

# JISC DEVELOPMENT PROGRAMMES

## Project Document Cover Sheet

### FINAL PROJECT REPORT: September 2007

#### Project

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

#### Document

<b>Document Title</b>	Draft final project report		
<b>Reporting Period</b>	June 2005- September 2007		
<b>Author(s) &amp; project role</b>	Richard Green Project Manager		
<b>Date</b>	2007-10-23	<b>Filename</b>	final-rep-v10.doc
<b>URL</b>			
<b>Access</b>	<input type="checkbox"/> Project and JISC internal	<input checked="" type="checkbox"/> General dissemination	

#### Document History

<b>Version</b>	<b>Date</b>	<b>Comments</b>
0.1	30/01/2007	Draft version
0.2	26/06/2007	Updated draft for e-Learning survey
0.3	28/08/2007	First draft for NJ
0.4	31/08/2007	Draft submitted to NJ
1.0	23/10/2007	Submitted version



# RepoMMan Project

---

Final Project Report

June 2005 - September 2007

Richard Green, Chris Awre, Ian Dolphin, Simon Lamb, and  
Robert Sherratt

October 2007



## The RepoMMan Project

<b>Project Director:</b>	Ian Dolphin, Head of e-Strategy, University of Hull (i.dolphin@hull.ac.uk)
<b>Project Manager:</b>	Richard Green (r.green@hull.ac.uk)
<b>Technical Lead:</b>	Robert Sherratt (r.sherratt@hull.ac.uk)
<b>Repository Domain Specialist:</b>	Chris Awre (c.awre@hull.ac.uk)
<b>Lead Software Developer:</b>	Simon Lamb (s.lamb@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.

## Table of contents

Executive summary	5
Background	6
Aims and Objectives	7
Methodology	7
Implementation	10
Outputs and Results	16
Outcomes	17
Conclusions	19
Implications and recommendations	20
Appendix 1: Glossary of acronyms and terms	22
Appendix 2: Summary of RepoMMan reports and documentation	25

## Acknowledgements

The Repository Metadata and Management Project (RepoMMan) at the University of Hull was funded by the JISC Digital Repositories Programme. The project was carried out by the University's e-Services Integration Group (e-SIG) within Academic Services working with sub-contractors Richard Green (IT consultant) and Warwick Bailey (Icodeon Limited).

We acknowledge with thanks the help and assistance given us by the digital repositories community world-wide. In particular we should like to acknowledge the contributions made by the Fedora team (especially Thornton Staples, Ross Wayland and Bill Niebel at the University of Virginia, also Sandy Payette and Chris Wilper at Cornell University), Eddie Budgen (Active Endpoints Inc), David Groenewegen (Arrow Project, Monash University), Walt Howard (UC Riverside, University of California), Christiaan Kortekaas and Matthew Smith (Fez Project, University of Queensland), Leslie Johnston (University of Virginia Library), Peter Murray (OhioLINK), Chi Nguyen (Macquarie University), Matthias Razum (eSciDoc Project, FIZ-Karlsruhe), Ryan Scherle (Indiana University), Andrew Treolar (Monash University), Paul Vander Griend (formerly UC Riverside, University of California).

We also acknowledge particular help and assistance provided by other members of staff at the University of Hull: Carl Barrow, Richard Garbutt, John Higham, Mike Park, Gary Thompson, and Wayne Thompson in the e-Services Integration Group, and David Jarvis, Alex Sharaz, and Craig Stephenson in Computing Services.

## Executive summary

The RepoMMan project was funded by the JISC and carried out by the e-Services Integration Group at the University of Hull between June 2005 and September 2007. The project set out with plans to better understand how an institutional repository might fully support the needs of a university in all its activities; establish generalised workflows for the development of digital objects in a repository and to develop a browser-based tool for managing these; and to investigate ways of automating the production of metadata to support repository content with a view to improving metadata quality for the purposes of discovery. These objectives were substantially met despite a range of obstacles to development work. During the project the project team established strong links with the academic repository community world-wide and have benefited from, and contributed to, a range of work being undertaken within it.

The most tangible output from the project is the so-called 'RepoMMan tool'. This is a browser-based interface that allows a user to interact with a closed, private space in their local digital repository. This space can be used in the manner of a digital vault whilst materials are being developed; by materials we mean almost anything that can be stored in digital form - texts, images, databases, data sets, multimedia files, and so on. Files are securely stored in the University storage area network and are available from anywhere that the user has access to a browser. Future versions of the tool will allow a user to share these materials with a small number of other users for the purposes of collaboration.

When materials are ready to be put into the digital repository so that they can be discovered by others (perhaps the general public, or perhaps a more restricted group of repository users) the RepoMMan tool can be used to add metadata to them. Metadata forms the basis for subsequent search and discovery by others. Crudely, the better the metadata the better the chances of the material being discovered and used by others. In the past, many repositories have had trouble getting their authors to provide good metadata to go with their outputs. The RepoMMan project set out to try and automate this process.

The tool, as delivered to the JISC, extracts information about the author from the secure system in which (s)he is working to deposit his materials. Because (s)he has logged in, the system can supply such details as the author's name and it is unnecessary to ask them to type it again. The tool is also capable of analysing text objects saved in pdf format and extracting from them metadata describing the subject matter. The author has the ability to change this metadata should they so wish. The project has also investigated the production of technical metadata and metadata related to long term digital preservation of these materials; these features are not required by the tool developed for the JISC but will be added into future versions which deal with publication and long-term management of authors' materials.

The RepoMMan team have produced a range of reports about their work during the course of the project and these are available from the project website at [www.hull.ac.uk/esig/repomman](http://www.hull.ac.uk/esig/repomman). These cover technical aspects of the work, the extensive user needs analysis undertaken and work undertaken to understand better the possibilities offered by the Fedora repository software that underpins the project's work. The team has contributed to a number of conferences world-wide and has written articles for a number of publications.

This document describes in some detail the methodology brought to the project, its implementation, and its outputs and results. It ends with a number of conclusions and implications.

## Background

From Callimachus to Dewey, the Library community has attempted to provide easy, or at least organised and rational, access to scarce resources. Whilst changes in printing and publishing have certainly modified the degree of scarcity, the artefacts a library houses remain necessarily limited. They are limited by physical bounds, by financial considerations, and frequently by the context in which they may be used.

By whatever shorthand the emerging "information age" is referred, it applies a fundamentally different set of parameters to issues surrounding academic and administrative publishing and access. The information age brings the challenge of organising access to abundant, rather than scarce, resources. In the current exploration and take-up of digital repositories by universities, this change of emphasis has required considered attention. Certain solutions, which applied in an earlier age of scarcity, may not be particularly suitable when set in the context of information abundance. Institutions must be prepared to combine the ability to learn from previous experience with an appropriate measure of challenging assumptions.

The growing interest in, and development of, a variety of local, regional and national repositories is one response to the abundance of digital materials. Much of the promise of institutional repository services, and a large part of the business case for such services, lies in their enabling the maximum use and re-use of rapidly growing digital asset collections. Population of these repositories comes about through a sequence of events, a workflow, of which the deposited digital asset is but an end result. It is important to maintain the perspective, however, that development of institutional repositories forms part of an emerging and connected national (and potentially international) repository infrastructure. Given this perspective, and the significant number of institutional repository uses which may require interfaces to disparate systems, flexibility and support for interoperability standards and common specifications are not an optional extra, but essential criteria for a repository framework. This point is further emphasised as the general movement towards Service Oriented Architectures within Higher and Further Educational Institutions gathers pace.

Whilst there is little shortage of technical and process-oriented obstacles to repository take-up and use, few feature as strongly in end-user comment as those surrounding the generation of metadata. Effective metadata underpins effective discovery, location and potential re-use. The creation of metadata, however, is both time consuming and laborious. This factor has stimulated the growth of a range of project and development activity aimed at the automation of metadata creation. Generally, these methods revolve around the extraction of key information from digital content itself. Although much of this activity might be classified as "R&D", some of these methods have escaped from the laboratory to the production environment.

Users of institutional systems should not need repeatedly to enter information regarding their role or preferences, particularly where this is already stored as part of an enterprise directory or portal deployment. Whilst this objective is frequently identified in business process redevelopment, it is less frequently applied to the creation of metadata. After authentication, applications should ideally access and communicate such information as is required to provide a user with a seamless experience tailored to their need, role and preferences. Once stored, this information should be available to be consumed by other services, such as those underpinning an institutional portal, Virtual Research environment (VRE), Virtual Learning Environment (VLE), or a desktop application forming part of a Personal Learning Environment (PLE). By effectively "warehousing" person and role information from a variety of back-office systems in a portal profile or enterprise person directory, a potentially rich source of contextual metadata is created. This information may range from the simple name/email/organisational unit of the creator to enabling a choice between modules taught, potentially each with its own keyword profile. Applying this information to the creation of metadata after authentication to an institutional system such as a portal, VLE or PLE, by, for example, pre-population of editable fields, has enormous potential to both increase accuracy and decrease repetition in metadata creation.

It is against this background that the RepoMMan project was conceived.

## Aims and Objectives

RepoMMan set out with the following aims and objectives:

- to gather data supporting greater understanding of user-needs, locally and potentially from a wider audience, requirements, processes and behaviour across a range of potential institutional repository uses. A particular focus would be on how users currently interacted with systems and data that might be considered repositories in nature if not by name. The data gathered by these means at regular intervals during the lifecycle of RepoMMan would both underpin development, and form an essential component of the IDEAL project management process that we used. In addition the data gathered would provide a valuable resource for the Higher and Further Education community.
- to take a combinatory approach to automated metadata creation. Where robust technology exists to extract descriptive metadata from simple digital objects (for example, office documents and certain types of digital image), RepoMMan would deploy and validate its use. This would be combined with a novel approach to the creation of contextual metadata, which would draw on the experience of the University of Hull institutional portal deployment.
- carefully to evaluate available contextual information elements which might meaningfully be drawn together in this manner, and suggest a number of implementation routes based around the creation of a range of profiles, including role based "Personal Metadata Profiles". RepoMMan would suggest implementation routes suited to institutions both with and without portal implementations.

Not listed, but rather implied, in the Project Proposal and the Project Plan is a fourth objective which might usefully be made explicit here:

- to establish generalised workflows for the development of digital assets in the spheres of research, learning and teaching (L&T), and administration and to develop a browser-based tool for managing these.

## Methodology

### General

Technology-related projects have a disturbing tendency to seek information regarding user-needs and requirements after the event of software development. The University of Hull has placed great emphasis on acting to reverse this. By a combination of traditional interviews and an innovative web-based survey a considerable amount of user requirements data has been gathered for a range of purposes. The CREE Project survey, for example, which solicited detailed information regarding user search behaviour, obtained over 2,000 responses from the HE and FE community.<sup>1</sup> This user-needs driven approach was continued in the RepoMMan Project.

The University of Hull has continued iteratively to develop a General Information Strategy, following its initial introduction as part of a JISC initiative in the late 1990s. An iteration, in the summer of 2004 shortly before the RepoMMan Project started, identified the development of repository infrastructure as a critical strategic architectural element. Subsequent investigation and planning identified four overlapping repository development areas or strands:

---

<sup>1</sup> Cree Project Survey At: [www.hull.ac.uk/esig/cree](http://www.hull.ac.uk/esig/cree)

- Learning objects
- Images
- Research artefacts and research outputs
- Meeting the needs of the Freedom of Information Act

RepoMMan undertook investigative work across all these four areas or strands, identifying user requirements based on (then) current and potential use in these areas, and validated the applicability of drawing contextual metadata from a range of institutional systems against the elements of common metadata schema. The RepoMMan tool that was subsequently developed is potentially applicable to repository development and use in all the areas listed above. Work was also carried out to prototype the deployment of the tool in a Collaborative and Learning Environment (C&LE), which can be used to support research and L&T collaboration (as a VRE or VLE, respectively). The Universities of East Anglia and Hull have deployed such an environment (the Sakai C&LE) to meet the collaborative needs of Humanities research under the JISC VRE Programme. The alignment of these two distinct activities will bring further substantial benefits for the community.

As may be inferred from the initial deployment priorities indicated above, key institutional criteria included considerable flexibility and support for standardised interfaces (based around Web Services specifications). Three repository systems were considered prior to the start of the RepoMMan project. Despite the considerable strengths of DSpace, particularly as an institutional repository solution 'out of the box', lack of versioning together with questionable scalability effectively precluded its use. EPrints was similarly rejected due to the lack of proven flexibility for use with the wide range of institutional assets and scenarios identified and the technical platform. The considerable scalability offered by Fedora, together with ingest and export of information in standardised XML formats and documented Simple Object Access Protocol (SOAP) and Representational State Transfer (REST) interfaces at both object and repository level made it a logical choice as a basis for initial investigation and deployment. RepoMMan validated this judgement and provided a report on its experiences with Fedora after the first year of the project.<sup>2</sup> Although Fedora presented a flexible repository framework, its successful deployment for any of the purposes indicated above relied on the provision of an equally flexible framework to support workflow. This would enable the repository framework to be adapted to meet the requirements of existing University processes, together with future needs highlighted by the user-needs analysis and process mapping activities.

## Research strand

In considering the potential use and benefits of repositories within an institutional context it is important to recognise that many of the activities that will and can be undertaken within a repository already take place on an everyday basis. Users already interact with a range of data stores, even if they are not always aware of this or how the data is managed and/or structured behind the scenes. File stores, local library and archive collections, image collections, network resources etc are all accessed by users as part of their learning, teaching, research or administration activities. This existing use can help to inform the nature of how an institutional repository should be presented to users, in order to facilitate its take-up and use alongside existing services. An institutional repository also, of course, has the potential to offer additional functionality and services and it is essential to identify user requirements above and beyond what they can currently do to understand how such services could be implemented.

A user requirements analysis was undertaken to discover what types of data file were used at Hull and elsewhere. A web-based survey was carried out to gather information from potential users in the research arena.<sup>3</sup> This was used to inform initial repository development, and specifically the implementation of research use cases. The survey was complemented by a series of detailed interviews with a cross-section of research staff to identify current and

---

<sup>2</sup> See: <http://www.hull.ac.uk/esig/repomman/downloads/D-D8-fedora-exp-v10.pdf>

<sup>3</sup> See: [http://www.hull.ac.uk/esig/repomman/downloads/R-D3-research\\_survey\\_data\\_11.pdf](http://www.hull.ac.uk/esig/repomman/downloads/R-D3-research_survey_data_11.pdf)



potential uses of repositories more closely.<sup>4</sup> At a later stage of the project a similar requirements analysis was undertaken with members of the L&T community and staff who had an administrative role.<sup>5</sup> This user requirements work was eventually synthesised into a single document.<sup>6</sup>

Metadata can take various forms, including technical metadata about the object itself, administrative metadata for contextual and/or local management purposes, and descriptive metadata describing the content. Manual generation of this metadata has revealed difficulties in ensuring that a full and accurate metadata record is stored with the object, limiting future use. At the beginning of the project there was considerable interest in automatic metadata generation within the repositories community, and in the use of metadata profiles that could be automatically attached to an object when it is deposited; this interest continues. The use of locally held information can offer a wide range of preset, and quality-assured, metadata and RepoMMan has investigated what requirements there are for this and how such profiles might be created.

The main thrust of work related to 'personal' metadata has focussed on the feasibility of extracting it from the user's context: generally from the environment variables available within a portal or C&LE. Additional, but less detailed, work considered how this might be extended to make use of an Identity Management System and/or additional 'mini-profiles' generated and stored within the user's repository space. These mini-profiles would allow storage of research-project-specific information which could be brought into metadata in the absence of a central Research Management System. This approach would easily be adapted for users in L&T or administration who have multiple roles. The study included consideration of data protection issues.

The production of any data or document automatically generates a range of data integrity factors that address the validity and authoritativeness of the data itself. All data has associated rights and the ongoing management of those rights is needed to ensure protection of usage; all data comes from somewhere and this provenance is key to a user being able to trust it for use in their study and/or work; and different data will have different levels of quality, and it is imperative that a user is aware and can be assured of the quality of what they are using. In considering metadata profiles, and what metadata might reasonably be stored, the project has addressed all three of these areas.

A literature review was carried out toward the end of the project to ensure that the most up-to-date developments in DRM, provenance and trust were taken on board.<sup>7</sup> Specific emphasis was given to the issues that arose in these areas from the user requirements analysis and investigation into personal metadata profiles.

## Technical development strand

Workflow issues have been recognized for some time as part of the articulation of Service Oriented Architecture. Central to this concept is the ability to address the orchestration of a number of discrete Web Services providing a coherent interface for the end user. These issues are currently being addressed by a number of standards development activities. Working on similar problems in a different domain space (e-Learning), the JISC ASSIS project<sup>8</sup> identified the Oasis managed Web Services for Business Process Execution Language (WSBPEL, or just BPEL) as an appropriate standard on which to base development activities and it was this that the RepoMMan Project adopted.<sup>9</sup>

A tool was developed to separate user experience from repository services, shielding the user from the complexity required to complete requests, and providing a simple but flexible task-

---

<sup>4</sup> See: <http://www.hull.ac.uk/esig/repomman/downloads/R-D4-rsch-int-data-11.pdf>

<sup>5</sup> See: <http://www.hull.ac.uk/esig/repomman/downloads/R-D12-TLA-user-needs.pdf>

<sup>6</sup> See: <http://www.hull.ac.uk/esig/repomman/downloads/R-D14-full-user-needs1.pdf>

<sup>7</sup> See: <http://www.hull.ac.uk/esig/repomman/downloads/R-D15-DRMreview.pdf>

<sup>8</sup> See: [http://www.jisc.ac.uk/whatwedo/programmes/programme\\_edistributed/deletassis.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_edistributed/deletassis.aspx)

<sup>9</sup> See: [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=wsbpel](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel)

based interface. In responding to user actions a 'workflow engine' coordinates requests with services defined in the workflow. The workflow process can deal with the invocation of any of a number of web services, combining them according to need.

The Fedora architecture has been designed to allow for individual repository functions to be called via its APIs. Both the Fedora Management API (API-M) and Access API (API-A) can be accessed using WSDL definitions over SOAP. The creation of a workflow engine could therefore use BPEL to define calls to one, or a number of, repository functions using Web Services standards.

The adaptation and deployment of a workflow engine of necessity adopted a layered model with separate tiers for presentation and process control. The presentation layer is created as Java servlets, allowing deployment of the workflow engine on Java application servers, and JSR-168 portlets that can be used with any conformant portal framework or JSR-168 conformant container (the University of Hull uses uPortal 2.5<sup>10</sup>). Adobe 'Flex' is used to dispatch requests and responses between the presentation and process control tiers. (The use of Flex is further considered on page 13.) Finally, the process control layer uses the Active Endpoints WSBPEL runtime environment to call services and process responses according to the pre-defined workflow definitions.

Technical work around metadata population focused in two areas. The implementation of pre-populating object entry forms with a contextual metadata profile has been provided via both JSR168/WSRP conformant portal framework, and with the Sakai Collaborative & Learning Environment. The core of this work is based around personal information warehoused within a portal profile mapped to an appropriate metadata scheme, Dublin Core (DC) for the purposes of demonstration. RepoMMan also conducted a summary investigation on the feasibility of deploying existing solutions to the extraction of metadata from simple objects, both technical and descriptive. A number of possible solutions exist for deriving technical metadata, tools to provide descriptive metadata proved more elusive. Ultimately this work centred around 'Data Fountains' produced as a collaboration between the University of California (Riverside) and the National Science Digital Library in the US.<sup>11</sup> Again, for the purposes of demonstration, this metadata was mapped into a DC representation. This work is further considered in the next section.

The technical strand of RepoMMan adopted a test-led approach to development. This ensured that the testing of software was an integrated part of the overall process and also led to better conformance with standards for interoperability. External evaluation of outcomes has been achieved where possible by releasing findings and tools to the community, especially the Fedora community, for review. Generally the feedback from the community has been extremely positive.

## Implementation

As noted above, the RepoMMan Project began with three fundamental decisions already taken: that the project would be user-needs driven, that the work would be based around the Fedora repository software, and that this would be combined with service orchestration using BPEL. Whilst the first two of these decisions were straightforward, the third was not. It was necessary to start the project by establishing the appropriate BPEL engine for the project to adopt from a number available, some free of charge, others quite expensive. The available candidates were considered and a decision was taken to adopt the Active Endpoints BPEL engine.<sup>12 13</sup> With the benefit of hindsight it is good to be able to report that this was a sound choice. The Active Endpoints software has served us well and the company itself has been very supportive of what we have been trying to do. This despite the fact that our workflow

---

<sup>10</sup> See: <http://www.uPortal.org>

<sup>11</sup> See: <http://dfnsdl.ucr.edu>

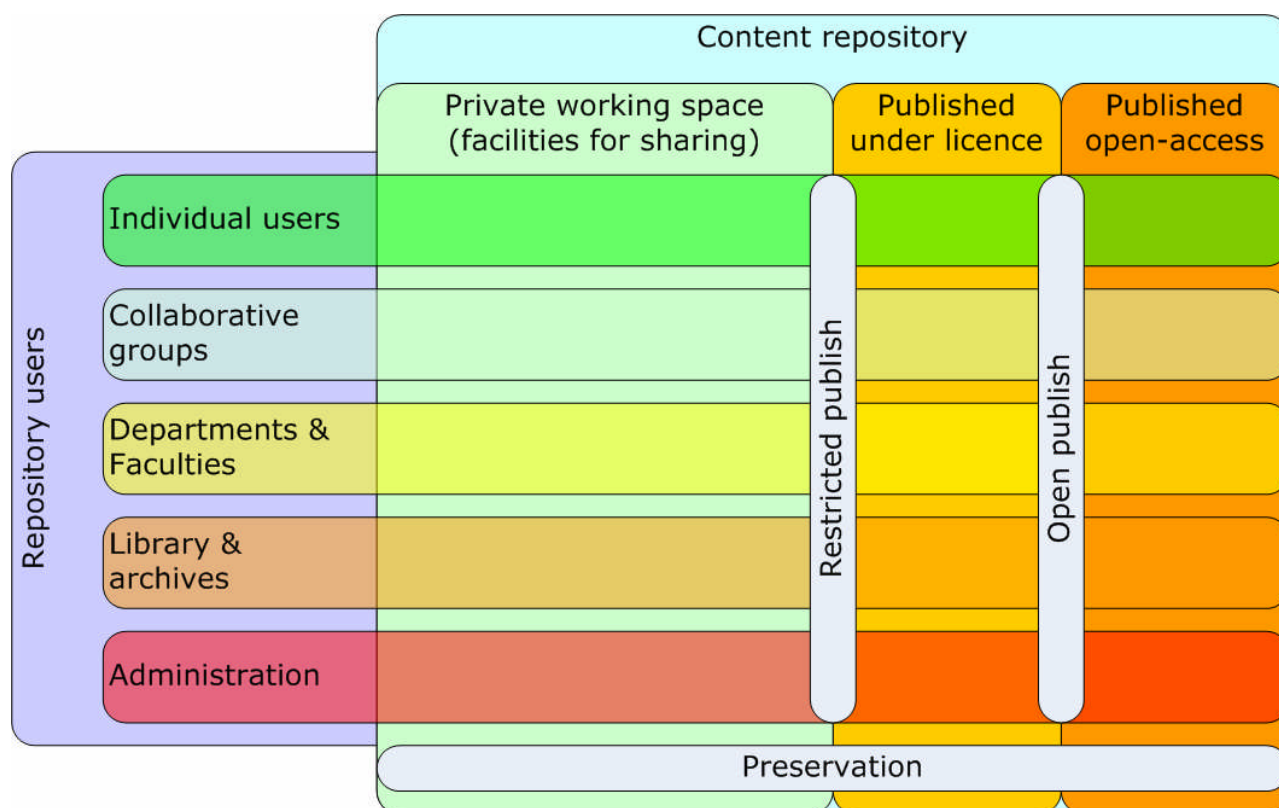
<sup>12</sup> See: <http://www.hull.ac.uk/esig/repomman/downloads/D-D1-BPEL.pdf>

<sup>13</sup> See: <http://www.active-endpoints.com/active-bpel-engine-overview.htm>

tool is based on the open source implementation of their work, rather than the commercial 'for profit' version.

The project's initial user-needs analysis was aimed at potential research use of a repository; it took two parallel tracks. On the one hand, a web-based survey was developed; this was initially deployed for respondents from the University of Hull only and then opened out to the Web. On the other hand, a set of user-needs interviews were arranged with a cross section of researchers from the University. These interviews took typically slightly over an hour, were recorded and transcribed verbatim. From the transcripts a set of generalised user-needs scenarios was developed. The on-line survey investigated the way that researchers currently 'did research' and how they managed their work and materials. The interviews covered the same ground but went into greater depth and additionally considered how researchers might use a repository were it available to them. This last point considered how a repository might be used during the production of research materials rather than just as a place to deposit finished work.

The outcome of the interviews was a clear understanding for the project of the repository facilities that might usefully be provided at Hull to support research. These facilities covered the development as well as the exposure of research work. A diagram was developed which has since been used as an integral part of presentations at conferences around the world:



Richard Green 16/01/07  
 ©2007 RepoMMan Project,  
 e-services Integration Group,  
 University of Hull

The vertical axis envisages that the repository may be used by individuals, collaborative groups and a number of organisational teams. (Researchers are clearly part of the first two groups and may also be part of the organisational teams.) The repository itself would have two working areas: a 'private' working space where materials are developed by an individual or team, and a 'public' space where materials may later be exposed either in a limited way, perhaps due to licensing or copyright issues, or more generally. The level of access in the public area would be distinguished by permissions associated with the object in metadata. Underlying all this is an understanding that preservation of digital materials in the repository is an important consideration. Given the wide-ranging nature of this view, an Institutional

Repository Liaison Group was formed, which could, throughout the project, offer a broader perspective on development issues than might be present in the Project Team alone.

The RepoMMan tool has been designed primarily to interact with the repository as represented in the left-hand half of this diagram; in other words, the tool facilitates interaction with the private space. Some researchers could see clear benefits in using a repository as a development tool; others were content to work as they had in the past but were generally open to the idea of depositing the finished materials. Those that were keen to use the repository for development identified four main areas of benefit. (Bear in mind that the interviews were conducted with University of Hull staff which inevitably introduces an institutional bias.)

- **Storage:** the repository could be used as a digital vault where materials were put for safe keeping at the end of a day. A major advantage of this would be that the materials would be part of the University's normal backup regime and thus researchers could be less concerned about the need for personal backups. (Historically, the University has offered only very limited storage space on its network to staff with the effect that many staff find it inadequate to their needs and ignore it.)
- **Access:** materials stored in the repository could potentially be downloaded and used at any location with web access (perhaps at a conference, in a lecture theatre, at home, or ...)
- **Management:** the repository software can manage issues such as versioning of materials (all materials are automatically versioned and kept), and sharing with a small group of collaborators (including the potential for some kind of locking to prevent conflicting edits)
- **Preservation:** all materials are kept, unless deliberately deleted, and thus a complete audit trail is maintained. When, if, the materials are published all versions are available and a considered decision can be made about which should be kept. The materials then move into a process which includes consideration of formal records management and preservation.

A small number of those that we interviewed (both in this first round, with researchers, and in a subsequent round with members of the L&T and administrative communities) could see that there was a further potential advantage to using the repository: that, at the time of making the material generally available, the RepoMMan tool could automatically generate considerable metadata with which to pre-populate the repository's digital object.

Administrators, in particular, were very interested in the potential for management and preservation (amongst them the University's Records Manager involved in developing systems in response to the Freedom of Information Act). This interest and the discussions that flowed from it and within the Liaison Group were in part responsible for the University submitting a funding proposal to the JISC which resulted in a related project, REMAP<sup>14</sup>

The user-needs analysis provided the background against which the RepoMMan tool would be developed and allowed an initial outline design to be developed. The design criteria that we developed have been described at a number of national and international events and have generally been well received; this has given us another, informal, level of validation for the work.

Whilst the user-needs work was going on, the development side of the project was coming to terms with the Fedora software and the implications for the University of using it. Fedora is a powerful and flexible system which, if it is to be used effectively, requires quite a high level of planning. In addition, the first stages of BPEL deployment were being developed with a view ultimately to creating a 'thin vertical slice' of functionality through the then proposed three-tier model of presentation, MVC and process control. All this work came together in the early Spring of 2006 as we started on the 'thin vertical slice'.

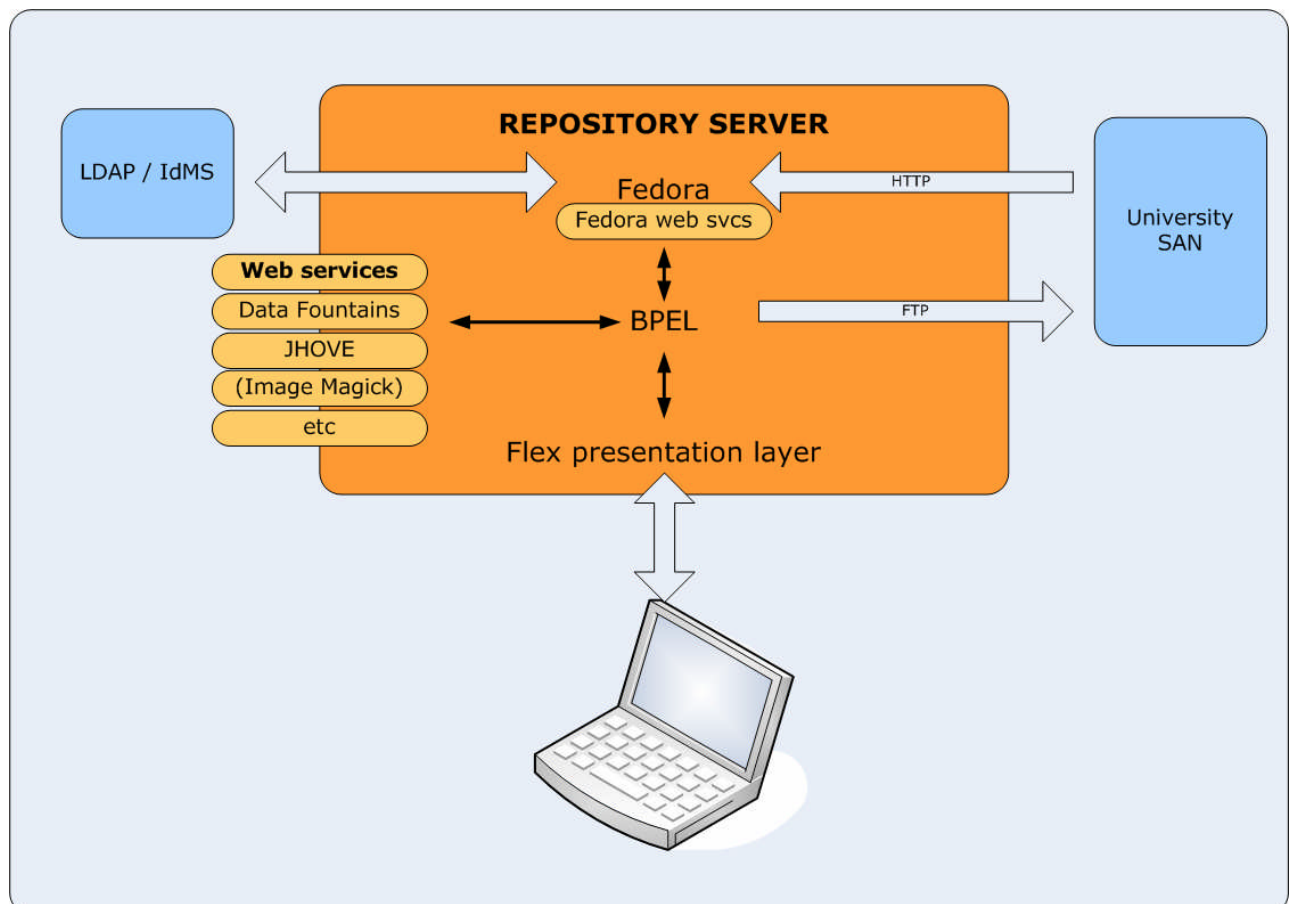
---

<sup>14</sup> <http://www.hull.ac.uk/remap>

As things transpired, we at Hull were the first group worldwide to try to use Fedora’s Web Services API extensively in a standards compliant fashion; the work brought together Fedora, Apache Axis and the Active Endpoints BPEL engine. This novel combination could not be made to function reliably together. We spent some time involved with the Fedora development team and the Fedora community more generally to try and solve the problems. A company in London, Rightscom Limited, managed to replicate our problems with Fedora and the Fedora development team decided that the solution was to rewrite their Web Services. The version we had been working with had been developed using the rpc/encoded convention; the new version would use the now more commonly deployed document/literal style. The new system solved our problems and enabled us to create our first instance of a working three-tier system.

Expressed like this, in ten or so lines of text, it may seem that these problems were relatively easily solved. In fact it was a long process involving many people world-wide and it had the effect of delaying our development work by several months. We should note that our active involvement with the Fedora community world-wide helped speed the resolution of the problem; in general, both the Fedora development team and members of the Fedora community generally (through the Fedora discussion lists) are prompt and helpful with responses to queries.

During this work it became apparent that the project would be better served by having a different delivery model to the one envisaged in the project plan and tested in the ‘thin vertical slice’. The project team decided that a richer browser interface was required than was available through a traditional forms-based approach. Following an evaluation of possible software solutions and existing in-house skills, the RepoMMan interface moved to an implementation in which Adobe ‘Flex’ is used to manage the user’s display and to interact with the BPEL engine and subsequent service calls. Further development would easily allow the use of other web presentation technologies such as JSP or Velocity. The diagram below gives an outline of the interactions between the various elements of the RepoMMan work and the broader repository infrastructure:



Whilst this development was going on, a parallel effort was planning the structure of the repository at Hull and investigating the facilities that Fedora provided for managing the security of digital objects within it. Once again, we found ourselves at the forefront of Fedora's use when we were the first to try and use its (then new) XACML security implementation. XACML is a declarative access control policy language implemented in XML; at the time of writing, version 2.0 is current, ratified by the OASIS Standards Organisation,<sup>15</sup> version 3 is under development.

There were essentially two elements to this parallel work: developing and describing a structure for the repository which included the range of objects that we might encounter, their internal structure and the relationships between them; and secondary to that attaching security to various elements in that structure, at object level or within objects, in order to control how, when and to whom they might be exposed. Much of the structural work related primarily to the 'public' area of the repository which strictly was beyond the remit of the RepoMMan project, however it was necessary to understand this structure so that materials developed with the RepoMMan tool could easily be moved into it. Digital objects within the 'private' area have simpler security needs and potentially a much simpler structure.

Draft content models were developed for text and image objects within the University of Hull repository. These models describe the various datastreams that are included in the Fedora objects, both for their 'digital payload' and for associated metadata. Considerable research time was given over to metadata in order to establish what type(s) of metadata would need to be held and what schemas would be used to represent it. On the basis of this work, the University took a decision in principle to use a metadata schema developed by the University of Virginia (UVA) which has the advantage that it has a wide compass but can easily be mapped down onto other schemas, most notably the Dublin Core. The RepoMMan project worked only with this Dublin Core representation, as proof of concept. We shall return to the subject of metadata shortly.

Having established draft content models, we tried to use Fedora's XACML engine to manage security constraints around them. This testing revealed that the system behaved somewhat unpredictably; it transpired that a number of the example scripts shipped with Fedora were mal-formed and our attempts to use them, or variants on them, could never work. This problem was eventually identified and resolved, however not without the expenditure of considerable effort and time.

With the basic XACML apparently in place we extended our work to use user information from local systems with the XACML security. This involved having Fedora authenticate a user against a local LDAP system, derive any user attributes from it, and then test these attributes against the security model for the material the user was trying to access. On the surface of it, Fedora 'out of the box' supported LDAP authentication; in the event, we found we were, again, the first people to try a deployment in this way. This may seem a curious statement, but it can be explained by considering that most Fedora repositories then in operation dealt only with materials made available to 'the public' in some way and thus required relatively little security. There could potentially be one 'public' user and a number of administrators with greater privileges. This small group of users would easily be dealt with in Fedora's basic 'Tomcat' security module. At Hull, on the other hand, we were envisaging several thousand users of our 'private' space, hence the need for LDAP and a move into relatively new territory as far as Fedora was concerned.

At the UVA, the Fedora installation used LDAP in this way: primary authentication was done using the 'Tomcat' module, control then passed to the LDAP module where the already authenticated user was checked for additional attributes. In other words, LDAP was not the primary authentication mechanism. In the event, after extensive interaction with the Fedora team, it was agreed that the LDAP module did not work correctly, could not easily be repaired to work in the way we needed, and that a better solution was to wait for the release of Fedora 2.2, in January 2007, where a completely new LDAP security implementation would address our needs. In the event the new LDAP filter did not work either. In June 2007 a working

---

<sup>15</sup> See: <http://xml.coverpages.org/xacml.html>

patch was produced. By this time, however, the RepoMMan team had discovered and was testing a different filter developed for Fedora as part of the DRAMA Project in Australia. This filter fits into a much broader scheme which should provide a much more flexible security system for Fedora than that currently shipped as an integral part of the software; it is likely that Hull will adopt the DRAMA work if it performs to specification.

Fedora cancelled plans for versions 2.3 and 2.4 and is moving directly to 3.0 anticipated before Christmas 2007. This means that multi-ownership of objects (sharing), which was to be introduced in these intermediate versions, will not now be possible within the timespan of the RepoMMan Project and the functionality will need to be added later.

The next area of focus for the project was metadata. As we have already described, it was the intention of the RepoMMan project that its repository tool should be able to draw on two types of source to pre-populate the metadata for a digital object at the point an author decides to make it more generally available. The first of these types is contextual metadata, derived from the University systems and environment; the second is descriptive and technical metadata automatically derived from the data payload of the object itself.

Once the basic tool was deployed in a portal environment it became relatively easy to extract contextual personal metadata and supply it to the metadata screen. The user is logged into the portal and this has already drawn down a range of personal information from other University systems. The project has briefly considered the possibility of storing additional personal metadata (related to specific research projects, say, or to specific taught courses) as objects within the user's repository space to be called on as necessary. The idea is that the user would provide this information once and then be able to draw on it again at will through automatically generated drop-down menus on the metadata screen.

The RepoMMan team spent some considerable time looking for tools that could provide metadata relating to the digital content of a repository object. On the one hand this meant somewhat technical metadata (eg file size, image dimensions, ...) and on the other descriptive metadata relating to textual objects. The search for a tool that would fulfil this last requirement has been likened to the search for the Holy Grail.

Useful technical metadata can be extracted from most objects by a tool called JHOVE from Harvard University.<sup>16</sup> This will be used as appropriate within the RepoMMan environment. It should be noted that JHOVE is widely used for providing preservation metadata and this will be followed through in the new REMAP Project.

Considerable time and effort went into tracking down a tool that could provide effective descriptive metadata. This involved desk-research and not a small number of conversations with colleagues at various conferences worldwide. Eventually the search was narrowed down to two possibilities. First found was 'Kea' from the New Zealand Digital Library.<sup>17</sup> This is clearly a useful tool but requires a controlled vocabulary to be really effective. Hull's vision of a repository used for development work across many uses and disciplines did not sit well with this approach. Later we were introduced to 'Data Fountains' which was produced as a collaboration between the University of California (Riverside), the National Science Digital Library in the US, and the US Institute of Museum and Library Services. One component of the Data Fountains toolset, the iVia metadata tool proved to be closer to our needs and was used as proof-of-concept for our descriptive metadata population work.

A brief evaluation of the metadata quality from the iVia tool is being undertaken jointly with the DEST-funded Arrow Project in Australia.

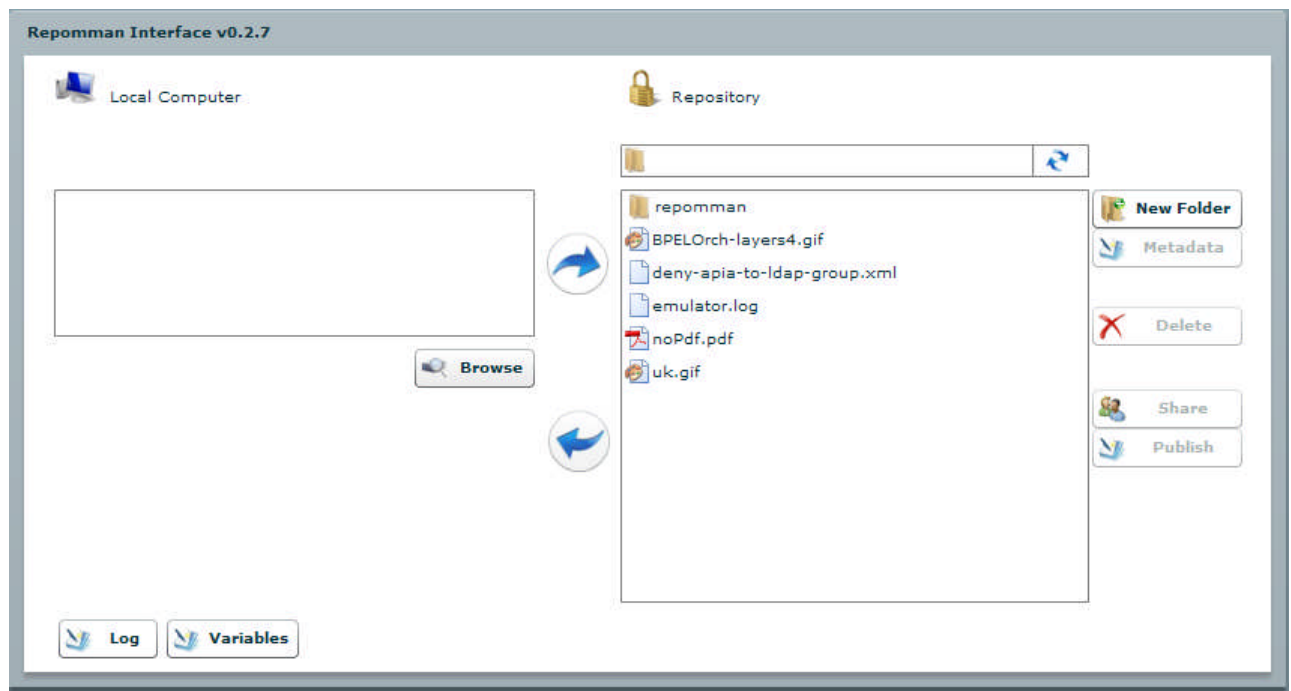
---

<sup>16</sup> <http://hul.harvard.edu/jhove/>

<sup>17</sup> <http://www.nzdl.org/html/projects.html>

## Outputs and Results

The resulting RepoMMan tool looks like this:



The left-hand side of the screen gives a browse capability onto the user's computer. The right-hand side of the screen represents the user's private repository space presented as a conventional file store. This can contain folders (in fact, repository collections) and files (in fact, digital objects containing the file transferred). Buttons in the centre of the screen allow for upload and download. Unlike a file store, double clicking an object in the repository space provides a complete list of the versions available and the user can manipulate these as appropriate (download, delete etc.)

Quite deliberately, in many ways the tool appears to function in the manner of an ftp client, an environment that many users will be familiar with.

The metadata button allows basic metadata to be associated with an object. The form is pre-populated with contextual, technical and/or descriptive metadata where appropriate but the user can amend or add to this metadata as required. For the proof-of-concept work in the RepoMMan Project this metadata has been limited to basic Dublin Core; for deployment as a University service, a richer schema will eventually be used as noted in the previous section.

The 'Share' button will allow users to share a specific file with named other for the purposes of, say, joint authorship or peer comment. As noted earlier, limitations of the current Fedora release mean that this functionality will need to be added after the formal end of the project; rather it will now be considered an integral part of University's own repository development work.

Users are unaware of where their data is being stored. The University's implementation of Fedora is somewhat unusual in that the data payloads of the digital objects are stored not on the Fedora server, but on the University's central storage network. Management of this approach is handled by the BPEL processes. The effect is that the repository data storage capacity is not limited by the constraints of the Fedora server.

Not yet implemented in practice, and beyond the remit of this project, is a 'Publish' button with which a user will be able to place a digital object into a queue for publishing. If metadata has not already been added, it will be required at that stage. The user's digital object will be cloned and the new object will be 'owned' by the repository: the author(s) will lose editing



rights on it. The new object will be checked by repository staff before being made available to more general repository users, taking into account any licensing restrictions.

The RepoMMan tool has largely achieved what we set out to do. However, work undertaken during development of the tool has had a significant impact. Two areas deserve specific mention: reports produced for the project and influence on the development of the Fedora software itself. The reports are available in the 'documents' section of the RepoMMan website.<sup>18</sup>

Our early report comparing available BPEL engines (D-D1)<sup>19</sup> was one of the first covering this topic and attracted considerable attention: we understand that a number of projects worldwide used it as a starting point for their own survey.

As we developed our Fedora installation we became aware that the documentation supplied with the software was not, in places, terribly user friendly. We decided to write our own Guide based on our experiences. This was very well received in the community and, although now slightly out of date, it is still being used as a starting point for new Fedora users around the world. We called it "D-D4 Development of Fedora Materials".<sup>20</sup> Perhaps inevitably it has become known as our "Dummies' Guide".

Our user-needs work was documented in a number of reports which were later summarised as R-D14 "Full User Needs Analysis Report".<sup>21</sup> A number of the contributory reports, in particular the results of the on-line survey (R-D3),<sup>22</sup> have been quoted in documents and presentations produced by other projects.

The RepoMMan repository tool will be further developed following the end of the formal project. This work will take two, interlinked, forms. Firstly the tool will be deployed at the University of Hull to serve as the main link between staff and their 'private' area of the repository. In order to do this, functionality that was not completed during the project (object sharing) or that was not formally part of the project (publishing) will need to be completed. In addition, the RepoMMan work will be extended into the areas of records management and digital preservation (RMDP) through the JISC-funded REMAP Project, which will run to March 2009. All this work will be shared with the community and the University will be happy to provide assistance to other institutions who may wish to take the Fedora Web Services and BPEL approach in order to integrate their repository with other institutional systems.

## Outcomes

The RepoMMan Project Proposal listed a number of benefits that could arise from the project. These are listed below with commentary:

### *Particular benefits to JISC and HE*

- *User requirements gathering methodologies and results will be applicable for use at other institutions, and assist them in assessing their current state of development and examination of local user requirements*

As noted above, our user requirements work has been quoted by a number of other projects. Subsequent to our on-line survey of researchers, we cooperated with the CD-LOR Project to use our software for a similar survey of those in the learning & teaching community. We ourselves used these results and they have been quoted elsewhere.

---

<sup>18</sup> See: <http://www.hull.ac.uk/esig/repomman>

<sup>19</sup> *op cit*

<sup>20</sup> See: <http://www.hull.ac.uk/esig/repomman/downloads/D-D4-iterative-dev-v01.pdf>

<sup>21</sup> *op cit*

<sup>22</sup> *op cit*

- *Repository use cases will be fed back to the community and will be grounded in existing user activity, rather than purely hypothetical repository uses.*

Our use cases have been made available to the repository community. In particular they have been contributed to the collection amassed by the Repositories Research Team for UKOLN and JISC CETIS, and to the collection being amassed by Sayeed Choudhury and his team at Johns Hopkins University in the US.

- *Particular emphasis will be given in the investigation and report on DRM issues to boundary issues between personal and institutional repositories. This will add to the body of research in this area to the general benefit of the community.*

The DRM report (deliverables R-D13 and R-D15 combined) addresses the issue of personal and institutional repositories. Notwithstanding research in this area, there is still little clarity on where this boundary lies or should lie, not least because different systems are used for these respective collecting activities. Development of institutional repositories must take account of personal collections in order to capitalise on existing content and to help foster a personalised view onto institutionally-held collections.

- *The report on metadata generation covering specific requirements will be fed into the continued discussion on this issue. This will be of particular relevance to those working in the context of an institutional portal or collaborative environment. The conclusions reached will also be of benefit to any institution developing or deploying an Enterprise Directory*

The report on metadata generation (deliverable D-D13) has provided a clear view of what is possible in this area as of the end of the project in October 2007. As originally envisaged, we believe the report will be of benefit to those working in an institutional portal or similar environment.

- *Further experience of Service Oriented Architectures, with a particular emphasis on coherent presentation of discrete services for the end user.*

The RepoMMan tool involves fairly complex interaction between a user interface driven by Adobe Flex, a BPEL engine and a range of web services - some part of Fedora, others not. This work has given us a range of expertise that is as yet uncommon. We have willingly hosted a number of meetings with visiting individuals from all over the world to discuss this work.

- *Add to the growing body of experience in adopting or integrating generic workflow solutions to the needs of web services orchestration, thus contributing not only to requirements of the Repositories programme, but also the Distributed eLearning Programme (DEL), eLearning Framework (ELF) and Integrated Information Environment.*

Members of the RepoMMan team have given presentations at a number of national and international meetings about digital repositories. The most recent such meeting was hosted by the JISC to consider the eFramework; the team provided a workshop and an associated paper.<sup>23</sup> In addition, the team's work has been mentioned in a number of articles both about RepoMMan itself and other development work. The most recent, and most formal, article appeared in *OCLC Systems and Services* as part of a special issue devoted to Institutional Repositories.<sup>24</sup>

<sup>23</sup> Lamb S and Sherratt R (2007) Orchestrating Services with BPEL - In publication with JISC

<sup>24</sup> Green R, Dolphin I, Awre C and Sherratt R (2007) The RepoMMan Project: automating workflow and metadata for an institutional repository *OCLC Systems and Services* Volume 23 Number 2 pp210-215

### ***Particular benefits to the Fedora community***

- *Further, detailed investigations of how the Fedora framework can be integrated into other frameworks, such as Spring.*

The project has certainly shown how Fedora's web services can be integrated with BPEL and Flex. An important result of the RepoMMan work is the reworking of Fedora's web services for the benefit of the Fedora community world-wide. We have also been at least partially responsible for aspects of improvements to Fedora's security work.

- *Experience and technical knowledge of how Fedora can be deployed in both an institutional portal framework and a collaborative environment*

Hull's approach to the provision of an institutional repository, that it should be a workplace as well as a showcase, has highlighted a number of issues with Fedora. In working these through the project has demonstrated how Fedora can be used as a highly flexible, scalable, repository solution.

- *A workflow engine addressing one of the recognised principle weaknesses of Fedora*

Whilst 'workflow engine' was not perhaps the best phrase, RepoMMan has provided a simple (to use) tool with which a user can drive fairly complex workflows in order to interact with a Fedora repository. The tool exhibits most of the features that the team originally proposed to the JISC.

## **Conclusions**

In his opening keynote speech to the Open Repositories 2007 conference held in San Antonio, Texas, James Hilton the Vice President and CIO of the University of Virginia noted that "puppies are not free". This was a key message of his talk "Open Source for Open Repositories" and was intended to remind delegates that whilst, like rescue puppies, adoption of open source software may appear to be "free" it comes with a long, and sometimes costly, commitment to maintenance and development.

The experience of the RepoMMan project has been that some of the open source software adopted by the team has required a great deal of work over a period of time to put in place successfully. In part this has been due to the almost inevitable bugs found in the first versions of a particular program or module, in part it has also been due to the - perhaps surprising - fact that the project has pioneered the use of some aspects of software with a rather longer development history.

Resolution of such problems has taken time and effort from the project team and from the software developers themselves. Our early decision to be active members of the open source community around the products we were using stood us in extremely good stead and meant that requests, for instance, for bug-fixes were often directed towards people with whom we already had a working relationship. The opportunity to meet some of these developers face-to-face at conferences has also been extremely important. Far from being offended by the stream of e-mails attempting to resolve bugs, met in person they have generally expressed their gratitude that someone is using the software seriously and keen to help them improve it. This very positive experience of the open source community world-wide led the project team to propose the establishment of a Fedora UK and Ireland user group; this loose confederation of software users has now met three times and a fourth meeting is planned. Those who have attended the meetings have found it very useful to be able to share and to discuss experiences and plans with others in the field.

Whilst discussing human interaction around the project it is worth emphasising the very positive aspects we in e-SIG find from an approach to software development which puts potential users at the heart of the development project. We noted in the section on

'Methodology' that some "technology-related projects have a disturbing tendency to seek information regarding user-needs and requirements after the event of software development." Our approach of researching user-needs prior to the event has stood us in very good stead and, in the context of this repository development work, is likely to do so for some time. A number of the potential users that we interviewed have been very keen to follow the progress of the project and are looking forward to being able to adopt the tool and the repository facilities that it accesses as soon as possible. Further, they themselves have suggested ways in which the repository (and to an extent the tool) could be further developed the better to serve their future needs. The results of such discussions have included a number of pilot projects to implement specific repository functionality and a (successful) proposal to the JISC for a follow-on project, REMAP, which will take the team's work into areas of records management and preservation. We anticipate that these enthusiasts will act as willing champions for Hull's repository facilities over time.

The project as a whole has received very positive comment at a number of international events where we have presented. This level of respect and our active contribution to the open source, and in particular the Fedora, community have enabled us to have a significant influence on future development work. Our problems with early versions of Fedora 2 were instrumental in having its Web Service provision completely rewritten. Other needs of the RepoMMan project, for instance in the areas of LDAP security and object ownership, should result in new and/or improved features in Fedora 3.0 when it is released later in 2007. Particular aspects of our work have also led us to forge very useful relationships with parts of the extensive DEST-funded Arrow Project in Australia, and with the work of Chi Nguyen and his team at Macquarie University as part of the DRAMA project. This latter relationship has led to e-SIG becoming an official test partner for the work they are doing to extend and improve Fedora's security model and to build a lightweight browser interface that exploits the new facilities. When, as seems likely, the University of Hull adopts this work the extended security features will, in principle at least, give our repository Shibboleth capability.

The RepoMMan project has largely fulfilled the work it set out to do and has provided a working demonstration of successful Web Services orchestration. In doing so, we believe that we have also demonstrated some of the principles and values of the JISC e-Framework. The principles of our work seem easily transferable to other institutions. Likewise the RepoMMan tool itself is relatively transferable providing, as it does, facilities for upload/download, delete, version handling and metadata provision; in other words commonly required functionality. However, institutions wishing to adopt our work and implement an installation of their own would need to provide orchestrated web services that address their own institution's service provision: building on the work we have done this may not be an onerous task.

## **Implications and Recommendations**

The initial and subsequent rounds of user-needs analysis undertaken by the RepoMMan team demonstrate that users see a clear benefit in being able to use a repository for development work as well as its being a showcase for finished materials. This view is shared by members of all three communities that were addressed by the project (research, learning and teaching, administration). The University of Hull will continue to develop its institutional repository with this wide functionality to serve all three groups. It will be interesting, over the next few years, to observe the actual use of facilities as they become available; the promise seems high. The active involvement of potential users in the RepoMMan project developed a level of broad 'ownership' in the work, which has stimulated further ideas for development and, indeed, a further development project. We commend this user-centric approach to project work such as this.

The repository model that we at Hull are implementing, and which the RepoMMan tool serves, identifies the repository clearly as part of the core university infrastructure rather than as an adjunct to it. We believe that this is a model that should be more widely adopted and that, if

this were done, repositories would come to be seen as a 'natural', integrated part of institutional provision.

RepoMMan's work to investigate the possibilities of automatic metadata generation has clearly shown that such an approach is feasible but further work needs to be undertaken to investigate the range of metadata that might usefully be held for various purposes and, in that context, the quality of metadata actually provided by the available tools.

Repositories have a key role to play in supporting long-term digital preservation, an area of added value that has sat well with our users. Within an institutional context, RepoMMan has shown that there are a wide variety of different potential roles for a repository, all of which lead to some form of preservation activity. The active involvement of a repository in preservation, and its precursor records management, led to the successful bid for the REMAP project, which recognised the institutional role a repository can play in this area. It is clear, though, that institutions cannot manage digital preservation on their own. REMAP will extend the work of RepoMMan in tandem with other JISC-funded projects to identify where the boundary sits between institutional and national preservation service provision.

## Appendices

### Appendix 1: Glossary of acronyms and terms

API	Application Programming Interface: an interface provided to support requests for services from another computer program or system
ASSIS	Assessment and Simple Sequencing Integration Services: a JISC (qv) Project that ran between September 2004 and March 2005
BPEL	Business Process Execution Language (also called WSBPEL): a business process modelling language that allows orchestration of a number of sequential or parallel computer processes
C&LE	Collaborative and Learning Environment: a particular form of internet-enabled working environment tailored towards education
CD-LOR	Community Dimensions of Learning Object Repositories: a JISC (qv) Project that ran from June 2005 to May 2007
CREE	Contextual Resource Evaluation Environment: a JISC (qv) Project that ran from February 2004 to July 2005
DEL	The JISC (qv) Distributed eLearning Programme
DEST	The Australian Department of Education, Science and Training
DRAMA	Digital Repository Authorization Middleware Architecture: a sub-project of the DEST (qv) funded RAMP (Research Activityflow and Middleware Priorities) specifically aiming to develop a web front-end for Fedora (qv) repositories using pluggable authentication and authorisation components
DRM	Digital Rights Management
DSpace	DSpace is an open source repository platform for accessing, managing and preserving scholarly works in digital form developed by MIT Libraries and Hewlitt Packard Laboratories
Dublin Core	Dublin Core is a set of 15 metadata elements widely used for resource description and discovery
e-SIG	e-SIG: the e-Services Integration Group at the University of Hull
ELF	ELF: the JISC (qv) eLearning Framework
EPrints	EPrints is an open source repository platform capable of dealing with open access research literature, scientific data, theses, reports and multimedia developed by the University of Southampton
ETD	Electronic Dissertations and Theses
FE	Further Education
Fedora	Fedora is an open source repository platform capable of dealing with a wide range of content, developed by Cornell University and the University of Virginia

HE	Higher Education
IDEAL	A framework for project management adapted from the Carnegie-Mellon Software Engineering Institute Capability Maturity Model. (Initiating, Diagnosing, Evaluating, Acting, Learning)
JHOVE	The JSTOR/Harvard Object Validation Environment: a software tool for extracting technical metadata from a digital object and for validating the declared structure of the object
JISC	The Joint Information Systems Committee
JISC CETIS	Formerly just CETIS: the JISC Centre for Educational Technology
JSR-168	A portlet specification enabling interoperability between portlets and portals
L&T	Learning and Teaching
LDAP	Lightweight Directory Access Protocol: a widely used security protocol for authentication and authorisation
MVC	Model-View-Controller: an architectural approach used in software engineering to separate the data (model) from the user interface (view) by introducing an intermediate component, the controller, to decouple data access and management from data presentation and user interaction
OCLC	The Online Computer Library Center
Open source	A term applied to the source code of software that is available to the public with relaxed or non-existent intellectual property restrictions
PLE	Personal Learning Environment
Portal	A web site that provides personalised access to users and which use distributed applications, middleware and hardware to provide services from a number of sources
Portlet	A pluggable user interface component managed and displayed in a web portal
QA	Quality Assurance
REMAP	A JISC-funded project being undertaken by the University of Hull in partnership with Spoken Word Services at Glasgow Caledonian University investigating aspects of records management and preservation in the context of digital repositories
RepoMMan	A JISC-funded project undertaken by the University of Hull described rather fully by this document.
REST	A style of software architecture for distributed hypermedia systems such as the World Wide Web which outlines how resources are defined and addressed
Sakai	The Sakai Project is a community of academic institutions, commercial organisations and individuals working together to build an open source Collaboration and Learning Environment (C&LE qv)

Service Oriented Architecture	Service Oriented Architecture employs both distributed computing and modular programming to build applications from software services
SOAP	Simple Object Access Protocol, then Service Oriented Architecture Protocol, now just SOAP: a protocol for exchanging XML-based (qv) messages over computer networks
UKOLN	A centre of expertise in library management based at the University of Bath
uPortal	A particular brand of portal (qv) software
UVA	University of Virginia, Charlottesville, US
VLE	Virtual Learning Environment: a software system designed to manage educational courses on-line
VRE	Virtual Research Environment: a software system designed to manage collaborative research on-line
WSBPEL	See BPEL
WSRP	Web Services for Remote Portlets: a protocol standard for accessing remote portlets
XACML	eXtensible Access Control Markup Language: a declarative access control policy language implemented in XML (qv)
XML	eXtensible Markup Language: a general purpose markup language used to facilitate the sharing of structured data across the internet

A number of the definitions above have been adapted from Wikipedia (<http://en.wikipedia.org>)



## Appendix 2: Summary of RepoMMan reports and documentation

The RepoMMan website is located at [www.hull.ac.uk/esig/repomman](http://www.hull.ac.uk/esig/repomman). The documents itemised below can all be downloaded from the 'documents' section of the site.

D-D1	Available BPEL runtime environments: evaluation criteria and evaluation results
D-D2	BPEL Deployment notes
D-D3	Familiarity with BPEL authoring tool
D-D4	Development of Fedora materials (aka the "Dummies' Guide")
D-D8	Report on experiences with Fedora during the first year
D-D11	Report on methods of accessing personal metadata from portal frameworks, enterprise directory or similar
D-D13	Report on feasibility of automatic extraction of object metadata
D-D15	Full systems documentation
D-D16	Full user documentation
R-D1	Criteria and toolkit for on-line survey
R-D2	Criteria for research user interviews
R-D3	Report on research user requirements survey
R-D4	Report on research user requirements interviews
(R-D9)	Report on administrator and learner user requirements survey data This was an 'if needed' output from the project and was not undertaken. Instead the project drew on a survey jointly undertaken with the CD-LOR Project: Margaryan A (2006) Report on Personal Resource Management Strategies, CD-LOR Project, Glasgow Caledonian University
R-D11	Report on metadata needs for the University of Hull Digital Repository (this incorporates earlier work at R5, 6, 8 and 10)
R-D12	Teaching & Learning and Administrator user needs analysis
R-D14	Full (research, learner and administrator) user needs analysis report
R-D13 /	Literature review and investigative report on Digital Rights Management
R-D15	
(combined)	

Green R, Dolphin I, Awre C and Sherratt R (2007) The RepoMMan Project: automating workflow and metadata for an institutional repository *OCLC Systems and Services* Volume 23 Number 2 pp210-215