



MIDESS Project

Final Report

Michael Emly
University of Leeds Library

November 16th 2007

Table of Contents

Acknowledgements.....	3
Executive Summary.....	4
Background.....	5
Aims and Objectives.....	5
Methodology.....	6
Implementation.....	7
Preparatory phase.....	7
Repository implementation.....	8
Metadata, digital preservation and IPR.....	8
Enterprise architecture and repository services for learning and teaching.....	10
Resource discovery, content sharing and re-use of digital objects.....	11
Outputs and Results.....	12
User requirements (WP3).....	13
Functional and technical requirements (WP2).....	13
Metadata (WP4 A & B).....	14
Preservation issues (WP5).....	14
IPR (WP7).....	14
Enterprise architecture (WP6).....	15
Repository implementation and population (WP8).....	15
METS and content packaging (WP4 & WP9).....	16
Metadata harvesting (WP10).....	16
Outcomes.....	17
External evaluation.....	17
Dissemination.....	18
Local outcomes.....	19
Conclusions.....	19
Implications.....	20
Recommendations.....	21
Glossary and acronyms.....	22
References.....	23

Acknowledgements

The MIDESS Project was funded by the JISC within its Digital Repositories Programme, with additional funding from the Consortium of Research Libraries.

Partners in the project were:

University of Leeds (lead)
London School of Economics
University of Birmingham
University College, London

We would like to thank all who have contributed to the work of the project, particularly:

- Steve Charles (project manager)
- Tracey Stanley (project director until February 2007)
- The project officers: Ilaria Corda, Mary Cordiner, Adrian Dover, Tom Grahame, Matthew Hoskins, Marie Lagerwall, Matt Martin
- Members of the local Working Groups which provided input and advice throughout
- Members of the Project Steering Group

Project website: <http://www.leeds.ac.uk/library/midess/>

Contact: Michael Emly (Project Director)
m.emly@leeds.ac.uk

Executive Summary

The MIDESS Project was funded by the UK JISC within its *Digital Repositories Programme* - with supplementary funding from the Consortium of Research Libraries - to explore the use of repositories for the management of multimedia content within the HE sector. The project ran for 2 years (2005-2007) and brought together the libraries and information services of the London School of Economics, University College London, the University of Birmingham and the University of Leeds.

The objectives of the project were:

- To create 3 demonstrator repositories and populate them with sample content
- To explore the options for sharing and re-using content between institutions
- To investigate the main IPR issues for this type of content
- To establish appropriate metadata standards for this content and the potential for using METS and IMS
- To explore the role of the repository within the institutional information architecture
- To investigate how repositories can encourage the wider exposure and sharing of content across institutions through metadata harvesting services.

In short, the project set out to create a set of suitable platforms, populated with content, in order to examine the issues and validity of implementing full digital content management services within each institution and the potential for sharing and re-use of content between institutions.

Three repositories were implemented, using Fedora, DSpace and DigiTool respectively. Material suitable for ingest was identified and the dialogue with academic partners in each institution helped clarify not only the complexity of the interactions required but also the potential value of the repository in supporting learning, teaching and research. A major output from the project therefore consists of a series of specifications which set out the key requirements of a multimedia repository:

- User requirements specification
- Functional and technical requirements specification
- Metadata requirements specification
- Digital preservation requirements specification
- Requirements for integration with enterprise infrastructure
- IPR and multimedia requirements specification

Having established three repositories, the project then explored how multimedia content could be exchanged and shared between the repositories, using OAI-PMH and METS as transport mechanisms. This phase of the work proved very frustrating because of a lack of consistency and compatibility across the platforms available. It became clear that METS provides little more than a broad structure for packaging digital objects, complete with their metadata and significant investment is required to develop a range of application profiles which will facilitate data exchange between different systems. The use of OAI-PMH within the context of multimedia repositories can also be somewhat problematic: in many systems, the handling of metadata for harvesting is far from ideal and the use of the protocol for object sharing even less certain.

One very important outcome from the project was the positive response of the academic partners to the possibilities opened up by using a multimedia repository. The scenarios explored within MIDESS won a significant degree of acceptance within the user community and the project's work has emphasised the need to fully integrate repositories into the institution's enterprise architecture and content management services. A number of issues are identified which require consideration in the implementation of a multimedia repository service.

Finally, there emerge some clear questions about how multimedia repositories should be incorporated into the broader national and international information infrastructure. MIDESS found that the value of the multimedia repository lies, not least, in its ability to make research materials available to support subsequent learning and teaching. The implications of this for content aggregators and for the national information architecture require further investigation.

Background

A number of university libraries in the UK have engaged in substantial digitisation programmes, and a large amount of content has been generated as a result of these initiatives. However, few institutions have, as yet, seriously explored the management of digitised content through a digital content infrastructure. Digitised content typically tends to be available through unstructured or semi-structured html pages, or managed through proprietary systems which do not provide for adequate exposure, sharing or re-use of materials. Awareness and use of digitised collections created by libraries – whether for teaching or research – is, consequently, quite low. Issues such as metadata creation and management, effective searching and retrieval, re-use of content in an e-learning context and digital preservation are also poorly understood, and not widely implemented.

Digital content is also being created by academic staff within institutions, primarily to support learning and teaching. Examples might include digitised film clips used in a classroom situation, and medical slides used in a laboratory environment. Digital content is often embedded into a virtual learning environment, or made available to students through a course web page. These activities are often quite widespread within an institution, but opportunities for the sharing and re-use of this type of content – which is often expensive to create – might be lost because there is no institution-wide facility for the shared management and discovery of the content.

These issues extend beyond any one institution and there is a need for joined-up thinking about the creation and management of digital content at a national and international level. With suitable infrastructure in place, institutions could avoid duplicating effort by digitising material that has already been digitised elsewhere, and can maximise the benefits to end-users by working together to create complementary collections.

Recognising these problems, the MIDESS Project set out to explore the management of digitised content in an institutional and cross-institutional context through the development of a digital repository infrastructure. It intended to address how support can be provided for the use of digital content in a learning and research context in an integrated manner and also to explore how the use and management of digital content can be joined up in a national context.

Project partners were the libraries of the London School of Economics, University College London, the University of Birmingham and the University of Leeds. The project ran for 2 years, 2005-2007, and was funded by the JISC (with supplementary funding from the Consortium of Research Libraries) as part of its *Digital Repositories Programme*.

Aims and Objectives

The aims and objectives of the MIDESS Project were stated in the project plan as follows:

- To build functional demonstrator digital content repositories at three of the partner institutions, providing a set of suitable platforms to examine the issues and validity of implementing full digital content management services.
- To populate these repositories with digital content which has already been created, or is currently under creation, by the partner institutions. This includes digitised images from slides and manuscripts and multimedia materials such as video and audio materials. A range of subject areas will be included, to include some complementary collections.
- To explore opportunities for the sharing and re-use of collections across institutions through the active involvement of academic staff at each partner institution.
- To establish metadata standards and develop metadata for each collection added to the repositories. The METS standard will be explored for bringing together various metadata elements – for example, IMS for e-learning materials. METS is an overarching standard which can be utilised to repackage and structure existing metadata.

- To explore the role of digital content repositories within the institutional information architecture, with a particular focus on interoperability with enterprise content management architectures.
- MIDESS will also aim to establish how distributed digital content repositories could encourage the wider exposure and sharing of content across institutions through an evaluation of requirements for centralised metadata harvesting services.
- MIDESS will seek to pilot an infrastructure which could serve as a model for future distributed national digitisation activities.

In short, the project set out to create a set of suitable platforms, populated with content, in order to examine the issues and validity of implementing full digital content management services within each institution and the potential for sharing and re-use of content between institutions.

In the second year of the project, it became clear that there would be a small underspend in the budget. A request was therefore made to JISC – and agreed – to extend the project by 3 months. This would permit further work on content sharing and re-use, specifically on

- Use of the repository to support learning and teaching.
- Re-use of recordings digitised under the ERA licence for learning and teaching.

Methodology

The workplan envisaged delivering these objectives via a set of inter-related and overlapping work-packages. Six of these work-packages consisted of a series of specifications setting out the key requirements for the implementation of the actual repositories, based on the local needs at each of the partner sites. A further work-package, informed by these specifications, allowed for the implementation of the repository systems and for their population with sample content, and provided an environment in which the specifications could be tested and refined. These repositories then provided the infrastructure on which to carry out an in-depth investigation of the possibilities for metadata harvesting, content sharing and re-use. Ongoing project management, evaluation and dissemination were undertaken within the framework of two additional and overarching work-packages.

Activity in the initial phase of the project was focused on understanding the requirements of a multimedia repository by drawing up 2 key documents:

- A functional and technical requirements specification
- A user requirements specification

The user requirements specification was based on a survey of stakeholders, identifying their requirements for depositing material in a multimedia repository. In contrast, the functional specification was compiled by a mixture of desk research and feedback from the user requirements specification, and included a detailed comparison of software platforms identified as providing repository functionality. An initial set of digital collections for inclusion in the repositories was also identified.

Work undertaken on these 2 specifications then allowed Birmingham, Leeds and LSE to each select and implement a suitable repository platform. Work on the initial set of digital collections, combined with the experience gained from implementing the actual repositories and with further desk research, allowed 4 further specifications to be compiled, viz.

- Metadata requirements specification
- Digital preservation requirements specification
- Requirements for integration with enterprise infrastructure
- IPR and multimedia requirements specification

With repositories installed and populated, the project had planned to move into a final phase which would explore how digital content in repositories can be made available for resource discovery and re-use. It was planned to explore the potential of standards and services such as OAI-PMH and METS,

together with the implications for metadata harvesting services. It was also intended to build collections of related materials in all 3 repositories in order to engage with academics and explore the possibilities for re-using digital materials in an inter-institutional context.

In practice, this final phase of the project – exploring resource sharing and re-use - proved problematic for a number of reasons, explored in greater detail later in this report. The actual implementation of the repository platforms proved more difficult than anticipated at both Leeds and LSE, absorbing more staff time than had originally been anticipated. Project staff also encountered unexpected obstacles in their attempts to share resources via METS and OAI-PMH. Detailed technical work was undertaken in both these areas in an attempt to overcome the problems, but it was never possible – within the project timeframe - to move to a situation where end-users could be invited to engage in testing the repository services or the potential for object re-use. However the engagement with interested parties within the academic community did permit an in-depth exchange of ideas about the role of repositories in supporting learning and teaching. This provided a significant impetus to refining earlier ideas about functional requirements for repositories and their place within the broad information architecture of the institution.

Implementation

Preparatory phase

In preparing the functional and technical requirements specification, the team used the repository survey carried out by the CURL Task Force on Digital Content Creation and Curation (Digital Resource Management Sub-Group) as its initial starting point¹. Although this survey was primarily geared towards digital preservation functionality, it provided a useful starting point for identifying key systems as it focused on products which could be used to support the management of digital image collections. The team identified four further systems which should be considered for evaluation and an initial analysis of the products was undertaken at a high level, focusing on the core technical requirements for the project. This narrowed the list of products to 6 and these were then evaluated against a more detailed set of requirements drawn from the results of the User Needs Analysis. A “Repository Evaluation Day” was held in December 2005 at which all 6 products were the subject of a 45 minute presentation, plus written responses to a set of questions. Those present were invited to score the products, and outstanding questions were followed up with the software suppliers and/or users. Because there was a mix of commercial and open source products, it was a little more difficult to achieve a fair and objective comparison, but the comparative tables which formed a major output of this work-package were used to inform the choice of repository at Birmingham, Leeds and LSE.

The user requirements specification was clearly key to the development process in order to align MIDESS development work with the needs of the user communities in the participating institutions. Drawing up the specification also provided an opportunity to identify individuals with content which might be used in later stages of the project. An initial list of stakeholders was drawn up, based on existing contacts. One to one interviews with 17 individuals followed to develop a set of case studies on current and potential use of digital content.

At the same time, an [on-line questionnaire](#) was constructed and publicised throughout Leeds, Birmingham and LSE (UCL did not take part in the survey since they were specifically assigned the copyright workpackage for the project). The questionnaire asked how respondents were creating, storing and using digital material. Section A of the questionnaire was designed for staff such as academics, researchers and administrative staff who were actively involved in creating and using digital multimedia material to support either their learning and teaching or research. Section B of the questionnaire was designed for university staff involved in supporting the creation and use of digital media, although these staff would not normally themselves be directly involved in creating the digital material. Typical staff here might include library, IT staff, media services staff and e-learning support staff. Some people in the partner universities have a dual role of both creating digital material while also assisting and supporting other colleagues and these respondents usually completed both sections of the questionnaire. 205 responses were received and the replies to each question were collated.

Based on the information gathered from both interviews and questionnaire, it was possible to identify a broad set of technical and functional requirements which needed to be addressed. It was also clear that the repository platform needed to be able to hold very large collections in a wide range of different formats, so flexibility would be critical. The ability to manage access permissions also emerged as essential for many collections.

Repository implementation

Using the information gathered during the project's first phase of preparatory work, it was possible to move forward to select and implement a digital repository platform at three of the partner institutions, Birmingham, LSE and Leeds. Each of these sites had local priorities and requirements, and each had identified collections of digital materials which could be used during the initial ingest phase. Equally, in order to facilitate sharing and re-use of objects at a later stage in the project, some similar collections were identified for inclusion across all 3 sites including images of coins and of medieval manuscripts and digitised recordings of lectures.

At the University of Birmingham, the project was seen as an opportunity to create a pilot for an institutional repository and thereby explore the long-term requirements and issues. Support for open standards was seen as essential, and there was particular interest both in storing and making available the outputs from the Library's digitisation programme and also in holding learning objects which could then be made available through a VLE. DSpace was selected as an open-source product which fitted well into the existing technical infrastructure at Birmingham. Sample files in various formats were loaded, including digitised images derived from the University coin collection, playscripts, maps and video. An attempt was also made to load learning objects, complete with IMS/LOM metadata. However the facilities for metadata handling within DSpace did not make this a viable option.

LSE chose to install Fedora, another open-source product. Fedora seemed to offer a powerful and very flexible structure within which to hold digital objects, together with multiple metadata datastreams where required. It also offered support for METS, for version management and for a wide range of protocols which offered the possibility of easy integration into the broader information architecture. LSE concentrated on migrating two major collections into Fedora: one comprised archival photographs associated with Malinowski's ethnographical fieldwork, the other a collection of digitised recordings of television programmes to support learning and teaching within the institution. In both cases, rich metadata was available in an external database, and METS was used to create a datastream for ingest into Fedora.

Unlike Birmingham and the LSE, the University of Leeds had secured additional funding in order to establish a multimedia repository service for the University. This permitted installation of a commercial product, offering the advantage of richer in-built functionality, including a fully-developed interface to support resource discovery. Endeavour Curator was selected because of its easy-to-use public interface and its support for a wide range of metadata standards including MODS and METS. In the event, both Endeavour and the rival software company Ex-Libris were purchased by a private equity company late in 2006. As a result, development work on the Curator platform ceased and Leeds was obliged to migrate to DigiTool in the first half of 2007. This had a severe impact on project timescales, and delayed the loading of live data; however a wide selection of test data – digitised images, audio and video files - was successfully loaded into both Curator and DigiTool.

Metadata, digital preservation and IPR

The work on populating the repositories ran in tandem with work on metadata, digital preservation and IPR issues.

One thread in the work on metadata was a broad survey of pertinent metadata schemas and related issues in order to inform subsequent work. From this came a summary document which forms Part A of Work-package 4 and has proved very useful in providing background information for staff who had little previous experience of work in this area. Equally, at each site, it was necessary to examine in

detail the metadata requirements of the collections identified for ingest. For this work, collaboration was necessary between project staff and other individuals in the institution who had a detailed knowledge of metadata. Working with multimedia formats across a wide range of subject areas, various issues arose which took time and energy to resolve.

At Birmingham, the project very quickly ran up against the limitations of DSpace, which uses qualified Dublin Core (qDC) for holding metadata internally. This in itself was a significant limitation and eventually excluded any serious work on holding learning objects as there was no way of mapping from the IMS CP standard to qDC, although substantial effort was invested in exploring this possibility. Birmingham also hoped to ingest images from a coin collection and the metadata was more successfully mapped onto qDC in this case. An attempt was made to adhere strictly to the standard, but this was not easy because of the widespread recommendation that where an object has been digitised, many fields (such as publisher and format) should contain details of the digital object rather than those of the original. This tends to compress most of the metadata relating to the original into the description, with a consequent loss of structure and implications for subsequent indexing. In the case of the coins, a further issue was the need to distinguish between information on the recto and verso of each coin. Non-standard qualifiers were used to achieve this, but with the unintended consequence that such data was invisible when the records were harvested.

LSE, working with Fedora, also encountered significant issues with metadata. Although Fedora is able to hold any number of metadata datastreams associated with the digital object, the software requires a datastream in simple Dublin Core (DC) for its internal indexing routines, and this is the only metadata used to drive resource discovery. Once again, the loss of specificity for resource discovery purposes was a significant disadvantage. LSE built complex pre-ingest routines which took existing metadata – both in EAD and a local, non-standard format – and created a parallel DC datastream for ingest to satisfy this requirement in Fedora. METS was used as a content packaging standard to bind together the object and the multiple metadata datastreams for ingest into Fedora. Since LSE had a substantial collection of digitised recordings of television broadcasts for ingest, consideration was given to using the European Broadcast Union's EBU Core Application Profile² (actually drawn up in the context of radio rather than television broadcasts). Mappings were successfully constructed from the previous metadata schema for many of the elements, but certain differences in structure and content meant that complete compliance was neither possible nor desirable.

Many of the collections at Leeds had very simple metadata and DC or qDC seemed completely appropriate. However a key objective was to load digitised page images from medieval illuminated manuscripts. An MS Access database held very full metadata in VRA Core 4.0 format. Unfortunately, Curator did not support VRA Core, so the decision was taken to migrate the metadata to the MODS schema. This was a complex operation, but there was considerable benefit in using MODS because of its ability to "nest" data, so that a single record could hold structured information in separate sections of the record about the individual digitised page and also about the manuscript from which it was derived. A mapping onto MODS 3.0 was constructed, at which point work on the Curator system ceased and Leeds was obliged to migrate to DigiTool. At that time, DigiTool did not support MODS, though support was promised for the next release in June 2007. Installation of that release revealed that support only extended to MODS 2.0, so all the mappings had to be modified, with some loss of granularity and specificity. In consequence, during the MIDESS project only a limited number of images were able to be loaded for this collection, and they were accompanied by a slimmed-down DC record. This did however permit testing of the functionality in DigiTool which permits multiple objects (in this case pages) to be linked to a single "parent" record (in this case, the manuscript from which various illuminated pages have been digitised). It was also useful to discover that this linking cannot be undertaken automatically at ingest but has to be undertaken manually within the staff interface.

Digital preservation is a key concern for all repositories, and the implications for MIDESS were explored in Workpackage 5. The range of formats to be handled by a multimedia repository had already been identified as part of the user requirements specification, as had the widespread concern regarding long-term preservation. Desk research identified the key issues to be considered and noted standards such as PREMIS and Z39.87 which could provide a framework for addressing them. An approach to digital preservation was then suggested for each of the principal media types identified. The approach to digital preservation embedded in the software platforms used within the project is also noted but the scope of the project did not include any attempt at implementing a preservation regime.

Work-package 7 comprises a comprehensive survey of intellectual property and related issues in institutional and cross-institutional multimedia repositories, including areas such as copyright, licences relevant to multimedia materials and international conventions and directives. The work undertaken was intended to be practically-oriented and capable of guiding those responsible for implementing repositories. Therefore, a series of case studies are considered, based on the list of media types from the user requirements specification and there is a substantial section detailing recommendations for good practice in managing multimedia collections.

Enterprise architecture and repository services for learning and teaching

The study of enterprise architecture in work-package 6 was primarily conducted via a case study of the MIDESS implementation at the University of Leeds. The choice of a commercial software platform at Leeds and the intention to move as fast as possible towards provision of a full repository service available across the institution meant that Leeds was better placed than the other MIDESS institutions to consider and address these issues. By considering the organisation of information systems at the University of Leeds, the relationship of these information systems to the digital repository could be explored and it was possible 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. Issues covered include large-scale storage via a SAN (Storage Area Network), streamed media, authentication, resource discovery and connectivity with portals and the VLE. Integration with the enterprise architecture at system level (use of the SAN and the streamed media server for storage and use of Shibboleth for authentication) was achieved by the end of the project but service integration was delayed by the switch from Curator to DigiTool which prevented the transition to a live service within the time-frame of the MIDESS project.

As the project progressed, further discussions with academics at both Leeds and LSE permitted scenarios to be drawn up illustrating how multimedia materials might be exploited for learning and teaching and the consequent implications for VLE connectivity. At Leeds, a partnership with the HEFCE-funded ALPS CETL (a Centre for Excellence in Teaching and Learning which is focused on Assessment and Learning in Practice Settings) proved particularly useful and illuminating. Most medical and healthcare students are required to spend extended periods in a clinical setting, often at some distance from their home university. The ALPS CETL is working in partnership with the local health services, practice networks and professional bodies to improve student knowledge and skills and ensure a good fit with the professional environment within which these students will work once qualified.

Through much of the project, ALPS staff worked with MIDESS staff on the storage of their learning objects / digital material in the DigiTool repository at Leeds. This material needs to be restricted to ALPS members across the partner institutions because of its potentially sensitive nature. The material consists primarily of image and video content captured on mobile devices by staff or students on clinical placement within the National Health Service, which is later used for training purposes by other medical staff and medical students. Sample objects were loaded, which clarified many of the issues involved, but the system migration and project time-scales prevented any service delivery to the students themselves.

The ALPS requirements are particularly complex because:

- The ALPS-CETL is a collaborative programme between five independent higher education institutions, each with its own infrastructure.
- Staff and students additionally require access to resources available through the National Health Service, particularly those included within the content management platform being developed for the NHS using EMC Documentum
- Some material is confidential and/or sensitive and so it is necessary to enforce complex access restrictions in a cross-institutional framework.
- Staff working within the project framework would potentially benefit from access to learning objects created in other HE institutions or available through national services such as JORUM.

Overall, the ongoing contact with these and other academics was invaluable in clarifying long-term objectives for a digital repository service and also, incidentally, in “selling” the service to senior managers within the University. However it did identify various deficits in current functionality:

- A basic requirement is for the module administrator within the VLE to interface seamlessly to the repository in order to provide students with access to materials held within the repository. Whereas Leeds’ original Curator system included a VLE integration module which permitted seamless execution of key functions, its replacement, DigiTool, lacked such functionality. It would be possible to use the VLE and DigiTool APIs to replicate this (at some cost), but there is no such facility built-in.
- The ability to download materials into the VLE for incorporation into new learning objects is also problematic since the public interface to many repository systems is designed solely for access by the end-user. Typically repository software launches an application which will permit the object to be viewed or otherwise used; there is generally no uniform way of bypassing this in order to access the raw datastream.
- Finally, storing learning objects in the repository requires support for IMS and/or SCORM – a feature lacking in many repository platforms.

Resource discovery, content sharing and re-use of digital objects

Once some material had been ingested into the repositories in 3 different institutions, the MIDESS Project was in a position to explore how such materials could be shared and re-used within an inter-institutional context. Work focused in particular on the possible use of the OAI-PMH protocol for resource-sharing and on the use of METS as a possible transport mechanism. On both counts, the experience was unfortunately somewhat negative.

Since OAI-PMH is an established protocol and widely deployed for metadata harvesting across many platforms, it was hoped to use this as a vehicle for resource discovery and content sharing. Unfortunately, availability of a harvesting facility had not been considered at the project planning stage and neither Curator nor Fedora included a facility for harvesting records from external repositories. In contrast, a harvester plug-in was available for DSpace and it was therefore planned to use the DSpace installation at Birmingham as a platform for harvesting records from the repositories at Leeds and LSE. However, efforts to get this plug-in operational were unsuccessful, and it became clear that it was not compatible with version 1.4 of DSpace, which was the version installed at Birmingham at that time. No Java programming expertise was available to update the plug-in, so a different approach was required.

It would have been possible to try to install a free-standing open-source harvesting package such as PKP. However, since the particular interest of the MIDESS project was in how digital objects could be shared between institutions, it was decided to write a simplified harvesting routine to run on the DSpace server. This routine would collect records from the other repositories and save them on the server. They could then be reformatted into DSpace-compatible XML format for subsequent loading into DSpace using the standard import procedures. This harvesting routine used a bare minimum of OAI-PMH syntax in order to obtain records from the remote repository.

Even this simplified “harvester” encountered problems initially when tested against the LSE and Leeds repositories. Firewall issues were encountered which prevented data being passed on the ports specified by the target repositories. And further problems were encountered at Leeds which at that time was running a beta version of the Curator software, with a bug present which prevented access. However metadata was eventually harvested successfully from both Leeds and LSE, and it was possible to explore what opportunities this presented for resource sharing.

It was a surprise that not all the software platforms deployed within the MIDESS project provided an implementation of OAI-PMH which exposed adequate metadata for harvesting. One fundamental issue was that the oai_dc records harvested from Endeavor Curator and from Fedora did not include the oai_dc:identifier element which would normally contain a URL linking back to the originating repository so that the user can access the digital object. Without such a link, the user who searches an aggregator service or portal and discovers the existence of a relevant object is obliged to navigate to the holding repository and initiate a new search there in order to access or use the object. This constitutes a significant impediment to effective exploitation of the underlying architecture.

Further and more complex issues were encountered for those objects whose associated metadata was held in a schema other than simple Dublin Core. Whereas Curator handled this situation by mapping the individual elements onto an appropriate element in the oai_dc schema, the project found that many DSpace records were harvested with only title, format, type and language elements present even though the internal qualified Dublin Core record was rich in metadata. It is technically possible in DSpace to expose additional metadata schemas over the oai channel but this would have required the creation / modification of multiple internal modules of the DSpace software. This issue of mapping between metadata schemas has proved less troublesome for eprint repositories, which tend to contain fairly homogeneous objects and with metadata commonly encoded in Dublin Core or qualified Dublin Core. For multimedia repositories as in the MIDESS Project, the issue is more problematic: the varied nature of the objects leads to a much wider range of metadata schemas including MODS, EAD, MARCXML and locally-tailored variants of qualified Dublin Core. Customisable mappings are essential if key metadata elements are not to be lost in harvesting.

Towards the end of the project, the ExLibris DigiTool software came on stream at Leeds. This included an integral module for harvesting and the opportunity was taken to further investigate OAI-PMH functionality, even though time was limited. A number of new issues were encountered (shortcomings in the logs and other reporting facilities, configuration issues for harvesting specific collections and a bug in the ingest process which meant for a time that records could be harvested but not ingested into DigiTool), none of them insuperable, but all demonstrating the complexities involved in setting up an OAI-PMH service provider facility.

The MIDESS Project wished to go beyond resource discovery and harvesting to explore how digital objects could be shared and re-used within a cross-institutional context. Since METS (Metadata Encoding and Transmission Standard) was designed *to facilitate the interoperable exchange of digital materials between institutions*³, it offered the obvious transport mechanism for achieving this. Moreover, LSE had successfully used METS to create ingest packages for Fedora and the project officer there had developed expertise in handling METS objects. However at an early stage, it was discovered that a METS file exported from DSpace would not load into Fedora – the format of a METS object exported from DSpace differed substantially from those created by LSE for local ingest. Software from the JISC Repository Bridge project⁴ – who had encountered the same problem - was installed, but without success. And the project lacked the technical/programming skills to explore the reasons for this failure and rectify it.

Curator also proved unable to ingest the METS objects from DSpace. However in this case, the problem seemed more likely to be caused by a bug in Curator, which was still in a beta release. After the migration to DigiTool further attempts were made to transfer a METS object from DSpace. Again, the initial outcome was failure, but, it proved possible to isolate the problem to a few key areas of incompatibility with DigiTool and, after manual editing of the METS object, a successful ingest resulted. One key edit was to remove the entire section containing administrative metadata (the amdSec element), where the lack of common metadata standards between DSpace and DigiTool was clearly a problem – in particular the fact that DSpace encoded a text string in base64 format. Although both simple and complex digital objects could be transferred, upon condition of this manual intervention, such a procedure is clearly impractical in an operational context. A secondary problem was also identified concerning the metadata formats embedded in the METS object. For the object to transfer successfully, it is clearly essential that the ingesting system should be able to handle all the metadata schemas present within the METS object (not only those relating to descriptive metadata, but also those for administrative and technical metadata).

Outputs and Results

Taken as a whole, the outputs from MIDESS should be able to provide something akin to a road-map for those seeking to implement a multimedia or institutional repository. Although inextricably rooted in the user communities and institutional context of the partners, the reports on each work-package indicate very clearly the key issues which need to be considered, suggest ways of addressing many of these, and identify potential pitfalls for others again. Indeed precisely because this work is rooted in a real-life context, it carries particular value for others facing a similar challenge. Although there is much experience within the JISC community of implementing eprint repositories to hold research

outputs such as journal or conference articles, working papers and similar textual materials, there are far fewer instances which can offer guidance in the implementation of a repository to hold a range of multimedia materials.

Each of the work-packages includes detailed tables, summaries and recommendations to which the reader is referred for more detail. However in this final report, an attempt will be made to draw out key themes and findings which are of general significance to those working with repositories in general, and multimedia repositories in particular.

User requirements (WP3)

Know your users and their requirements! This has to be the most important lesson from MIDESS. The survey of users uncovered a wide range of material suitable for ingest into the repository. Analysing this material and how it would be used enabled appropriate decisions to be made in all subsequent work-packages. It determined functional and technical requirements (and therefore choice of repository platform), metadata schemas, connectivity to other campus systems, preservation, access, resource discovery, and more.

It also revealed the quantities of digital material which exist on campus which would benefit from robust and secure management within a repository framework. Significant institutional assets are scattered on a wide variety of media, some more secure than others and are therefore very vulnerable. They are also, in the main, invisible to others, and would benefit from being made visible on a central system.

The range of such material is very varied, and so too are the technical formats in use. Many are not optimal, the quality of digitisation may indeed be poor compared with what can be achieved, but they are of value for research and/or learning and teaching.

Many users have also identified further materials which they wish to digitise. In many cases, there is a lack of resources and/or expertise to undertake this. A repository must have a clear policy of which areas of service are included and excluded. Relevant issues can include actual digitisation, ingest, metadata, ongoing maintenance, long-term preservation, IPR, re-use of materials.

Functional and technical requirements (WP2)

Although the tables comparing functionality for a range of software platforms will quickly become out-of-date, the list of functional and technical requirements used for the evaluation do identify the key areas which need to be considered. Each implementation will set its own priorities, but these are likely to include the ease of adding digital material to the system, the requirement for long-term storage, the ability to password-protect or restrict specific digital collections, the ability to search across collections, the effective management of consistent metadata schemas, and the provision of easy access to the digital material. Connectivity and the ability to interface to other systems both on and off campus is also likely to be critical, so it is important to specify relevant standards which can enable this.

The 3 case studies describing how each site chose a software platform can also inform other institutions needing to reach a similar decision, particularly as they cover both commercial and open-source products. In both cases, it is important to adopt an approach which predicts costs over the whole life cycle of the product. The experience of MIDESS would also suggest that the costs of implementing and tailoring the chosen system should not be underestimated, particularly for open-source products. More generally, the lack of experience within the community in implementing multimedia repositories means that staff costs are likely to be higher as solutions will have to be tailored to meet local needs.

In the second year of MIDESS, a further document entitled "MIDESS – institutional buy-in case studies" was prepared, briefly outlining how each repository had bedded in to the institutional

structures. It is important not to lose sight of these broader and less technical aspects of any implementation.

Metadata (WP4 [A](#) & [B](#))

Part A of this work-package provides a useful overview of metadata standards for the non-specialist. This can be of particular use because of the very varied requirements when dealing with multimedia materials.

Metadata is absolutely critical to subsequent discovery and use of the digital objects held in a repository. Metadata creation is expensive, but an object without metadata is for most purposes invisible and effectively lost. Therefore, the requirements of each collection need to be expertly assessed, taking into account factors such as

- Intended use and longevity of the material
- The relationship of the material to other collections both internal and external to the repository
- Whether any metadata already exists; if so, in what format; if not, the options for automatic capture or manual input
- Technical limitations of possible metadata solutions e.g. failure of the repository software to adequately support the desired/existing metadata schema

Adopting a well-supported metadata schema has advantages, but in some cases the richness, complexity or special requirements of the metadata can make it desirable to implement a less common schema. Consideration should be given as to how the metadata schema chosen will map onto indexing rules within the repository and how the public interface will be presented. In all this, staff expertise is required in order to complete a successful metadata implementation.

One clear lesson from MIDESS is that even where established standards are being followed, this is not sufficient to guarantee compatibility and interoperability. Application profiles can help in this regard, often supporting a very specific and tailored approach to metadata, deriving recognised elements from the name-spaces specified in order to meet the needs of a particular community. This can be particularly valuable where data is to be shared across institutions.

Preservation issues ([WP5](#))

MIDESS did not focus particularly on preservation, particularly in its practical work, but it did survey the key issues affecting multimedia files, summarised the main issues and options and looked at the degree to which the 3 software platforms provided support for preservation. Strategies for digital preservation are still far from common, but every repository should be working towards developing one.

IPR ([WP7](#))

One of the key outputs from the MIDESS Project is the survey of intellectual property and related issues for multimedia repositories that forms work-package 7. With its case studies and recommendations of good practice, this is key reading for any repository manager, particularly as many academic staff have only a limited understanding of the issues involved. Legal compliance is crucial if the host institution is to avoid embarrassment and prosecution. It should be noted that the report is a snapshot illustrating good practice, not a definitive legal statement, and some parts may be rendered obsolete by subsequent legislation (the Gowers review⁵ of intellectual property was published in December 2006, just after the MIDESS work-package had been completed!).

Two case studies from the MIDESS partners can illustrate some of the issues arising from widely used licences such as ERA and CLA.

LSE wanted to digitise recordings of television broadcasts made by UCL under the ERA licence and make them available to LSE students. The ERA licence prohibits digital transmission of such material between institutions, so it was necessary for UCL and LSE to maintain copies of all the broadcasts in both repositories. Moreover, the licence states that recordings may only be viewed by users who are physically on the institution's premises, so the repositories must be configured to enforce these access controls.

Similarly, Leeds wanted to use the repository to store and make available course readings digitised under the CLA Photocopying and Scanning Licence. However, the terms of the licence place tight constraints on access by any students except those enrolled on the specific module for which the material has been registered, and this restriction extends to public availability of the metadata. Access controls at this level imply full integration of authentication and authorisation mechanisms for the repository with other relevant campus information systems.

These two case studies – and further examples relating to data subject privacy – illustrate the importance of identifying requirements and establishing mechanisms for controlling access when setting up a multimedia repository.

Enterprise architecture ([WP6](#))

The work undertaken shows 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.

Leeds is used as a case study and the issues addressed include:

- Storage
- Streamed media servers
- Authentication
- Resource discovery
- Integration into the broader information architecture, including OPAC, portals and VLE

This work follows on directly from the specification of functional and technical requirements.

Repository implementation and population ([WP8](#))

The implementation phase of MIDESS resulted in 3 functional repository platforms, each containing a mix of material to test how different types of multimedia content can be ingested, stored and made available.

For some repository managers, the detailed descriptions of 3 very different implementations will provide a useful resource. This is particularly true for those working with Fedora, because LSE have fully documented the process of ingesting into Fedora audio and video files which were previously held in an external database. The METS manifest, the Fedora commands and post-ingest screenshots from Fedora are included.

More generally, several key themes and lessons emerge from the combined experience:

- For open-source products, another site may have succeeded in implementing a particular feature, but this does not imply that it will install easily at your own repository. This can result from differences in version, in hardware platform, or simply in having local expertise to tweak the software module.
- Even with a commercial product, bugs and other unpredictable behaviour are likely to emerge because the products tend to be fairly new to the marketplace and complex in what they are trying to deliver. Comparison can usefully be made to the experience of Library Management Systems during the previous decade.

- Particularly for multimedia and/or unusual collections, it is unwise to underestimate the staff effort involved in the range of processes required to achieve a successful ingest. Not only staff time is required, but also the particular expertise necessary to resolve the issues concerned – and these can be very varied.
- Following on from this, nearly all repository implementations require technical expertise to be available locally if a full range of services is to be delivered.
- The fact that a product supports a particular protocol does not necessarily mean that implementation will be easy. Version numbers and interpretations may differ, impeding integration and interoperability.

METS and content packaging ([WP4](#) & [WP9](#))

The failure of the project to transfer digital objects successfully between any of the repositories via METS content packaging, even after installation of the Repository Bridge software, suggests that there remain significant obstacles to the use of METS for this purpose. In practice, METS provides little more than a broad structure for packaging digital objects, complete with their metadata, and the schema provides so much flexibility that it cannot in itself guarantee a successful transfer between two divergent systems. Potential issues can include:

- divergent use of metadata schemas
- differences and restrictions in the way that the various systems implement METS
- non-compliant METS files produced by the METS export facilities from certain products.

However, within strictly controlled parameters, METS is capable of offering a robust and flexible standard for content packaging: its use by LSE to ingest complex objects into Fedora is demonstrable proof of that and is in line with the experience of other implementations such as the Oxford Digital Library⁶. If METS is to be used in order to facilitate data exchange between different systems, then it is imperative to ensure that common standards are implemented at both ends. This could be achieved by using an application profile – although there is currently little or no built-in support for application profiles within many software platforms.

The work which MIDESS undertook with IMS CP was similarly lacking in success. Although IMS and SCORM are both widely implemented for content packaging of learning objects, none of the repository platforms implemented within the project provided integral support for these standards, and attempts to map onto schemas which were supported were unsuccessful. In selecting and implementing a repository, it is therefore most important to recognise the divergence between products oriented towards learning objects and those which address more generic requirements.

Metadata harvesting ([WP10](#))

Although OAI-PMH is widely used for metadata harvesting, MIDESS encountered a surprising number of problems in implementing it for resource discovery. Any prospective service needs to consider carefully the following issues:

- Are there firewall or related issues which will obstruct the operation of the harvesting operation?
- Capacity of the repository into which the metadata is to be harvested. If this is commercial software with a licence charge based on the number of objects in the repository, then any significant harvesting could be problematic on grounds of cost.
- How is the service user going to access the object described? Not all repositories provide a link back to permit access to the object through that repository,
- What precisely is to be harvested? If the intention is to harvest only selected objects in the repository, then that repository needs to support harvesting subsets via OAI-PMH identifiers.

- Is adequate metadata going to transfer successfully? This depends on the nature of the implementation at both ends of the service. If oai_dc is to be used within the harvest operation, then the data provider must provide appropriate mappings for other metadata schemas (including qualified Dublin Core!). If other schemas are to be used, then the service provider must have a means of indexing and displaying all pertinent metadata in order to provide an appropriate service to the end-users.

MIDESS had hoped to move beyond metadata harvesting and also investigate whether digital objects could be shared by using the OAI-PMH protocol as a transport mechanism for METS objects. Although such a transfer is valid within the framework of the protocol, it is not widely supported by repository platforms, and a successful implementation would additionally depend on compatibility in the way that both repositories implemented METS.

Outcomes

External evaluation

As part of the project evaluation, an outside evaluation of the MIDESS Project was commissioned from an individual with significant experience of both repository implementation and of JISC projects. In the brief given for the evaluation, 2 main questions were to be addressed:

- Where does the MIDESS project fit into the broader picture of work on repositories in the UK (particularly that funded by JISC) and what are its distinctive features?
- What has the project delivered and how has this advanced our understanding of how to implement a working repository?

The resulting [report](#) adopts a threefold approach in its evaluation:

- A review of the aims and objectives of the project as stated in the project plan
- A review of the intended project outcomes to assess whether the project has been able to bring these about
- A review of each workpackage, covering the area addressed within these and the lessons learned from carrying out the work

The report also identified additional issues arising from the project and makes recommendations regarding issues and lessons learned that would benefit from wider dissemination. Finally, conclusions are given on the overall impact of the MIDESS project.

The report provides a comprehensive assessment of the MIDESS project, and is of particular importance in evaluating the project since the various difficulties encountered in the implementation of the 3 repositories meant that it was not possible to obtain feedback from actual users, as originally planned. The following quotation seems therefore to accurately summarise the real value of the project to the wider community:

It is acknowledged that staff and technical problems have prevented the project from achieving all that it had hoped ... The work undertaken though, dealing with issues that all too often occur at an institutional level, has produced a body of information that can sit alongside the outputs from the projects listed above to help flesh out the developing picture of how repositories can be used effectively within institutions, taking into account links with national services. There is no doubt that there is some way to go before repositories become an established part of institutional infrastructure: projects at the institutional coal-face like MIDESS provide real experiences that others can learn from.

It is to be hoped that the availability on the MIDESS web-site of detailed accounts of the work undertaken in each work-package will contribute towards this. However, the project has also responded to the evaluation report by summarising the work and key findings in a brief document intended to provide guidance to those implementing multimedia repositories in other institutions.

Dissemination

The project has also played a significant part in raising awareness of the role which repositories can play in supporting learning, teaching and research, not only within the partner institutions, but also within the wider academic community. The exploration of issues around information architecture and the work on scenarios describing how a digital repository can support learning and teaching in a research-intensive university have the potential to shape opinion in this developing area. The following are the key channels through which the MIDESS partners have been able to share and disseminate their experience.

1. Dissemination within the partner institutions has helped raise awareness of best practice in many areas relating to the management of digital materials, and will inform future decisions.
 - The project Steering Group and local Working Groups brought together colleagues from a number of different backgrounds, who in turn reported back to interested colleagues in their own area.
 - The user requirements survey and subsequent contact with academics around specific collections has helped develop an understanding in each institution of the role to be played by repository developments.
 - At each of the partner institutions, presentations were made to management teams and advisory boards of Library/Information Services and to colleagues generally within the services.
 - Presentations were also given to senior academic staff such as at Learning and Teaching Advisory Board.
 - Presentation to be given at the University of Leeds Learning & Teaching Conference, 11 January 2008
2. By sharing results of the project within the repository community MIDESS should permit other repository managers to benefit from the experience gained. Channels have included:
 - MIDESS Dissemination Day, Leeds, May 4th 2007, publicised within the UK repository community and beyond, at which principal findings of the project were presented and discussed
 - [Presentation](#) at the JISC Conference *Digital Repositories: Dealing with the Digital Deluge*, Manchester, 5-6 June 2007
 - Notification of work-packages availability on the JISC-Repositories listserv.
 - Liaison with other JISC projects, including CHERRI, CLiC, RepoMMan, Repositories Bridge, SAFIR, WM-SHARE
 - Liaison with VSM Portal Demonstrator project with a view to making content available via the VSM Portal
3. Dissemination activities such as conference papers have also reached a wider audience of academics and information managers and may therefore play a part in shaping wider policy and development.
 - [Presentation](#) given at the ALISS Conference 16 August 2007. Pitman, L. [IPR and multimedia in institutional repositories: lessons from the MIDESS project](#); subsequently published as a paper in: [ALISS Quarterly](#), vol.3 (1) Oct. 2007, 24-27
 - Poster presentation at ALT-C, Nottingham, 4-6 September 2007.
 - Paper given at Digital Libraries and Learning 2007 Conference, Riga, Latvia, 19 November 2007. Emly, M. & Charles, C. [Multimedia repositories in learning and teaching – lessons from the MIDESS Project](#)
 - Poster presentation at 3rd International conference on Automated Production of Cross Media Content for Multi-channel Distribution, Barcelona (Spain), 28-30 November 2007.
4. CURL – which provided additional funding for the project – will receive copies of the final report and other MIDESS outputs, and the lessons learnt will be shared with other members of the consortium.

Local outcomes

At the beginning of the project, Birmingham, LSE and Leeds all understood the value of establishing a multimedia repository and saw MIDESS as offering an opportunity to develop a pilot service and explore the issues involved. The experience has been invaluable in identifying local requirements, clarifying the issues critical to success and exploring the potential of the chosen software platform. All 3 institutions now feel far more confident in moving forward, although budget constraints limit the options at Birmingham and LSE.

For Birmingham, DSpace proved in practice to have severe limitations and Information Services is currently considering the next steps in terms of software platform, function and support for multimedia materials. A pilot repository for research publications (University of Birmingham Research Archive - UBIRA), previously running on Eprints 2, is being expanded and relaunched as a live service using Eprints v.3, with separate silos for research papers, for other research material and for e-theses. Work has therefore started on trialling a parallel repository for multimedia materials, running on the same server.

The LSE experience with Fedora is somewhat similar. Fedora has been shown to require a substantial investment of resources and a high level of technical expertise to produce a working repository. Equally, significant enhancements have been made to the Eprints software, improving its ability to manage image and multimedia files. Development work on the Fedora platform has therefore ceased for the present and LSE will be exploring the possibility of using Eprints to manage multimedia and images.

Leeds always intended to use the MIDESS work to establish a live repository service, and this is embedded in the Library's strategic and operational plans. During the final year of the project, a bid was submitted for ongoing staffing to take forward the work begun through the MIDESS Project, arguing the case for a comprehensive repository service which would support learning, teaching and research, fully embedded within the University's core processes. Interim funding has been identified, and work is continuing on the DigiTool platform.

Conclusions

When MIDESS began in 2005, the majority of repository implementations were concerned with eprints and other research outputs. A few repositories specifically for learning objects had also been established, often closely linked to a VLE implementation. There was little experience of the issues which arise in handling multimedia objects. More than anything else, the work of the MIDESS Project has provided a series of indicators which can usefully guide any repository implementation which will include a large number of multimedia files. The project identified a number of factors critical for success including:

- Clarity about the organisational and service drivers which underly the need for a repository and ensuring these are embedded in both planning and in ongoing management processes
- The need to understand user requirements, drawing on both those who create the data and those who will use it
- The benefits of whole life cost modelling of competing systems
- The importance of integrating the repository into the broader institutional information architecture in order to avoid it becoming a specialist silo, ignored by all but a few enthusiasts
- The need to consider and plan for digital preservation
- Ensuring that IPR is taken into account and suitable controls are embedded in the repository management processes
- Devoting sufficient resource to defining and controlling metadata, identifying appropriate schemas and – most importantly – application profiles. Not only descriptive metadata but also administrative and technical metadata and content packaging standards must be taken into account
- Establishing whether there is a requirement for harvesting, either internal to the institution or in a broader framework, and the practical implications which follow on from that.

This broad planning framework should lead to appropriate decisions being made in the areas of:

- Hardware and software platforms - understanding the limitations of the chosen software platform is particularly important
- Staff resource required - managing multimedia materials, even within a well-established context, is more demanding than may be imagined
- Service delivery - user requirements can be very specific and also very exacting
- Service integration – maximum value will be derived from the investment if an integrated approach is adopted which recognises the value of digital multimedia for learning, teaching and research on an ongoing basis.

Implications

Although MIDESS succeeded in exploring many of the issues surrounding multimedia repositories, it was very disappointing that none of the partner sites was able to progress the work to the point where testing with live users became possible. The response of the academic partners to the possibilities opened up by using a repository was extremely positive, the scenarios explored within MIDESS won a significant degree of acceptance within the user community and it would have been very beneficial to have moved forward to test working models for the delivery of digital objects within and across institutions. If this work can be undertaken, then the broad educational community will derive benefit.

Questions to be answered include:

- MIDESS identified a divergence between repository systems oriented towards learning objects, those towards research outputs and those with a general remit which includes multimedia objects. What are the implications of these differences in functionality and what measures can be taken to overcome these limitations?
- Following on from this, if an institution requires more than one platform, how can these be integrated in a way which delivers a satisfactory service to the users?
- In managing a repository, is there any way of reconciling the long-term stability and archival preservation requirements of some objects with the flux and re-use that is demanded for others?
- How can the ingest process be simplified in order to permit rapid and easy processing of large quantities of material (whether occasional large ingests of many similar objects or frequent ingests of single and perhaps disparate objects)? This is an area which the IncReASE Project⁷ is attempting to address, at least in part.

Similarly, the technical problems encountered prevented a full exploration of metadata harvesting and the sharing and reuse of digital objects. There therefore remain many unanswered questions from these more technical areas of the project's work. In particular, it would be useful for further development work to explore:

- The practical issues which arise when multimedia objects are harvested by a service provider (the VSM Portal Demonstrator⁸ should be able to address this area)
- Whether application profiles can be used to improve the exchange of digital objects between repositories using METS
- Whether current repository functionality is able to adequately control access to digital objects without impeding their legitimate use
- Whether new protocols and methods, such as proposed by the OAI-ORE Project⁹, can facilitate the exchange and re-use of digital objects.

The project's experience of using Fedora and DSpace does open up a number of questions about whether open-source repository products are suited to the complex requirements of multimedia objects, at least in the typical HE institution. It could be useful for a standard implementation to be developed which would address some of the deficiencies in functionality which MIDESS encountered in the default installation of these products, and thereby also make them a less costly option in terms of staff time and expertise.

Finally, there are some clear questions about how multimedia repositories should be incorporated into the broader national and international information infrastructure. Many services already exist which harvest and provide access to research outputs, and services such as JORUM have been established

to facilitate resource discovery and re-use of learning objects. But where do multimedia repositories fit into this emerging framework? MIDESS found that the value of the multimedia repository lies, not least, in its ability to make research materials available to support subsequent learning and teaching. The establishment of VSM seems to imply a third type of service provider, specifically for multimedia. Is format sufficient reason for segregating access to multimedia objects in a third silo? The implications for the national infrastructure and information architecture need to be explored.

Recommendations

1. If institutions within the HE sector are to manage their digital resources effectively, they need to plan for the implementation of a comprehensive digital repository service which offers support for materials in a wide range of formats. JISC should continue to encourage institutions to engage in this process, taking account of the following factors:
 - a. For most institutions this service should address the needs of both research and of learning and teaching, also recognising the interaction between these two fields of activity. In some cases, this repository service may integrate more than one repository platform.
 - b. Academics should be encouraged to safeguard the digital materials they create by making use of the repository infrastructure.
 - c. In planning the repository service, it is vital to plan for its integration into the broader enterprise and information architectures at the levels of both infrastructure and service provision.
2. Those responsible for implementing repositories need to follow best practice and draw on the experience already gained through projects like MIDESS. Initial planning of all aspects of the implementation is crucial for success. JISC should consider whether any additional measures can be taken to assist in this.
3. None of the repository platforms used within MIDESS provided adequate support for a comprehensive range of metadata schemas. The failure to support standards commonly used for content packaging of learning objects was a particular issue. The HE community needs to exert pressure on developers of commercial and open-source products to broaden their functionality in this area so that more inclusive repository services can be implemented. Pressure will need to be exerted through a number of channels including professional associations, national bodies and individual institutions in their commercial negotiations.
4. MIDESS has highlighted potential issues in implementing OAI-PMH metadata harvesting within the context of multimedia repositories, particularly the need for metadata harvested to include a link back to the object and the failure of some repository platforms to map between schemas in order to expose adequate metadata for harvesting. It would be helpful, within the framework of OAI-PMH, to agree minimum standards in both these areas, and establish mechanisms for certifying software compliance. JISC is well-placed to take such work forward.
5. At present the potential of METS as a content packaging standard for transferring material between repositories is very limited because the schema provides so much flexibility in implementation. The creation of a range of METS application profiles and inclusion of support for these within repository software could significantly improve interoperability. Again, this is an area where JISC is well-placed to provide a lead.

Glossary and acronyms.

CLA – Copyright Licensing Agency

Curator – digital repository software package from Endeavour

DigiTool - digital repository software package from Ex-Libris

Documentum – document management system which can fulfil many of the functions of a digital repository

DSpace - digital repository software package (open-source)

Dublin Core – metadata schema with 15 basic elements, widely used for all kinds of digital material; often abbreviated to DC. See also qDC

EAD - metadata schema devised particularly for archival materials (Encoded Archival Description)

Eprints - digital repository software package (open-source)

ERA Licence – a licence available to UK HE institutions which permits broadcast material to be recorded in analogue or digital format (Educational Recording Agency licence)

Fedora - digital repository software package (open-source)

IMS – a suite of specifications relating to learning technology applications

IMS/LOM – metadata specification for educational materials sponsored by IMS and the IEEE

IPR – “Intellectual Property Rights”

JORUM – a UK repository service for learning and teaching materials for further and higher education

MARCXML – metadata schema expressing the library MARC21 metadata standard in an XML schema

METS – XML schema for content packaging (Metadata Encoding and Transmission Standard)

MODS – metadata schema often used for library materials (Metadata Object Description Schema)

OAI-PMH – protocol widely used to allow aggregator services to harvest metadata for objects held in a number of dispersed repositories or other data sources (Open Archives Initiative Protocol for Metadata Harvesting)

PREMIS – metadata schema for digital preservation

qDC – qualified Dublin Core: metadata schema which expands and qualifies the 15 basic elements of the Dublin Core schema. See also Dublin Core

SCORM – content packaging standard used for educational materials

VLE – Virtual Learning Environment

VRA Core – metadata schema for describing visual cultural works

Z39.87 – ANSI/NISO standard for technical metadata for digital still images

References

-
- ¹ <http://ahds.ac.uk/preservation/curl/DPSoftwareSurvey.pdf>
 - ² EBU core metadata set for radio archives. 2001. http://www.ebu.ch/CMSimages/en/tec_doc_t3293_tcm6-10494.pdf
 - ³ METS: Metadata Encoding and Transmission Standard : primer and reference manual. Version 1.6. 2007. p.10. <http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf>
 - ⁴ <http://www.inf.aber.ac.uk/bridge/>
 - ⁵ Gowers review of intellectual property. 2006. http://www.hm-treasury.gov.uk/media/6/E/pbr06_gowers_report_755.pdf
 - ⁶ <http://www.odl.ox.ac.uk/>
 - ⁷ <http://eprints.whiterose.ac.uk/increase/>
 - ⁸ <http://edina.ac.uk/projects/vsmportal/>
 - ⁹ <http://www.openarchives.org/ore/>