# JISC DEVELOPMENT PROGRAMMES

# Project Document Cover Sheet

# PROJECT REPORT: January 2007

## 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<br>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** | August 2006- January 2007 | | |
| **Author(s) & project role** | Richard Green<br>Project Manager | | |
| **Date** | 31/01/2007 | **Filename** | mgt-rep-0701-v10.doc |
| **URL** | | | |
| **Access** | ☑ Project and JISC internal | | ☐ General dissemination |

## Document History

| Version | Date | Comments |
|---|---|---|
| 1.0 | 30/01/2007 | Submitted version |
| | | |
| | | |


# RepoMMan Project

---

Project Management Report

August 2006 - January 2007

Richard Green

January 2007

## 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 deployment of workflow engine in uPortal framework
2.2 deployment of workflow engine in collaborative environment
2.3 report on investigation of methods to access user profile data from portal framework and other sources such as enterprise directory
2.4 implementation of methods to automatically populate metadata fields for users of both portal and collaborative environments
2.5 criteria and associated materials for administrator and learner interviews
2.6 report on administrator and learner user requirements
2.7 identification of feasibility and requirements for, and source of, personal metadata for research
2.8 identification of feasibility and requirements for, and source of, personal metadata for learning
2.9 identification of feasibility and requirements for, and source of, personal metadata for administration
2.10 combined report on 2.7-2.9

The majority of these targets have been met. However, due to delays experienced with work on the web services element of the project, discussed in detail in the last Management Report (July 2006), we are some months behind with technical elements of the work.

# 3. Overall approach

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

# 4. Project outputs

D-D9 Deployment of workflow engine in uPortal framework

The (then) current version of the workflow tool was successfully deployed into uPortal 2.5.3 at the end of November. This is a fully JSR-168 compliant implementation. It should be noted that the implementation does not function correctly with uPortal 2.5.1 due to a bug in that software release, a bug that we believe was fixed in 2.5.2. However, the uPortal 2.5.2 'quickstart' tool is no longer available on the web to perform this test.

Work on the developing tool and its portlet implementation will now take place in parallel. The portlet work has been undertaken by our sub-contractor Icodeon and we are making use of their on-line development environment to keep the two elements synchronised.

D-D10 Deployment of workflow engine in collaborative environment

The delays in developing the workflow tool itself have meant that this work, scheduled for August, has not yet taken place. It is anticipated that this will now be undertaken

during February and March. An amount of preliminary, investigative work has been undertaken.

D-D11 Report on investigation of methods to access user profile data from portal framework and other sources such as an enterprise directory

This work, scheduled for August, is now complete and the document is attached. More extensive work with the portal (see D-D12) has not yet taken place because successful integration of the workflow tool with the portal framework (D-D9) is a recent event. The urgent development need is to complete the user interface as it relates to deposit in a private space. Work has been done with a test LDAP server at the University of Hull and this part of the report was completed in the autumn, however the work identified a number of issues with the current (2.1.1) version of Fedora that will need to be resolved by its development team before an LDAP enterprise directory could become the entire basis of our authentication and authorisation needs. Fedora version 2.2, which purports to deal with these issues, was released on January 19th.

D-D12 Implementation of methods to automatically populate metadata fields for users of both portal and collaborative environments

This work, scheduled for November, has been delayed to await the implementation of the basic deposit tool. It is anticipated that it will take place for the portal implementation in February 2007 with the collaborative environment implementation following.

R-D7 Criteria and associated materials for administrator and learner interviews

These criteria were discussed and mapped out during the early autumn.

R-D12 Report on administrator and T&L user requirements interviews

This report was due for completion in January 2007 and is attached.

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

This work has been undertaken as planned, see R-D11.

R-D8 Identification of feasibility and requirements for, and source of, personal metadata for teaching and learning

This work has been undertaken as planned, see R-D11.

R-D10 Identification of feasibility and requirements for, and source of, personal metadata for administration

This work has been undertaken as planned, see R-D11.

R-D11 Combined report on D6, D8 and D10.

This work has been undertaken as planned. The report is attached. It has been somewhat expanded to look beyond the needs of the RepoMMan Project itself to the eventual use of the tool with a repository at the University of Hull.

R-D13 DRM literature review examining issues raised by the user requirements analysis and metadata profile generation

This work is in hand but will not be completed by its due date in the Project Plan of January 2007. With the benefit of hindsight, this date was unrealistic; the report depends on R-D12 (due and completed January 2007) which is only just available. R-D13 is now anticipated in late February or early March.

# 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 (and the previous) report and is only recently in a position to try and make up lost time.

## 5.1 Research strand

Work on the research strand during this reporting period can effectively be divided into two sections: work related to use of the tool in terms of user requirements and subsequent metadata needs, and work related to RepoMMan workflow tool and the use of Fedora.

### 5.1.1 User requirements and metadata

Work was undertaken at the end of 2005 to survey potential 'research' users of the RepoMMan tool in order to identify user needs. During this current reporting period, we have undertaken similar work with members of the 'teaching and learning' (T&L) community and with University administrators. (In saying this, we recognise that many university staff at Hull and elsewhere belong to two or even all three of these communities.)

During 2006 we worked with the team from the CD-LOR project to undertake an on-line survey of those involved in T&L at a wide range of universities[1]. This, and our own survey of researchers last year, produced interesting information but nothing wholly unexpected. In view of this, we made the decision not to attempt an on-line survey of administrators. This decision is in-line with the RepoMMan project plan which indicated that surveys of these two groups would be undertaken only "if necessary". In addition to the survey, we have undertaken detailed interviews with a number of representative members of staff here at Hull from the T&L community and from administrative staff. The report summarising the outcomes of these most recent interviews (R-D12) is attached.

### 5.1.2 RepoMMan workflow tool and Fedora

During the autumn, we spent a considerable amount of time working with 'Fez', a user-interface for a Fedora repository developed at the University of Queensland (UoQ) in Australia. The team at Hull was interested in Fez inasmuch as it offered the possibility of providing a relatively lightweight front-end to our Fedora repository which could be used to develop a demonstrator with which to show off the results of work with the RepoMMan tool. We found the then current release (1.2) to be very 'buggy' but in working with the UoQ developers to resolve matters we were able to undertake a number of pieces of work which will be of use to RepoMMan.

[1] Margaryan A (2006) *Report on Personal Resource Management Strategies* CD-LOR Project, Glasgow Caledonian University

We investigated the use of LDAP for authentication with Fez. The Fez code (written in PHP) assumes a Microsoft LDAP server whilst the University of Hull operates an LDAP server that conforms to 'Open LDAP' protocols which are significantly different. This being the case, we re-wrote the Fez LDAP code to work with our own installation and thereby effectively undertook part of the work required for our own D-D11 deliverable. the rewritten code was fed back to the UoQ developers. We then attempted to extend this work to use Fedora's LDAP provision. In doing this we identified serious problems with the Fedora code. This issue is addressed in section 5.2.

In attempting to get Fez to work in Hull we had problems with two further areas of Fez code. The first was related to the identification of MIME type. The Fez code uses a Unix call to identify the MIME type for a file and this call is not implemented in Solaris (the platform on which our server runs). In finding a work-round we were introduced to a product called MIME Magic which we are now using in the RepoMMan tool to identify the MIME type of files before ingest to the repository. The product works across all major platforms and, in particular, provides a useful way to type Mac files without extensions. (We are now aware of DROID, from the UK National Archives but have not yet had the time to compare its functionality with MIME Magic. When a web-service version is available we may use DROID in preference.) The second area of code involved invoking a tool called ImageMagick which is used to generate surrogate images (a thumbnail, for instance) from a JPEG (say) file that is being ingested. This will be a useful tool in developing specific RepoMMan functionality for image ingest.

We have since abandoned work with Fez itself for three reasons: firstly, the very 'buggy' nature of the software had turned into a major issue for us, secondly the security mechanisms that Fez uses are non-standard for Fedora and cannot easily be invoked except by Fez itself, and thirdly a better alternative was identified.

In the very early days of the project, perhaps even pre-project, one of the team looked at a product called 'ELATED' which offered a simple front-end to a Fedora repository. The version seen then did not appear to offer the sort of functionality that we needed and the search moved on. In November, we decided to take a quick second look at ELATED and found that the current version was a close fit for our 'lightweight front-end' requirement. However the code had been developed for an earlier version of Fedora and did not work with a secured Fedora repository. We have now re-written this part of the Java code and have a version of ELATED running with a secured Fedora 2.1.1. This code is to be offered back to the community even though ELATED itself is no longer being developed or supported by its authors. ELATED will be used as the basis for demonstrating RepoMMan-ingested digital objects. The test ELATED interface is at http://repository.hull.ac.uk

In parenthesis, we have recently discovered that the Fedora FIRE client, which was dropped as a project, was dropped only by Fedora. The NSDL team have continued to develop it and a first (beta) version of 'OnRamp' is due for release in March 2007. In addition, a product called 'Topaz' is under development which will first be deployed for the US Public Library of Science.

Looking forward a little to the potential for automatic population of descriptive metadata, we have continued our quest for a tool which will help with this. We have identified a tool called 'Data Fountains' out of UC Riverside and NSDL. This is a tool which takes an unseen text file and attempts to determine a sensible set of metadata for it including title, description, keywords, and language. Our first, very preliminary, experiments with this (December 2006) look very promising. It has the advantage that it does not explicitly use a controlled vocabulary to determine keywords, something that we feel is important in a wide-ranging repository where identifying the appropriate reference vocabulary might be a problem. We have agreed a cooperation with the DEST-funded Arrow Project to jointly test the software and this work will be reported on in due course.

It was mentioned in the last Management Report that we had held private discussions in Charlottesville with University of Virginia (UVa) staff about metadata. The basis of these discussions was that UVa has developed its own metadata schema based on the 'best' from a

range of emerging standards - accepting that no single one of them is yet adequate to their needs. We have spent some time considering their schema in the light of our own needs, identified through past an ongoing user-needs work, and have decided that we should adopt it. The schema has the advantage that it is relatively easily mapped onto Dublin Core. The needs of the University of Hull in this regard are, of course, outside the scope of the RepoMMan Project. It seems likely that we shall develop the release-candidate for the RepoMMan tool in such a way that any 'automatic' metadata generation takes place only in the context of a Dublin Core representation. However, deliverable R-D11 on metadata has been expanded to look at the wider picture.

### 5.2 Technical strand

In the last Management Report we noted that the technical side of the project had been significantly delayed by problems with Fedora's web service interface, but that a patch had recently been delivered which we hoped would allow rapid progress.

In the event we discovered that the patch solved all the web service issues bar one, and that the remaining problem - predictably? - was central to some of our development work. Notwithstanding this issue, work on the RepoMMan tool went ahead to develop an ingest routine which would support standard Fedora objects but storing the data 'payload' external to the repository ('externally referenced content' in the Fedora jargon). Whilst Fedora supports native versioning of 'internal' content it does not do this for 'external' and we needed to develop mechanisms for doing this using BPEL to orchestrate all the component parts of the process.

The first working implementation of this process created the XML template for a Fedora object, requested a persistent identifier for it from the repository and then ingested it. At this stage, the object had no content as such. The content was then taken from the user's local storage, time-stamped and transferred to the user's personal area in the repository's external storage using ftp. (The user space is created by the process if it does not already exist; at Hull, the file store will be a dedicated area of the University SAN.) Once the file was transferred, the embryo Fedora object was modified to reference the stored file. (This part of the process invokes MIME Magic to determine the correct MIME type of the file regardless of the file extension, if any.) It is this first version of the RepoMMan tool that has been tested in a uPortal environment.

The remaining problem area associated with Fedora's web services (in this case not a problem with Fedora *per se*) was that Java Axis threw an error when attempting to list items in the repository. This meant that any operations which involved knowing what was already stored were impossible. At the end of November the Fedora team sent us a further patch that allowed the 'findObjects' repository call to work properly.

Thus it was possible to modify and further develop the RepoMMan code so that when a user attempts to upload a file to the repository, the storage is first checked to see if it already exists; if it does, the file is uploaded as a new version within the existing object rather than as a brand new object.

This latest patch also makes work on a 'proper' user interface possible because we can now interrogate the repository for existing content and structure. This first incarnation of the user interface is being designed deliberately to mimic Windows file dialogues and draws on a commercially tested design for transfer between two 'file systems' listed on screen.

As this work progresses we are being careful to design with potential users other than the University of Hull in mind. Where appropriate, parameters that may be different in other institutions are being abstracted from the code into a configuration file.

As noted in the 'research' section, work to try and use Fedora's (apparent) support for LDAP authentication came to a standstill when we discovered that the process simply did not work in Fedora 2.1.1. E-mail exchanges with the Fedora team resulted in them testing the LDAP

interface and agreeing that it did not work. As Fedora 2.2 (scheduled for release in January 2007) uses a quite different approach to LDAP authentication, it was decided that work to fix the problem could not be justified. Thus RepoMMan's work with LDAP authentication for Fedora is on hold until we have access to the new system.

It should be noted here that during the late autumn of 2006 we were visited by Chi Nguyen, Manager of the DEST-funded RAMP Project. RAMP is developing a new approach to Fedora authentication and authorisation which, as well as improving Fedora's native capabilities, will also allow access via a Shibboleth federation. This work has the support of the Fedora core development team and it seems likely that Hull will eventually adopt it, though it is not likely to be complete within the timespan of the RepoMMan Project (see 6.3).

# 6 Stakeholder analysis

## 6.1 JISC

There have been no formal JISC events during the period covered by this report. However on 24th November we were visited by John Robertson, our JISC Repositories Research Team contact.

Richard Green has been, and continues to be, involved in the work by the Digital Repositories team to develop a generic repositories deposit API (CRIG).

## 6.2 Other 'Digital Repositories' projects

RepoMMan presented a paper at Open Scholarship 2006 held in Glasgow 18-20th October. We were also instrumental in arranging the meeting of the Fedora UK and Ireland User Group that took place on 18th October (see below).

RepoMMan presented a poster in the non-platform-specific time at Open Repositories 2007 (OR07) (see next section).

## 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.

Five meetings are worthy of particular comment:

On 2nd October 2006 we were visited by Andrew Treloar and David Groenewegen from the DEST-funded Arrow (and related) Project(s) in Australia. We spent a number of hours discussing their work, our work and areas where there was potential overlap. This relationship was further developed at OR07.

On 18th October 2006 the Fedora UK and Ireland User Group met at Glasgow Caledonian University immediately prior to the Open Scholarship 2006 meeting. We reported back on events in Charlottesville earlier in the year and a useful exchange of views took place about ongoing and potential future projects. At least one discussion initiated there resulted in a bid into the Circular 04/06 Repositories and Preservation funding.

On 20th November 2006 Chi Nguyen, the RAMP Project Manager from Macquarie University, Australia, visited us on the way home from meeting the Fedora team in the US. We discussed at some length his plans for a new security schema for Fedora and the part we might play in testing it. It seems to be an excellent extension to Fedora's current functionality and fits well with our plans for the University of Hull Institutional Repository. His work has the active support of the Fedora development team.

On 23rd November 2006 a representative of the team went to London to meet Thornton Staples, Co-Director of the Fedora Project. The meeting was not so much about Fedora as about our intention to adopt the University of Virginia metadata schema for use at Hull, a matter discussed above. That said, a number of matters of concern to the RepoMMan Project were aired.

Finally, in January 2007 (23rd-26th), the RepoMMan Project took part in Open Repositories 2007 held in San Antonio, Texas. We presented a summary of our work to date during the Fedora-specific sessions where the interface and its design received enthusiastic comment: in particular many people voiced their support for our 'KISS' approach to interaction and the idea of embedding the repository as part of the creation process.

In the main part of the conference we presented a poster on RepoMMan as a part of a Service Oriented Architecture. We were disappointed not to be allowed a presentation on this but, in the event, one of the other speakers devoted a couple of minutes to praising our approach and work! (Matthias Razum, eSciDoc Project, FIZ/Max Planck Society.)

During a BOF session in the Fedora specialist time, initiated by Richard Green with others, he was asked to form and chair a Fedora Working Group on Disseminators and Content Models. This is process is now in progress.

Chris Awre, similarly, initiated a BOF on 'User Interfaces for Fedora' so that we could share our experience in this area with others and learn from them.

### 6.4 Potential users at the University of Hull

In previous management reports we have noted the meetings of a working party which meets from time to time to promote interaction between the RepoMMan team and other potential users in the University, particularly from Academic Services. This met again in the autumn and discussed a range of issues. The group requested access to some of RepoMMan's internal discussion papers and some written deliverables that were, at the time, in development. This was arranged.

The last management report noted the possibility of piloting an access point for past undergraduate examination papers using the emerging repository. This work has progressed on two fronts. Firstly, a demonstrator has been set up using ELATED with some older examination papers; secondly, discussions are ongoing with the Department of Geography about hosting their examination papers for the last several years and linking them back to Geography's own website. This second initiative will allow us to demonstrate the use of links to the repository in other web sites. We have an agreement in principle that staff from the Department of Geography will, in due course, test the RepoMMan 'product' as a mechanism for putting the past papers into the repository.

A further initiative has produced the main elements of a demonstrator for images within the repository. An image repository is a service needed at the University of Hull and thus it makes an ideal candidate for a demonstrator. The initial work has been done with ELATED but the process has proved useful in terms of establishing features that should be incorporated into a version of the RepoMMan tool which makes specific provision for work with images.

The project is in discussions with the Graduate School at the University about the possibility of moving to electronic theses stored on the repository. The School has officially minuted its desire to see the University launch a repository for theses as soon as possible and we are

discussing the provision of a private area in the repository, accessed with the RepoMMan tool, for graduate students to develop their work.

Our user-needs interviews with potential repository users at the University have stimulated a good deal of interest in its use. We are following up on this interest.

### 6.5 Potential users elsewhere

A report by Chris Awre and Richard Green of the summer Fedora User Group Conference in Charlottesville, VA, was published in the July 2006 issue of Ariadne.[2]

At the OR07 Conference (see 6.3) we held a private meeting with Andrew Treloar and David Groenewegen from the Arrow Project (DEST, Australia) and reached an agreement with them to undertake collaborative work to investigate the Data Fountains metadata extraction software. This software is of considerable interest to both projects.

The RepoMMan team has had an article accepted for the OCLC Systems & Services International Digital Libraries special issue on institutional repositories.

## 7 Risk analysis

The risk analysis in our Project Plan identified one element as being "Implementation in Fedora is more problematic than thought". As noted in section 4 and detailed in section 5, 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, and although some work is behindhand this involvement has minimised the resulting 'slippage'.

## 8 Standards

There is nothing to report under this heading.

The Digital Repositories Programme wiki page is correct in terms of our use of the standards listed there.

## 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.

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

[2] http://www.ariadne.ac.uk/issue48/fedora-users-rpt/


## 12 Project management

There have been no changes to the project staff.

It will be recalled that our software developer, Simon Lamb, was appointed some time after the start of the project whilst the budget accounts for two full years of employment. We shall be contacting the JISC to request that we retain the funding allocated to those first few months in order to keep his services after the official end of the two-year project in order to continue development work on the RepoMMan tool.

# 13 Programme support

We have no immediate needs in terms of project support from JISC.

On 24th November we were visited by John Robertson, our JISC Repositories Research Team contact, and spent some hours talking through the project and its wider implications.

# 14 Programme synthesis

We have checked the Digital Repositories Programme wiki entries in respect of this project and have updated them.

# 15 Budget

The budget expenditure template is attached as Appendix A.

# 16 Workpackages

**Current reporting period**

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

D-D9 Deployment of workflow engine in uPortal framework

A fully JSR-168 compliant portlet using the first RepoMMan tool version was successfully tested in uPortal 2.5.3. Development of this portlet is now proceeding in parallel with development of the tool.

Due: June 2006 First iteration successfully tested: December 2006

D-D10 Deployment of workflow engine in collaborative environment

Work on this part of the workpackage has been significantly delayed because of problems with Fedora's web services provision. Now that these have been remedied it can shortly be undertaken.

Due: August 2006 Anticipated: February 2007

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

D-D11 Report on investigation of methods to access user profile data from portal framework and other sources such as an enterprise directory

Due August 2006 Completed: October 2006 (slight revisions - v1.1 - January 2007)

D-D12 Implementation of methods to automatically populate metadata fields for users of both portal and collaborative environments

This element of the workpackage is significantly affected by knock-on delays from the development strand.

Due: November 2006 Expected: March 2007

**Workpackage R1: User requirements analysis**

R-D7 Criteria and associated materials for T&L and administration interviews

Due: September 2007 Completed: October 2007

R-D9 Report on T&L and administration user requirements survey (if needed)

The RepoMMan team contributed to the CD-LOR survey of the T&L community and is drawing on the report resulting from this. As noted above we shall not undertake an on-line survey of the administration community.

Due: November 2006 Completed: August 2006

R-D12 Report on T&L and administration user requirements interview data

Due: January 2007 Completed: January 2007

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

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

This work was completed on time.

Due: July 2006 Completed: July 2006

R-D8 Identification of feasibility and requirements for, and source of, personal metadata profile for T&L

This work was completed on time.

Due: September 2006 Completed: September 2006

R-D10 Identification of feasibility and requirements for, and source of, personal metadata profile for research

This work was completed on time.

Due: November 2006 Completed: November 2006

R-D11 Combined report on above

This report is completed in its first version. Its scope has been significantly expanded to encompass the University of Hull Repository's metadata requirements. This expanded document will continue to evolve for some time yet.

Due: December 2006 Completed: January 2007

### Workpackage R3: Investigation into digital rights management issues

R-D13 DRM literature review examining issues raised by the user requirements analysis and metadata profile generation

This work is underway but was slightly delayed awaiting the final version of the report from R-D12

Due: January 2007 Expected: February 2007

## Next reporting period

### Workpackage D1: Development of three-tier workflow engine

Objective

- Completed workflow engine with full QA

### Workpackage D6: Investigation of descriptive metadata extraction of object metadata

Objective

- Report on feasibility of automatic extraction of object metadata

### Workpackage D7: Final version of systems and user documentation

Objectives

- Full systems documentation including details of testing
- Full user documentation describing how tasks are completed

### Workpackage R1: User requirements analysis

Objective

- Final report on user needs analysis identifying current repository usage, possible future usage and mapping of user repository process

**Workpackage R3: Investigation into digital rights management issues**

Objective

- Investigative report on DRM implications of the issues raised by the user requirements analysis and metadata profile generation, with particular emphasis on the boundary of institutional and personal use. Identification of potential solutions.

# 17 Evaluation plan

The considerable delays encountered in producing a working software tool have delayed evaluation work. In the final months of the project 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 - most notably the DEST-funded Arrow Project, to form a considered opinion as to the effectiveness of the tools available to us.

The notes that we gave last time for 'evaluation one year on' still seem appropriate.

# 18 Quality assurance plan

Report D-D8, recording our experiences with Fedora in the first year of the RepoMMan Project was sent to Thornton Staples, co-director of Fedora for comment. He was happy that the report presented a fair picture of his team's work.

A number of reports, either recently completed or due to be completed in the next few months require peer review and this is being arranged:

- D-D11 Report on extraction of personal metadata from portal etc.
- R-D11 Final report on contextual metadata
- R-D14 Final report on user needs analysis
- D-D13 Report on feasibility of automatic extraction of object descriptive metadata
- R-D15 Investigative report on DRM implications
- D-D15 Systems documentation
- D-D16 User documentation

The workflow tool itself will be unit tested using J-Unit.

# 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 and wiki |
| October 2006 | Fedora UK & Ireland group meeting, Glasgow |
| October 2006 | Open Scholarship 2006 conference, Glasgow |
| January 2007 | Open Repositories 2007, San Antonio. The project presented a paper and a poster session |

A report by Chris Awre and Richard Green of the summer Fedora User Group Conference in Charlottesville, VA, was published in the July 2006 issue of Ariadne.[3]

[3] http://www.ariadne.ac.uk/issue48/fedora-users-rpt/

The RepoMMan team has had an article accepted for the OCLC Systems & Services International Digital Libraries special issue on institutional repositories.

During the final few months of the project we shall continue our ongoing dissemination work. This will include following up contacts that have recently been made at OR07 in San Antonio. We shall also continue to contribute to events in the UK, JISC-arranged and otherwise. These will include:

| | |
|---|---|
| February | Content packaging workshop in Bristol |
| February | JISC meetings on SoA and workflow in Birmingham |
| March/April | Third meeting of the Fedora UK & Ireland Group in Oxford |

The two things we should like to point up in our presentation at the proposed Programme Closing event in June are:

- the RepoMMan tool itself, its functionality and use within a University SOA framework
- the applicability of a repository in underpinning a wide range of institutional activities

## 20 Exit/sustainability plan

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

The following projects in the Repositories and Preservation programme may be particularly interested in some of our work:

- Complex Archive Ingest for Repository Objects (CAIRO)

  CAIRO may well be interested in our experiences with automated metadata and in the basis of the RepoMMan tool itself.

- Defining Image Access

  The DIA project may be interested in the image content models that are being developed for the University of Hull and the idea of "asset actions" being developed in the Fedora community.

# Appendix A: Budget June 2005-May 2007

| | YR1 Budget | YR 1 Spend | YR1 Balance | YR 2 Budget | YR 2 Spend | YR2 Forecast | Budget Total |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Staff** | | | | | | | |
| Technical developer 1.0fte | | | | | | | |
| Project manager 1.0fte | | | | | | | |
| Technical consultant 18+7d | | | | | | | |
| | | | | | | | |
| **Travel & Subsistence** | | | | | | | |
| **Equipment** | | | | | | | |
| **Dissemination** activities | | | | | | | |
| **Evaluation** activities | | | | | | | |
| **Other** | | | | | | | |
| Consumables + | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| **Total from JISC** | | | | | | | |