Multimedia repositories in learning and teaching – lessons from the MIDESS Project

Michael Emly & Stephen Charles
University of Leeds

Abstract

The MIDESS project brought together 4 UK universities to explore the management of digitised content through the development of a digital repository infrastructure. The project focused on multimedia materials in particular and looked at how support can be provided for their use in a learning and research context and how resources can be shared both within and 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 value of the repository in supporting learning, teaching and research. Having established a repository platform within each institution, the project then explored how multimedia content could be exchanged and shared between the repositories, using OAI-PMH and METS as transport mechanisms.

This paper will summarise the project's main findings. In particular, it will address how a multimedia repository might fit into the information architecture of the university, the likely requirements for integration into an inter-institutional or national framework and some of the obstacles which can impede such integration. Scenarios will be presented illustrating how student learning can benefit from such a repository within a research-intensive university and the relationship between the repository and the VLE will be discussed.

Background to the MIDESS Project

In recent years, the ease with which material can be digitised and the comparatively low costs associated with the necessary equipment has led to an explosion in digitisation projects within the higher education sector. A number of university libraries in the UK have embarked on ambitious digitisation programmes, while individual academics have also undertaken digitisation – often on a piecemeal basis – in support of their teaching and research. There is also an increasing quantity of material "born digital" to add to this array of digital content. However the infrastructure for managing this digital content has lagged behind the ability to create it. In many institutions, such materials tend 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. Issues such as metadata creation and management, effective searching and retrieval, re-use of content in an e-learning context and digital preservation are poorly understood and not widely implemented, mainly because there is no institution-wide facility for the shared management and discovery of this content. VLE's – although implemented within most institutions – provide only limited tools, and those targeted principally at course delivery to students. Many institutions have also set up repositories to capture research outputs – in particular eprints (preprints and postprints of journal articles and conference papers published by their staff) – but the focus of these repositories tends to be narrow and their functionality restricted to what is required for the exposure and exchange of research papers.

It was within this context that, in 2005, the UK JISC funded the MIDESS Project (with supplementary funding from the Consortium of University and Research Libraries) as part of the JISC *Digital Repositories Programme*. This 2 year project brought together the libraries of the London School of Economics, University College London, the University of Birmingham and the University of Leeds to explore the use of repositories for the management of multimedia content within the HE sector.

The objectives of the project were:

- To create demonstrator repositories at each partner site 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 to explore 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.

Project Workplan

The workplan envisaged delivering these objectives via a set of inter-related 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. These documents were as follows:

- 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

A further work-package allowed for the implementation of the repository systems and for their population with sample content. This provided the infrastructure on which two further work-packages could carry out an in-depth investigation of the possibilities for content sharing and re-use:

- Resource discovery and shared services
- Metadata harvesting services

Finally, an evaluation work-package assessed the lessons learnt throughout the project and compared the original specifications to the final outcomes.

The outputs from each of the work-packages, together with other supporting documentation, are available from the project web-site[1]. This paper will focus in detail on how a multimedia repository can support student learning within a research-intensive university, how the repository might fit into the information architecture of the university and the relationship with the VLE. Possible drivers for integration into an inter-institutional or national framework will also be discussed, along with some of the obstacles which can impede such integration.

Repository Infrastructure

After undertaking an initial joint evaluation of repository software, three of the partner institutions selected a software platform for implementation and began to identify suitable content for loading.

At the University of Birmingham, the project provided an opportunity to create a pilot 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.

The London School of Economics (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 Malinowksi'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.

Resource discovery and resource sharing

With repositories established and populated with multimedia materials within 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 a surprise that only half of the software platforms deployed within the MIDESS project provided an implementation of the protocol 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 Dublín Core. This can be explained most simply by describing how Endeavour Curator handles harvesting in this case. For each metadata schema in the system (including qualified Dublin Core), a mapping file is created which maps each element containing data to be harvested onto an appropriate element in the oai_dc schema. For example, the qualified Dublin Core element Date Created (dcterms:created) would normally be mapped onto oai_dc:date. In this way, full metadata can be made available for harvesting regardless of schema with relatively little effort. However in the case of DSpace, the project found that many 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 require 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 homgeneous 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, and customisable mappings are essential if key metadata elements are not to be lost in harvesting.

The MIDESS Project wished to go beyond resource discovery and explore how digital objects could be shared and re-used within a cross-institutional context. METS (Metadata Encoding and Transmission Standard) was designed *to facilitate the interoperable exchange of digital materials between institutions*[2] and LSE had successfully used METS to create ingest packages for Fedora. 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 project[3] – who had encountered the same problem - was installed, but without success. Initial attempts to load the same METS object into DigiTool also failed. However, it was 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).

There is, then, a clear need for interoperability standards to be developed and agreed in this area too. METS provides little more than a broad structure for packaging digital objects, complete with their metadata. Significant investment is required to develop a range of application profiles which will facilitate data exchange between different systems. Equally, repository software must recognise and be able to handle those application profiles, preferably with some degree of tolerance so that – at least under some circumstances - the inability to handle one particular schema or feature within the METS record might generate a "warning" rather than a complete failure to process the object.

The role of repositories in supporting learning and teaching

In the course of the project, there was substantial engagement with academic staff in all three institutions. This focused not only on the different types of material held locally or centrally within the institution which might appropriately be housed within the repository but also on how those academics saw the repository service functioning, how it could support learning, teaching and research within the institution, and the broader framework within which it needed to be integrated.

Academics were very clear that their own research benefits, and often relies on, access to digital materials in a number of formats, including images and audio and video materials or, in other disciplines, datasets derived from experimental procedures. In many cases, these represent the primary resource on which their research is based. Working within a research-intensive university, which aims to translate excellence in research and scholarship into learning opportunities for students, it can be important to expose students to these primary materials. This is indeed what already occurs with many print-based and textual sources, which are available through the university library and through access to electronic journals.

As blended learning becomes the norm for many university modules, academics therefore wish to pull in illustrations from the primary materials which underpin their own research. Sometimes these will be repackaged with commentary or other material into learning objects and stored in the VLE. But more commonly, it suffices to show 3 or 4 images or a short clip from a video in the course of a lecture by way of example, and then link to these - or to a wider set of examples - from within the VLE.

For both these scenarios –the desire to expose students to a "library" of digital objects to encourage student-centred independent learning and the use of specified digital objects to support learning and teaching – the repository infrastructure provides a strategic way forward.

At Leeds, an example of this approach is in the School of Fine Art where an academic has built up a collection of slides of art and cultural objects from SouthEast Asia, representing one of the primary sources on which their academic research is based. Some of this material is digitised, but the images are only readily available to that one individual academic. There is a need to digitise all this material and make it available to students via the repository so that they can browse it and base their own work on a rich collection of appropriate images.

Similar needs have been expressed by academics from a large number of other schools including Medicine, English, Music, Education, Healthcare and Modern Languages and all these have identified relevant material which they are keen to put into a centrally supported service. Potential applications are as diverse as pathology slides, dance videos, political speeches, English dialect recordings and photographs of historical scientific equipment as well as a wide variety of textual materials and their modern podcast equivalents.

Although over-simplified, Figure 1 presents something of the vision which emerges – where the repository provides access to a rich collection of digital objects to support independent learning, in the same way that the Library does for non-digital formats.

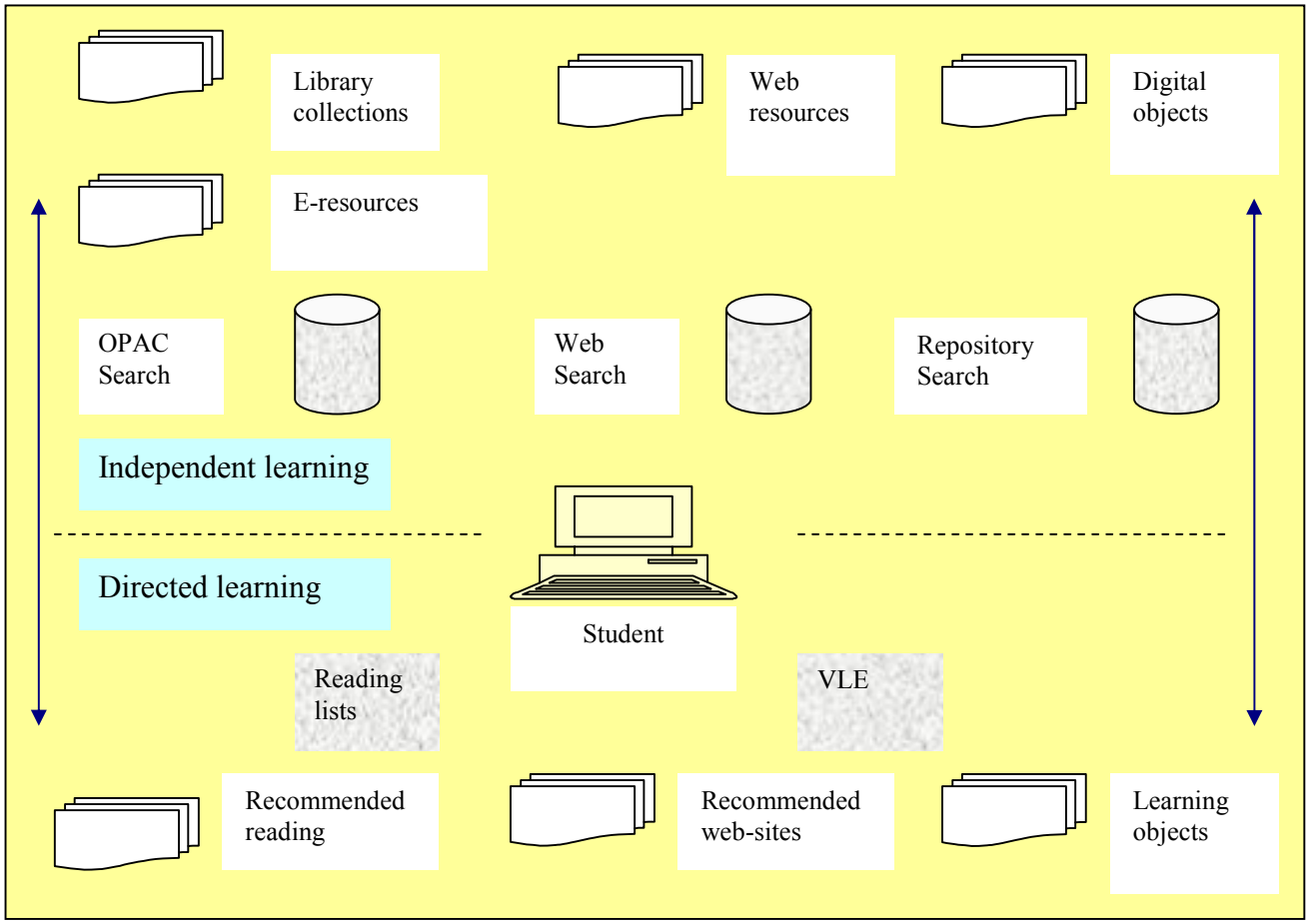

Figure 1. How a repository can support independent, research-led learning

The discussion so far has focused on the place of a single institutional repository within the broader architecture. However in reality, the situation is more complex. Firstly, the University of Leeds already has an operational repository for research outputs (mainly articles), using the Eprints platform. Clearly the information architecture must integrate both repositories in a way which does not confuse the end-user. It may be that the federated search portal can provide the necessary integration for most purposes. An alternative which is being actively considered would be to use the OAI-PMH protocol to harvest metadata from all repositories (and possibly other data sources such as the VLE as well) into a single searchable silo.

A somewhat different scenario has arisen within the Faculty of Medicine, which is a partner in the ALPS CETL (part of a national network of Centres for Excellence in Teaching and Learning and with a particular focus on **A**ssessment and **L**earning in **P**ractice **S**ettings)[4]. 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.

ALPS staff have been working with the MIDESS project to investigate the storage of their learning objects / digital material in the DigiTool repository at Leeds University. This material needs to be restricted to ALPS members across the partner institutions because of its potentially sensitive nature. The material will consist 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. Although no sensitive medical material will be displayed for the ALPS pilot, there is the potential that material unsuitable for open access could be stored in the repository and that this material may need to be shared. This material will therefore need to be restricted to authorised users from among the ALPS partners and authentication will be provided via Shibboleth[5].

The ALPS information architecture is particularly complex for a number of reasons:

- 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
- 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[6]

Figure 2 shows the IT architecture of the ALPS project – centred on the staff member or student working within the framework of the VLE of their home institution. Material captured from mobile devices is passed into the VLE and/or the university repository. Equally, the various repositories provide a rich information backdrop to directly support student learning.

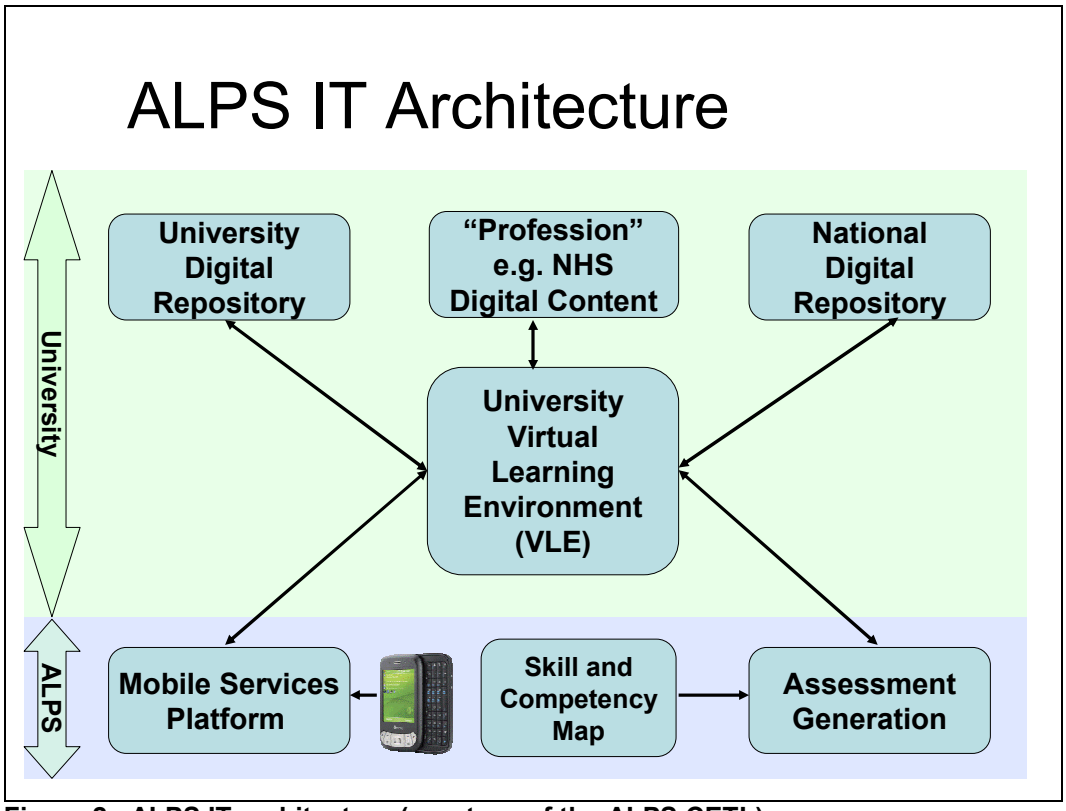

Figure 2. ALPS IT architecture (courtesy of the ALPS CETL)

Interaction between the Repository and the VLE

The two scenarios described have broad acceptance within the relevant communities and provide a convincing – if somewhat oversimplified - architecture from the student perspective. However, the academics operating within this environment would need to undertake certain "maintenance" functions which involve a degree of interaction or integration between the VLE and the repository. As a minimum, it is necessary to have functionality which permits a VLE module administrator to:

- Create a link from the VLE to a specific object or collection within the repository.
- Create a link from the VLE to a search box within the repository into which the student can then enter criteria for the repository search.
- Pass a specific search string from the VLE as a search term for the repository (pre-defined search).

Whereas the Curator system included a VLE integration module (compatible with both BlackBoard and WebCT) which permitted seamless execution of these functions, DigiTool at present has no similar functionality. It is possible to use the VLE and DigiTool APIs to replicate this, but there is no such facility built-in. In particular, the DigiTool resource discovery interface does not offer a straightforward way of establishing the permanent URL for a specific object or collection.

Further issues arise when considering re-purposing of objects held in the repository, for example the incorporation of a digital object such as an image or video file within a new learning object. For this to occur, the academic must be able to access and transfer into a different environment (such as the VLE or RELOAD) the actual file or files which constitute the original digital object. However, the public interface to many repository systems is designed solely for access by the end-user, and typically launches an application which will permit the object to be viewed or otherwise used; there is generally no way of bypassing this in order to access the raw datastream. This is a clear limitation in DigiTool. By contrast, within Fedora, download is possible provided that access is given to the "item index" as well as the "dissemination index".

Additionally, none of the software platforms used within the MIDESS Project offered any specific support for holding learning objects within the repository. Support for IMS-CP and/or SCORM would be essential if it is desired to fully integrate a repository and a VLE so that learning objects can be flexibly transferred and exchanged between the two environments. There are indeed repository software platforms which offer such functionality, but they are typically designed for this specific purpose and, while offering full support for IMS and SCORM, fail to provide the functionality which is required in a more generic multimedia repository (e.g. support for metadata schemas and other standards in common use within the library/museum/archives domain). A further difference is that learning objects tend to have a fairly limited life-span whereas long-term access, secure storage and preservation are major considerations for repositories holding multimedia or cultural objects. This difference in purpose has led to differentiated products and currently forces a stark choice on those wishing to implement a repository.

One possible solution to this dilemma may be to view the VLE as a mini-repository in its own right, appropriate for holding complex learning objects for their (limited) lifetime while individual digital objects (which may in some cases become the raw building blocks for a learning object) are held in an external repository which can offer functionality appropriate for long-term storage and use. This might also provide a way of addressing another issue, namely the need within some contexts to maintain dynamic collections of digital objects, either because the lecturer needs to update the object at frequent intervals, or in scenarios where students are themselves invited to contribute digital objects for viewing and use by their peers, as can occur within the ALPS project previously discussed.

Creating a national framework

From the work of the MIDESS Project, it is quite clear that users need and expect an integrated framework for resource discovery of digital objects, one which is capable not only of providing access to resources held locally, but also of identifying relevant objects held elsewhere within the professional, regional or national/international community. This probably seems obvious within the context of academic research, but our work suggests that it is necessary at all levels of tertiary education if students and other learners are to receive a rich and relevant learning experience. This conclusion should not be a surprise to either information or educational professionals when we see so many students using Google to leverage online information wherever it resides. This paper will therefore conclude by considering some of the key building blocks for the framework required.

With multiple repositories, even perhaps within a single institution, cross-searching and/or aggregation of content is clearly a key factor. Experience of implementing Z39.50 to permit cross-searching of OPACs within the Library community suggests that cross-searching can be very effective with a small number of compatible targets, but that aggregation of records within a single database provides the best solution where a larger number of data sources needs to be searched. Within the repository community OAI-PMH provides a mechanism for such aggregation, and there are several successful deployments in place. Within the United Kingdom, JISC has funded a service Intute Repository Search[7] which provides a single point of access to over 80 institutional repositories. The content available currently consists mainly of academic research papers, reflecting the predominant content of university repositories at this time, but there is no reason why the scope should not broaden in line with repository content. A similar service exists for the Netherlands[8] and various other countries/user communities, while the DRIVER Project[9] is attempting to prototype a pan-European aggregation of research materials. Of equal interest is the Nereus service[10] which offers an integrated portal to multiple information services relevant to economists, including institutional repositories, library catalogues and bibliographic databases. Again, OAI-PMH is used to harvest from the repositories; significantly, a "converter layer" is used to ensure compatibility between the various data sources.

The experience of the MIDESS Project suggests that the level of interoperability which can be achieved via OAI-PMH is not always as good as these successful implementations might suggest. Firstly, it is bizarre that not all software offering a data provider service via OAI-PMH includes in the metadata to be harvested a URL permitting direct access to either the digital object or to its metadata record within that repository. Absence of this element presents a major barrier to seamless access for the end-user. Secondly, support for harvesting specific subsets or collections within the repository and compatibility with access restrictions enforced through the native resource discovery interface are often poorly implemented. However of greatest significance for resource discovery is the poor handling of metadata held in schemas other than simple Dublin Core, resulting in inadequate metadata exposed for harvesting by some software platforms. The facility to map between metadata schemas is essential for effective harvesting.

Within specific user communities, much can be achieved by the adoption of common standards. However, it would be extremely useful to agree minimum standards and expectations for metadata harvesting if broader interoperability is to be achieved; and equally, these standards need to be embedded within the core functionality of all software platforms. If such standards could extend to the actual delivery of the identified resource – and this is an objective of the OAI-ORE Project[11] – then so much the better.

Conclusions

In conclusion, the work already undertaken on repositories has yielded some understanding of the architecture and level of service required and some working models which suggest how

these may be achieved. Within the narrower field of eprints archives, there is indeed much experience to draw on. However, as repositories are becoming more established, they are also broadening their scope. They are now being used to manage material in a wide range of formats and for a variety of purposes – and this trend will increase as institutions realise the importance and value of their digital assets.

It is essential therefore that individual higher education institutions adopt a proactive approach to these developments, implementing appropriate repository solutions and embedding them within a broader information architecture which allows both students and staff to discover and access the information which they require, regardless of its format. Key interactions between the repository and the Library Catalogue, Portals and the VLE need to be addressed, but these should be considered within the context of broader information provision. Various protocols such as Z39.50, SRU/SRW and OAI-PMH are available to enable cross-platform resource discovery. Even so, the implementation of connecting services can often be difficult and only partial. However the sector is investing substantially in service integration and we would expect substantial progress to be made in this area at the same time that repository software matures and provides the functionality necessary for such integration. Such developments will need to encompass a broad range of issues, including authentication and authorisation.

At a national level, an informed and cohesive approach is required which will facilitate resource discovery across these local repositories by establishing aggregator gateways (what OAI-PMH calls *service providers*). In some countries, the initiative will probably come from the national library; in others a different national agency, a single institution or a consortium may take the lead. Finally, at an international level it is imperative that this shared vision should lead to effective action to improve interoperability through agreed standards and application profiles, alongside a commitment by developers – both open source and commercial – to delivering compliance in the major software platforms.

References

[1] http://www.leeds.ac.uk/library/midess/index.html
[2] 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
[3] http://www.inf.aber.ac.uk/bridge/
[4] http://www.alps-cetl.ac.uk/
[5] http://shibboleth.internet2.edu/
[6] http://www.jorum.ac.uk/
[7] http://irs.ukoln.ac.uk/
[8] http://www.darenet.nl/en/page/language.view/search.page
[9] http://www.driver-repository.eu/
[10] http://www.nereus4economics.info/
[11] http://www.openarchives.org/ore/