

JISC DEVELOPMENT PROGRAMMES

Project Document Cover Sheet

PROJECT REPORT: July 2006

Project

Project Acronym	RepoMMan	Project ID	
Project Title	Repositories, metadata and management		
Start Date	June 2005	End Date	May 2007
Lead Institution	University of Hull		
Project Director	Ian Dolphin		
Project Manager & contact details	Richard Green r.green@hull.ac.uk		
Partner Institutions			
Project Web URL	www.hull.ac.uk/esig/repomman		
Programme Name (and number)	Digital Repositories Programme		
Programme Manager	Neil Jacobs		

Document

Document Title	Project report		
Reporting Period	February 2006- July 2006		
Author(s) & project role	Richard Green Project Manager		
Date	31/07/2006	Filename	mgt-rep-0607-v10.doc
URL			
Access	<input checked="" type="checkbox"/> Project and JISC internal	<input type="checkbox"/> General dissemination	

Document History

Version	Date	Comments
0.13	13/07/2006	Working version
1.0	26/07/2006	Submitted version

RepoMMan Project

Project Management Report

February 2006 - July 2006

Richard Green

July 2006



The RepoMMan Project

Project Director:	Ian Dolphin, Head of e-Strategy, University of Hull (i.dolphin@hull.ac.uk)
Project Manager:	Richard Green (r.green@hull.ac.uk)
Technical Lead:	Robert Sherratt (r.sherratt@hull.ac.uk)
Repository Domain Specialist:	Chris Awre (c.awre@hull.ac.uk)

The Repository Metadata and Management Project (RepoMMan) at the University of Hull is funded by the JISC Digital Repositories Programme. The project is being carried out by the University's e-Services Integration Group (e-SIG) within Academic Services.

1. Grant statement

The RepoMMan Project is being conducted under the terms agreed with JISC in the letter of grant and the JISC Terms and Conditions attached to it.

2. Aims and Objectives

The aims and objectives of the project remain as set out in the Project Plan dated August 2005.

The main targets for the period of this report were

- 2.1 to produce Java code and a Spring configuration which will allow integration between the MVC and BPEL layers
- 2.2 to develop a Java servlet providing a user interface to the three-tier system underlying the workflow tool
- 2.3 to develop a Java portlet providing a user interface to the three-tier system underlying the workflow tool
- 2.4 to continue developing a 'Beginners' Guide' to working with Fedora
- 2.5 to report on RepoMMan's experiences with Fedora in the first year of the project
- 2.6 to deploy the workflow engine in the University's uPortal framework
- 2.7 to identify the feasibility and requirements for, and source of, a core personal metadata profile
- 2.8 to identify the feasibility and requirements for, and source of, a personal metadata profile for research
- 2.9 to promote RepoMMan's work within and beyond the JISC Digital Repositories community
- 2.10 to promote RepoMMan's work within the Fedora community

The majority of these targets have been met on time. However some of those relating to the development of the workflow tool and dependent upon an interaction between BPEL and Fedora web services have not. The delays in meeting these remaining targets are discussed in detail in the remainder of this report.

3. Overall approach

There has been no significant change in our approach to the project.

4. Project outputs

D-D4 Further work on the 'Beginners' Guide' to using Fedora

Work on this document has continued and the latest version of it is submitted with this report. The document has attracted favourable comment from the Fedora community and work on it will continue as appropriate for the duration of the RepoMMan project. In particular, the Fedora Project's new Outreach Working Group has requested permission to advertise the document freely. The document has been used, to our knowledge, by a number of people coming new to the software, most recently a member of staff in BT's research division.

D-D5 Java code and Spring configuration enabling integration between the MVC and BPEL tiers of the three-tier stack within the workflow tool.

A 'thin vertical slice' through the three-tier stack, establishing that communication between a web service, BPEL, Spring and a jsp page could be implemented was completed in March 2006. This work was then used to inform the early development of a stack that specifically interacted with a Fedora web service.

D-D6 Java servlets for user interface

A first Java servlet for the user interface was completed in June 2006, some three months later than planned. The interface was not developed prior to June because of the problems encountered in implementing the interaction between Fedora and BPEL. This issue is discussed in the next section.

D-D7 Java portlet for user interface

This interface, scheduled for May 2006, has just been commissioned from our sub-contractor Icodeon. It has been delayed by the problems referred to under D-D6.

D-D8 Report on experiences with Fedora in the first year of the project.

This report was published in July 2006 and is attached. It has been submitted for peer review by Thornton Staples, co-director of the Fedora project.

D-D9 Deployment of the workflow engine in the University's uPortal framework

This work, scheduled for July 2006, awaits the development of the portlet referred to under D-D7.

R-D5 Identification of feasibility and requirements for, and source of, core personal metadata profile

This short report was completed in June 2006. The University of Hull is in the process of commissioning an Identity Management System (IdMS) from Sun Microsystems and much of the metadata associated with work package R2 will eventually be derived from it. The IdMS work at the University is slightly behind schedule by a matter of a few weeks and so RepoMMan deliverables which rely on it may be similarly delayed. Once Sun's proof-of-concept work is completed at the end of July, an LDAP server will be freed up which RepoMMan can use to further the practical elements of this work. This will be a minor delay and does not represent a source of concern to the project team.

R-D6 Identification of feasibility and requirements for, and source of, personal metadata profile for research

This report was due in July 2006 but is slightly delayed for the reasons described in the last paragraph. Again, we do not see this as a cause for concern.

5 Project outcomes

The research elements of this project are running approximately to timetable; however, the technical strand of the project suffered a major delay during the period covered by this report and is only recently in a position to try and make up lost time.

5.1 Research strand

Most of the work undertaken in the research strand during this reporting period has been spent investigating the capabilities of Fedora and coming to an understanding of how it might serve the repository needs of an institution such as the University of Hull, exemplified thus far by work with its research staff.

The investigative work has entailed a great deal of time working with the product core, coming to an understanding of how objects and collections can be built and how they might interact, and with some of its other features, such as XACML security.

Whilst the work has been ongoing, a number of items of documentation have been developed initially in draft and then, in some cases, in a publishable form:

- a range of possible content models
- a range of possible disseminators to work with the above content models and associated XACML security scripts
- a 'Beginners' Guide' to working with Fedora
- a document dealing with RepoMMan's experience with Fedora over the first year (approximately) of the project

The results of this background work are now being fed into the technical strand to inform the software development surrounding the management of objects using the embryo RepoMMan workflow tool. A number of small, on-line demonstrators have also been developed.

An important part of the research development has been interaction with developers and projects elsewhere, and with the Fedora community. The exchanges of ideas that have taken place have been important in informing the plans for a repository at the University of Hull and, by extension, the RepoMMan development work. In addition, this interaction has flagged up a number of possible developments for Fedora which we are thus able to plan for. Section 6 of this report details further some of the collaboration that has taken place.

In late spring, we worked with staff on the CD-LOR project to adapt the on-line survey that we had used with researchers to investigating the needs of teachers producing learning objects. At the time of writing, the findings of this work have been made available to us in draft and should shortly be published on the CD-LOR website.^{1,2} This survey will help in our own research of teaching and learning needs.

Most recently, the research side of the project has started to investigate in some detail the project's needs for automatic population of metadata fields.

At the University of Hull we have been following the proof-of-concept tests for a Sun Identity Management System which will ultimately be the source of much of the metadata relating to an individual author at the University. This work is very slightly behind schedule whilst we wait for the release to us of an LDAP server in use for the tests.

Beyond the University we have been looking for solutions to our problem of automatically populating metadata fields for new objects. The JHOVE metadata tool, developed by Harvard University, will go a long way to providing us with basic technical metadata for an object; a solution is less clear in the field of descriptive metadata. Here there are a number of tools that rely for their efficacy on controlled subject vocabularies or some sort of ontology. The very wide spectrum of likely content in Hull's repository argues for a more flexible tool. We intend to try a tool to be released at the end of July developed jointly by the National Science Digital Library (NSDL) in the US and the University of California, Riverside. Our understanding, at the time of writing, is that this may be able to take 'unseen' text and do a good job of producing a keyword list. We acknowledge here the help of the Natural Language Processing Unit at the University of Edinburgh as their advice has been very useful in aiding our understanding of the possibilities in this field and the concepts involved. It should be possible to report on progress in this area when the next JISC report becomes due.

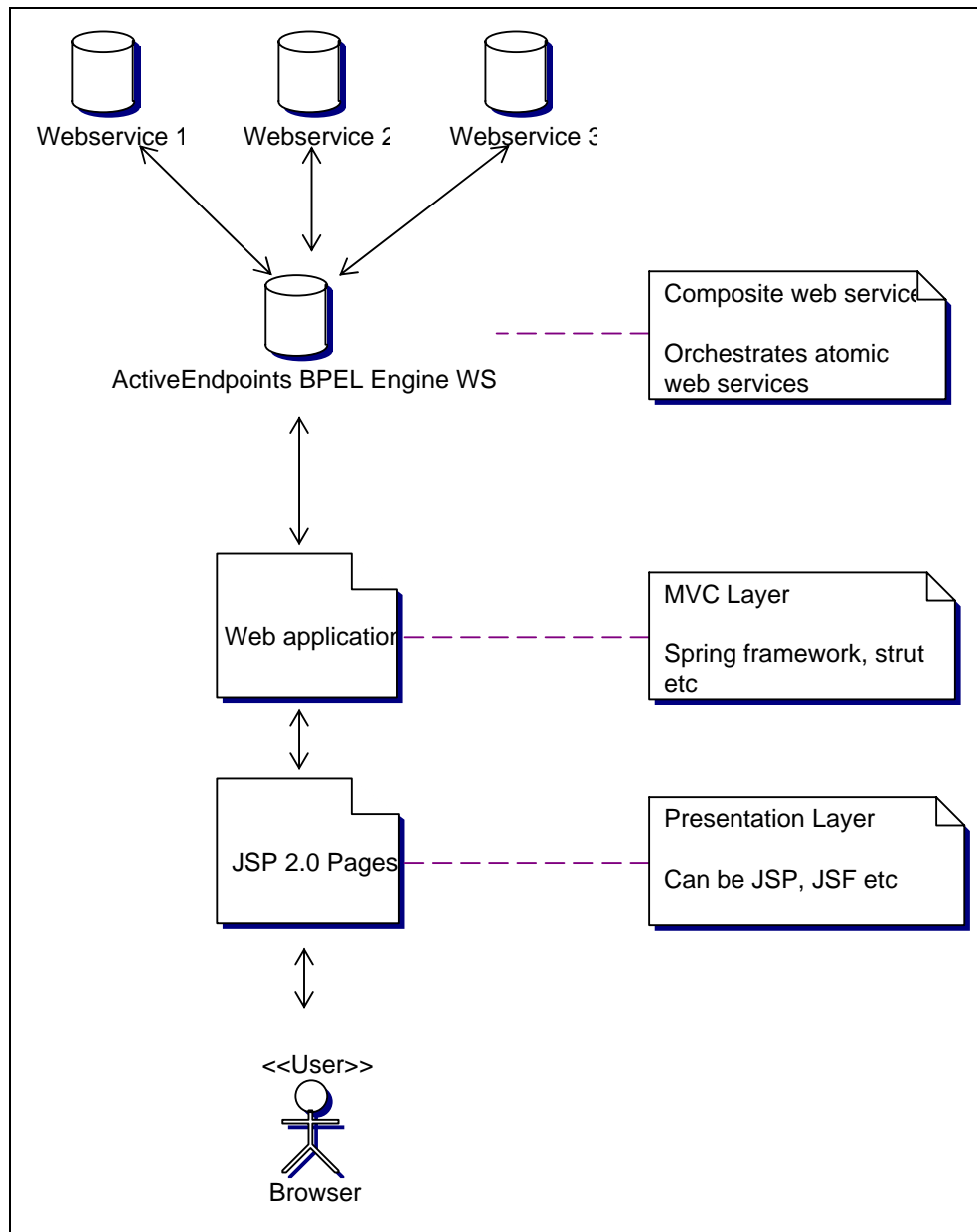
5.2 Technical strand

After the first exploratory work with BPEL, the RepoMMan team came up with the idea of developing a 'thin slice' through the three-tier stack that we eventually planned to produce.

The diagram overleaf shows the three-tier approach that we had decided upon.

¹ Margaryan A (2006, in preparation) *Report on Personal Resource Management Strategies* CD-LOR Project, JISC

² <http://www.ic-learning.dundee.ac.uk/projects/CD-LOR/>



The web-services at the top of the diagram represent just some of the services that RepoMMan’s workflow tool will be required to access; Fedora’s web services form part of this set, others will include services for the auto-generation of metadata and so on.

The BPEL engine orchestrates these services as necessary and interacts with the user through a model view controller (MVC) layer which we have decided to provide with Spring. The MVC layer interacts with jsp pages in the user’s browser.

The ‘thin slice’ tested out one vertical strip through this diagram. A browser, communicating through the MVC layer with BPEL which, in turn, communicated with a single, simple service. This was essentially proof-of-concept that such an integration could be made to work and to serve our eventual purpose.

This work was completed satisfactorily in early Spring 2006.

Not surprisingly, our first attempt, in April, to combine BPEL and Fedora in any serious way centred around trying to ingest a simple object into Fedora 2.1. To do this, we exported an existing object from the repository in Fedora's own FOXML format and attempted to orchestrate a process to re-ingest it as a new object.

We quickly came across two problems:

The first was that Fedora's management WSDL (Web Services Description Language) file would not import into the BPEL design tool, throwing a '401 error' (unauthorised). An exchange of e-mails with the Fedora team revealed that the file itself was the subject of http security, rather than the more usual arrangement where the file was accessible to all but the service endpoints were secured. The software we are using for development does not support a process of authentication for access to the WSDL file. This problem was overcome on a temporary basis by taking a local copy of the file so that authentication was not needed to access it.

Having made Fedora's management WSDL accessible to our software, work was started to use it. However, this proved impossible due to a second problem. The Active Endpoints BPEL software declared the WSDL file to be an illegally formed document. To check this, the file was imported into Altova's XML Spy, and into Sun's Java Studio Creator. Both agreed that the WSDL was, indeed, illegally formed.

Fedora has two WSDLs, one for management functions and the other for access. The access WSDL was tried in XML Spy with similar results. By comparison, both the WSDLs from Fedora 2.0 validated and it was possible to invoke operations using them.

E-mails to the Fedora-user list generally got some sort of response from the development team, but nothing that solved our problems.

Over the next few weeks a number of different approaches were tried to the 'WSDL problem'. We switched to what we thought would be a more straightforward process, that of retrieving information about the datastreams in an object and trying to set them all to 'inactive'. Again a range of problems was encountered with the Fedora WSDLs and another series of e-mails to the Fedora-user list provoked a response from staff at Rightscom, in London, who were starting to explore similar areas of Fedora 2.1 web services and encountering similar problems. The Fedora UK and Ireland group meeting in Hull provided the opportunity to discuss matters and, as a result, over a period of days we compared notes and put together a detailed, joint e-mail to Fedora which was sent, not to the mailing list, but to a range of key people within the project. The e-mail pointed out that the difficulties with the 2.1 WSDLs were effectively preventing either of our organisations from making any progress.

This joint e-mail was taken seriously by the Fedora team who produced a patch to fix the http security problem within a couple of days. Having looked into the issues with the contents of the WSDL files themselves they acknowledged that there was a problem. The 2.1 files use the rpc/encoded message style which a number of major software products, they accept, "have problems with". The team conceded that they should move to the document/literal style which they now see as best practice. They quickly set to work to produce a further patch.

During development, the patch was tested at Hull and at Rightscom; within the space of ten days or so a new WSDL had been produced which seemed to work as expected.

At Hull work re-started to try and develop a process that would ingest a Fedora FOXML object and this was quickly achieved. That process will now be 'grown outwards' to encompass more and more of the functionality that will be required for the RepoMMan workflow tool. Given all the time previously expended in this area, it is hoped that progress will be fairly rapid.

A concern at the back of our minds had always been that the Fedora Admin Client uses web services to communicate with the repository. Why, then, did this work in Fedora 2.1, whilst our constructions did not? This puzzle was solved at the Fedora Users' conference where it

was admitted that the Fedora Admin Client 'knowing that everything works' does not attempt to validate anything.

6 Stakeholder analysis

6.1 JISC

The RepoMMan project has been represented at the following JISC events:

JISC Digital Repositories meeting, Warwick
API meeting, London (family illness prevented attendance also at Warwick)

6.2 Other 'Digital Repositories' projects

The RepoMMan project has been represented at, and contributed to, the following events hosted by other Digital Repositories Programme, or related, projects:

RIVER project meeting, University of Oxford
VERSIONS project meeting, LSE, London

In addition, RepoMMan has collaborated with the CD-LOR Project in undertaking an on-line survey based, in part, on its own on-line survey undertaken at the end of 2005. The results of the new work will be integrated with the old data to provide a larger sample, where this is appropriate. It will also be used as part of the RepoMMan research into Teaching and Learning needs. Further information is given at 6.5.

6.3 The Fedora Community

Work with the Fedora community, most especially here and in the US, has been an important – indeed essential – aspect of our work over the last six months.

Throughout the period the team have been in contact with key members of the Fedora community through the Fedora mailing list and by private e-mail. These contacts have been instrumental in developing our ideas about repository architecture and essential in terms of furthering the technical work on RepoMMan's workflow tool.

At the Fedora Users' meeting at the National Library of Wales (NLW) last autumn, the RepoMMan representatives volunteered to form a Fedora UK User Group. This we have done and the group met first in May. In the event, it became the 'UK and Ireland' group as the result of a contact from Dublin. The formation of the UK&I group was done with the full knowledge and support of the Fedora project team. The group does not maintain a website of its own but, instead, uses pages on the Fedora wiki; it has, on the other hand, formed a JISCmail mailing list for communication about domestic events. The second meeting of the group will be held in Glasgow in October, immediately before or after (tbd) the Open Scholarship 2006 conference. The Fedora UK&I group will work in cooperation with the newly formed Fedora Outreach Workgroup.

At the NLW meeting it was also suggested that the Fedora community would benefit from having access to a wiki, in addition to the mailing list. This was implemented in spring 2006 and the RepoMMan team have contributed extensively to it.

The RepoMMan team was one of only two to contribute two papers to the Fedora Users' Conference at the University of Virginia, Charlottesville in June 2006. It was the only team to present a paper in both the 'technical' and 'general' strands. The event was very well worthwhile; our papers were well-received and we learned much from other presentations

and meetings that were arranged. In particular, the presentation about 'Fez', a private meeting about metadata schema, and a conversation about the possible automatic extraction of descriptive metadata from texts are likely to have long-term, beneficial, effects on the RepoMMan work. A full report on the conference has been submitted to JISC³ and a shortened version of it will appear in the July 2006 edition of Ariadne.⁴

6.4 Potential users at the University of Hull

The last six-monthly report noted that a working party had been set up bringing together members of the RepoMMan team and staff, primarily from Academic Services. This has met twice more to discuss areas of mutual interest and to begin to develop use case scenarios which can inform the development of aspects of the RepoMMan workflow tool. We have also given a presentation to the Library Systems and Services Working Group which draws on a wider group of Library staff.

The project has made a formal presentation to the Academic Affairs Management Group about a repository for the University of Hull and RepoMMan's work in that context, and a similar talk to the Information Services committee. Both these events are seen as important in the process of advocacy for Hull's digital repository. In addition, the project submitted a briefing paper to the University's Research Committee.

At the time of writing, discussions are underway about the possibility of using a new repository to set up a student access point for past examination papers, and the possibility of providing some support for the Research Assessment Exercise 2008. Whilst neither of these initiatives is directly a part of the RepoMMan project they may represent significant stepping stones to the acceptance and integration of the repository for which RepoMMan is developing support. That said, the experience of these initiatives may well prove important to informing RepoMMan's work with the administration community.

6.5 Potential users elsewhere

The RepoMMan team have worked with staff on the CD-LOR project to undertake an on-line survey of practices relating to the development of learning objects. RepoMMan provided the Flash tool, the web server and undertook preliminary analysis on the results before passing them on to the CD-LOR team. The fully processed results have just been returned and we shall integrate them with the results of our own, earlier, survey as appropriate. They will go some way to informing our own requirements for work with the Teaching and Learning community.

RepoMMan has been asked to contribute an article to the OCLC Systems & Services International Digital Library Perspectives special issue.

7 Risk analysis

The risk analysis in our Project Plan identified on element as being "Implementation in Fedora is more problematic than thought". As noted in section 4, this has turned out to be true. The risk management strategy "ensure involvement with the Fedora community to assist with any problems as they arise" has worked extremely well.

8 Standards

There is nothing to report under this heading.

³ Awre C and Green R (2006) *The Fedora Users Conference, University of Virginia, 19-20th June 2006*

⁴ <http://www.ariadne.ac.uk/>

9 Technical development

There is nothing to report under this heading. that has not already been dealt with in Section 5.

10 Intellectual property rights

There is nothing to report under this heading.

11 Project partners

There have been no changes to the project partners (none) or to the list of subcontractors.

12 Project management

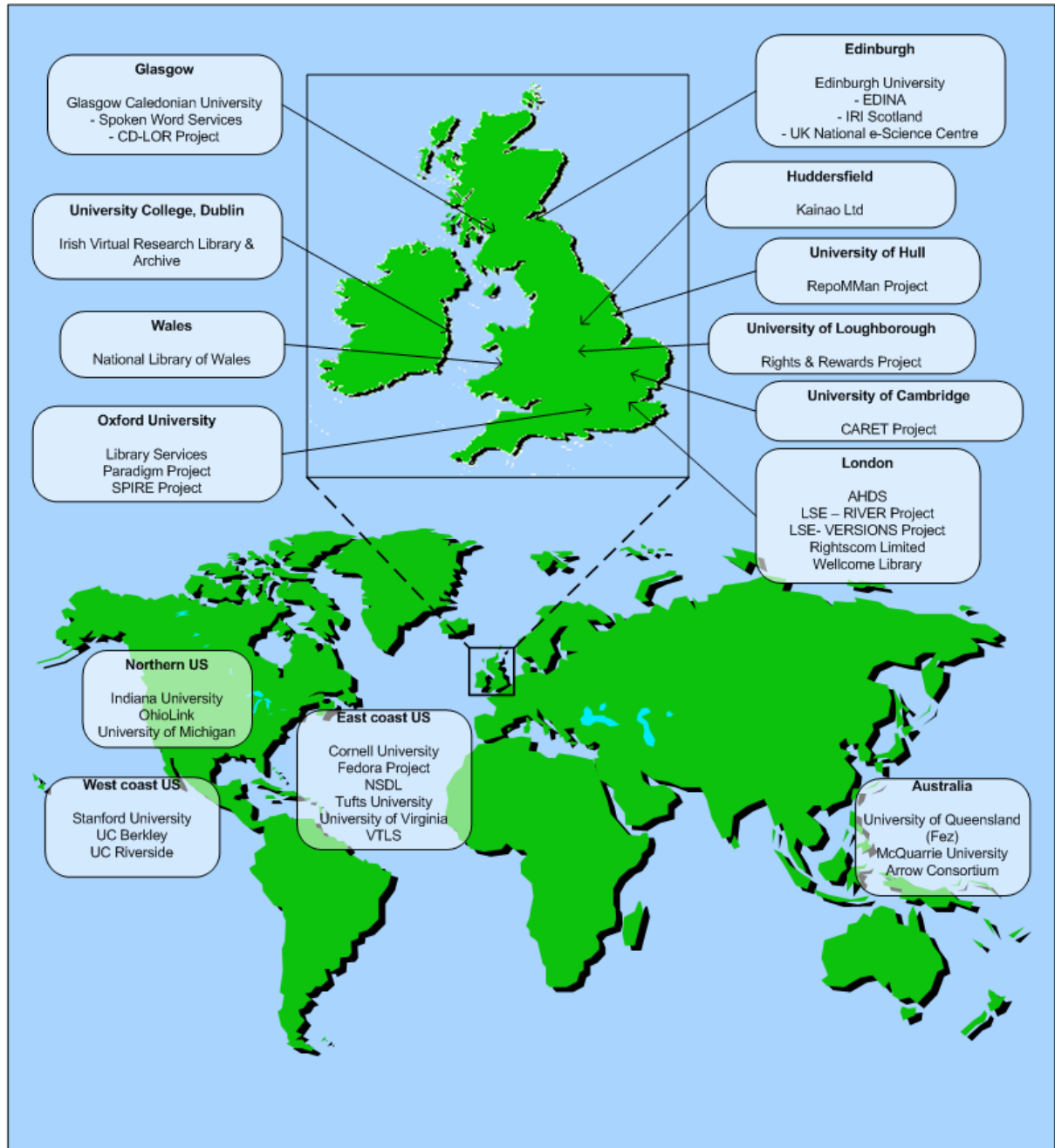
There have been no changes to the project staff.

13 Programme support

We have no immediate needs in terms of Project Support from the JISC. However, as we have noted before, matters of IPR, DRM and licensing are a common concern to many Digital Repositories Programme projects and we wonder if it would be timely to make this a major theme of the proposed January conference. Projects which have already looked carefully at this field as related to the DR programme could brief those of us who are further behind. Perhaps JISC Legal might contribute too. Such a theme might enable RepoMMan, as an example, not to waste time reinventing the wheel when it comes to our own deliverable on DRM.

14 Relationships with other project activities

RepoMMan has working links with a number of projects and organisations worldwide:



15 Budget

The budget expenditure template is attached as Appendix A.

16 Workpackages

Current reporting period

Workpackage D1: Development of three-tier workflow engine

Deliverables

D-D2 Deployment of BPEL environment

This document was completed in a first version in October 2005 but not completed until the project had the BPEL engine running on a dedicated Solaris-based server in March 2006.

Due: 10/2005 Completed 10/2005 (v0.9), updated and augmented 3/2006 (v1.0)(attached)

D-D5 Java code and Spring configuration enabling integration between MVC and BPEL layers

Due: 02/2006 Completed: 03/06

D-D6 Java servlet user interface

Work on this servlet was delayed because of the problems discovered trying to integrate the Fedora 2.1 service calls with the BPEL engine. With the resolution of those problems a basic interface was quickly developed.

Due: 04/2006 Completed: 07/06

D-D7 Java portlet user interface

This interface has recently been commissioned from our partner, Icodeon. Clearly, it could not be commissioned until the problems integrating the Fedora 2.1 service calls with the BPEL engine were resolved.

Due: 05/2006 Anticipated completion: 08/06

Workpackage D3: Fedora evaluation

D-D4 Iterative development of research use case

By agreement with the Programme Manager, this deliverable was repurposed to become a beginners' guide to Fedora and its delivery was retimed to 'summer 2006'. As it has progressed this document has been welcomed by the Fedora community and accordingly it will continue to be developed during the lifetime of the RepoMMan project. The document is being published at key stages as a 'work-in-progress'. Two such interim releases have taken place.

Due: 'Summer 2006' Interim releases: 02/2006, 04/2006 (attached)

D-D8 Report on experiences with Fedora in project first year

As noted elsewhere in the report, work with the web services side of Fedora has been considerably delayed. Accordingly the timespan of this deliverable has been deliberately extended by a few weeks to encompass both the period in which the problems were overcome and the Fedora Users' conference at the University of Virginia in the second half of June.

Due: 05/2006 Completed: 07/2006 (attached)

Workpackage R2: Feasibility and requirements study of the use of contextual metadata for the identified institutional use cases

R-D5 Identification of feasibility and requirements for, and source of, personal metadata profile

This document is complete but, as noted elsewhere, we have not yet tested the project's preferred solution against available systems. This is likely to take place in late July and early August. This document will form part of a larger, integrated report on metadata requirements in due course; it is therefore not attached but can be supplied on request.

Due: 05/2006 Completed 06/06

Workpackage D4: Workflow integration with institutional portal framework and collaborative environment

D-D9 Deployment of workflow engine in uPortal Framework

This work has been delayed by the difficulties encountered in Workpackage D1.

Due: 06/2006 Anticipated completion: 08/2006

Next reporting period**Workpackage D4: Workflow engine integration with institutional portal framework and collaborative environment**

Objectives

- Deployment of workflow engine in uPortal framework
- Deployment of workflow engine in collaborative environment

Workpackage D5: Implementation of metadata population from 'personal metadata profile'

Objectives

- Report on investigation of methods to access user profile data from portal framework and other sources such as an enterprise directory

- Implementation of methods to automatically populate metadata fields for users of both portal and collaborative environments

Workpackage R1: User requirements analysis

Objectives

- Criteria and associated materials for administrator and learner user interviews
- Report on administrator and learner user requirements survey data

Workpackage R2: Feasibility and requirements study of the use of contextual metadata for the identified institutional use cases

Objectives

- Identification of feasibility and requirements for, and source of, personal metadata for research
- Identification of feasibility and requirements for, and source of, personal metadata for learning
- Identification of feasibility and requirements for, and source of, personal metadata for administration
- Combined report on the above

17 Evaluation plan

There are no elements of our Evaluation Plan relating to this reporting period.

During the next reporting period we shall be testing the developing workflow engine with potential users and seeking their opinions on its usability. These comments will be fed back into the development process. We shall also be investigating the automatic population of metadata for objects created for the University repository and attempting, with the help of others, to form a considered opinion as to the effectiveness of the tools available to us.

In considering what metrics might be used to measure the success of the RepoMMan project one year after its completion, we suggest the following:

- the level of local (University of Hull) use of the workflow tool for uploading and access
- the level of interest in (and possible adoption of) the tool elsewhere
- the number of downloads of RepoMMan project documents

and more generally

- the level of take-up of BPEL in the repository community (including the possible integration of the Active BPEL Engine identified by RepoMMan as part of the official 'Fedora framework')

If JISC wish this evaluation actually to be possible we, as a project, need to know this urgently. For instance, we have made no attempt thus far to log document downloads from our website and it may prove impossible to gather this information retrospectively.

18 Quality assurance plan

Deliverable D-D8, Report on experiences with Fedora in the first year, is listed for quality assurance using peer review. This document has been sent to Thornton Staples, co-Director of the Fedora Project for appropriate comment. At the time this management report was submitted we have not had a formal response; should the feedback require it, a version 1.x of the document will be produced and published.

19 Dissemination plan

During the reporting period, a number of dissemination activities have been undertaken:

Ongoing	Continued, regular update of project website
Ongoing	Interaction with the Fedora Users' discussion list
February 2006	Involvement with JISC DR conference at Warwick; presentation of paper on 'Personal Resource Management'
May 2006 – ongoing	Contribution to, and interaction with, the Fedora Wiki
May 2006 – ongoing	Work with the Fedora UK & Ireland User Group
May/June 2006	Team involvement in Community Source Week, Vancouver, Canada
June 2006	Two papers presented to the Fedora Users' Conference at the University of Virginia, Charlottesville, US

20 Exit/sustainability plan

There are no issues within this section of the plan at the present time.

