available. Expertise in iRODS is available at STFC and CeRch. STFC liaises closely with the SDSC iRODS team.
Technical: a complete solution cannot be implemented within the project time constraints using the technologies selected.	2	3	6	The absence of a complete solution is not an indication of failure, as one aspect of the project is to investigate potential problems. The project report will address the issues that could not be resolved.
Failure to meet project milestones.	2	3	6	Produce project plan with clear objectives. Continuous project assessment and rescheduling when necessary.
NGS does not have Shibboleth access when testing starts	2	2	4	NGS has deployed prototype Shibboleth portals, e.g. in the ShibGrid project. We can reuse prototypes from this project.
Lack of engagement from user communities	1	4	4	Identify relevant communities early. Both STFC & CeRch work closely with diverse communities and have existing SRB users.

8. Standards

Name of standard or specification	Version	Notes

SAML	1.1	Security Assertion Markup Language
RDF Specifications[9]	Latest	W3C Recommendations
OWL	1	W3C Recommendation

9. Technical Development

Our design approach will be to develop additional modules that can be incorporated into iRODS rules without changes to the core software, following the approach of iRODS that uses a standard "micro-service" interface. In some cases, changes to the core software may be necessary, e.g. when capturing and storing Shibboleth attributes; however in any case a modular approach will be taken and changes to the core will be minimised. A key design aim is to decouple the services from the iRODS architecture to enable different implementations of authorisation or provenance services to be used by different iRODS systems. For example, some systems may wish to use PERMIS for access authorisation, others a different service; different systems may require provenance metadata to be recorded in different ways. In each case the architecture will facilitate the substitution of one implementation for another by using a common interface layer.

Within this framework, we will address a number of possible implementation architectures, from which our preferred solution will be selected. At the current stage, it is of course not possible to describe this architecture in any detail. However, it is likely to include the following components[10]:

- a Policy Information Point (PIP), which captures and stores user attributes from Shibboleth.
- a Policy Decision Point (PDP), which processes each access request, receives user attributes from the PIP, applies the access control policy, and determines whether to allow or reject the request. To accommodate the use of third party authorisation modules, this component is likely to be divided into two parts, a component that interfaces to iRODS, and a component (which may for example be PERMIS-based) that makes the decision.
- a Policy Enforcement Point (PEP), which enforces the decision made by the PDP.
- components for generating provenance metadata. Again, like the PDP, we will develop at least two components, one acting as a generic interface, another generating the metadata, so as to accommodate future modules satisfying different metadata requirements.

10. Intellectual Property Rights

IPR in all reports and other documents produced by the project will be retained by the authors and host institutions but made freely available on a non-exclusive licence as required by JISC.

All software created during the project will be made available to the community on an open-source basis on the GPL licence. We will respect the licence model of all third party software incorporated– as far as possible, only software made available under open source licences will be used in this way.

Any non-open source third-party software to be incorporated will be included here at a later date, along with how we will obtain permission to use them, and any legal implications for the project outputs.

All modifications and enhancements to the iRODS software will be made available to the iRODS development team for inclusion in their core releases.

Project Resources

11. Project Partners

The project partners are:

- Centre for e-Research, King's College London (lead institution), contact: Mark Hedges;
- Science and Technology Facilities Council, contact: Adil Hasan.

A consortium agreement has not yet been signed. It is expected to be signed before the end of March.

[9] See http://www.w3.org/RDF/#specs.
[10] The terminology is explained in http://www.terena.nl/events/tnc2006/core/getfile.php?file_id=753.

12. Project Management

The project team is as follows:

- Mark Hedges: Project Director, 0.25 FTE, 15 months. Contact details: mark.hedges at kcl.ac.uk. Mark is the Deputy Director of CeRch. He will be responsible for project management; technical management; production of use cases; project advocacy.
- Adil Hasan: Senior Technical Officer, 0.1 FTE, 15 months. Contact details: a.hasan@rl.ac.uk. Adil will be responsible for contributing to the use cases and technical architecture, and for supervising the STFC Technical Officer. He will report to the Project Director.
- Roger Downing: Technical Officer, 0.6 FTE, 15 months.
- TBD: Technical Officer, 15 months. 1.0 FTE at CeRch/KCL.

The Technical Officers will be responsible for the technical/research review, software development and testing, production of technical reports and other documentation. These activities will be broadly divided into: (i) access management, (ii) access audit/provenance. However, this is unlikely to be an even split between the 2 partners, and a more granular division of work will be made as the project progresses. In any event, close collaboration between the CeRch and STFC staff will be necessary.

The project will draw on the expertise of other staff within KCL and STFC, who will provide ad hoc consultancy: Dr Jens Jensen (STFC) on Shibboleth, Dr Tobias Blanke (AHeSSC) on arts and humanities applications, and Dr Simon Miles (KCL) on provenance.

13. Programme Support

The project would benefit from programme-level workshops allowing knowledge and experiences to be shared among the projects (including previous related projects in the programme).

14. Budget

The budget is in Appendix A. There has been no change to the budget put forward in the project proposal.

Detailed Project Planning

15. Workpackages

See Appendix 2.

16. Evaluation Plan

To ensure that the project outputs meet the needs of the research community, we will follow a user-driven approach to development, involving incremental cycles of implementation and evaluation in collaboration with targeted evaluation partners. These will include:

- Projects/researchers using the SRB services provided by STFC, both within and outside the NGS.
- Other NGS users, possibly including the UK e-Science Engineering Task Force.
- The JISC e-Infrastructure team, and projects arising from the JISC e-Infrastructure Community Engagement and Support Programme (see Section 3.1).
- Grid Technology Group at the STFC Daresbury Laboratory
- In liaison with AHeSSC, current and potential data grid applications in the arts and humanities.
- The iRODS development team at SDSC.

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Evaluation phase	Utility of software to researchers.	Researchers can use the prototype to manage access to their data, and to capture	Feedback from evaluator communities (see above	Positive feedback received from evaluator communities.

		provenance metadata.	for targeted communities).	
Evaluation phase	Flexibility and extensibility.	Possibility of replacing existing modules with alternative implementations.	Examination of software.	Architecture allows modules to be replaced without excessive difficulty.

17. Quality Plan

See Section 16 for a list of the communities that will be targeted for user testing. The testing process will be iterative, involving cycles of implementation and evaluation in collaboration with these users.

Output	**Software**				
Timing	**Quality criteria**	**QA method(s)**	**Evidence of compliance**	**Quality responsibilities**	**Quality tools (if applicable)**
Ongoing	Fitness for purpose	Review of code and testing.	Test results and code review.	Project Manager (or delegated staff member) to verify.	Trac, Subversion
Ongoing	Adherence to standards and good practice	Review of code.	Code review.	Project Manager (or delegated staff member) to verify.	Trac, Subversion
Evaluation phase	Fitness for purpose	Review of testing and evaluation documentation.	Test specifications and test results.	Project Manager (or delegated staff member) to verify.	Trac, Subversion, automated test scripts

Output	**Documentation**				
Timing	**Quality criteria**	**QA method(s)**	**Evidence of compliance**	**Quality responsibilities**	**Quality tools (if applicable)**
Ongoing	Comprehensiveness; usability by intended audience.	Review of documentation.	Documentation and review outputs.	Project Manager (or delegated staff member) to verify.	MS Word spelling and grammar checkers.

18. Dissemination Plan

Timing	**Dissemination Activity**	**Audience**	**Purpose**	**Key Message**
Ongoing	Conference papers and posters	Academic community; researchers; digital repository/archive communities; e-science & grid communities.	To encourage awareness of project outputs	The utility of the enhanced iRODS for data & access management.
Ongoing	Project website	ditto	To provide regular updates about the	Information about the project

			project and its findings.	
Ongoing	JISCmail lists, RODS-Chat list	JISC community	To inform about project outputs and outcomes	Raise awareness of key outputs of the project.
Ongoing	RODS-Chat list	iRODS community	To inform about project outputs and outcomes	Raise awareness of key outputs of the project.

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
iRODS enhancements and rules	Make all software developed available to the wider community on an open-source basis, in accordance with the *Policy on Open Source Software for JISC Projects and Services*. Liaise with OSS Watch as regards other ways of making the source available.	Make all enhancements available to the iRODS development team, and at other locations as suggested by OSS Watch.
Prototype iRODS data grid	Encourage further use of prototype after project completion.	STFC (at RAL) and CeRch (at KCL) will maintain the prototype for at least 12 months after project completion. Encourage creation of production service within NGS.
Experience of project team with iRODS.	Will contribute to implementation of campus iRODS data grid at KCL, and potential production services within NGS.	Start implementation of campus data grid at KCL, and cross-institutional grids in collaboration with other organisations. Assist with creation of production service within NGS.

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
iRODS enhancements and rules	Useful for other institutions and projects using iRODS to manage data.	Make all enhancements available to the iRODS development team, and at other locations as suggested by OSS Watch. Roll out a production iRODS campus data grid at KCL. Encourage roll-out of production iRODS service within NGS.	

		Encourage take up of enhanced iRODS by other institutions.	
Technical reports.	Will document practical experience with iRODS and Shibboleth.	Make available on project website and via JISC.	

Appendixes

Appendix A. Project Budget

Directly Incurred Staff	**Year <07-08>**	**Year <08-09>**		**TOTAL £**
Total Directly Incurred Staff (A)	**£ 23,303.05**	**£ 94,231.99**		**£ 117,535.04**
Directly Incurred Non-Staff				
Travel and expenses	£ 3,000	£ 14,000	£	£ 17,000.00
Hardware/software (new servers at CeRch – not charged to project)	£ 12,000	£	£	£12,000.00
Dissemination (not including travel)	£	£ 1,000	£	£1,000.00
Evaluation	£	£	£	£
Other	£	£	£	£
Total Directly Incurred Non-Staff (B)	**£ 15,000**	**£ 15,000**	**£**	**£ 30,000.00**
Directly Incurred Total (A+B=C) (C)	**£38,303.05**	**£109,231.99**	**£**	**£147,535.04**
Directly Allocated				
Staff	£	£	£	£
Other	£	£	£	£
Directly Allocated Total (D)	**£5,111.26**	**£ 20,702.83**	**£**	**£ 25,814.09**
Indirect Costs				
(E) total	**£20,088**	**£81,240.30**		**£101,328.30**
Total Project Cost (C+D+E)	**£63,502.31**	**£211,175.12**	**£**	**274,677.43**
Amount Requested from JISC	**£34,923.5**	**£143,974.78**	**£**	**£178,898.28**
Institutional Contributions	**£28,578.81**	**67,200.34**	**£**	**£95,779.15**
Percentage Contributions over the life of the project	**JISC 65.13%**	**Partners 34.87 %**		**Total 100%**

Nature of Institutional Contributions

Directly Incurred Staff				
Directly Incurred Non Staff				
Hardware/Software etc.	£12,000	£0		£12,000
Directly Allocated				
Staff, Estates etc.	£248.76	£995.24		£1,244
Indirect Costs				
Indirect Costs	£14,637.60	£59,436.40		£74,074

Total Institutional Contributions	**£28,564.76**	**£67,145.12**		**£95,779.15**

Appendix B. Workpackages

WORKPACKAGES	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
	M 08	**A**	**M**	**J**	**J**	**A**	**S**	**O**	**N**	**D**	**J 09**	**F**	**M**	**A**	**M**					
1: Project Management																				
2: Research/technology review																				
3: Use Cases																				
4: Software Architecture																				
5: Software Development																				
6: Testing and Evaluation																				
7: Final Reports																				

Workpackage and activity	**Earliest start date**	**Latest completion date**	**Outputs (deliverables in bold)**	**Milestone**	**Responsibility**
YEAR 1					
WORKPACKAGE 1: *Project Management* **Objective:** To manage and coordinate activities, to prepare and report as required, and to assess risks and opportunities as the project progresses.	**01-03-08**	**31-05-09**			
1.1 Develop a detailed project plan with timescales, deliverables, and milestones	**01-03-08**	**31-03-08**	**Project Plan Version 1a**		MH

1.2 Produce project summary on CeRch web site	**01-03-08**	**31-03-08**			MH
1.3 Biannual report for JISC, covering March to August 2008, with financial statement.	**01-03-08**	**31-08-08**	**Progress Report 1**		MH
1.4 Biannual report for JISC, covering Sept. 2008 to Feb. 2009, with financial statement.	**01-09-08**	**19-02-09**	**Progress Report 2**		MH
1.5 Final report for JISC with final financial statement and Completion Report.	**01-03-09**	**31-05-09**	**Final Report**		MH
1.6 Periodically update website as project progresses	**01-03-08**	**31-05-09**	Website updates		MH
1.7 Attend conferences to disseminate project results	**01-03-08**	**31-05-09**	Papers, presentations or posters		MH
1.8 Attend JISC programme level activities.	**01-03-08**	**31-05-09**			MH
1.9 Liaise with relevant external groups on technical matters: NGS, iRODS development team, OMII-UK, Grid Tools Liaison Group (comprising members of JISC, NGS and OMII-UK).	**01-03-08**	**31-05-09**			MH
1.10 Liaise with UK e-Science Engineering Task Force (ETF) regarding pre-production evaluation.	**01-03-08**	**31-05-09**			MH
WORKPACKAGE 2: Review of existing research and technology ***Objective:*** To review in brief existing research and technologies relevant to the project, to avoid duplication.	**01-03-08**	**30-04-08**			
2.1 Familiarisation of staff with the main technologies relevant to the project: iRODS, Shibboleth	**01-03-08**	**30-04-08**			All (depending on degree of

					prior knowledge)
2.2 Review of existing grid-Shibboleth integration projects.	**01-03-08**	**30-04-08**			MH/AH
2.3 Review of projects investigating provenance metadata, and tools for generating, browsing and otherwise using provenance metadata (e.g. Taverna, PASOA, Provenance Explorer).	**01-03-08**	**30-04-08**			MH/AH
2.4 Produce interim document.	**01-04-08**	**30-04-08**	Interim document.		MH/AH
WORKPACKAGE 3: Use cases ***Objective:*** To identify use cases and user requirements for managing access to data in a virtualised data grid environment.	**01-03-08**	**30-06-08**			
3.1 Consultation with stakeholders, including review of any existing use cases (e.g. those produced by the ESP-GRID[11] and PASOA[12] projects.	**01-03-08**	**30-04-08**			MH/AH/RD/TBD
3.2 Create use cases for control of access to data (e.g. role-based, fine-grained).	**01-03-08**	**30-06-08**			MH/AH/RD/TBD
3.3 Specify use cases for generation, capture and use of audit/provenance metadata. In particular, address issues relating to personal data in provenance.	**01-03-08**	**30-06-08**			MH/AH/RD/TBD
3.4 Produce interim document.	**01-06-08**	**30-06-08**	Interim document.		MH/AH/RD/TBD
WORKPACKAGE 4: Software	**01-05-08**	**31-08-08**			

[11] http://wiki.oucs.ox.ac.uk/esp-grid/UseCases; http://wiki.oucs.ox.ac.uk/esp-grid/GridRequirements
[12] http://eprints.ecs.soton.ac.uk/10269/01/pasoa04requirements.pdf; http://www.pasoa.org

Architecture ***Objective:*** To specify the overall architecture for the iRODS-Shibboleth integration and provenance enhancements, and to identify the approach to be taken regarding specific technical and design issues.					
4.1 Identify options for implementation, possibly developing small-scale prototypes.	**01-05-08**	**30-06-08**			MH/AH/RD/TBD
4.2 Specify overall approach to high-level architecture for toolkit, including breakdown into modular components.	**01-05-08**	**30-06-08**			MH/AH/RD/TBD
4.3 Detailed specification of modular components for iRODS-Shibboleth integration.	**01-07-08**	**31-08-08**			MH/AH/RD/TBD
4.4 Detailed specification of modular components for provenance implementation.	**01-07-08**	**31-08-08**			MH/AH/RD/TBD
4.5 Produce interim document.	**01-08-08**	**31-08-08**	Interim document.		MH/AH/RD/TBD
WORKPACKAGE 5: Software Development ***Objective:*** To design, develop and integrate the software components. This workpackage will be broken down in more detail at a later stage, when the overall architecture has been decided.	**01-06-06**	**28-02-09**			
5.1 Design and develop software components for iRODS-Shibboleth integration.	**01-06-08**	**31-12-08**			RD/TBD
5.2 Design and develop software components for provenance implementation.	**01-06-08**	**31-12-08**			RD/TBD

5.3 Integrate of software components.	**01-10-08**	**31-12-08**	Unit tested and integrated software components.		RD/TBD
5.4 Create documentation: • technical documentation. • installation, configuration and system management/administration guide • user documentation	**01-01-09**	**28-02-09**	Interim documentation		RD/TBD
WORKPACKAGE 6: Testing and Evaluation ***Objective:*** To test the software components produced by the project, both in isolation and in integrated form.	**01-03-08**	**30-04-09**			
6.1 Set up test-bed iRODS data grid at CeRch and STFC.	**01-03-08**	**31-03-08**	Test-bed data grid.		MH/AH
6.2 Module testing – individual components will be tested as separate modules. This may not be appropriate for all components.	**01-08-08**	**31-10-08**	Test documentation (specifications, logs, issues). Modified software.		RD/TBD
6.3 Integration testing.	**01-09-08**	**30-11-08**	ditto		RD/TBD
6.4 Internal system testing (at CeRch/STFC).	**01-10-08**	**28-02-08**	ditto		RD/TBD
6.5 Identification of and liaison with external evaluators. These may include: • Projects using the SRB services provided by STFC/NGS. • Other NGS users, e.g. the UK e-Science Engineering Task Force. • The JISC e-Infrastructure team, and projects arising from the JISC e-Infrastructure Community Engagement and Support Programme.	**01-09-08**	**31-12-08**	List of identified evaluators. Evaluation plan.		MH/AH

• The Grid Technology Group at STFC. • Current and potential data grid projects in the arts, humanities and social services. • The iRODS development team at SDSC.					
6.6 Set up a prototype iRODS data grid based at STFC/RAL, which is a core NGS site, with additional storage at CeRch, for user testing and evaluation. A Shibboleth SP will also be installed at RAL.	**01-11-08**	**31-12-08**	**Prototype data grid.**		RD/TBD, with assistance of NGS staff.
6.7 External testing – external users will use and evaluate the prototype data grid. As the system will be non-production, it can not be made available as an NGS service. Instead, it will be made available to targeted NGS users (this user base will be widened as the project progresses).	**01-01-09**	**30-04-09**	Test documentation (specifications, logs, issues). Modified software.		External users, coordinated by RD/TBD
WORKPACKAGE 7: Final Reports ***Objective:*** To produce final versions on the interim reports and documentation produced during the project.	**01-03-09**	**31-05-09**			
7.1 Technical documentation.	**01-03-09**	**31-05-09**	**Final versions of documentation.**		MH/AH/RD/TBD
7.2 Test documentation.	**01-03-09**	**31-05-09**	**Final versions of documentation.**		MH/AH/RD/TBD
7.3 User documentation.	**01-03-09**	**31-05-09**	**Final versions of documentation.**		MH/AH/RD/TBD
7.4 Final release of software.	**01-05-09**	**31-05-09**	**Final release of software.**		RD/TBD