UNIVERSITY OF BIRMINGHAM

MIDESS Integration with Enterprise Architecture Specification

Executive Summary

The MIDESS Project is a JISC project funded under the *Digital Repositories Programme*. MIDESS explores the management of digitised content in an institutional and cross-institutional context through the development of digital repository architecture. The project addresses how support can be provided for the use of digital content in a learning and research context, in an integrated manner. The partners in the project are the University of Leeds, University of Birmingham, London School of Economics (LSE) and University College London (UCL).

As part of work-package 6 of the MIDESS project, a study of the enterprise architecture at the University of Leeds has been carried out. The enterprise architecture describes the current organisation of information systems at the University of Leeds, and the relationship of these information systems with the digital repository. We have specifically focused on three areas: infrastructure, applications and integration with other systems.

The output of this workpackage is a case study of the enterprise architecture at the University of Leeds and is intended to show how a large institutional repository designed primarily for the storage of multimedia materials can potentially sit within the enterprise architecture of a typical large University.

1. Drivers for Digital Repositories and their Integration with Enterprise Architectures

Driver	Description	Research	Learning and Teaching	Administration
Information as a key Strategic Enabler	Information, particularly when it becomes useful as "knowledge", enables effective teaching, research, collaborative partnerships and the ability to make well-founded decisions.	√	√	√
Requirement for Increased storage of digital material.	Volumes of data are exploding at Universities (and elsewhere) with more data being produced in a variety of rich-media formats. Data volumes are currently often resident on portable devices (e.g. laptops, memory sticks, CD-ROM's) and the quantity of this data is growing rapidly with the potential risk of data loss if this digital material is not copied to some component of an enterprise storage system such as a digital repository.	√	√	√
Information Discovery and Access	The integration of content into enterprise applications such as digital repositories will enable more effective business processes and intelligent business decision making.	√	√	√
Information Sharing enables Collaboration	Initiatives involving the sharing of digital material require systems such as digital repositories that can enable or disable the sharing of material.	√	√	√
Information Lifecycle Management	Information stored as digital material needs a structure and data storage medium appropriate to its present importance, reflecting needs such as the accessibility and speed of access to the material.	√	√	√
Legislative Frameworks	Increasingly there are legislative factors that dictate how long digital data can (or must) be kept and			√

	how the data is disposed of. Digital repositories are structured to address these issues.			
Improved E-Learning is needed for students	To deliver good learning and teaching material requires the development of more rich-media material, incorporating sound, images and video which can be very storage intensive. Digital repositories are specifically designed to handle these different types of digital material. Rich-media material needs to be available and often stored in different resolutions to take into account location and resolution of the display device – ranging from large high resolution screens in campus clusters to (possibly) students' own iPods (that now support video). Digital material needs to be shareable across the campus to facilitate re-use and thus needs to be managed and controlled.		√	
Need to participate in national e-Infrastructure initiatives	The government has established the need for a persistent e-infrastructure for the country with future e-Science/e-Research funding to go beyond the support of grid activities and to encompass digital repositories and the preservations needs of the material stored in them. Digital Repositories can assist with this. Feedback from Faculties at the MIDESS partners indicates that some of them will pursue their own solutions if institutional solutions such as digital repositories are not forthcoming – this would lead to silos of information rather than centralised 'pools' of knowledge within Institutions.	√	√	

Analogue material needing digitising for ongoing preservation	The user needs analysis carried out by the MIDESS project has identified large amounts of analogue material. Much of this material is valuable and needs preserving into the future by conversion to digital media.	√	√	√

Table 1 *Drivers for Digital Repositories and the applicability to Research, Learning and Teaching and Administration.*

2. Enterprise Architecture

2.1 Storage

2.1.1 Direct Attached Storage (DAS) and the Storage Area Network (SAN)

Digital repositories are considered to be systems where the storage and long-term maintenance of digital resources can be undertaken. The storage is expected to support long term preservation of the material in the repository. Once the digital repository is in service the material in the repository is expected to be secure, error free and readily available. The physical storage within the repository is therefore a central part of the architecture of the digital repository.

The primary storage enterprise architecture at the University of Leeds is based upon the centralised Storage Area Network (SAN).

A SAN is one (or more) large banks of disks mounted in racks that provide for 'shared' storage space which is accessible by many servers/systems. Other devices, such as tape libraries can also be directly attached to the SAN.

Traditionally, data storage resides on hard disks that are locally attached to individual servers. This is known as Direct Attached Storage (DAS). Although this storage may be large (in the order of 100's of Gigabytes of data storage per server) the storage is usually only easily accessible from the server to which it is attached. As such, much of this disk space on the directly attached storage device remains unused and 'contingency' has to be built into storage needs when determining server specifications. In addition, if the server fails, access to the data held on those disks is usually lost.

A representation of the difference between Direct Attached Storage (DAS) and a Storage Area Network (SAN) is shown below.

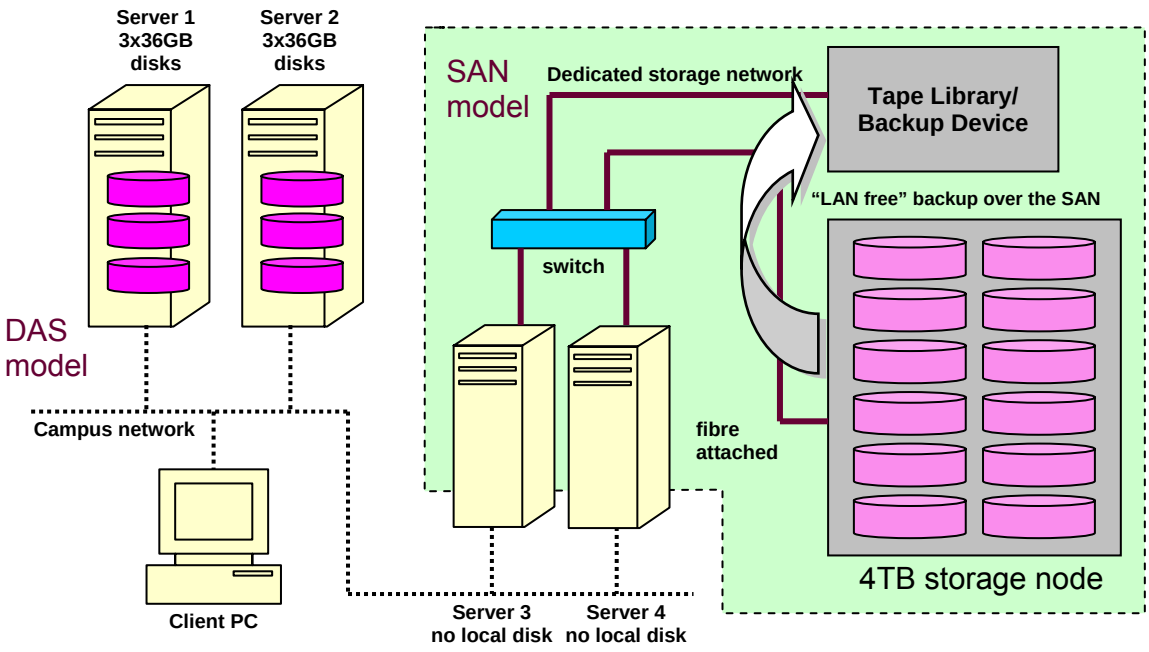

Figure 1 *A schematic illustrating the differences between Direct Attached Storage (DAS) and SAN models. The SAN model shows how data could be directly backed up to a large tape library/backup device across the SAN.*

	Storage Area Network (SAN)	Directly attached Storage (DAS)
Allows for easy upgrade of Storage facilities	√	Only by separate disk
Variety of types of storage High Performance (fast) or lower performance but high density	√	Not on same disk
Maximises use of available disk space	√	X
Additional storage can be added without restarting the system	√	X
Complies with Security and Disaster audit requirements	√	Varies
Good fault tolerance	√	Depends on type of disk
Expensive to install?	√	X
Automatic backup to tape	√	Varies
Supports Streaming Video	√	Usually use separate disk drive.
Centralised storage solution	√	X

*Table 2.Difference**s** between Storage Area Network (SAN) and Directly Accessed Storage (DAS)*

The SAN at the University of Leeds functions as a high-speed network similar to a conventional local area network (LAN) and establishes a direct connection between storage resources and the file server architecture. The SAN effectively acts as an "extended storage bus" using the same networking elements of a LAN including routers, hubs and switches.

SAN's are particularly useful for digital repositories, since if the data can be stored centrally it can usually be more efficiently managed. Attempting to centrally manage distributed digital information stored on individual departmental servers is more problematic since it is unlikely that the digital repository manager will have administrative control over individual servers distributed in different departments throughout the organisation.

The SAN provides a mechanism for storing "live" data that is frequently changing, with sufficient unallocated disk space so as to be able to quickly respond to changing needs such as the addition of large digital collections accessed via a digital repository. The SAN supports the playback of streamed video, an important feature for the digital repository at the University of Leeds.

The SAN also allows for 'pay as you grow' storage scalability, storage planning and non-disruptive growth and/or reconfiguration of the storage system. Extra disks can be purchased at any time and added to the total storage provision without taking the system down. Configuration utilities allow for that storage to be 'presented' to any host that requires space and free space can also be re-assigned as necessary.

The actual nature of the storage on the SAN can vary – for example the SAN may be subdivided into storage areas which either offer rapid access to data or alternatively high density storage for digital data accessed less frequently. This flexibility is again of particular use to digital repositories since many collections contain both digital material that must be accessed rapidly (such as graphics in the form of thumbnails) and also other digital material which is accessed less frequently (such a large high resolution images stored in TIFF format.

2.1.2 Storage Options for the Digital Repository at the University of Leeds.

The digital repository software (Endeavor Curator) at the University of Leeds enables digital material to be either:-

- Stored centrally on the SAN
- Stored physically on the same server as the digital repository.
- Hyperlinked to files stored on any accessible web server.

For the purposes of the MIDESS pilot project at the University of Leeds, the data was stored both directly on the digital repository server and links were created to media files on existing departmental web servers. A proposed upgrade to storing and accessing media files directly from the SAN will be considered once the digital repository moves from the pilot stage to becoming a fully functioning service. At the point at which the repository becomes a service it is expected that the shortage of storage space on the digital repository server will also become an issue. The digital repository software (Curator) along with much of the data is currently stored on the main MIDESS Sun Solaris server, This Solaris server mirrors the information stored within it. This is an example of the directly attached storage. The contents of the server are also automatically backed up to nightly.

This flexibility in the physical location of the digital files stored within the Curator digital repository software is useful since at the University of Leeds there is a variety of content and a range of locations where content could potentially be stored.

Storing large quantities of digital material on the SAN does have cost implications for departments at the University of Leeds since departments are expected to contribute financially for the storage space used on the SAN beyond the normal space allocated to them

by the computer services department. Currently this additional storage space costs £1000 per year for 250 GB of storage on the SAN. Storing digital material on local departmental servers is seen by many departments as a means of avoiding these additional fees charged by the computer services department, since the necessary funding for storage space on departmental servers will in many cases have already been allocated to individual departments and there may be space available on these departmental servers.

However storing digital material on departmental servers rather than the SAN does have disadvantages at the University of Leeds. These include:

- Bandwidth. While the SAN at the University of Leeds is directly connected to the University Internet backbone, it provides maximum speed of access to digital material stored within the digital repository. Access to digital material stored on departmental servers is usually slower due to the reduced bandwidth between the digital repository and the individual departmental servers throughout the university. Delayed responses in viewing material are therefore expected when numerous people attempt to view large video files via the repository in cases where the video files are physically stored on departmental servers widely dispersed across the University campus.

- Management of digital material stored on individual departmental servers scattered around the university is more problematic than management of material stored centrally on the SAN. Since the backup of individual departmental servers is the responsibility of IT staff within the individual departments, the frequency and accuracy of back-up is expected to prove variable and be dependant on the various backup strategies of the respective departments. Ensuring that all departmental servers which contain digital repository material are regularly backed up add to the workload of the repository manager.

- The digital repository manager has no direct control over the removal/maintenance of digital files stored on the individual departmental servers. Removing a digital file from a departmental server would result in a broken URL link when an attempt is made to view the file from within the digital repository.

In addition to the SAN, the University of Leeds is proposing an archiving strategy to ensure that the University can benefit from the value it holds within its data over the lifetime of that data. This capacity is known as Information Lifecycle Management (ILM). Here Information (content) is managed across its lifecycle, from creation, modification, delivery (publishing), filing, retention (preservation) and deletion. Data stored in this archiving system can be demonstrated to be unchanged over time, again to comply with appropriate legislation. This disk area is mirrored by the University of Sheffield ensuring that there are two disk based systems with the same data.

2.2 Authentication

As part of the pilot project, the University of Leeds has configured the digital repository to use the authentication provided via University of Leeds Microsoft Active Server and LDAP. The Lightweight Directory Access Protocol, or LDAP is a networking protocol for querying and modifying directory services running over TCP/IP. Practically this ensures that staff or students who have valid user accounts at the University of Leeds can be specifically included or excluded from viewing specific collections in the University of Leeds digital repository.

The MIDESS project also specifically wishes to consider Shibboleth as a method of authentication as part of the potential interoperability of collections with other institutions.

Shibboleth http://shibboleth.internet2.edu/, has emerged as the front-runner for the most widely adopted authentication standard across Europe. A number of commercial service providers are planning to create Shibboleth interfaces to their services or are already providing them.

Shibboleth does not carry out authentication itself. Instead Shibboleth defines a set of protocols for the secure passing of identity information between institutions and service providers. It relies on the institution to establish identity, and on the service provider to confirm access rights, given information about institutional affiliation. It is written in SAML (Security Assertion Markup Language).

Thus practically the Identity Provider (origin) provides attribute assertions about that user to the Service Provider (target) site which in this case is the University of Leeds. A trust exists between institutions, allowing each site to identify the other, and assign a trust level. Identity Provider sites are responsible for authenticating their users, but can use any reliable means to do this. Service providers need to be confident that the institution or organisation that the user belongs to has a robust and up-to-date authentication system in place.

This need for trust leads to the concept of federations which are groups of similar organisation such as universities who have agreed to a common set of policies. They are typically being established at a national level. JISC plans to establish a federation covering the UK's higher and further education sectors.

Another feature of Shibboleth is the emphasis it places on user privacy. The system devolves all responsibility for user authentication to the user's institution. The information passed to the service provider by the institution is of the form "this user is a member of our institution", or perhaps "this user is a member of the psychology department of this institution". In other words, the information passed back to the service provider is about status rather than personal identity.

Version 4 of the digital repository software at the University of Leeds (Curator) specifically supports Single Sign-On via Shibboleth.

When the University of Leeds digital repository has been configured to use Shibboleth, the user will be authenticated as they click the URL to access the digital repository system Curator. For other protocols such as LDAP, the user is authenticated when they click on the login button from within the Curator interface. For Shibboleth-enabled Curator users, the user will be rerouted to a standard service, Where Are You From? (WAYF), to start the authentication process. The Shibboleth service will return user attribute information to Curator. Curator uses this information to set the resources the user has access to within the application and thus is able to restrict users from other institutions to specific collections in the digital repository should the managers of the digital repository wish to do so.

The University of Leeds will be evaluating Shibboleth by attempting to link our Digital Repository to Leeds Metropolitan University's Web CT Virtual Learning Environment (VLE).

As part of this evaluation Learning Objects will be stored in the digital repository and restricted to Leeds Metropolitan University based upon authorisation via shibboleth. (see section 4.2).

3 Applications

3.1 Streamed Media

Central IT services at the University of Leeds currently provide a streamed media server for the delivery of digital sound and video.

Details of the streaming server at the University of Leeds are available at the following web address:

http://www.leeds.ac.uk/iss/digitaldelivery/

The streaming media server at the University of Leeds currently runs Microsoft Windows Media Services 2003 and any content must be encoded to the appropriate format for this system. Windows Media Server stores and streams content in Windows WMA and MP3 format for audio and WMV format for video. Streamed content must be viewed using a media player program that supports these formats. This would normally be Windows Media Player when accessing the material from a PC running Microsoft Windows.

For video, the size and quality of the displayed image determines the amount of network bandwidth required to stream the content. This is normally referred to as the bit rate of a streamed video clip. For ‘on demand' video, that amount of bandwidth is required for each individual stream to a target audience. This must be borne in mind when streaming high bit rate streams to large audiences, for example in a PC cluster. Under normal circumstances this should not be a problem but at present, no priority is given to this type of traffic on the University of Leeds campus network, so if network congestion was to occur, the media stream could currently be adversely affected.

The media server has also been configured as part of the storage space of the SAN (see section 2.1.1), thus the digital media files at the University of Leeds potentially has a large storage medium (in theory up the maximum space available on the SAN) from which to run.

Media content in the form of streamed video and sound files can be delivered to audiences either within the University of Leeds or off campus via the digital repository. Off campus applications can include remote teaching over the Internet and therefore the streamed server and thus the digital repository has the potential to be of use for distributing teaching material via distance learning. On campus, the use of streamed media files includes both teaching in student PC clusters and large screen projection in lecture theatre environments.

In order to store the files in the University of Leeds digital repository, the repository (Curator) needs to link to the files on the web enabled media server. An example of a streamed file running from the Windows Media streaming server is available at mms://iss-video.leeds.ac.uk/midess/Archive Test 1.wmv.

There are also plans by the computing services at the University of Leeds to purchase the Helix server software. The Helix software supports the streaming of a larger number of sound and moving image formats and includes support for RealAudio, Real Video, Windows Media, QuickTime , MP3, MPEG-4, 3GPP* (H.263 and H.264) from one server infrastructure. This will enable the digital repository to natively stream a wider variety of formats. It is expected that this Helix server software will be available as a service by the start of January 07.

3.2 Federated Search Tools

The goal of federated searching is to enable a user to search multiple independent, discretely mounted, data sources or databases through one search query. This avoids selecting a specific database, searching the database, collecting and evaluating results and then continually repeating the process[1].

3.2.1 Z39.50

Z39.50 is designed to enable communication between computer systems such as those used to manage library catalogues. Thus, provided that the digital repository and the library system are Z39.50 compatible then searches of the digital repository can potentially be made from the Library catalogue system using Z39.50[2].

By using Z39.50, the benefits of distributed data management are combined with the benefits of unified data access, which allows the user to submit a single search across multiple resources, regardless of their physical proximity to one another or to the user Like a large number of other applications of Z39.50, the technology is effectively hidden from the user in both these examples; and so far as the user is concerned they are simply searching one large database.

Z39.50 follows a client/server model, where one computer (the client or, in Z39.50 terminology, the 'origin') submits a request to another computer (the server or, in Z39.50 terminology the 'Target') which then services the request and returns an answer.

Potentially the University of Leeds digital repository can be linked to the University of Leeds library system via Z39.50. This would result in searches being made from the online library catalog system finding and then displaying matching material in the digital repository as well as records from the library catalog. To enable federated searching based upon the Z39.50 protocol would require that the University of Leeds purchase the additional module which supports Federated search via Z39.50 within Curator. If this were to be installed then the library system at the University of Leeds could search the digital repository and thus in addition to books, documents etc then other material such as the content of the digital repository could be viewed.

Practically the Digital Repository would be considered the Z39.50 client while the Library System would act as server. Thus when a query was carried out the matches from numerous clients (including the digital repository) would display digital material from the repository in the library search results.

3.2.2 Web Services

Web Services provide a model for communication between application programs via the web. The novelty of the Web service approach is that it builds upon established web technologies, complements existing web applications and it is relatively easy to implement.

Software applications written in various programming languages and running on various platforms can use Web Services to exchange data over the Internet in a manner similar to inter-process communication on a single computer. This interoperability (e.g., between computer languages such as Java and Python, or operating systems such Windows and Linux) is due to the use of open standards.

Web Services rely on the same architecture as the Web, and exchange XML documents over HTTP though a layer of protocols (such as SOAP), which are themselves based on XML. Simple Object Access Protocols (SOAP) combines XML with HTTP for accessing services, objects, and servers. SOAP defines a formal structure for specifying information that can be passed to a web application, and the same structure can be used to pass the results of the request back.

Web Services bring the concepts of Z39.50 into the web environment using web technologies.

The software for the Digital Repository at the University of Leeds (Curator) supports both and Search/Retrieve via URL (SRU) [3] and Search/Retrieve Web service (SRW) [4] .SRW is a web service for search and retrieval which provides a SOAP interface to queries to augment the URL interface provided by its companion protocol SRU.

Practically in order to add collectons from the University of Leeds digital repository one of the options within a specific collection is to enable searching of the database by selecting SRW/SRU database and typing a database name for the collection (see figure below) The results are returned in Dublin Core XML. The name (handle) for the collection needs to be passed to the federated search tool so that it can access the collection.

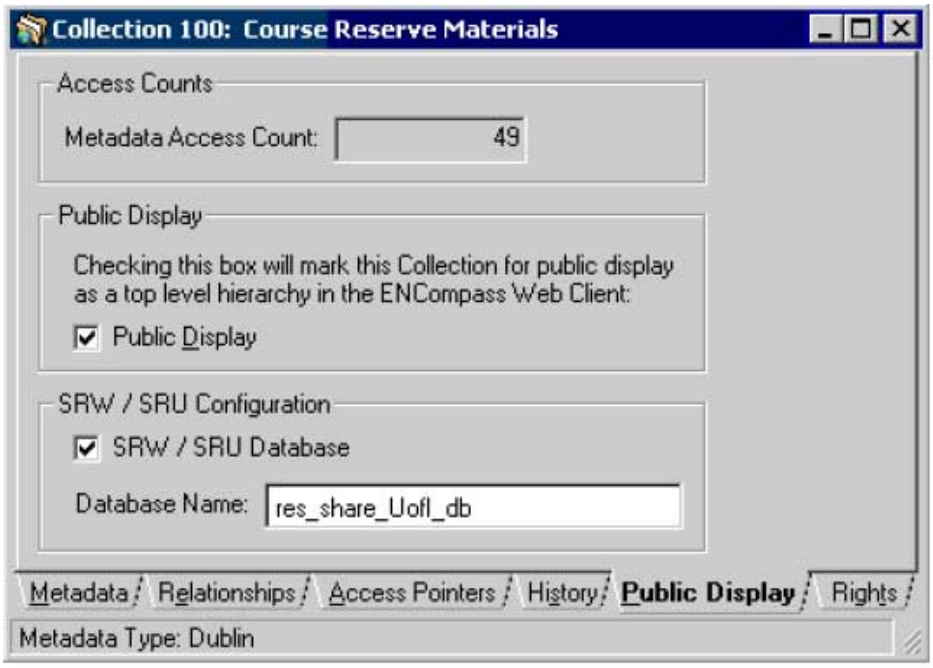

Figure 2 *Specifying the Database Name within the Digital Repository Curator software allowing search using SRW/SRU.*

4. Integration

4.1 Integration of the Portal with the Digital Repository.

The portal at the University of Leeds is based on Sungard Luminis software product[5] and can be accessed via the following web link http://portal.leeds.ac.uk.

The Sungard Luminis platform assumes that other Web-based systems (such as the digital repository) can provide significant functionality, therefore it provides a system that can be used as an integration platform and thus provides software (called a connector) for connecting to other applications. One of the first steps for integration into the portal is to define the attributes of authorised system users which the portal can then use to authorise users to use external systems such as the digital repository. Once this has been achieved in the portal, users can access all integrated Web-based systems via session and user identifiers transferred in the form of Application Programmer Interface (API) URL's. Thus the API URL's are transferred between the portal and the digital repository. This is done using the Campus Pipeline Integration Protocol (CPIP). Some examples of URL names used in this way are shown below.

URL Name	Description
1. GetConfigVersion(n)	Delivers the CPIP connector configuration to CPIP
2. Create User	Creates the user on the external system. Create User is not required if the external system already has the user records
3. Identify	Logs in the user with user identification only.
4. Authenticate Version 1: digest Authentication	Logs in the user with digest authentication level security (over ssl –https).
5. Authenticate Version 2: Split Credentials	Logs in the user with multi-request security (over ssl – https).
6. Pickup	A one time request used to associate the browser and the external system. Pickup is not required if cookies are not used.
7. Last Activity	Request for last activity on user
8. Deauthenticate User	Requests to logout the user.

Table 3 *URL names the portal uses to pass information between the portal and the external application.*

There are five components required that comprise the CPIP architecture.

- The Luminis System
- The CPIP servlet
- The CPIP protocol Specification
- A CPIP connector that utilises the protocol specification.
- An external application (in our case the digital repository).

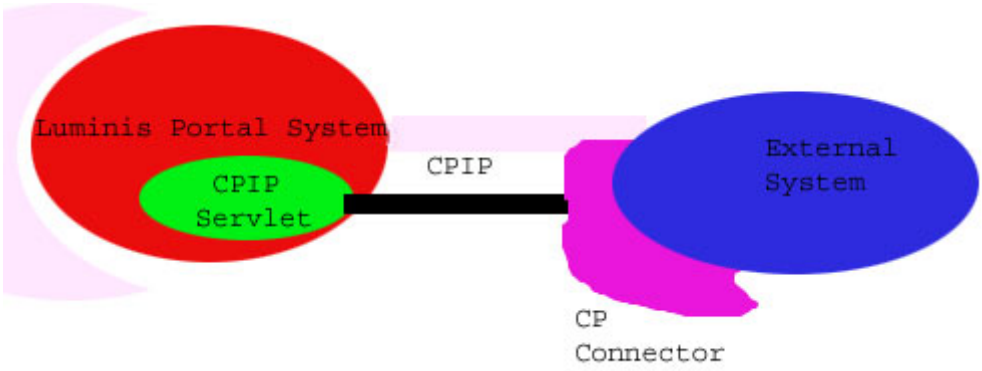

Figure 4 *Representation of the communication between the portal and an external system (in our case the digital repository)*

The Luminis portal at the University of Leeds uses the JSR 168 standard [6.] JSR 168 establishes a standard API for creating portlets, the integration component between applications such as the digital repository and portals that enables delivery of the application through the portal. Without this standard, each version of an application needs its own portlet API, and the portal would require that these portlets be specifically tailored for implementation through that portal.

Thus In order for the University of Leeds digital repository to work within the portal at the University of Leeds, a connector would need to be written for the portal to authenticate against the digital repository.

The University of Leeds has already enabled authentication against the Microsoft Active Directory using LDAP and our next course of action is to progress to developing the connection for linking the digital repository directly to the portal. This would enable students and staff who access university material via the portal to automatically have their access rights transferred to the digital repository. Without the authentication against the portal, users of the digital repository would be required to login directly to the repository to view material to which only they had rights to even though they may have already logged into the portal with the same username and password.

To use the digital repository from within the portal, the user will need to select a link at which time they will be automatically authenticated to the digital repository. Behind the scenes (i.e. not from within the browser) the portal server sends the user credentials over http and logs into the digital repository. Still behind the scenes, the session cookies are collected. The browser is then directed to the digital repository, but it needs the cookies setting first before it can do this. On the basis that students at the University of Leeds are now actively using the portal, we expect the number of hits the digital repository will receive to continue to increase significantly once its is “contained” within the portal.

There is also the ability to use Web Services (SRU/SRW) from within the portal. The functionality has been implemented as a channel in the portal.

The portal can conduct a SRU search preconfigured to a single database and can present the output in a variety of templates based upon XSL. Various XSL views that have been constructed at Leeds (details_view.xsl and search_view.xsl). The channel builds an SRU query and can retrieve the content directly from the digital repository. It then applies the XSL stylesheet to it to display the results.

The following configuration is required.

- The name of the SRU source to display in the channel.
- The SRU server.

- The namespace to associate with the prefix srw_dc
- The schema to request from the server. (Dublin Core can be used)

Given that different SRU services associated different namespaces with the prefix srw_dc it is suggested that for the repository that the Library of Congress namespace is used.

Organisation	SRU Base URL	Namespace used with srw_dc
Library of Congress	http://z3950.loc.gov:7090/voyager	info:srw/schema/1/dc-schema

Table 4. *Base URL used within the portal for SRU*

Thus this portal channel should enable the system to extract data directly to the portal using Web Services.

4.2 Integration of the Virtual Learning Environment (VLE) with the Digital Repository.

The University of Leeds has chosen Endeavor Curator as the digital repository of choice for the purposes of the MIDESS project. An additional component of the Endeavor product range is the Endeavor Course Content Integrator [7] which can integrate the Curator digital repository with either Blackboard or WebCT commercial VLE's. Unfortunately the course content Integrator product does not interface to the current University of Leeds VLE (Bodington) since Bodington is a bespoke VLE designed and developed at the University of Leeds. It is expected however that the Bodington VLE system will be replaced by a commercial VLE system within the next two years.

It has been decided that for the purpose of MIDESS, as an evaluation the University of Leeds will be linking the Curator digital repository to the WebCT VLE at Leeds Metropolitan University. As part the evaluation, Leeds Metropolitan University will have learning material directly stored in the University of Leeds digital repository.

The University of Leeds is also evaluating using Shibboleth as the method of authentication between the University of Leeds and the Leeds Metropolitan University thereby restricting the learning content material supplied by Leeds Metropolitan University to their staff and students..

The Course Content Integrator supplied by Endeavor integrates with both the Blackboard VLE Learning System release 6.0 using Blackboard's Building Blocks API and the WebCT Vista VLE version 2.1 - 4.1 using WebCT's PowerLinks API. Endeavor is also considering creating a similar level of integration with other VLE's if there is sufficient demand for this.

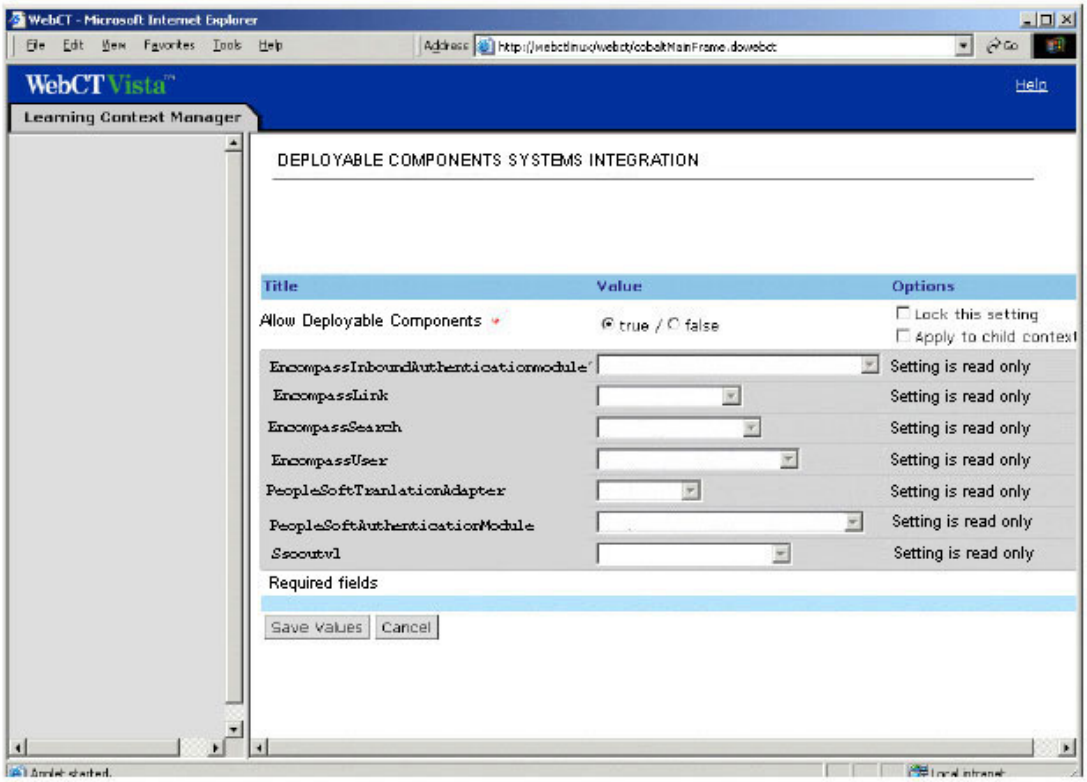

Figure 5 *Deployable components system integration fields required within the WebCT Interface.*

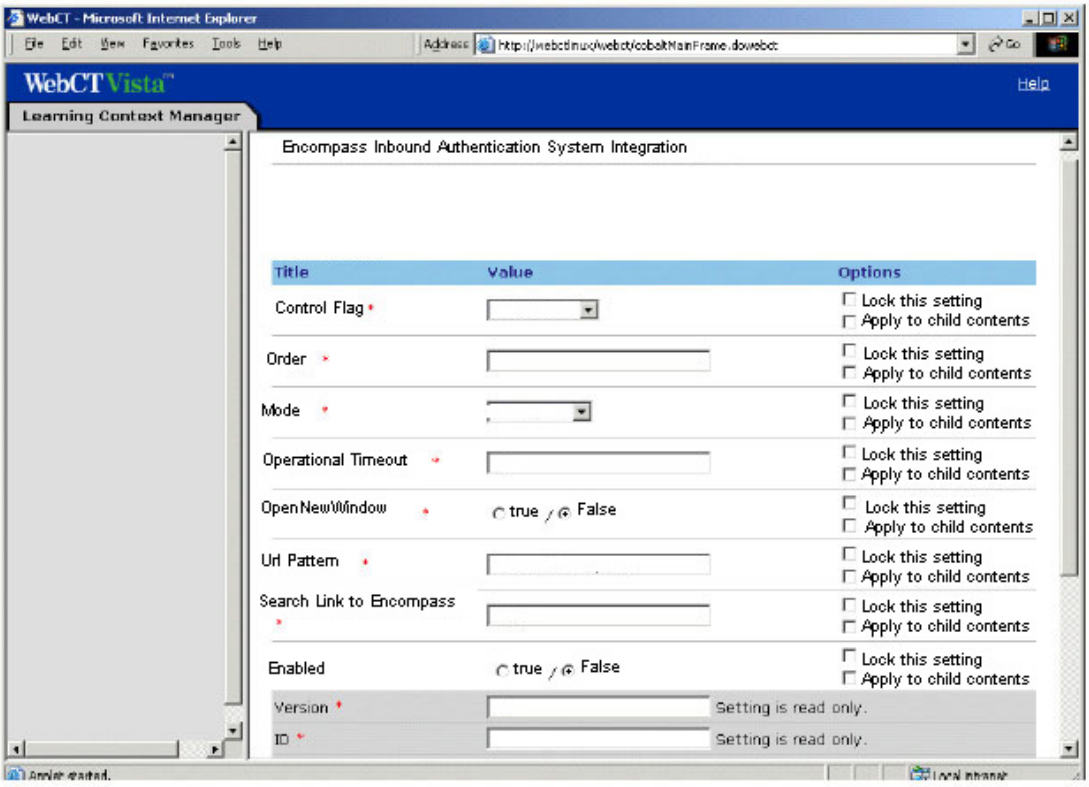

Figure 6 *Fields required for the enabling of the inbound authentication within WebCT.*

The Course Content Integrator can add the following digital repository functions to the WebCT or Blackboard VLE's without cutting or pasting:-

1. Link to a predefined search in the digital repository.
2. Link to a digital object stored in the digital repository.
3. Browse or search the digital repository from within the VLE.

1. Link to a predefined search in the digital repository. This facility uses the EncompassSearch function .The predefined search terms is created by the VLE course creator and records in the digital repository matching the predefined search are displayed within the VLE.

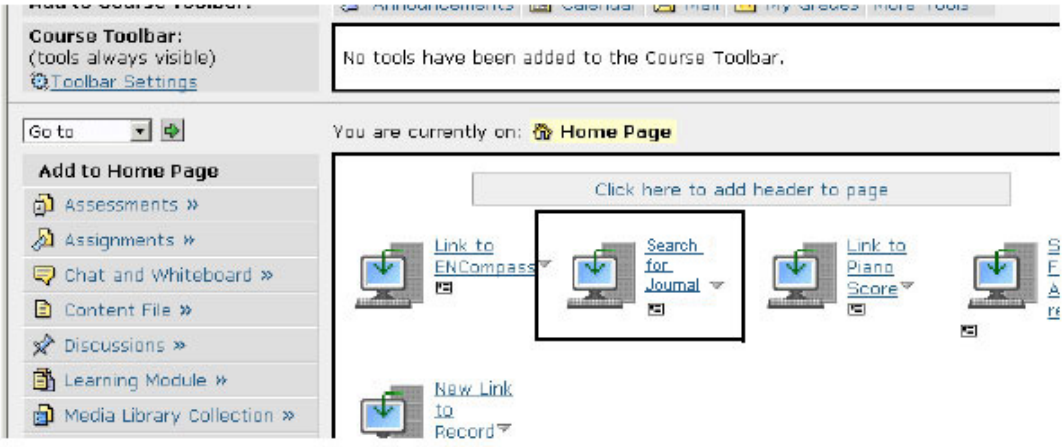

Figure 7 *Predefined search of the Digital Repository from within WebCT*

2. Link to a digital object stored in a Curator digital repository.

This links between the WebCT VLE and a specific digital objects stored in the University of Leeds repository. Thus the digital object is stored in the digital repository rather than the VLE. This not only avoids duplicating digital material in each course that uses that material it also allows more details to be determined about the digital object since it would be expected that the digital object would have extensive metadata associated with it.

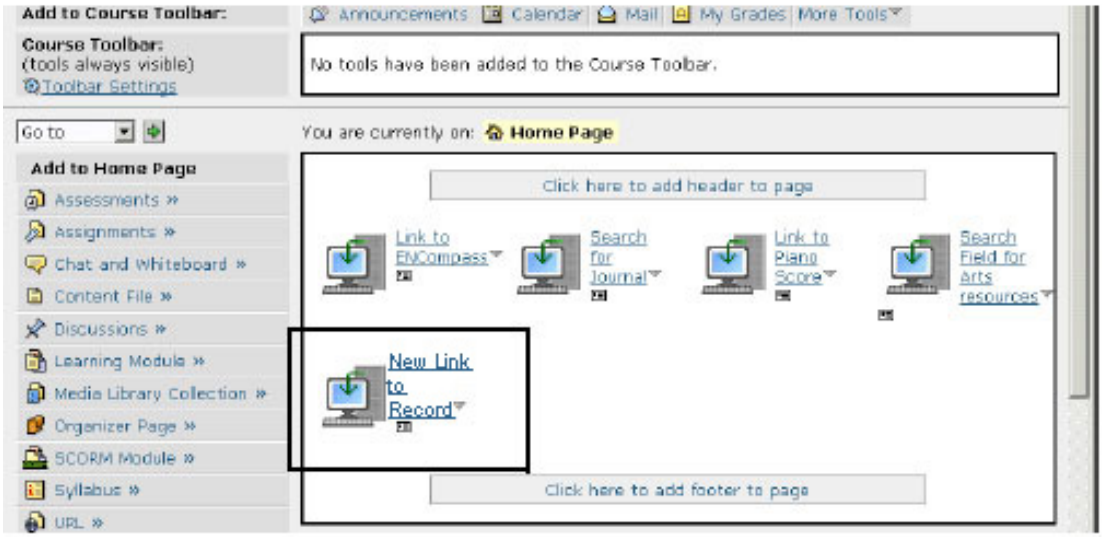

Figure 8 *Link to a digital object in the Repository within WebCT*

3. Access to the full Curator interface within the course management system interface, so that students may browse or search Curator resources on their own

The option allows the user to insert a search field into a course page enabling the user to search all of the resources to which they have access to.

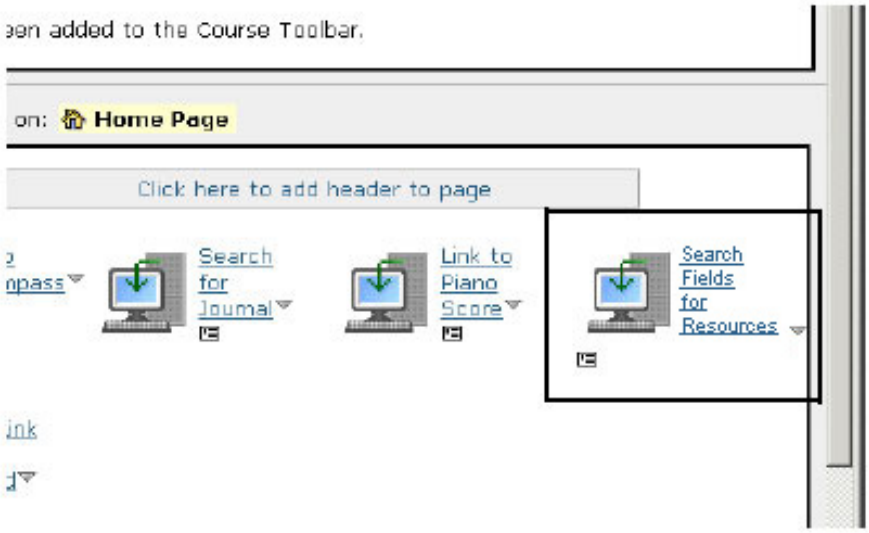

Figure 9 *Searching the Curator digital repository from within the WebCT.*

Instructors can deliver access to all of the resources managed by Curator, or to specific searches or defined objects. Which resources are available will be dependent on the rights allowed to the material via both LDAP and Shibboleth. To access course resources, students do not need to re-authenticate, as the digital repository uses trusted peer authentication.

Currently files cannot be created in the VLE and stored directly in the Repository. Thus files for the VLE will need to be either created via Web Forms which can input metadata and media files into the repository or created via the Curator staff client software. Information is not entered into the Digital Repository via the VLE rather the VLE access the information stored in the digital repository in various ways.

Students at many universities are familiar with how to use VLE's; however they are much less familiar with the use of digital repositories. The Endeavor content integration software links both systems and should ensure that multiple copies of the same digital material within the VLE are avoided wherever possible by linking each potential instance of the file in the VLE to the definitive copy in the digital repository.

The goal of the integration between the two systems is to be as seamless as possible thereby ensuring that the user is able to achieve their goals without becoming 'unsettled' but constantly moving between the VLE and the digital repository.

A modular diagram of the course content integrator is shown below:

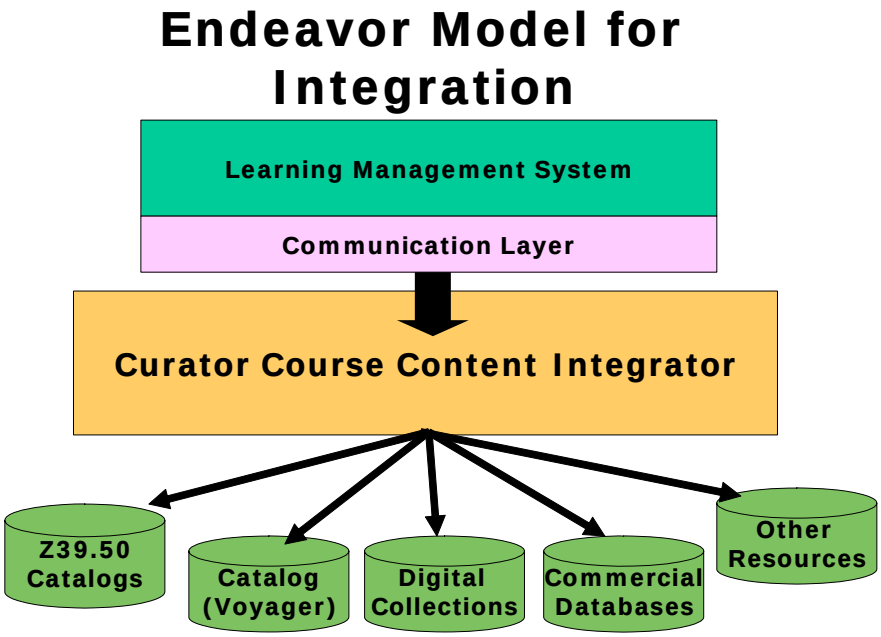

Figure 10 *The Endeavor Model of Integration between Virtual Learning Environments (VLE's) and systems such as Z39.50 Catalogues and Digital Collections.*

One of the potential difficulties with the Integration between the Digital Repository and the VLE is that although the digital repository supports Active Directory via LDAP and can thereby authenticate students based upon this combination, within many universities there is an additional layer of complexity.

The University of Leeds (in common with many other universities) uses a product called Banner from Sungard Luminis. The Sungard banner system is used to record which students are registered on which courses during the year.

In order to integrate the digital repository with Sungard Banner directly and thereby enable material in the digital repository to be restricted based upon a particular taught course (e.g. Java for beginners) rather than by year and department (e.g. 1st year Computer Studies, Student), then the University of Leeds will need a software connector to be developed between Banner and the Curator digital repository. This can be achieved by using the Sungard Luminis Data Integration Suite, which provides a software development Kit (SDK) specifying the transport and data format protocols need for third-party application software providers (such as Curator) to understand the individual course provided by Banner.

4.3 Integration of the Library System with the Digital Repository.

One of the major goals at the University of Leeds is to link the digital repository to the library service so that staff and students searching for material in the library catalogue will be able to find matching media material in the digital repository as a match to their search.

The University of Leeds is currently evaluating piloting a federated search tool called OVID Linksolver™ at the University of Leeds. This is currently available at: http://ssolver.ovid.com/muse/servlet/MusePeer

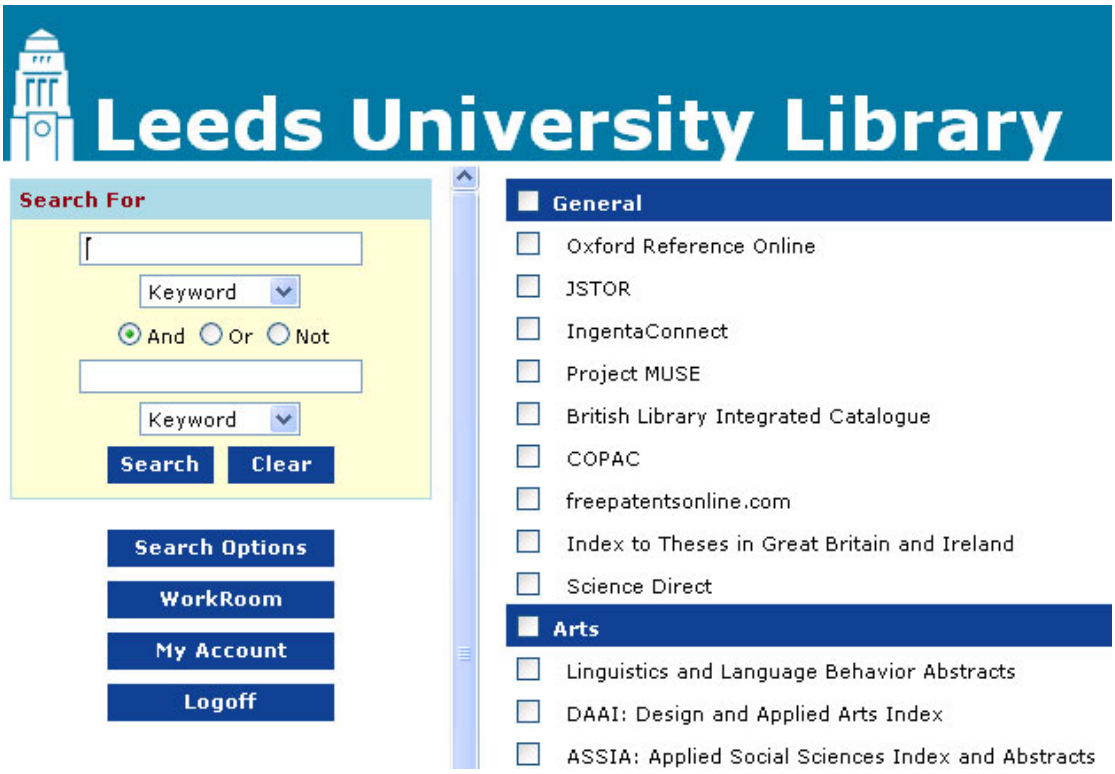

Figure 3 *Pilot Federated search tool currently being evaluated at the University of Leeds which would allow search of the digital repository via SRW/SRU.*

On the basis that the Curator digital repository software supports SRU/SRW Web Services and the OVID Linksolver™ federated search tool also supports SRU/SRW Web Services, then the digital repository at the University of Leeds can potentially be searched using this federated search tool. It is intended that this search tool is used to search both local (University of Leeds) resources and remote resources, and should enable students and staff to easily search a variety of both local and remote resources via an easy to use interface.

5.Recommendations for Integration of Digital Repositories with Enterprise Architecture.

Store digital material centrally where possible. Centralised storage architectures such as SAN's provide an excellent means of storing, accessing and preserving digital data. Where digital material is stored on individual departmental servers and linked to the digital repository, ensure that there is sufficient procedures are in place to safeguard and secure the digital material.

Consider linking to library resources via Z39.50 or Web Services. While viewing library resources from within the digital repository may be of limited use for many digital repository systems, the ability to access the digital repository from the library system and thereby view all institutional information (including image, audio and video material) related to a specific query would be of considerable interest to many library system users.

Consider integrating the digital repository with the Institutional VLE. Examine the functions provided by the VLE when integrated with the digital repository. Ensure the integration is as seamless as possible. Determine policies to determine whether digital material will be stored in the VLE or digital repository or both.

Examine the viability of storing media files on a streaming server and then accessed these files from the digital repository. Streamed media servers should provide a considerable improvement over both a central or departmental web server when viewing large video and sound files.

6. References.

1. WikiPedia. Federated Search. http://en.wikipedia.org/wiki/Federated_search

2. Ariadne Issue 21. September 1999. Z39.50 for All
http://www.ariadne.ac.uk/issue21/z3950/

3. The Library of Congress SRU Version 1.1 13th Feb 2004.
http://www.loc.gov/standards/sru/

4. The Library of Congress SRW Version 1.1 13th Feb 2004
http://www.loc.gov/standards/sru/srw/

5. Sunguard website portal software http://www.sungardhe.com/default.aspx?id=102.

6. JSR 168 Portlet Specification
http://developers.sun.com/prodtech/portalserver/reference/techart/jsr168/

7. Endeavor Course Content Integrator. Endeavor website (http://www.endinfosys.com/) at (http://www.endinfosys.com/software/course_content.html)