

MIDESS

Project Evaluation

An external end of project evaluation
of the JISC-funded MIDESS project

Chris Awre

August 2007

Executive summary

This external evaluation represents a part of the work that is being undertaken for workpackage 11 of the MIDESS project. It places the MIDESS project in the context of related repository activity, focusing on the two main areas covered by the work: the institutional requirements for a repository and the factors involved in implementing one; and specific issues relating to the management of images derived from digitised materials and other multimedia content. A third factor was the interest in sharing the content through the use of the OAI-PMH protocol between institutional partner sites.

The evaluation considers the original aims and objectives of the project, plus the intended project outcomes, before assessing the outputs from each workpackage in turn and additional issues that arose from the work undertaken. A list of dissemination priorities is proposed on the back of this assessment to help take further the lessons learned and experience from the MIDESS work.

In summary, the project has been of value in addressing the three factors described above and much can be learned from the project outputs by institutions looking to follow the same path. MIDESS suffered a number of technical difficulties that hampered the work, and this has impacted the project from achieving all that it had hoped for. The project has, though, capitalised on the technical problems to report on how these have been addressed and this has resulted in observations that would not otherwise have been possible.

Contents

Executive summary	1
Contents	2
Introduction and scope	3
Background and context to the MIDESS project	4
Review of aims and objectives	8
Review of project outcomes	11
Review of workpackages	14
Additional issues arising from the project	21
Dissemination priorities	23
Conclusion	24
Annex A: Documents reviewed	25

Introduction and scope

The MIDESS (Managing Images in a Distributed Environment with Shared Services) project was funded under the JISC 03/05 Digital Repositories Programme for two years from June 2005. Following a three-month extension, the project concluded at the end of August 2007.

As part of an extensive evaluation plan MIDESS included scope for an external review of the project on two counts. These are:

- 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?

This evaluation report is centred round addressing these two issues. The report is not an evaluation of the way the project was carried out or of the structure of the project, except where these factors have impacted upon the answers to the two questions posed. This external evaluation sits alongside internal evaluation activities and they will together form the output from workpackage 11 of the project.

The evaluation was carried out through desk research based on the documentary outputs from the project, which were supplied by the Project Director. A number of enquiries were also addressed via email by the Project Director to clarify issues arising from the documents themselves. The documents reviewed are listed in Annex A to this report and it is recommended that this evaluation is read in their context. As suggested by the first question above, this evaluation of MIDESS has been carried out in the context of the wider repository and educational landscape.

The approach taken to the evaluation has been threefold:

- 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

Additional issues arising from the project are raised and a recommendation of issues and lessons learned that would benefit from wider dissemination is also made: these may be taken up by the project in outstanding dissemination activities or by the JISC as part of its general evaluation of the Digital Repositories Programme.

Finally, conclusions are given on the overall impact of the MIDESS project.

Background and context to the MIDESS project

The MIDESS project was one of a wide range of projects funded under the JISC 03/05 Digital Repositories Programme. It was, though, the only one focused specifically around the management of images and related multimedia resources¹. Furthermore it was one of only a few that addressed specifically *institutional* repository infrastructure and the relationships between repositories at different institutions. These two characteristics epitomise the benefits the MIDESS project brought to the Programme. They also, though, drive home that MIDESS was a project, in many ways, working on its own, and which did not have the directly related support that the clusters of research data and e-learning related projects provided. The following projects within the Programme did cover areas overlapping with MIDESS and provide context for the period in which the project took place.

CLiC (Community Led image Collections)²

This study was funded alongside the Digital Repositories Programme, with a remit to identify models for the greater disclosure and awareness of the many collections of digital images – many of them freely available – in existence in the community, to facilitate their use within education and research. The study considered local institutional collections in the context of their relationships with subject and national collections and proposed a model that linked these through a visual directory.

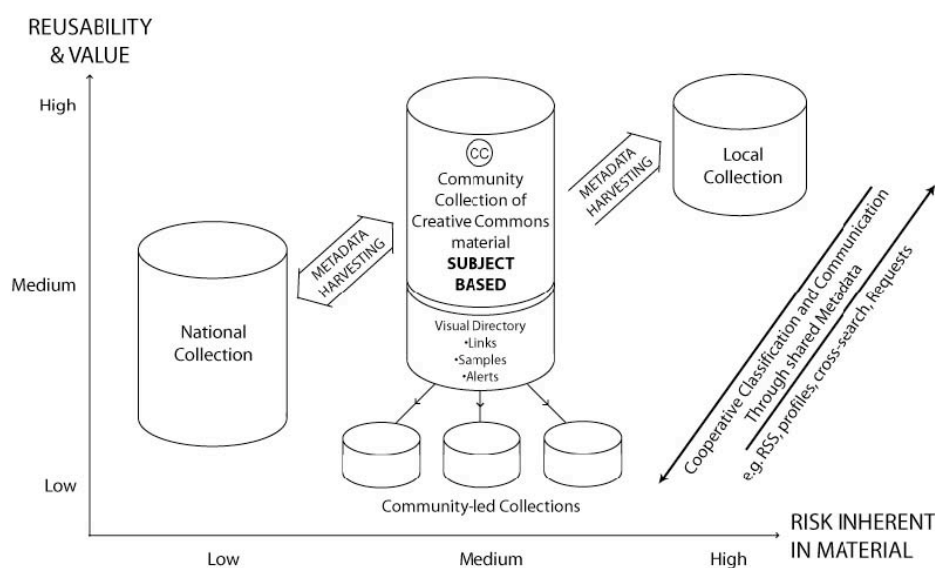


Figure 1. CLiC three tier model

The CLiC study highlighted that image collections that sit in isolation have limited value, and that there is value in bringing information about them together to facilitate use.

IRIScotland (Institutional Repository Infrastructure for Scotland)³

The IRIScotland project sought to develop an infrastructure that would allow all Scottish Higher Education institutions to establish a repository presence, either one

¹ A number of projects were also funded to examine the management of e-learning materials, which will undoubtedly include images. However, their emphasis and outcomes were generic to e-learning and did not focus on multimedia specifically.

² CLiC - <http://clic.oucs.ox.ac.uk/>

³ IRIScotland - <http://www.iriscotland.lib.ed.ac.uk/>

hosted locally at the institution or one provided for them by the National Library of Scotland. The emphasis on content collection has been research papers. Key to the success of the infrastructure was the development of a search service⁴ that would allow access across all Scottish repositories, thereby raising awareness of Scottish research and facilitating access to it. This search service has been developed using the OAI-PMH protocol and relies on harvesting metadata from across all repositories.

Project experience has indicated that even considering the specific development of an OAI harvester system to enable harvesting from the disparate repositories, the issue of consistent metadata has led to problems in what can be harvested and what can be effectively surfaced through the search service.

Repository Bridge⁵

The Repository Bridge project was a short project carried out jointly by the National Library of Wales (NLW) and the University of Wales sites at Aberystwyth and Swansea to build a means for transferring e-theses metadata to the National Library. This also involved the use of OAI-PMH for harvesting, and relied upon harvester software developed by staff at UW Aberystwyth and the NLW. Furthermore, due to preferences expressed by the Universities concerned, the metadata harvested was wrapped in a METS record, and provided an example of how alternatives to Dublin Core metadata can be harvested using OAI-PMH. The project had a successful outcome and the outputs from the project are currently in use at the participating sites. The use of OAI-PMH allowed the movement of metadata from DSpace into a Fedora system, thereby also demonstrating cross-repository interoperability. It is noted that the Repository Bridge software was used within MIDESS, although with some difficulty.

Rights and Rewards in Blended Institutional Repositories⁶

This project, based at Loughborough University, sought to identify both the rights issues revolving around sharing content and the rewards that come from this activity, particular for the support of teaching and research. A survey of academics carried out highlighted concerns that are held about using repositories, assuming that it is understood what a repository is. The rewards are not overcoming the rights issues yet, whilst time and organisational issues also affect whether academics chose to interact with their institutional repository. It is stressed that the repository needs to become part of existing practice to aid adoption and allow the rewards to become clearer.

Linking UK Repositories⁷

The Linking UK Repositories study was another project funded alongside the main Digital Repositories Programme. It looked at technical and organisational models for the development of end-user services across open access repositories. The main technical model proposed was an aggregation model, which encourages the creation of different types of aggregation upon which a wide range of services can be built. These could be formal aggregations generated through an OAI-PMH harvester, or more informal aggregations such as those generated through an RSS newsfeed aggregator. In both cases, dispersed sources of information can be brought together to enable services across them.

⁴ IRIScotland Project Pilot Cross-repository service - <http://cdlr.strath.ac.uk/iriscotland/>

⁵ Repository Bridge – <http://www.info.aber.ac.uk/bridge/>

⁶ Rights and Rewards in Blended Repositories - <http://rightsandrewards.lboro.ac.uk/>

⁷ Linking UK Repositories reports - <http://www.keyperspectives.co.uk/openaccessarchive/reports.html>

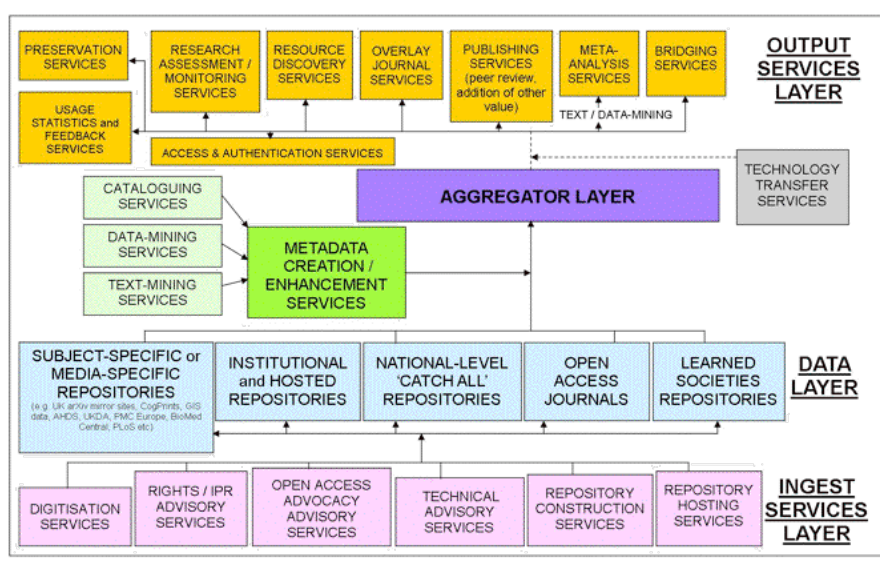


Figure 2. Linking UK Repositories Services Model

CHERRI (Common Healthcare Educational Recordings Reusability Infrastructure)⁸ CHERRI was a study funded alongside the main Digital Repositories Programme and acted as a companion piece to the CLiC study insofar as it dealt with how medical images, as well as other clinical recordings, can be deposited, shared, and re-used, an issue previously addressed by the BioMed Image Archive project⁹. The work found that whilst the law is generally on the side of patients, its application can vary and a common activity framework is required to ensure consistency of practice. The authors propose the use of GUIDs and a common content + licensing model for consideration by the community. There was active liaison between CHERRI and MIDESS during the former's lifetime.

The projects described here all have synergies with the work of MIDESS. There has been much activity generally in the use of images and other multimedia content within teaching & learning over the same period, with the introduction of the Education Image Gallery¹⁰ and Film & Sound Online¹¹ services at EDINA. There is the still images advisory service provided by TASI¹², whilst BUFVC¹³ continues to do the same for film and video. A JISC project to develop a video and sound materials portal is also currently underway¹⁴, and will deal with still images as well. There has also been huge investment by the JISC in digitisation of high quality multimedia resources through its Digitisation Programme¹⁵ that will be freely available for educational purposes. Looking beyond academic activity, the museums sector has long had to face the issue of how to deal with images of physical collections, and indeed see such resources as integral to expanding access to these collections: the Harvesting the Fitzwilliam¹⁶ and Accessing the Virtual Museum¹⁷ projects from the

⁸ CHERRI - <http://www.cherri.mvm.ed.ac.uk/>

⁹ BioMed Image Archive project -

http://www.jisc.ac.uk/whatwedo/programmes/programme_fair/fair_synthesisintro/fairsynthesis_biomed.aspx

¹⁰ Education Image Gallery - <http://www.edina.ac.uk/eig/>

¹¹ Film & Sound Online - <http://www.filmandsound.ac.uk/>

¹² Technical Advisory Service for Images - <http://www.tasi.ac.uk/index.html>

¹³ British Universities Film and Video Council – <http://www.bufvc.ac.uk>

¹⁴ VSM project - <http://www.edina.ac.uk/projects/vsmportal/>

¹⁵ JISC Digitisation Programme -

http://www.jisc.ac.uk/whatwedo/programmes/programme_digitisation.aspx

¹⁶ Harvesting the Fitzwilliam project - <http://www.fitzmuseum.cam.ac.uk/projects/htf/>

JISC FAIR Programme¹⁸ were two examples of early work on how OAI-PMH could be used for such purposes. The work of these projects in tandem with other JISC projects led to the development of position papers that act as helpful guides in this area¹⁹.

It is an ongoing challenge for institutions to work out how to juggle the provision of these national services alongside locally held collections. Whilst the CLiC model proposed a way of addressing this through the use of a third party directory, it requires all parties involved to reach a common understanding of how to implement the model, and for institutions to have in place the skills and technology that can enable this join-up. It is also noticeable that the majority of activities have been at the national level: indeed CLiC recommends that local collections should be taken out of institutions where their wider worth is recognised in order to enable their delivery at a national level independent of institutional issues. This highlights a lack of grant activity at institutional level to aid understanding of how to manage and support use of collections so that they *can* be made available effectively, which JISC or others could do well to address.

It is in this context that the MIDESS project has taken place. It is acknowledged that staff and technical problems have prevented the project from achieving all that it had hoped (an issue to be addressed more fully in the following sections). 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.

¹⁷ Accessing the Virtual Museum project -

http://www.jisc.ac.uk/whatwedo/programmes/programme_fair/fair_synthesisintro/fairsynthesis_avm.aspx

¹⁸ JISC FAIR Programme, 2002-5 - http://www.jisc.ac.uk/whatwedo/programmes/programme_fair.aspx

¹⁹ Images and museums cluster group position papers -

http://www.fitzmuseum.cam.ac.uk/projects/hf/mi_cluster/

Review of aims and objectives

Seven aims and objectives were specified in the project plan, covering both aims and objectives together. These are assessed for how they have been addressed and met in the course of the project's lifetime. Where individual aims and objectives cover areas of work in specific workpackages, further discussion is given in under that workpackage in a later section.

1. *MIDESS will 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.*

As workpackage 8 makes clear, this objective was achieved, the three sites being Leeds, Birmingham and the LSE, though not without difficulty. The implementation of the commercial Endeavour Curator repository software by Leeds had to be stopped when Endeavour was bought out by a private equity company, Francisco Partners, which also controlled Ex Libris. The Ex Libris repository system, DigiTool was the preferred platform in this merged company and Curator development was ceased in its favour. This led to a switch to DigiTool at Leeds causing technical and time delays in establishing a working repository. Notwithstanding this, the work carried out by Leeds has clearly produced some valuable results for the project, and the added bonus of the opportunity to compare and contrast two commercial repository products.

The adoption of Fedora at LSE was made on the understanding that Fedora itself is a complex system that requires work to develop a useful repository. As with many other Fedora sites, this realisation did not necessarily match the actual level of work required. Nevertheless, a working collection was established, and much experience gained in importing the objects for this from a separate database using METS. Although the developer has since left the LSE, the documentation that came out of the work will stand the LSE in good stead if they decide to pursue their interest in Fedora long-term.

The decision to adopt DSpace at Birmingham was made with clear objectives, and it is clear from resulting documentation that staff there were disappointed and frustrated by the system's inability to meet all requirements. DSpace did provide the most stable base for exposing metadata for harvesting using OAI-PMH, although the metadata harvested did not prove to be as valuable as it might have.

The project thus generated three workable repository demonstrators that could be used to carry out a number of tests. The experience gained from this work will have been valuable and it is to be hoped that the three sites do not lose heart in their aim to generate stable, long-term repository image collections given the frustrations met to date.

2. *These repositories will be populated 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.*

The range of collections held within the three partner institutions is impressive, and provided no shortage of materials with which to work. Indeed, it is the range and size of such collections that no doubt inspired the need for the MIDESS project in the first

place. It is also clear that thought went into the complementary nature of the collections in order to help make sense of subsequent harvesting activities and searches across these. It is notable that all the collections considered were of digitised resources, and not born digital resources, which merit a different level of attention.

3. *Opportunities for the sharing and re-use of collections across institutions will be explored through the active involvement of academic staff at each partner institution.*

The setting up of the local Working Groups at each site afforded the opportunity to involve academics in the project and gather their views on sharing and re-use. The level of engagement was high at Leeds and less at the LSE and Birmingham. The Groups seem to have resulted in little active discussion on sharing and re-use that has been reported, efforts being focused on getting repositories to a stage where they could be considered usable for the project (see above). The interest and involvement of the ALPS CETL group at Leeds and their interest in linking the repository with a VLE highlights the potential value of repositories in supporting teaching & learning, even if technical issues prevented this from being demonstrated within the project's lifetime.

This objective has thus not been met in the way it was anticipated, although there is nothing to suggest that the work of the project will prevent such explorations from taking place after the project's end. Any such work could usefully compare and contrast the relative value and success of delivering multimedia items through a repository when contrasted with the individual, probably less structured, means of access each collection currently has.

4. *Metadata standards will be established, and metadata developed, 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.*

The work undertaken within workpackage 4 on metadata is broad in its coverage of metadata standards and their application. The use of the phrase 'metadata standard' is potentially misleading in reporting the findings, though, as it is application profiles of metadata standards that need development, a topic acknowledged in the workpackage report.

Metadata for each of the collections ingested by the three sites was created, though it is not clear whether any specific metadata application profiles were generated. The need for such profiles led to a further bid to the JISC that was not, unfortunately, funded²⁰. The lack of profiles had its own impact on the project, in particular affecting the harvesting of collections from partners sites into a single resource for searching.

The identified value and examination of the benefits of the METS packaging standard are welcome, and the reporting of practical experience with this of value, in particular its comparison with IMS Content Packaging, its equivalent in the IMS stable of metadata standards. The need for an application profile for METS is particularly key for its effective use and the evidence from the project reinforces this need.

²⁰ Though note that JISC did anticipate activity to develop a range of application profiles in its repositories and preservation briefing paper for Circular 04/06 in September 2006 - http://www.jisc.ac.uk/media/documents/programmes/capital/repositories_and_preservation_briefing.pdf

5. *MIDESS will explore the role of digital content repositories within the institutional information architecture, with a particular focus on interoperability with enterprise content management architectures.*

One of the dangers of adopting a system like a digital repository is that it risks becoming yet another silo of information that users have to engage with. It is thus important that the repository can be integrated with existing routes for accessing content, and that this access should be in the context of how the content will be used. This objective has been addressed specifically through the work of workpackages 6 and 9, which provide a useful overview of how a repository can sit amongst other library and institutional systems.

6. *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.*

The premise within the MIDESS project, and addressed within this objective, was that sharing content between institutions allowed virtual collections to meet specific purposes and allow re-use as required. The project anticipates a flexible harvesting environment, with harvesting taking place effectively on demand as requirements arise, and with more than one harvesting requirement occurring at any one time.

The experience of the project in setting up a harvester unfortunately shows that we are still some way from achieving this original goal of the Open Archives Initiative. There have to be very clear agreements between the data providers and service providers on what is being harvested to ensure that the harvested metadata can be used in the way intended. The management overhead of enabling this flexibility and of managing a harvester is high and greater experience is required in building established OAI service providers before there is enough knowledge and capability to allow this goal to be met.

As such, MIDESS was able to meet this objective in principle if not entirely in practice. Capturing the harvesting requirements identified and disseminating these will be of value in increasing awareness of how such sharing can take place.

7. *MIDESS will seek to pilot an infrastructure which could serve as a model for future distributed national digitisation activities.*

This end goal of the project was ambitious, but certainly one to aim for. However, it is unclear to what extent the development of repositories and harvesting between them would have an impact on digitisation activity itself. It is recognised that in principle the ability to share metadata from collections, and possibly the content itself, would offer digitisation efforts an additional and potentially more effective means for allowing access to the content, and this is to be welcomed. But this connection is unproven and, unfortunately, MIDESS was unable to demonstrate a connection either way. It is to be hoped that repositories can stimulate digitisation in this way, and additional work following on from the experience of MIDESS is recommended to identify whether such a link exists.

Review of project outcomes

As required by the JISC project plan template, MIDESS had listed a number of project outcomes in addition to its aims and objectives within the project plan. The success in bringing about these outcomes is considered in this section. A number of the outcomes mirror the aims and objectives closely and the relevant discussion in the previous section referred to where appropriate.

1. *An evaluation of user requirements for the management, sharing and re-use of digitised content to support learning and research.*

The output from workpackage 3 has addressed this outcome. It is notable that the study of user requirements focused on users who are primarily creators and managers of the digitised content, rather than end-users. This is a useful distinction, and enabled the project to cover an area that repositories need to understand well if they are to be seen as valuable to content owners. It is a shame that the case studies do not delve deeper into the specifics of supporting learning, teaching, and research currently, other than hypothetically in the case studies from Leeds. The comparison between the current state of play and the difference a repository can make would be useful in demonstrating the added value a repository can offer.

2. *An evaluation report on the key functionalities and requirements for the management of digital content, including a comparison of open source tools such as Greenstone, FEDORA and DSpace against proprietary products such as Luna Insight.*

The output from workpackage 2 has addressed this outcome, plus some of the conclusions that derived from the user requirements investigation and analysis. Although undoubtedly originally intended, it is unfortunate that systems such as Luna Insight, with its specialism in supporting the management of images, were not included in the review of systems. A continuing issue that repositories have is their generic nature, their ability to be used for different purposes. Where specific systems exist to address one of these purposes, repositories need to show how they can support that purpose without detriment to the content at hand, and be able to support additional functionality through their ability to hold a broad spectrum of content.

3. *An analysis of the role for digital content repositories within the wider institutional information architecture – in particular, exploring the relationship with central storage services, and enterprise content management tools such as Documentum.*

The output from workpackage 6 has addressed this outcome, and focused on the situation at Leeds only. Documentum was also included in the original analysis of potential systems for managing images, though was rejected at an early stage in this process. The successful link-up between the repository and the central storage service at Leeds is noted, and represents an essential piece of institutional join-up that allows repositories to become embedded. There remains uncertainty about the exact relationship between digital repositories and content management systems like Documentum and greater analysis of this in the context of Leeds would have been welcome. Although there may not be any identified need for the two systems to be linked technically, they do require clear demarcation lines in terms of collection management policy.

4. *The development of a number of digital repository demonstrators, likely to be across a variety of open source platforms.*

The output described in workpackage 8 has addressed this outcome. Two of the three demonstrators were built on open source platforms, whilst Leeds decided to adopt a commercial system, for reasons highlighted within workpackage 2.

5. *A tested model for the management of consortial digitisation activities.*

As indicated in discussion under the last objective in the previous section, MIDESS was not able to reach a stage where it was able to test the intended model, or demonstrate the link between repositories and digitisation activities. This was in the main due to technical difficulties.

6. *The development of wider expertise in the use of the METS standard for metadata management, and recommendations for best practice in use, especially in relation to support for e-learning activities.*

The project clearly gained much experience and expertise in the use of METS, though it is also apparent from the documentation that there is still much that can be learned. The use of METS is clearly not straightforward, and careful planning and scoping of the purpose of use is required to ensure its successful adoption. The project achieved this outcome through recommendations on the use of METS, though was not able to fully address the issue of its potential use in e-learning.

7. *An evaluation of the relationship between METS and the IMS metadata standard, with recommendations for how an interaction between the two could best be utilised to support the re-purposing and re-use of digitised content in an e-learning context.*

Part of the work undertaken as part of workpackage 4 has addressed this outcome. The investigation of the two packaging standards was detailed, though the hoped for outcome of an easy means for moving metadata and content between the two was not easy to achieve. In addition to the sources of information listed in workpackage 4's report, the RAMLET initiative²¹ has also sought to enable interoperability between these two standards, as well as MPEG-21 DIDL and ATOM, though with limited success to date. As concluded, and given the metadata standards that tend to be used within these packaging standards it may need to be accepted currently that there will be loss of information when translating between the two.

8. *An evaluation of options and requirements for sharing digital collections within and between institutions.*

The outputs from workpackages 9 and 10 have addressed this outcome. The two reports highlight a number of ways in which collections might be shared, and describe activities undertaken to test these (predominantly through the use of OAI-PMH). In considering 'sharing' of collections, it is useful to be clear about whether it is the metadata that is being shared, or whether the digital objects themselves are being shared. The former offers ways of alerting users situated elsewhere to become aware of available collections they might not otherwise have known about. The latter has a number of additional implications about the IPR, hosting, provenance and management of the copied objects and is not straightforward. Some

²¹ RAMLET - <http://www.ieeeltsc.org/working-groups/wg11CMI/ramlet/Pub/>

of these implications are being examined by the current OAI-ORE initiative²² and it remains to be seen what conclusions are reached.

9. *An assessment of IPR and other issues in relation to the re-use and sharing of digital content, especially in relation to audio-visual and multimedia materials.*

The output from workpackage 7 has addressed this outcome. The report output is a very useful description of the current state of play regarding IPR and copyright, and how these apply in relation to audio-visual and multimedia materials. The sections on good practice, recognising that these areas of law can be flexible if all parties want them to be and what to do when they are not, plus the questions to ask when considering sharing collections are valuable contributions to a complex field.

10. *An evaluation of the potential for exposure and searching of distributed collections, utilising OAI-PMH, Z39.50 and Web Services.*

As with outcome 8, the outputs from workpackages 9 and 10 have addressed this outcome: the output from workpackage 6 also examined the use of Z39.50 and SRU/W. The project has clearly given close attention to the relative merits of OAI-PMH and Z39.50, though without a very positive outcome in either case. The underlying aggregation mechanism of OAI-PMH seems to be favoured, and is in keeping with the recommendations of the Linking UK Repositories report. A consideration of Web Services is limited to SRU/W, though the lack of uptake of these standards in the repository systems used prevented a proper assessment of their capability.

11. *An evaluation of the requirements for a centralised metadata harvesting service to support exposure and searching of distributed collections.*

The outputs from workpackages 9 and 10 have addressed this outcome. This report considers the complex environment in which a repository will be situated and envisages the repository at the centre of this environment, serving a range of different needs. A dilemma the project faced was the lack of an obvious candidate or location for its central harvesting service. It is debatable to what extent repositories can or should act as harvesters themselves, or whether a separate piece of infrastructure, closely linked to the repository, should be preferred. Serious thought needs to be given as to how institutions should structure this architecture so as to enable the repository to do what it does best.

12. *An evaluation of the digital preservation requirements for distributed digital image collections*

The output from workpackage 5 has addressed this outcome. This report clearly demonstrates that there has been much work carried out in the community to consider the preservation requirements for images and related multimedia formats, though it is a shame that the project itself could not put much of this into practice (generally due to systems lagging behind theoretical concepts). Preservation standards and practice are a developing area, though, and this needs to be recognised alongside taking on board current best practice. The need to build in preservation planning when establishing a repository, and to educate content creators to how they can assist with this is key to ensuring this aspect of repository functionality will succeed. It is tricky to put this into practice, but worth pursuing.

²² OAI-ORE - <http://www.openarchives.org/ore/>

Review of workpackages

Previous sections of this evaluation have focussed around the original intentions of the MIDESS project, as laid out in the project plan, and have examined the extent to which the project has achieved these, with reference to workpackage activity and wider context where relevant. This section considers each of the workpackages in turn and looks closer at the activity that did take place and which has been reported in the outputs from the project. The workpackages are considered in numerical order with the exception of the first two reported, as the user requirements analysis informed the development of the functional and technical requirements specification. Workpackage 1, project management, is not considered here, as it is not in the remit of this evaluation.

Workpackage 3: User requirements analysis

The key outcomes listed in the executive summary of this workpackage's report provide a very useful overview of the needs that arise when considering the digitisation and management of multimedia materials. In particular, the list in key outcome 1 and the support issue in key outcome 3 are important points that all institutions implementing an institutional repository need to take on board. Along with the extensive list of requirements drawn up in the recommendations section, the user requirements analysis is a valuable reference tool.

The report is also a fascinating insight into the number and variety of image and multimedia resources that are held within institutions, often independently and probably in ignorance of each other. These collections appear to be in various stages of digital availability, though the potential of a repository to assist in their management is clear. There has been a mixture of routes followed to enable access to the collections, from Microsoft Access to a media database to the proposed virtual museum. Given this clear example of a museum it would have been valuable if the project as a whole had drawn more evidence from the museums and wider cultural sector; this sector has had long experience of managing images and other forms of multimedia. In venturing down the path of such management within universities, the academic sector could learn much from this existing experience.

Where applied, the wide variety of metadata schemas used is also noted, and it is welcome in many ways to see the effort given in using these²³. As is described in a later workpackage, repository systems are not always flexible enough to cope with this breadth, and the demands of OAI-PMH require a level of dumb down mapping that is not always helpful. This is a challenge for repositories, to ensure that the richness built up in catalogued collections is not lost when transferred into the repository: a loss that could compromise the access and preservation benefits that use of a repository can bring.

The special attention required for medical images and videos both here and in other workpackages, and the liaison with the CHERRI project, is noted and welcomed. There is no simple answer to effectively dealing with medical images, but a clear statement and understanding of the issues at hand is a necessary step towards developing a solution and avoiding potentially costly errors.

²³ SNOMED and MeSH, though, as mentioned in Leeds case study 2, should be considered classification schemes rather than metadata schemas, though the noted difficulty of expressing the appropriate level of richness within Dublin Core is appreciated.

It is noted that the questionnaire was largely focused on those involved in creating content, whether it be from an academic or support perspective, rather than end-users of the materials. Given the nature of the project in examining the implementation of repository infrastructure to manage and share image content this is understandable and, as noted elsewhere, useful, though a fuller picture will need to be drawn at some future point that encompasses end-users and how they interact with the repository: this is likely to have its own impact on how the image content is stored that it would be useful to take on board at an early stage.

Notwithstanding this, the findings and conclusions from the questionnaire are a valuable insight in institutional motivations and issues surrounding the development of image and multimedia collections. Key points from within these conclusions are:

- That the purpose of creating the images was to enhance local learning & teaching, but that there was little consideration of sharing the materials beyond this local use, mainly for copyright or technical reasons or due to lack of time. There was perceived value in sharing if this could be facilitated, though.
- Much image content has been generated because academics couldn't find alternatives elsewhere and were not aware of external copyright-cleared collections. Technical difficulties in generating these materials were common, and external support would be welcomed.
- There are institutional issues around where digitised images are stored, who controls this, and who should pay for it, plus who owns the IPR in the images wherever they are stored. There is a clear need for appropriate policies and licences to help foster wide agreement on these matters, and enable a more considered strategic approach. The effort involved in this and in supporting academics should not be underestimated, and the benefits made clear when any case for funding is being made.

The list of requirements given in conclusion is a valuable one, though a clear distinction between functional and policy requirements would be a useful addition.

Workpackage 2: Functional and technical requirements specification

Many of the requirements needed for the specification of appropriate repository software were concluded from work undertaken in workpackage 3, on user requirements. This workpackage benefited from that analysis, but has focused on a comparison between repository software systems the choices made by the three partner sites in the project rather than further defined a specification itself. The comparison itself is of value, though a more detailed assessment of what is required, building on the user requirements analysis would have been welcome.

Comparison tables are always a useful means of highlighting similarities and differences between systems – there is no exception here – and it is particularly useful to see commercial systems compared against open source ones. There is a need in both comparison tables to specify the version of the software being assessed and when the comparison took place. It can be unclear, otherwise, exactly what is being compared. The detailed assessment in the second table also reflects a difficulty in comparing commercial and open source systems: commercial providers have sales teams that will always be 'positive' in their assessment ("it is under consideration for the next version"), whilst open source respondents are likely to be realistic, especially if they are users. It is unfortunate that the developers of the open source systems could not have been invited to give their input. Early dissemination

of the outputs of such comparisons is also necessary. A technology analysis of repository software by Johns Hopkins University in 2005-6 highlighted the difficulties such work has, with many of the findings out of date shortly after the results were released due to software updates.

Web Services are mentioned in the table, though this term needs to be scoped to avoid being the catch-all it is sometimes used as²⁴. Although not mentioned in the list of core technical requirements, which do not mention Web Services, SRW/U is listed alongside Z39.50 in the first table. Workpackage 6's discussion of Web Services centres round SRW/U, which certainly can be considered Web Services in their use of SOAP and REST for communication, respectively. Consistency in these descriptions across workpackages would avoid confusion.

The summary analysis contains some useful pieces of information about each package, though direct comparison is difficult, as the information does not report against, for example, the core requirements listed earlier. The decisions reached by each site on the software to adopt were clearly well informed as a result of these comparisons, though additional local factors also played their part.

There is high value in assessing the whole-life costs of a system²⁵, and the charts displayed provide a useful insight, highlighting amongst other attributes that the up front costs of a commercial system do not always make it more expensive over five years when put against the costs of employing a developer for an open source system. In contrast, the expected lifetime of the commercial software, and the extent to which additional staff are required to help manage the commercial system over this lifetime are questions for the other side of the coin. It is to be hoped that this example of whole-life costs will encourage others to do likewise.

Workpackage 4: Metadata requirements specification

As indicated previously, the work carried out in understanding the metadata requirements for the project was clearly broad and detailed, particularly in the assessment of METS and the attempts to get this to work to meet the needs of the project. This workpackage report starts from the most useful point, a consideration and statement of why to use metadata. This is not often stated clearly enough and usefully sets the scene for the work reported.

The review of metadata schemas is useful if not necessarily novel. As suggested in a previous section, there is a lack of comparability between schemas such as Dublin Core and MODS, and classification schemes like SNOMED and MeSH, though they are treated as equals in the report: clarification of the difference would have been useful. The reason for covering SNOMED and MeSH is clear, linking back to one of the likely sources of image content for the repository, the University of Leeds Medical School. Taking into account the metadata used in some of the other suggested collections, both at Leeds and at partner sites, coverage of VRA Core, ISAD(G) and UKLOM would also have been beneficial, though this is addressed to a degree in Part B of the report. Nonetheless, it is recognised that the primary aim of the

²⁴ It is acknowledged that there is no absolute definition of Web Services, which can encompass many different mechanisms of communicating over the Web: it is valuable, though, to scope how the term is used to aid wider understanding.

²⁵ Though care needs to be taken in the comparisons made: the green line representing a price for a repository of over 100,000 objects seems incongruous on a chart where one of the axes is number of objects.

exercise was to consider the metadata to be used within the repository, and which these other schemas would have to map to.

The review of available packaging formats is useful, if perhaps not as detailed as might have been useful, particularly on AICC, which is not commonly mentioned in this field. The comparison between METS and IMS and work to put this into practice is, as previously mentioned, valuable, and highlights the silo boundaries that repositories seek to cross. A greater understanding of how to enable interoperability between these two standards will be essential if library-based repositories are to effectively serve VLE and related e-learning systems.

The conclusions reached are a useful set of guidelines for others deliberating over their own requirements. The reports also highlight two key aspects of implementing metadata: the complexity of mapping between different metadata schemas, and the complexity of implementing packaging formats containing these schemas.

Workpackage 5: Digital preservation requirements specification

The clear recognised need for preservation planning, one of the key components within the OAIS reference model, by all parties involved or connected to the development of the repository is the main strength of this workpackage's outcome, though it does not appear to have been a message taken on board by the developers of the Curator software, at least in the version used within the project. It would have been interesting to see how the promised Kronos product would have made such planning, and implementation of it, easier, but the future of this product seems unpromising since the sale of Endeavour.

The benefits of a national agency such as the AHDS examining this area and making recommendations to save institutions time and effort is also clear from the report. Such advisory work will continue to be needed in a field that is not yet fully understood. In the context of MIDESS it is interesting to note that whilst the preferred preservation format for images is reasonably clear – uncompressed TIFF – there are no equivalent, or possibly too many, formats for video or sound.

The advice from the AHDS builds on the experience of the metadata workpackage in examining, and recommending, an application profile built from a number of different schemas. Clarity in stating the application profile used is then a fast route to enabling understanding and interoperability.

Workpackage 6: Integration with enterprise architecture specification

The report from this workpackage represents an extended University of Leeds case study of a consideration of the issues and needs in integrating the repository into the rest of the institution through links to existing enterprise systems. It starts with a list of very useful drivers for why a repository is important and how it can support research, learning & teaching, and administration. This valuable list will be most useful for others in institutions where information is recognised as a strategic issue; it will not always be a case of preaching to the converted and linking the importance of a repository to the value of information in general will assist in gaining acceptance for this technical approach.

The discussion of links into enterprise infrastructure through storage and authentication are key to recognising that a library cannot run a repository in isolation, and that working with colleagues in computing services is a necessity. The

point regarding the advantages of central storage is also well made in an extensive consideration of storage issues. Whilst looking to the future and the use of Shibboleth, greater information on the current method for authentication would have been useful for sites unclear what the use of LDAP implies (or its advantages). Nevertheless, the promotion of Shibboleth is welcome and will undoubtedly need to be part of any future development.

The issue of streamed media is essential to any planning of a repository that will include video. Less certain is the value and current benefit of including the repository in federated search access, though this is clearly one, and probably the best current, way of enabling access to the repository alongside commercial and other sources of information: this is an area that warrants further investigation.

The links into the portal and the VLE are examples of how to manage the linking between different commercial systems, and the methods used, whilst being of interest to sites with a similar infrastructure, have a proprietary nature. The focus on integration is on access by end-users rather than deposit. The latter is presumed to be focused around the repository's own interface, though there is reason to believe that integrating deposit elsewhere may assist take-up of the repository as well²⁶. Information on the ability to integrate deposit by commercial repository systems would be interesting to gather.

Workpackage 7: IPR and multimedia requirements specification

This area is reported under the previous section discussion of project outcomes. The report is an excellent overview of IPR and copyright as it refers to images and multimedia, and will be a useful reference document. As indicated at the start, the information does not constitute legal advice. Taking this on board, application of the information the report contains will require consideration of local institutional guidelines and advice very much as an addendum to what is described here.

Workpackage 8: Repository implementation and population

This workpackage represents a summary of the practical activity undertaken to implement the outcomes of the work in other workpackages, and put it into practice. It is a valuable set of case studies of repository implementation that could benefit from wider dissemination to assist others in their understanding of the requirements involved. Key skills in metadata and XML are noted that a number of institutions may not have, or at least may not have easy access to.

The Leeds case study also offers an interesting contrast between two commercial systems that is an unexpected outcome from the work. Although this will be of limited usefulness now that Endeavour's new owners have deprecated Curator development, it could be of relevance to others selecting commercial packages as a guide to the possible limitations and benefits to look out for.

Workpackage 9: Resource discovery and shared services

There is a close link between the work of this workpackage and that of workpackage 6, appropriately acknowledged, particularly in the area of Z39.50 searching. Given the interest in SRW/U driven by the availability of an interface onto Curator using this

²⁶ RepoMMan project user needs analysis - <http://www.hull.ac.uk/esig/repomman/downloads/R-D14-full-user-needs1.pdf>

protocol, it is unfortunate that this interest could not be pursued with the change to DigiTool at Leeds. Nevertheless, the report makes some useful points on the basis of testing Z39.50 access, particularly related to the lack of full LMS functionality when using the library system's Z39.50 client, and the user expectations when searching across a very large catalogue and a relatively small repository. This latter potential mis-match of resources is an issue faced in general by meta-search tools: it will be interesting to see how the JISC study on meta-search addresses this and the conclusions it reaches²⁷.

The section on moving to a more complex environment is well titled, and the issues at hand are usefully described. They move from the relatively simple issue of resource discovery to the ability to capture an object itself, rather than just the metadata, and re-use it in a different context. There are parallels here with the role that the JORUM²⁸ is playing at a national level. There are good reasons why an individual institution would wish to capture content in this way to enable its re-use, though where this involves repositories outside the home institution there will always be IPR issues that will require careful attention: the JORUM itself has had great difficulties reaching a position where it can be used effectively without compromising the wishes of institutions. Where such agreements are on a smaller scale, such as between the like-minded partner sites in MIDESS, perhaps, there is potentially greater scope for their success. Case studies of such agreements would be valuable in demonstrating how sharing resources can take place, though would themselves need to be carefully set up.

The exploration of requirements for linking the repository into ALPS is a useful exemplar of how a learning & teaching unit might use a repository, and it is to be hoped this interest is pursued and reported further on implementation.

The issue of multiple repositories across institutions, and within a single institution, is also well made, as this can complicate resource discovery. Creating a single search across all repositories will alleviate this, and the potential of harvesting using OAI-PMH over federated search is one that the project has subsequently explored in the following workpackage. The University of Glasgow in their JISC-funded DAEDALUS project²⁹ also proposed having a single point of access to their two repositories through a search interface onto a harvested aggregation of both. When harvesting across repositories at different institutions, and using METS to transport the content as well as metadata, a copy of the original resource will be made. Subsequent harvests may also make additional copies. The presence of multiple copies may not be of concern to the parties involved, but they do raise the following issues for end-users:

- Authenticity – is this the real thing?
- Integrity – is this copy the same as the original?
- Identification – how do I know which copy is which?
- Provenance – where did this object originate?

These issues are being investigated as part of the OAI-ORE initiative, as described in the workpackage report. Identification issues are also being addressed within the JISC-funded RIDIR project³⁰, and will consider the CHERRI proposals for a GUID as

²⁷ JISC ITT: Towards optimal information retrieval -

http://www.jisc.ac.uk/fundingopportunities/funding_calls/2007/04/info_retrieval.aspx

²⁸ JORUM - <http://www.jorum.ac.uk/>

²⁹ DAEDALUS project - <http://www.lib.gla.ac.uk/daedalus/>

³⁰ RIDIR project – <http://www.hull.ac.uk/ridir/>

a means of enabling such linking between copies amongst other aspects. Both projects will report in 2008 and add direction to the situations MIDESS explored and would like to pursue.

The report on testing the exchange of METS files is welcome for its honesty and also in its highlighting of the many small details that can impact on the success of exchange. A greater appreciation of the issues involved in exchanging metadata in this way may have assisted the project in moving forward more quickly, though the technical difficulties met in the respective partner platforms prevented this in their own way. The DSpace/Fedora incompatibility is one that the Repository Bridge project addressed, though the problems faced in reproducing this indicate that transfer of a solution from one place to another is not always straightforward. The APSR project³¹ in Australia has also successfully addressed interoperability between these two systems, but it remains to be seen how widely their solution can be implemented. Heterogeneity of metadata is an ongoing issue. The EU DRIVER project³² is promoting guidelines that seek a level of homogeneity to enable successful services over aggregations, whilst RDF and Topic Maps offer potential routes for encompassing heterogeneity without losing its richness.

The experience with METS is valuably reported as a case study, though an experimental approach of reporting outcomes against different tests would have been helpful. A UKOLN workshop on packaging formats in February 2007³³ reached the conclusion that although it is frustrating that METS and similar schemas are permissive, this is also their strength, as they can then be used for many different purposes. The main requirement for using them is thus agreement on their application.

Workpackage 10: Metadata harvesting

The report from this workpackage represents a good overview of the issues pertaining to metadata harvesting and case studies of its implementation across the three partner site repositories. The decision not to implement central harvester software, but write code that allowed the bare bones of OAI-PMH to be demonstrated, is understandable given the problems encountered by the project, though potentially introduced its own variables. The introduction of DigiTool at Leeds and testing using its harvester module allowed both approaches to be undertaken, and an ultimately reasonably successful outcome.

Particular issues of note arising from this workpackage are as follows:

- That the Curator software almost forced the use of sets, a much underused, but valuable component of OAI-PMH that can be used for selective harvesting
- That OAI-PMH can be used to harvest many different metadata types, even if one of the metadata streams must be Dublin Core
- That persistent identifiers are essential to make sense of what is harvested and allow a link from the metadata to the resource³⁴. This conclusion matches the recommendation of the ePrints UK project³⁵.

³¹ Australian Partnership for Sustainable Repositories - <http://www.apsr.edu.au/>

³² DRIVER project - <http://www.driver-repository.eu/>

³³ Content packaging for complex objects workshop - <http://www.ukoln.ac.uk/events/content-packaging/>

³⁴ Though note, what is really required is a persistent locator, which can be derived from the identifier. In many systems the two are the same, but in Fedora the locator needs to be derived from the identifier in the record, often by translating the identifier into a Handle.

Issues arising from the project

The main issues arising from the project have been addressed in previous sections. Additionally, a few issues not directly related to any of the workpackages are considered here. Generally, those issues that have potential for dissemination for the benefit of the community are listed in the following section.

Within the project itself the pragmatism of establishing repository infrastructure has seemed to sometimes obscure the nature of the content being examined. Images and multimedia content are dealt with, and very usefully in places, though a project such as MIDESS could have chosen to report outcomes from this perspective more. This is partly the impact of reality, where attention needs to be given to the infrastructure, and it is to be hoped that as infrastructure matures more content-focused reports can be made.

In doing so there will need to be clarity in the definitions used to describe different content types. Across the project's report outputs there have been references to images, multimedia, video, sound, audio-visual, content (in its generic sense). Whilst these all refer to the same broad grouping of materials, a more specific vocabulary will be needed in the future to aid understanding of the detail of how repositories deal with different content types. Fedora has the concept of content models, and definition of these will be helpful in understanding requirements.

There are also differences between digitised and born digital materials. Up front they are the same, files of a particular format. Managing metadata about them, though, can be very different, and care needs to be taken that they are not confused with one another. The emphasis within MIDESS has been on collection of information about image and related content that has been digitised. A future companion piece could usefully focus on born digital materials and factors affecting these when shared between partners.

Involvement of university photography services would be an interesting collaboration to explore. As such services move into digital they will themselves have a need to manage the images taken, which may prove to be a valuable link for libraries to establish. The pragmatism of this institutionally focused project could be expanded in this direction.

Such pragmatism poses an intriguing dilemma. JISC projects quite rightly insist on a high level of institutional commitment and buy-in for projects to aid subsequent embedding of the outputs and sustainability. The institutional buy-in document produced by the project was a useful clear statement of the state of play at each partner site in this regard. Projects, though, have their own demands, and there can often be a conflict between institutional and project priorities. Within MIDESS there is a sense that the institutional priorities tended to win. This has its own benefits, allowing real experience to come out and directly benefit the community in a way that straightforward project outputs sometimes don't. The balancing act will always be a tricky one: but it is a distinction the JISC should be aware of in their funding and management of projects that are so closely linked into institutional activity.

Dealing with commercial systems also brings the distinction home. Communication with commercial vendors often suggests a willingness to participate in JISC projects,

³⁵ ePrints UK project guidance on full-text jump-off links - <http://eprints-uk.rdn.ac.uk/project/docs/encoding-fulltext-links/>

as it allows the company to test out functionality in a real world scenario. But such companies work to relatively fixed development schedules that don't always match the project's timetable. Early engagement and involvement is advised for the best outcome. The issues faced by Leeds within MIDESS were unforeseeable, and the project has done its best to make the best of a difficult situation, with some unintended bonuses at times.

Focusing on the repositories in the end, as the aim of this project has been to do, an issue that many repository systems do not always seem comfortable with, and seen within MIDESS, is the distinction between internal and external metadata. Internally, repositories can have the freedom to use whatever metadata schema they see fit for best practice. Externally, they need to have the flexibility to present and exchange a range of different metadata schemas according to need. And there needs to be the ability to map between the two for this to happen. Repository systems to date, particularly commercial, have limited internal infrastructures, and often lead to the internal and external metadata needing to be the same. It is to be hoped that increased flexibility becomes part of the repository software stable, allowing libraries to manage what they need and attend to access without one have a detrimental effect on the other.

Dissemination priorities

The following list is an indication of those parts of the project it is considered are of wide benefit to the community and which merit attention in dissemination opportunities, as these arise. The list is undoubtedly not comprehensive, as it is based on documentation solely and project staff will have their own views. It is hoped, nevertheless, that the list will be of value to the project and to readers of the evaluation. The list is not given in any sense of priority other than as they were met whilst writing this report.

- The fact that there is an institutional need for understanding the infrastructure required to manage image content and effectively use this both within and between institutions. This is of benefit to the institutionally held collections and there is much local benefit from exposing resources that might otherwise have remained on limited access.
- The application of metadata and the use of packaging formats requires close definition of how they will be applied, potentially leading to the development of appropriate application profiles.
- The issues experienced when implementing a harvesting process, involving the architecture to be used, within and between institutions, the metadata to be used, and the clear and unambiguous use of persistent identifiers/locators to enable links from metadata records to the full content.
- The importance of considering the needs of both content creators/managers and end-users when planning a multimedia repository.
- The legal report in workpackage 7, particularly the questions in the Appendix.
- The potential for conflict between systems that default to using METS as their core packaging format, often repository systems, and those that default to using IMS CP, usually VLEs. This cross-silo issue is one that will benefit from further attention in enabling repositories to best serve e-learning.
- The importance of incorporating preservation planning into interactions with content creators/managers to embed this practice at source.
- The value of gathering user requirements, for both deposit when dealing with content creators/managers and access when dealing with end-users.
- The need for a clear policy framework and appropriate policies to guide proper use of the repository within the institution.
- The benefit of whole life cost modelling in assisting the selection of appropriate institutional systems.
- The summary lists reflecting on why it is important to use metadata and what the institutional drivers are for having a repository.

Aside from traditional means of dissemination, it is recommended that dissemination through the UK Council on Research Repositories (UKCoRR) is considered in order to reach repository managers directly.

Conclusion

The MIDESS project has produced a series of reports based on work that fills a niche in the development of the repository landscape, addressing both institutional and image-related issues. It has done this in the face of a series of technical difficulties that have hampered what the project had hoped to achieve, but which have been honestly reported: the outcomes from the project are both tainted and enhanced by these difficulties and having to try and overcome them.

The project could potentially have extended its risk management strategy to take into account previous experiences elsewhere more than it perhaps did. Drawing on greater external assistance to help solve the technical difficulties may also have been helpful. Nevertheless, the project has successfully completed the majority of its aims and objectives and provided much case study evidence that will be of benefit to others following the same path. It is to be hoped that the partner sites in particular can build on their experience as they seek to develop stable repository instances.

Annex A: Documents reviewed

The following documents were consulted in the course of completing this evaluation.

MIDESS Project Plan (and iterations of this)

MIDESS workpackage 2 – Functional and Technical Requirements Specification

MIDESS workpackage 3 – User Requirements Specification

MIDESS workpackage 4 – Metadata requirements specification, parts A, B, and Appendix

MIDESS workpackage 5 – Digital Preservation Requirements Specification

MIDESS workpackage 6 – Integration with Enterprise Architecture Specification

MIDESS workpackage 7 – IPR and Multimedia Requirements Specification

MIDESS workpackage 8 – Repository Implementation and population

MIDESS workpackage 9 – Resource discovery and shared services

MIDESS workpackage 10 – Metadata Harvesting

Institutional buy in from MIDESS Partners paper

Steering and Working Group minutes

Presentations given by project staff