



UNIVERSITY OF BIRMINGHAM



MIDESS Workpackage 9: Resource Discovery and Shared Services

Executive Summary

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

This work-package sets out some key drivers for resource discovery and explores various ways in which a repository can be integrated into the broader context within which it operates. A model which permits cross-searching of the Library OPAC and the multimedia repository is presented, and then elaborated as part of a more complex and user-centred information architecture which would permit access to both local and national level. The requirement for integration with the institutional VLE is discussed, and various issues identified which can impede the flow of data between the two systems.

Although users may benefit from awareness of digital objects held in a remote repository, they will generally desire access to the object itself for use or re-use. The MIDESS partners

MIDESS Workpackage 9: Resource Discovery and Shared Services

attempted to transfer material electronically between their respective repositories using METS, but with mixed results. It was concluded that while schemas such as METS can provide a mechanism to allow the transfer of material between repositories, there are a number of significant compatibility issues around this. These issues include divergent use of metadata schemas, differences in the way that the various systems implement METS and non-compliant METS files being produced by the METS export facilities. Installation of the software written as part of the Repository Bridge project still did not permit successful transfer of digital objects. At present, therefore, the use of METS for this purpose seems to be viable only under very specific circumstances. However recent work by the OAI-ORE Project to facilitate the sharing of digital objects may improve the situation in the future. The creation of a range of METS application profiles and inclusion of support for these within repository system software is also identified as a key development which could improve interoperability.

Sharing broadcast transmissions recorded under the ERA licence was another area of project activity. This work illustrated very clearly the issues and constraints which can arise in a real-life context, and the implications for metadata content and for the transfer of digital recordings between institutions are discussed. Rights issues were also at the fore of joint discussions with the JISC CHERRI Project, which looked at methods for ensuring compliance with privacy and data protection concerns when images and recordings originate in a clinical context.

MIDESS Workpackage 9: Resource Discovery and Shared Services

| | |
|--|-----------|
| Introduction – the need to share data | 4 |
| Resource discovery – the role of the repository within the institutional information architecture | 5 |
| Cross-searching using Z39.50 | 5 |
| Moving towards a more complex environment | 9 |
| Sharing Digital Repository Content..... | 13 |
| Interoperability of digital content using METS..... | 14 |
| Initial attempts to transfer digital objects | 15 |
| Tests using Repository Bridge software..... | 15 |
| Further tests using DigiTool | 16 |
| Current Developments in Improving Digital Repository Interoperability | 22 |
| Sharing video content under the ERA licence | 22 |
| Medical recordings and liaison with the CHERRI Project | 24 |
| Conclusions | 25 |
| References | 27 |
| | |
| Appendix 1 | 28 |
| Appendix 1a. UKETD_METS | 28 |
| Appendix 1b. METS | 32 |
| Appendix 2 | 35 |
| 2a. XML data from Repository Bridge in MODS format | 35 |
| 2b. XML data from Repository Bridge changed to Dublin Core format..... | 36 |
| Appendix 3 | 37 |
| 3a. dc.xml file supplied by Birmingham exported as METS from DSpace..... | 37 |
| 3b. Modified dc.xml file ingested into DigiTool..... | 40 |
| Appendix 4 | 44 |
| 4a dc.xml METS file from Birmingham before modification | 44 |
| 4b modified METS file from Birmingham for ingest into DigiTool. | 46 |
| Appendix 4b-1..... | 47 |
| Appendix 4b-2..... | 48 |

Introduction – the need to share data

Digital repositories are now coming of age – but within an increasingly complex information environment where the user requires to know about and access many different types of resource, in both digital and non-digital formats. Within an HE institution, the Library OPAC and A&I databases such as Web of Science or Medline have traditionally been key tools in this resource discovery process, while Google and other search engines revealed less formal internet information resources. However VLEs, local and national repositories, national information services such as Intute and subject-oriented or specialist services such as arXiv and AHDS are simply some of the alternative places where critical information may reside.

Even within a single institution, multiple repositories may co-exist, one specialising in research outputs, another supporting learning and teaching (perhaps as part of or alongside the VLE), while a third may be run by a single faculty or department to meet its own specific requirements.

In this complex environment, there are three key activities which have to be addressed:

1. **Resource discovery:** the user must be able to identify what resources are available, and where, quickly and with minimum effort, no matter what their format or location
2. **Access:** having located a pertinent resource, the user will probably desire to access and view/use it. In many circumstances, this will require authentication and authorisation.
3. **Re-use:** in some cases, the user will want to take a digital object and re-purpose it, either making it accessible within a new context or incorporating the original into a new digital object such as when a digital image is reused as part of a more complex learning object. It is normally imperative to establish whether this is permitted and rights issues become crucial.

Each of these activities can be considered from a number of different and overlapping angles. More importantly, each community of users will bring its own requirements, its own priorities and perhaps its own shared model of how this information environment should operate. For example, a group of academics undertaking research in a single subject discipline may wish to emphasise common access to shared research materials whereas an undergraduate will probably focus on the ease with which s/he can identify and gain access to books, articles and digital objects which directly support the learning process.

The complexities of this situation were recognised at an early stage of the MIDESS Project, which was funded to explore the management of digitised content in an institutional and cross-institutional context through the development of digital repository architecture. Two of the original project objectives were to:

- Explore the role of digital repositories within the institutional information architecture
- Explore opportunities for the sharing and re-use of collections across institutions

In order to address these questions, the Project investigated a number of different approaches, including:

- The use of the Z39.50 protocol to provide a single interface to the repository and to other information resources on campus
- The use of METS to enable the transfer of digital objects between repositories
- The use of OAI-PMH to enable interoperability between repositories, including its potential as a transport mechanism for transferring digital objects between repositories. The work on OAI-PMH is reported in MIDESS Work-Package 10.

MIDESS Workpackage 9: Resource Discovery and Shared Services

Within the framework of the MIDESS project, three different digital repository platforms were installed at the University of Leeds (originally Endeavour Curator, subsequently migrated to Ex-Libris DigiTool), the University of Birmingham (DSpace) and the London School of Economics (Fedora). This provided a good testbed for exploring the effectiveness of both METS and OAI-PMH for resource discovery and exchange of digital objects in a cross-platform and cross-institutional context.

Resource discovery – the role of the repository within the institutional information architecture

Resource discovery is fundamental if the users' needs are to be met and if maximum benefit is to be derived by the institution from the (not inconsiderable) costs of setting up a repository. The precise interactions which are required – and therefore the means by which they are to be achieved – will vary from institution to institution. Within the MIDESS project, Leeds in particular has given some consideration to these issues and Workpackage 6 (MIDESS Integration with System Architecture Specification) has already set out some of the key interactions.

The sections of Workpackage 6 on interactions with the Library OPAC, the Federated Search Tool and the VLE set out the basic requirements as understood at the beginning of the project and informed the work subsequently undertaken on the broader information architecture. However, some of this preliminary work was rendered obsolete by the replacement of the Endeavour Curator system by an Ex-Libris DigiTool repository in the first half of 2007 (consequent on the purchase of both companies by Francisco Partners). The new system, as purchased, included functionality for a Z39.50 target but had no SRU/SRW capability, nor a Z39.50 public client interface. Integration with the Portal was therefore not pursued at this stage.

Cross-searching using Z39.50

The library OPAC is a key component in discovering what resources are owned by the institution; a key question must therefore be how, within the overarching information architecture, digital resources held in the repository can be discovered with the same ease as print and other non-digital resources which are held in the library. Since the DigiTool repository at Leeds supports remote queries via the Z39.50 protocol, it was decided to test cross-searching with the library OPAC and thereby explore some of the issues which would arise in developing a live service between these two systems. The most accessible client around which to base these tests was the web-based client which came with the Innovative library management system, although this client has never been deployed in a live context at Leeds.

In order to cross-search 2 targets, it is necessary to configure Z39.50 use attributes which are broadly supported by both targets. Whereas most OPAC's offer multiple indexes for searching and are therefore able to support a variety of use attributes, much of the metadata within a repository is usually less complex in structure. In the case of DigiTool, although the Z39.50 configuration is very configurable, the underlying system indexes are designed to support keyword searching and to differentiate creator, title and subject searching, or a global keyword search. Since all of these are also supported by the Innovative OPAC (they were added in order to comply with the Bath Profile), it was decided to configure for searching by the following use attributes:

MIDESS Workpackage 9: Resource Discovery and Shared Services

| Label | Use | Relation | Position | Structure | Truncation | Completeness |
|---------|------|-------------|----------|---------------|-------------|-----------------|
| Keyword | 1016 | 3 Equals | 3 any | 6 wordlist | 100 None | 1 Incomplete |
| Title | 4 | 3 Equals | 3 any | 6 wordlist | 100 None | 1 Incomplete |
| Author | 1003 | 3 Equals | 3 any | 6 wordlist | 100 None | 1 Incomplete |
| Subject | 21 | 3 equals | 3 any | 6 wordlist | 100 None | 1 Incomplete |

ISN searching was omitted at this stage, since no repository metadata records contain ISNs, but it might be desirable to include this search type in the future to optimise on OPAC searching.

Access to the cross-search facility can be provided in the form of a standard URL:

<http://lib.leeds.ac.uk:211/z39m/Digitool%3AGEN01%2Cxmleeds%3AINNOPAC>

This results in the following screen (fig.1), from which a particular search type can be selected (this is the default Innovative behaviour). No effort was made to standardise or “beautify” the page e.g. by replacing the buttons with the icons used in the standard OPAC.

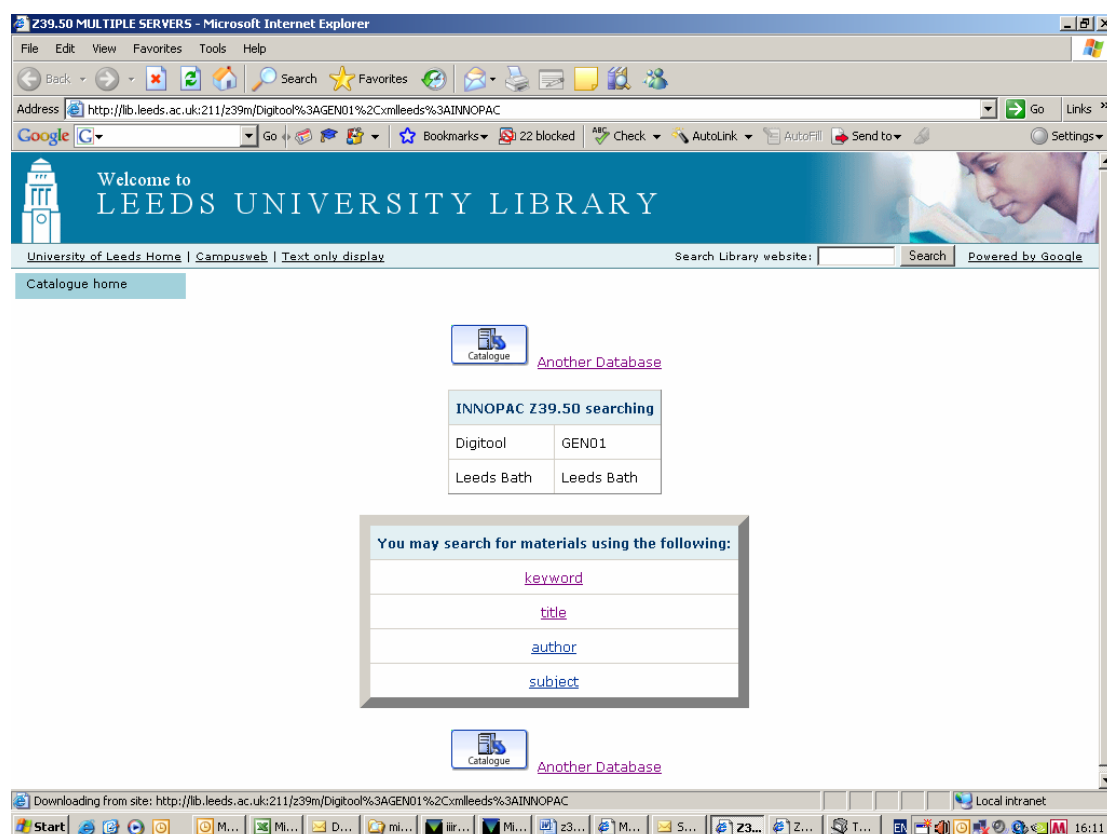


Figure 1. Initial screen to cross-search the OPAC and DigiTool via Z.39.50

Selecting a search type causes a box to be displayed into which the search word or words can be entered. Executing the search causes a results screen to be displayed from which it is

MIDESS Workpackage 9: Resource Discovery and Shared Services

possible to select either the results from a single target, or a combined list of all results from both targets (fig.2-3).

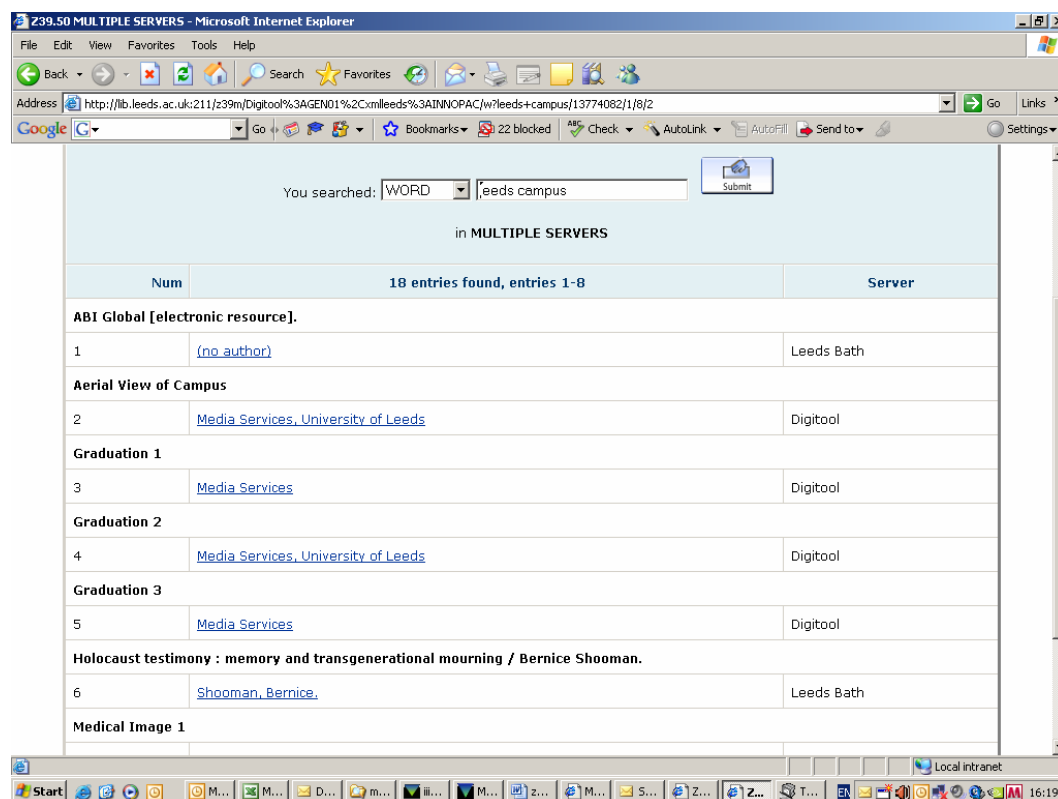


Figure 2 Results returned from a cross-search of the OPAC and DigiTool

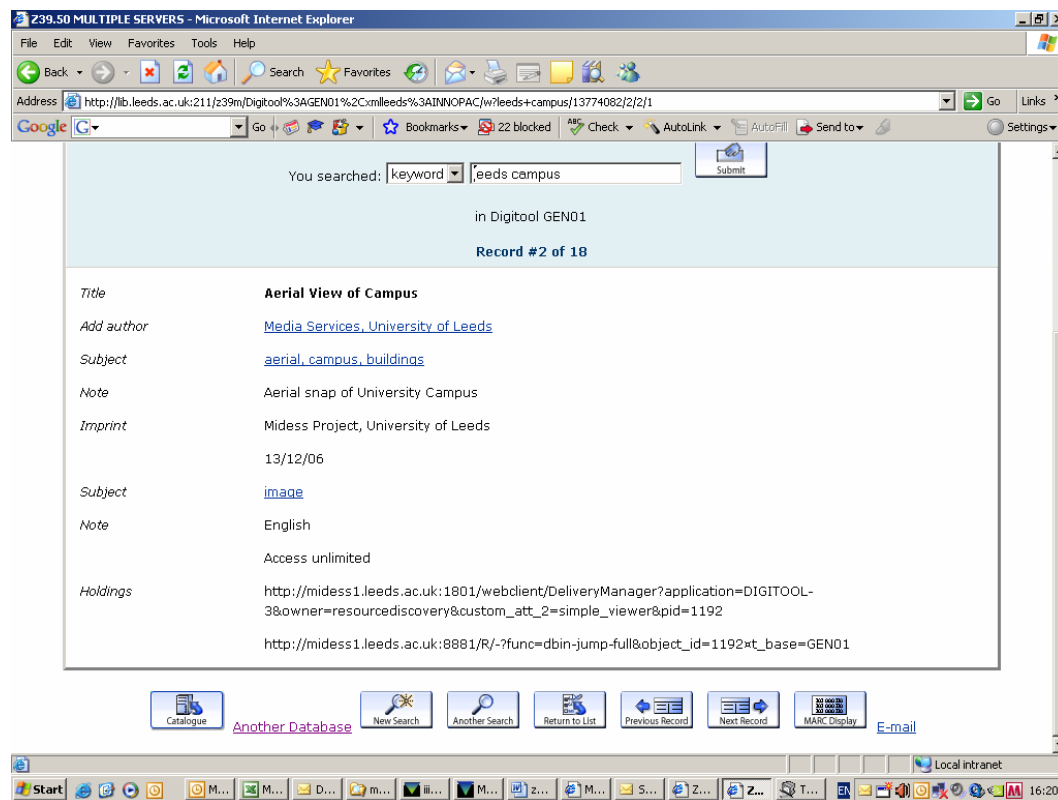


Figure 3. Elements in a record from a Z39.50 search of DigiTool

MIDESS Workpackage 9: Resource Discovery and Shared Services

In testing this cross-searching facility, a number of questions and issues arose:

1. The lack of full LMS functionality when records from the OPAC are displayed. Since these records are displayed exclusively through the web client software which applies the same rules to records from remote and local databases, the functionality is much reduced compared to the standard OPAC e.g. it is not possible to tailor screen displays to the same degree, to tailor field labels, to display full information about the circulation characteristics of the item, to place hold requests, etc. In federated search facilities, this is normally overcome by providing a link through to the native OPAC interface; however the Innovative Z39.50 client does not offer this.
2. It was decided to request records from DigiTool in MARC21 format since this is one (amongst several) formats which both systems supported. This resulted in a broadly comprehensible record display, although the metadata field labels are necessarily very vague and often fail to describe the real nature of that metadata element. More seriously, the Innovative client failed to render the URL (passed in MARC tag 856) as a clickable link. It is suspected that configuration options in the Innovative system may have been able to rectify this anomaly, which greatly reduces the usefulness of the cross-searching facility, since the only way of accessing the digital object would be through a manual cut and paste to a new browser tab/window. However time constraints prevented the detailed investigation necessary to correct this feature. Again, this would not typically be an issue if using federated search software.
3. User expectations in constructing searches. Whereas the Leeds OPAC contains over 1.4 million bibliographic records, the repository in the medium term is only expected to contain tens of thousands of records, for which, however, the metadata terms and their context can be very specialised. This can lead to a search which is sensible within the repository context (e.g. leeds campus video) retrieving thousands of entries from the OPAC which bear no relation to the material retrieved in the repository search. This is somewhat analogous to the issues in federated search facilities where A&I databases are cross-searched with OPACs and other monograph-oriented databases. Perhaps users will find ways of adapting to and working around such situations, perhaps more sophisticated ranking algorithms will provide solutions; at present we need to recognise the difficulties users face in obtaining meaningful results from certain types of searches and assist them in whatever ways we can while working towards long-term solutions.

Testing cross-searching through the Innovative Z39.50 client was always seen as a way of providing proof of concept. At the time, the Leeds University Library was bringing on stream a federated search facility using the OVID SearchSolver software. Implementation within that context (as illustrated in fig.4) would overcome the first two issues discussed above and it is likely that Leeds will seek to offer DigiTool as a target offered through SearchSolver once a larger number of objects with associated metadata has been loaded into the repository.

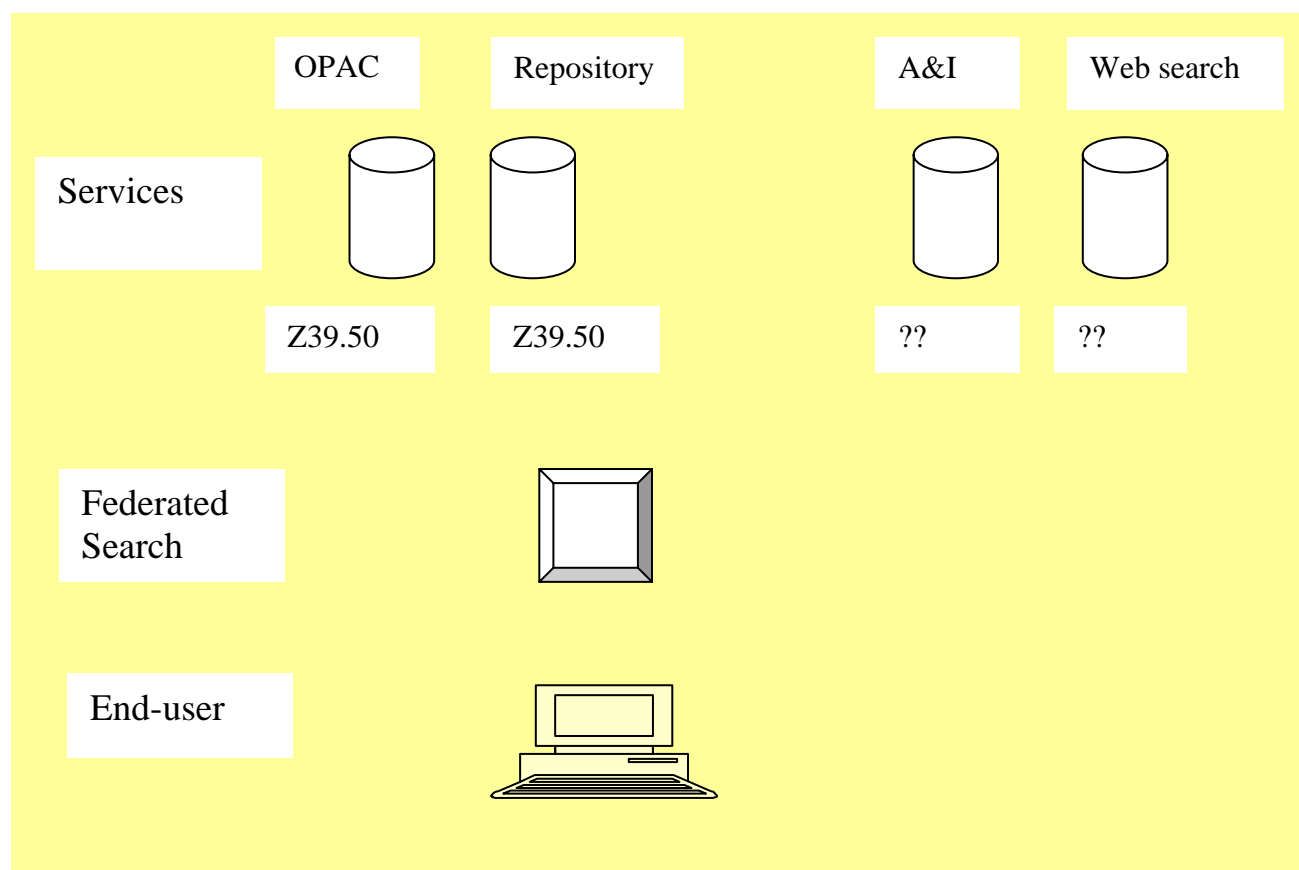


Fig.4 Simplified information architecture with a federated search service

Moving towards a more complex environment

Whereas the previous section on Z39.50 concentrated on the integration of diverse resources, of which digital objects held in a digital repository represent just one type, the implementation of a new VLE at Leeds raised other fundamental issues about where precisely digital objects should be stored, and why. Digital objects can of course be stored within the actual VLE framework, but many VLEs also include an option for a separate learning objects repository. The tightness of the integration with the VLE and full support for standards such as IMS and SCORM which are the accepted standards for content packaging of learning objects can make a dedicated repository of this kind a very attractive option. In contrast, many general-purpose repository solutions do not offer full support for these two standards. In some cases, it is possible to ingest objects in these formats, but the ability to handle complex objects of this nature and to access and edit the associated metadata is often limited or non-existent. The limits on the use of DSpace for learning objects was a key reason why Birmingham decided not to develop their repository into a live service once the MIDESS project finished.

At Leeds neither Curator nor Digitool offered support for IMS and SCORM. However the situation was further complicated by the fact that, whereas Curator had included a module which facilitated interaction with a VLE (either BlackBoard or WebCT), no such support was available in the standard Digitool system. Use of the Digitool API to create seamless routines to support the desired interactions was found to be too costly in development time, as well as demanding a good knowledge of the VLE toolbox - which was not readily available. Instead it is planned to support core workflows by exploiting standard Digitool functionality (including facilities for deep linking) and embedding these, where necessary, in very simple BlackBoard building blocks, using Shibboleth authentication services in both environments to provide seamless access. The core functions to be supported are:

MIDESS Workpackage 9: Resource Discovery and Shared Services

1. To carry out a search of the repository from within the VLE
2. To embed within the VLE environment a link to an object (or group of objects / pre-specified search) within the repository

With shared authentication services, it will also be relatively simple:

3. To derive a digital object from the repository for embedding within a learning object in the VLE
4. To upload (deposit) an object currently stored within the VLE environment into the repository

Although it is important to identify mechanisms whereby interaction between the VLE and the repository can be supported, even more important is the underpinning vision of how the repository can support learning and teaching within Leeds. As a research-intensive university, Leeds aims to translate excellence in research and scholarship into learning opportunities for students, and the repository provides one of the ways in which this can be done.

In the course of the MIDESS project, discussions with academics from various disciplines have identified how their research benefits – and often relies on – access to digital materials in a number of formats, including images and audio and video materials. In many cases, this material represents the primary resource on which their research is based. However, whereas traditional research papers are readily available to students via the University Library – usually today via electronic publishing – students do not often have access to a similar body of material in these other formats.

Moreover, as blended learning becomes the norm for many University modules, academics 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 may only be necessary in the course of a lecture to show 3 or 4 images or a short clip from a video by way of example and then link to these same 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. The project’s experience was that this vision found widespread acceptance amongst academics; contact with students was more limited, but those who attended demonstrations of the repository were equally enthusiastic about its possibilities.

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 this rich collection of appropriate images.

Similar needs have been expressed by academics from a large number of other schools including the School of Medicine, the School of English, the School of Music, the School of Education, the School of Healthcare and the School of 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 (there is detailed discussion of some of these collections in Workpackages 4 & 8).

MIDESS Workpackage 9: Resource Discovery and Shared Services

Although over-simplified, figure 5 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.

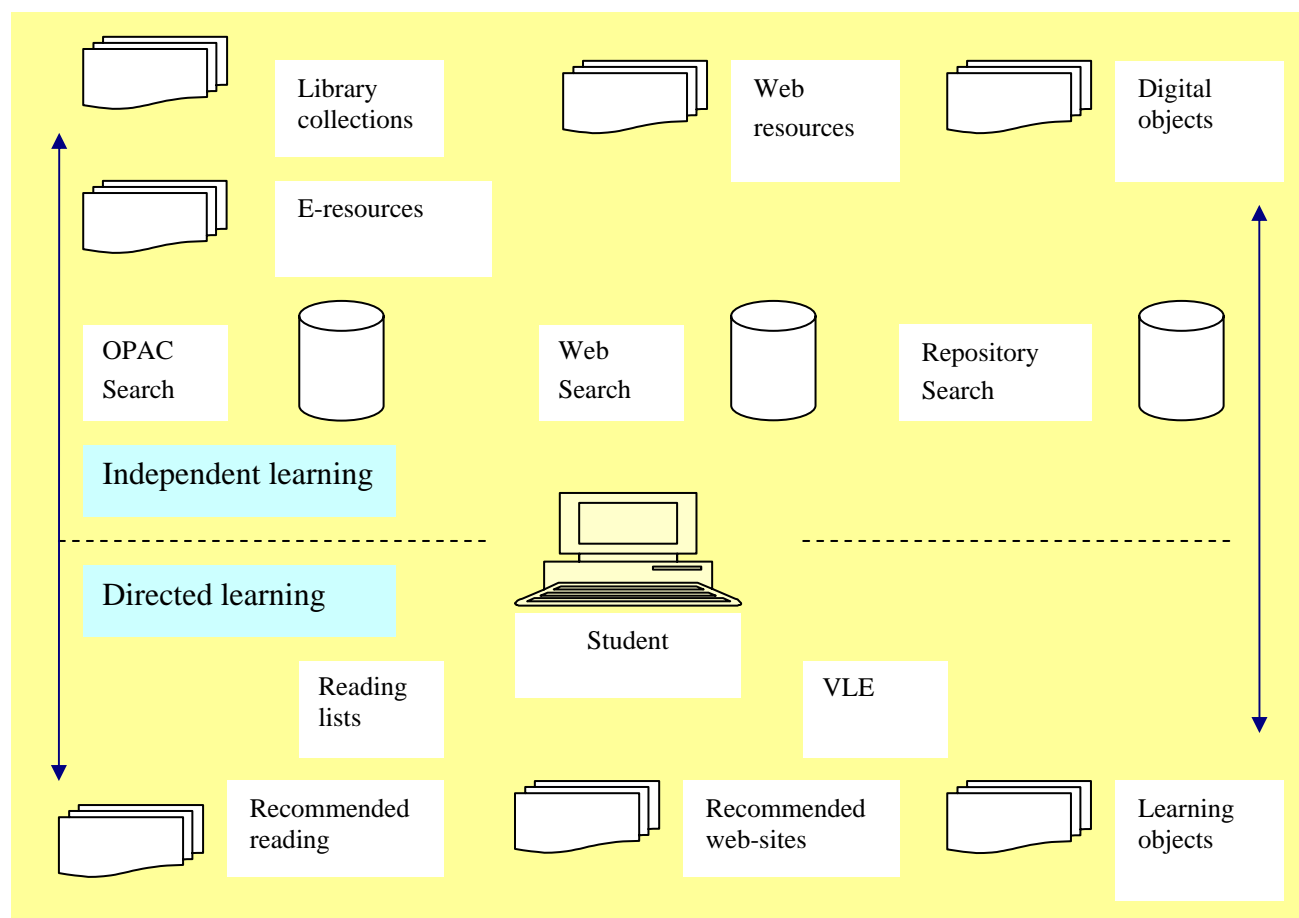


Figure 5. How a repository can support independent, research-led learning

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 than this. 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. Workpackage 10 describes the work undertaken on metadata harvesting.

A somewhat different scenario has arisen within the Faculty of Medicine, which is a partner in the HEFCE-funded ALPS CETL (a Centre for Excellence in Teaching and Learning which is focused on Assessment and Learning in Practice Settings). 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 Yorkshire and Humberside NHS, 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 the learning objects / digital material in the DigiTool repository at Leeds University. This material

MIDESS Workpackage 9: Resource Discovery and Shared Services

needs to be restricted to the ALPS members across the institutions because of the potentially sensitive nature of the material. The material will primarily consist of image and video content captured on mobile devices by staff or students on placement within the NHS, which is then 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 a repository and 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.

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

- The ALPS-CETL is a collaborative programme between five independent HE institutions, each with its own infrastructure.
- Staff and students additionally require access to resources available through the NHS such as the digital repository being developed on a Documentum platform
- Staff working within the project framework would potentially benefit from access to learning objects created in other HE institutions or available through services such as JORUM

Figure 6 shows the IT architecture of the ALPS project – centred on the staff 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.

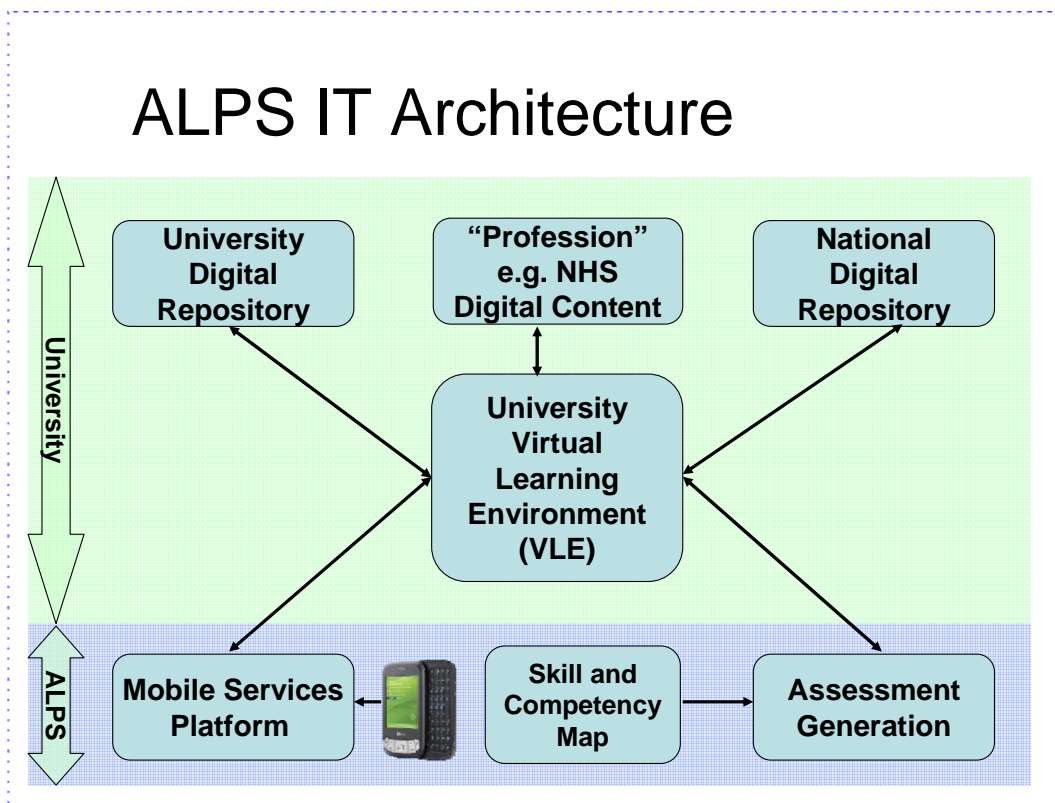


Figure 6. ALPS IT architecture (courtesy of the ALPS CETL)

The synergies for collaboration between ALPS and the MIDESS project were identified at an early stage, and an outline work programme drawn up. Unfortunately, the transition from Curator to Digitool at Leeds seriously delayed implementation, and the absence of the VLE Integration Module within Digitool proved a further obstacle. However, the use of Shibboleth to provide seamless authentication across institutions is now at an advanced stage of implementation and some test objects have been loaded into Digitool. Three requirements for repository-VLE integration have been specified, viz.

MIDESS Workpackage 9: Resource Discovery and Shared Services

1. Ability to link from the VLE to a specific object or collection within the DigiTool repository.
2. Ability to link from the VLE to a search box within the repository into which the user can then enter criteria for the repository search.
3. Ability to pass a specific search string from the VLE as a search term for the repository (pre-defined search).

This list aligns with the more general requirements for VLE-repository integration within Leeds, and it is planned to take the work forward on this basis.

Sharing Digital Repository Content

The example of the ALPS-CETL demonstrates the requirement which academics have to access digital materials stored in repositories outside of their own institution. As the number of digital repositories around the world increases, so there is a growing requirement to provide resource discovery tools and services which unlock the contents of these digital repositories. Two very divergent approaches are possible to address this need.

It would be possible to cross-search multiple repositories at point of need. This is the approach often adopted for discovering resources across multiple platforms within a single institution, but it does not scale well. In the late 1990's, a number of JISC-funded projects explored the use of Z39.50 for cross-searching library catalogues. It is clear that this is able to address the needs of particular communities (e.g. libraries within the London area) but does not provide the flexibility required by the numerous disciplines, research groups and interests within an HE institution. The MIDESS Project did not explore this avenue, not least because no Z39.50 target was installed at LSE or Birmingham.

Instead the repository community has moved in the direction of metadata harvesting, whereby one or more central harvesting services automatically derive data from individual repositories and make it available for searching. The OAI-PMH protocol provides a widely-implemented standard for harvesting and Workpackage 10 describes how this has been explored within the MIDESS Project.

However, resource discovery is usually only the first stage of the process: the user then requires access to the digital object and will perhaps desire to make a copy to store locally for further exploitation and/or re-use. The MIDESS Project deliberately identified similar materials and collections in each of the partner institutions in order to explore how this can be implemented in practice. It was clear that a number of issues would need to be addressed including:

- the ease of transfer of material between repositories using different software
- the different metadata schemas that can be used to describe the same digital object
- divergences in implementation which could hinder interoperability, even where common standards have been agreed.

If both the metadata and the digital object could be harvested and stored, this would provide immediate access to the user. However this is not a common scenario. Instead it is usual for metadata to be harvested and stored centrally, leaving the digital objects (datastreams) in their original locations. The harvested metadata contains a link to the original objects stored in the digital repository. A mechanism is then required for making the object available to the user.

METS provides just such a mechanism, and was explicitly designed as a format for the exchange of information between repositories, a format suitable for a Submission Information Package (SIP) or a Dissemination Information Package (DIP).

Interoperability of digital content using METS

METS (Metadata Encoding and Transmission Standard) provides an XML schema which allows a digital object to be packaged, together with appropriate descriptive and administrative metadata, an indication of the structure of the object and information regarding what is required to use it. A METS object potentially comprises 7 sections:

1. METS header <metsHdr> - containing information about the METS object per se.
2. Descriptive metadata <dmdSec> - this describes the object and is encoded in one or more appropriate XML schema(s).
3. Administrative metadata <amdSec> - including technical information about the object, IPR information, provenance, preservation metadata, etc. Again, any schema(s) may be used.
4. File section <fileSec> - this lists the files which make up the object and, where appropriate, groups them by version.
5. Structural map <structMap> - this is the heart of the METS record: it provides a hierarchical structure of the digital object and links the elements of that structure to the files and metadata which pertain to each element.
6. Structural links <structLink> - this can record hyperlinks between different elements in the structural map (e.g. when a website is being described).
7. Behavior <behaviorSec> - this indicates how the METS object can be used e.g. software requirements to run/use it.

The METS object can physically contain the files which make up the digital object, or alternatively can specify an external location where each file may be found. Similarly, metadata can be encapsulated within the METS object, or held externally and referred to from the appropriate METS section. Only the structural map section is obligatory in a METS object: all the other sections are optional. METS is therefore extremely flexible and yet provides a powerful mechanism for

- indicating the complex links that exist between the various parts of a digital object
- encoding descriptive and administrative metadata for that object
- indicating how the content structure and the metadata interrelate

Staff at the LSE built up METS expertise early in the project because METS provided the easiest way of handling the bulk ingest of metadata from pre-existing databases such as the CLT database of audio and video files and the CALM Archive system which held records pertaining to the Malinowski Archive. For the purposes of ingest into Fedora, METS files must adhere to certain restrictions specific to Fedora implementations and the experience of working with these for the CLT database migration is documented in a separate paper: [CLT media database to fedora ingest.pdf](#)³

These METS files each consist of:

1. A **CLT** (metadata) datastream,
2. A default **DC** datastream,
3. A **RELS-EXT** datastream describing every object's relationship to the CLT media collection object
4. One or more datastreams which redirect to the video or audio files on the CLT server.

Initial attempts to transfer digital objects

With this experience of using METS for Submission Information Packages, it was hoped that exchange of data with the other repositories would be straightforward, but this did not prove to be the case. Not least, it was found that one of the major issues is that different repositories can support one or more metadata schemas. Under these circumstances, moving data from one repository to another not only needs to address the issues of compatibility in the METS wrapper and contents, but also the ability to be able to handle and address the potential changes in metadata schemas that may be necessary between repositories as well. Clearly this is a particular issue where one of the repositories is not able to natively handle the necessary metadata schema.

Of the three repositories, the metadata schema issue is was particularly important for DSpace. Both Fedora and DigiTool are more flexible in the metadata schemas they support. DSpace's internal item importer/exporter is primarily aimed at backup/restore and wholesale repository moves rather than importing external items.

In theory, LSE METS files can produce files which contain oai_dc. DSpace's import translator would then use this oai_dc section to create the metadata for DSpace. This it does by its import translator creating a separate import item in DSpace's own internal format based on the data held within the METS file and, having done this, fetches any files referenced in the file section of the METS file. The standard DSpace internal item importer is then used to ingest this item into DSpace.

However, problems were encountered at a very early stage. Structurally the files created for export by Dspace are quite dissimilar from the files that can be imported into Fedora. Further complications arose even when the same metadata schemas were used since, in the case of Dublin Core, there were inconsistencies in the contents of the elements within the Dublin Core metadata schema which led to problems in interoperability.

Tests using Repository Bridge software

In order to enable better interoperability between DSpace (Birmingham) and Fedora (LSE), software from the [Repository Bridge Project](#)⁴ was installed. This project had examined the interaction between a regional theses repository based at the National Library of Wales and pilot institutional repositories which were under development at UW Aberystwyth and UW Swansea within the EThOS project. It attempted to develop repository software and tools allowing for full interaction between the DSpace and FEDORA open source systems to migrate and store items in a persistent manner. The issues involved are described in a particularly useful paper entitled [Using OAI-PMH and METS for exporting metadata and digital objects between repositories](#) by Jonathan Bell and Stuart Lewis⁵

The software produced by this project enabled potential transfer of digital material from a DSpace repository to a Fedora repository by providing a metadata 'bridge' between the two systems. The original workflow in the Repository Bridge project was for data to be exported from DSpace at the National Library of Wales, imported into Fedora at UW Aberystwyth, and then re-exported from Fedora to provide a feed to the EThOS project. An application profile UKETD_METS was defined to support this workflow, both to provide METS compatibility and also to enable metadata transfer between all 3 systems. This latter objective was to be achieved by including a qualified DC metadata datastream (the format for internal storage within DSpace) in parallel to the descriptive metadata in MODS (which is the DSpace default for export). This qDC metadata is in the UKETD schema¹ (UK Electronic Theses and Dissertations). Output of METS objects in this schema from the DSpace repository was

MIDESS Workpackage 9: Resource Discovery and Shared Services

achieved via a modification to the OAI-PMH harvesting routines, using the metadataPrefix parameter.

However, even when the Repository Bridge software was installed at both Birmingham and LSE, Fedora still refused to accept METS records which had been output from DSpace. It was not possible to identify the precise cause of this failure. What was remarked was that the METS objects still only included a single datastream of descriptive metadata, and that in MODS. This could be because the UKETD qDC schema was not implemented at Birmingham. Certainly the output routines failed to collect the qDC metadata which was available.

Before installing the repository bridge software, the Curator system at the University of Leeds was not able to import files from DSpace even though Curator included support for MODS. However after installation, Curator was able to request files in the UKETD_METS Schema. Comparing both the METS schema without the repository_bridge software installed and the UKETD_METS schema, the primary differences which enabled the METS file to be imported in Curator was that the UKETD_METS removed the / character and replaced it with an _ character in various ID statements e.g. within elements dmdSec, amdSec, etc. See Appendix 1a and 1b for a printout of the same record in UKETD_METS and METS schemas.

An experiment was also conducted to see if the output from the repository bridge program could be modified to hold metadata in Dublin Core instead of MODS. The MODS section was deleted and the oai_dc output from the standard oai output was copied and pasted in its place, with references to MODS changed to oai_dc. Details of the modifications are shown in Appendix 2. It was found that Curator could again import the modified file containing Dublin Core.

The only problem here was a bug in the Curator software which allowed the file to be input, but directly a user started to work with the file then this crashed the system. Endeavor were aware of this problem and it was scheduled for a bug fix in the next release (which never happened because the product was replaced by the ExLibris DigiTool product).

Tests were also carried out by Birmingham and Leeds on the METS files produced by Fedora. Fedora has the ability to output more than one metadata schema at a time. Potentially compatible systems such as Curator/DigiTool and DSpace however only have the ability to import one of these schemas at a time (usually Dublin Core). Files exported from Fedora could be directly imported into Curator via the Dublin Core metadata schema contained in the METS file. MODS files could also be ingested into Curator.

DSpace from Birmingham was only able to use the OAI_DC section of the Fedora file from LSE to create the metadata for Dspace. DSpace doesn't import a METS file directly instead the import translator creates a separate import file in DSpace's internal format based on the data held within the METS file and then fetches any files referenced in the file section. The internal item importer is then used to ingest this item into DSpace.

Further tests using DigiTool

When Leeds migrated from Curator to DigiTool, further testing was carried out on the import of METS objects. Currently within DigiTool, references to the digital object are only supported from the METS fileSec. However it is understood that, from the next service pack (DigiTool service pack number 17), it will be possible to add references from the structMap in METS as well.

The University of Leeds attempted to ingest a simple METS file created by DSpace at the University of Birmingham, containing the JPEG reproduced in Appendix 3b. It was found that the METS object required certain changes to be made to enable a successful ingest.

MIDESS Workpackage 9: Resource Discovery and Shared Services

The dc.xml file initially supplied (**dc.xml in Appendix 3a**) by the University of Birmingham was modified (**dc.xml Appendix 3b**) and then ingested into the DigiTool repository at the University of Leeds. The specific changes that were made to the original file METS file from Birmingham to convert it into a form suitable for ingest into the digital repository at Leeds (Appendix 3b) were the following:-

- Xmlns for xlinking was changed:
from `xmlns:xlink="http://www.w3.org/1999/xlink"`
to `xmlns:xlink="http://www.w3.org/TR/xlink"`
- A non-compatible character was removed in the original XML file.

Greenwoods

The character following the d in Greenwood's was a non Unicode character which needed to be removed before the file could be ingested into DigiTool.

- FileGrp USE attribute had to be changed from ORIGINAL to REFERENCE for DigiTool to accept it
- The entire amdSec element and children had to be removed for DigiTool to accept the METS file to be ingested into the repository. This element had a child element called binData that contained a 2448 character long string which DigiTool was not able to process.

The last of these changes was because DSpace appears to encode its administrative security data in base64, although the data thus encoded is in fact an encoding of English text. DigiTool does not accept data in this format and this is potentially an ongoing problem for interoperability between DSpace and DigiTool.

Once these changes had been made, the object could be successfully ingested. The ingest actually created two objects with different pid numbers, one as a part of another – the “pid” or “persistent id” is DigiTool's unique number for the object. In DigiTool's public interface, both the METS object and the JPEG image it contains are shown as separate objects, one containing the other, as illustrated in figure 7. Clicking on the JPEG results in the image being displayed in the usual manner.

Testing continued with a second, very similar, METS file, also imported from DSpace. Again two modifications had to be made to the METS file before it would ingest in DigiTool (see appendix 4a and 4b). These were identical to two of those made for the previous file - the change to the fileGrp USE attribute and removal of the amdSec – confirming this as an obstacle to interoperability. Figure 8 shows the two ingested objects as viewed in the staff interface. The METS object (pid 58174) and the image it contains (pid 58175) are discrete objects. Figure 9 shows an expanded view of the METS object, and how administrative metadata and other details of the XML object are held, including the “include” relationship to pid 58175. Figure 10 shows a similar view of the image object, with a “part_of” relationship to pid 58174.

Finally, a third, more complex object containing 2 zipped files was imported from the same source. Again, once the FileGrp attribute and the base64 data had been removed, ingest was possible. The ingest file is displayed in Figure 11 and shows the METS object (pid 58179) and the 2 zipped files (pids 58180 and 58181). Because no default delivery application is available for files with extension .zip, a question mark displays against the column headed delivery.

MIDESS Workpackage 9: Resource Discovery and Shared Services



Figure 7. METS file from DSpace ingested and displayed in DigiTool's Resource Discovery

Find: By: Admin Unit:

Search: 'ing1153', Display: 1 - 2 of 2

| # | Delivery | Metadata | PID | Label | Note | Usage Type | Entity Type | Mime Type | Created By | Creation Date | Modification Date |
|---|----------|----------|-------|-------------------|--------------------------------|------------|-------------|------------|-------------------|---------------|-------------------|
| 1 | | | 58175 | | | VIEW | | image/jpeg | creator: DIGITOOL | 07/25/2007 | 07/25/2007 |
| 2 | | | 58174 | null: DSpace Item | DSpace at the University of... | VIEW | METS | text/xml | creator: DIGITOOL | 07/25/2007 | 07/25/2007 |

Figure 8. METS file from DSpace – ingest results

MIDESS Workpackage 9: Resource Discovery and Shared Services

Pid id 58174 consisted of the following:

| null: DSpace Item (58174) | | | | | | | | | | |
|---------------------------|--|-----------------------|------|------------|---------------------|---------------------|----------------|------------|---------------|-------------|
| Control | | | | | | | | | | |
| Label: | null: DSpace Item | | | | | | | | | |
| Note: | DSpace at the University of Birmingham CUSTODIAN | | | | | | | | | |
| Entity type: | METS | | | | | | | | | |
| Usage type: | VIEW | | | | | | | | | |
| Creation date: | 2007-07-25 15:21:05 | | | | | | | | | |
| Creator: | creator:DIGITool | | | | | | | | | |
| Modification date: | 2007-07-25 15:21:05 | | | | | | | | | |
| Modified by: | creator:DIGITool | | | | | | | | | |
| Admin unit: | LUD01 | | | | | | | | | |
| Ingest ID: | Ing1153 | | | | | | | | | |
| Ingest Name: | Tom_METS_test8_250707 | | | | | | | | | |
| Preservation Level: | any | | | | | | | | | |
| Stream Ref | | | | | | | | | | |
| File name: | mets.xml | | | | | | | | | |
| File extension: | .xml | | | | | | | | | |
| Mime type: | text/xml | | | | | | | | | |
| Storage id: | 1000 | | | | | | | | | |
| External type: | -1 | | | | | | | | | |
| External type: | -1 | | | | | | | | | |
| Relations | | | | | | | | | | |
| Type | Pid | Label | Note | Usage type | Creation date | Modification date | File extension | Mime type | External type | Entity type |
| include | 58175 | | | VIEW | 2007-07-25 15:21:05 | 2007-07-25 15:21:05 | .undefined | image/jpeg | 1 | |
| Metadata | | | | | | | | | | |
| Name | Type | Mid | | | | | | | | |
| mets_section | structMap | 59386 | | | | | | | | |
| mets_section | metsHdr | 59387 | | | | | | | | |
| mets_section | fileSec | 59388 | | | | | | | | |
| mets_custom_dmd | UNSPECIFIED | 59385 | | | | | | | | |

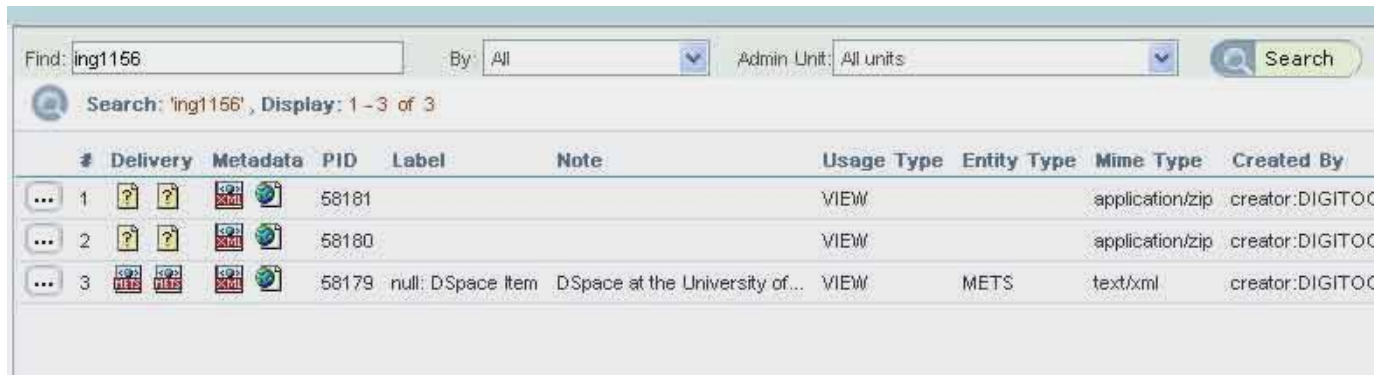
Figure 9. Details of the METS object, PID 58175. The contents of links 59386, 59387, 59388 and 59385 on this screen are specified in Appendix 4b-2

MIDESS Workpackage 9: Resource Discovery and Shared Services

| (58175) | | | | | | | | | | | |
|---------------------|-----------------------|-----------------------|--|------------|---------------------|---------------------|----------------|-----------|---------------|-------------|--|
| Control | | | | | | | | | | | |
| Usage type: | VIEW | | | | | | | | | | |
| Creation date: | 2007-07-25 15:21:05 | | | | | | | | | | |
| Creator: | creator:DIGITool | | | | | | | | | | |
| Modification date: | 2007-07-25 15:21:05 | | | | | | | | | | |
| Modified by: | creator:DIGITool | | | | | | | | | | |
| Admin unit: | LUD01 | | | | | | | | | | |
| Ingest ID: | ing1153 | | | | | | | | | | |
| Ingest Name: | Tom_MET5_test8_250707 | | | | | | | | | | |
| Preservation Level: | any | | | | | | | | | | |
| Stream Ref | | | | | | | | | | | |
| File name: | undefined | | | | | | | | | | |
| File extension: | undefined | | | | | | | | | | |
| Mime type: | image/jpeg | | | | | | | | | | |
| File id: | METSID-1 | | | | | | | | | | |
| Storage id: | -1 | | | | | | | | | | |
| External type: | 1 | | | | | | | | | | |
| External type: | 1 | | | | | | | | | | |
| Relations | | | | | | | | | | | |
| Type | Pid | Label | Note | Usage type | Creation date | Modification date | File extension | Mime type | External type | Entity type | |
| part_of | 58174 | null: DSpace Item | DSpace at the University of Birmingham CUSTODIAN | VIEW | 2007-07-25 15:21:05 | 2007-07-25 15:21:05 | .xml | text/xml | -1 | METS | |
| Metadata | | | | | | | | | | | |
| Name | Type | Mid | | | | | | | | | |
| mets_custom_dmd | UNSPECIFIED | 59385 | | | | | | | | | |

Figure 10. Details of the digital image object, PID 58175. The contents of link [59395](#) on this screen are .specified in Appendix 4b-1.

MIDESS Workpackage 9: Resource Discovery and Shared Services



Find: By: Admin Unit:

Search: 'ing1156', Display: 1 - 3 of 3

| # | Delivery | Metadata | PID | Label | Note | Usage Type | Entity Type | Mime Type | Created By |
|-----|----------|----------|-------|-------------------|--------------------------------|------------|-------------|-----------------|------------------|
| ... | 1 | | 58181 | | | VIEW | | application/zip | creator: DIGITOO |
| ... | 2 | | 58180 | | | VIEW | | application/zip | creator: DIGITOO |
| ... | 3 | | 58179 | null: DSpace Item | DSpace at the University of... | VIEW | METS | text/xml | creator: DIGITOO |

Figure 11. METS file from DSpace containing multiple digital objects – ingest results

Current Developments in Improving Digital Repository Interoperability

Although metadata within Digital Repositories can be effectively searched and harvested through techniques such as OAI-PMH, SRU/SRW etc., there are no current widely adopted standards for exchanging/transferring the actual digital objects (datastreams) attached to the metadata (images/video/sound etc). This is however being addressed, and foremost amongst those addressing this issue is Herbert Van de Sompel working at the Los Alamos National Library in the USA on the OAI-ORE project.¹ Other interested parties include Microsoft, the Andrew Mellon Foundation, the coalition for Networked Information, the Digital Library Federation and JISC.

One of the major objectives is to develop standards that can be used for the interoperability of digital objects between digital repositories. The current problem is that across the digital community, there is a different perspective on the design and management of digital repositories. Repository systems use different software platforms, serve different communities and have different policies for deposit, storage, preservation etc resulting in a lack of cross-repository interoperability. Such standards would help the MIDESS project enormously if they were implemented for all digital repository software, since interoperability would then be available across the three MIDESS partners who are all using different digital repository software (Birmingham-Dspace, LSE-Fedora, Leeds-Curator (and then DigiTool).

The stated goal of OAI-ORE is:-

To reach a common understanding regarding a data model and a limited set of core protocol-based repository interfaces that would allow services and downstream applications to interact with heterogeneous repositories in a consistent manner. Such repository interfaces include interfaces that support locating, identifying, harvesting, obtaining and depositing compound digital objects.

This would imply the development of a cross repository interoperability layer and the support for a data model for digital objects that is supported across repositories.

Sharing video content under the ERA licence

Another strand of content sharing within the MIDESS project focused on a collection of off-air recordings held in the UCL School of Slavonic and East European Studies (SSEES) Library. These were recorded under the ERA licence and many of them relate to the Cold War – a topic which is relevant to courses taught at LSE. Since Workpackage 7 had reported on IPR issues for multimedia content, it was decided to explore how this material – once digitised – could be made available to staff and students within the two institutions via their respective repositories. The LSE Fedora system already held a substantial quantity of similar material – including some recorded under the ERA licence - as part of the CLT ingest. UCL were in the process of implementing a DigiTool repository, though in all other respects outside of the framework of the MIDESS project.

Discussions between library and academic staff in both institutions confirmed that the programmes on the Cold War were indeed of interest and a small selection of 66 recordings were identified for digitisation. This was to be undertaken at SSEES using a Canopus A-D converter, Adobe Premiere Pro CS3 and Autodesk Cleaner XL to produce files in Windows Media format.

MIDESS Workpackage 9: Resource Discovery and Shared Services

Metadata describing each programme existed in a listing on the SSEES web-site; although quite detailed, it was held as flat HTML. It was decided to manually convert the relevant records to qualified Dublin Core, following the same approach as had been established for the CLT database, broadly based on the EBU Core application profile (see Workpackage 4 for a detailed discussion). However it was possible to revise one or two of the decisions which had been dictated by the structure of the original CLT database and the need to accommodate that data within the qDC schema.

The work undertaken by MIDESS on IPR also resulted in a very clear vision of what rights information needed to be stored, and how. The ERA licence includes the following instructions:

Always label or mark up the cassette or jewel case plus the tape or disc itself with the four following items:

- *Date (when the recording was made)*
- *Name of the broadcaster*
- *Programme title*
- *The wording 'This recording is to be used only for educational and non-commercial purposes under the terms of the ERA Licence'.*

Recordings stored digitally on a server should include the above four items as a written opening credit or webpage which must be viewed or listened to before access to the recording is permitted.

Delivery at LSE will be from a streamed server, via an intermediate webpage (see figure 12) which displays the information specified by the ERA licence and to which the repository record refers.

The four required data elements were also coded into the metadata as follows:

- Date of recording > date.issued (dcterms:issued)
- Name of the broadcaster > audience.mediator (dcterms:mediator)
- Programme title > title (dc:title)
- Licence statement > rights (dc:rights)

Additionally, a link to the ERA licence online was included in the rights.license element (dcterms:license). Although it would have been possible to be more precise and use dcterms:accessRights for the ERA licence statement, it was felt preferable to use the unqualified dc:rights element instead as this would avoid any risk of the statement disappearing if repository software failed to transform or display qualified Dublin Core correctly – a significant risk given the experience of the MIDESS project!

MIDESS Workpackage 9: Resource Discovery and Shared Services

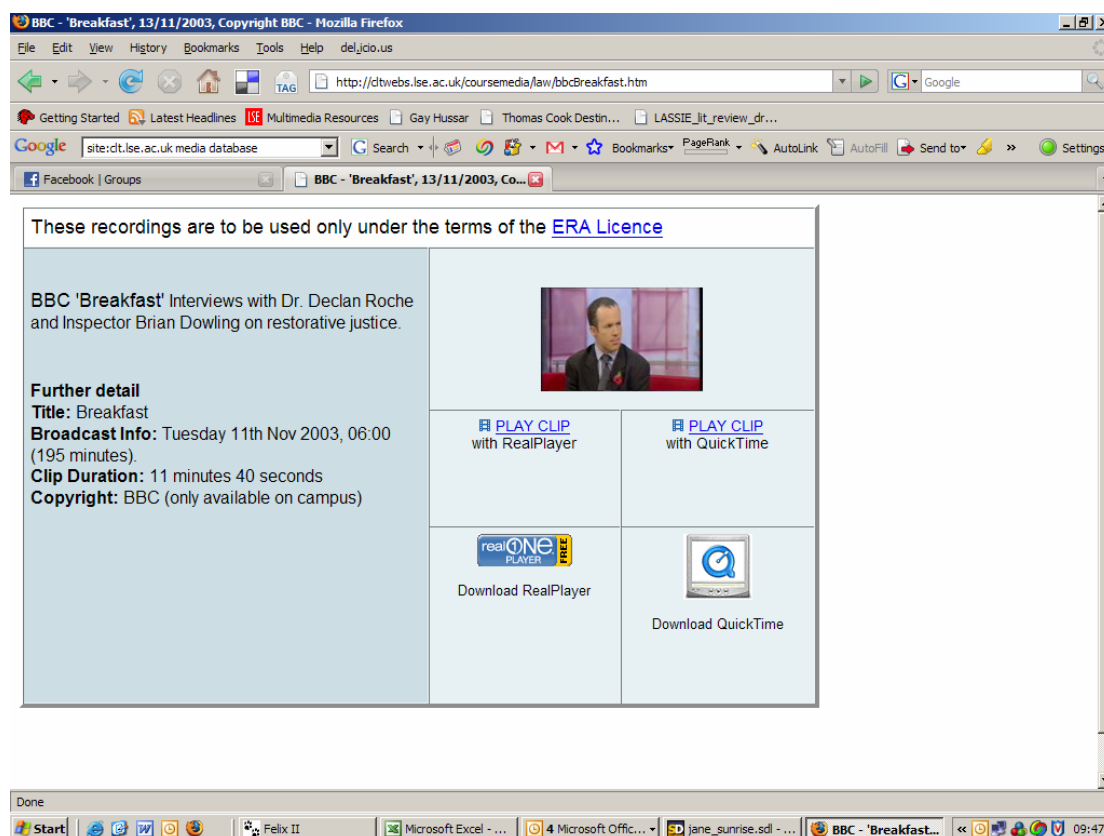


Fig.12 Typical LSE credit page for an item recorded under the ERA Licence

More depressing were the conclusions drawn on precisely how the digital files could be shared between the two institutions. The ERA licence permits the digitisation of analogue recordings and “electronic communication of licensed recordings within an educational establishment” though this only extends to “on-site” terminals (this restriction is lifted from August 2007 for institutions subscribing to the ERA Plus licence). It also permits an establishment to copy a recording for another licensee. These terms clearly excluded delivery of a recording held in the UCL repository to any LSE staff or students, even with strict authentication and authorisation mechanisms. However the licence also explicitly forbids online delivery or electronic transmission of a recording to another institution, even if this is only to enable the second institution to mount the recording locally. The regrettable conclusion was that, once the recordings had been digitised, they had to be written onto DVDs and manually transferred to LSE for loading into the FEDORA repository. Work has therefore progressed on this basis.

Medical recordings and liaison with the CHERRI Project

Working with material recorded under the ERA licence illustrates one type of rights and access issue which has been encountered within the MIDESS project. Liaison with the ALPS project and other collections from the medical /healthcare area has equally demonstrated the importance of access and use restrictions for reasons of privacy and data protection. In order to explore further this second strand, the project has been in active liaison with the CHERRI (Common Healthcare Educational Records Reusability Infrastructure) Project, which is currently examining the issues when a digital recording (such as an image or video) is taken of a patient by a medical physician.

MIDESS Workpackage 9: Resource Discovery and Shared Services

For digital material requiring explicit consent by the patient, CHERRI recommends that the use of such material should be potentially tracked via the use of a shared global unique identifier (GUID) which encodes the identify of the patient (and potentially any copyright information) and is permanently linked to the digital object. The GUID could then be maintained securely by the healthcare provider with an expectation that medical staff viewing the material understand the conditions under which the digital material can be used. The absence of this attached information could result in the digital material being used in unauthorised ways without the consent of the original patient.

CHERRI also suggests for all circumstances visual tagging of the materials and associating such materials with a license which describes the copyright and usage conditions. This attempts to protect the interests of patients by ensuring that whenever the digital material is to subsequently used, there is also a reminder to the user of the specific licence conditions. CHERRI suggests this could be achieved through a title frame or sequence and/or incorporation of rights metadata somewhat similar to the recommendations of the ERA for off-air recordings.

Although the MIDESS repositories support persistent URL's, there is currently no support for the GUID identifier methods (although this could potentially be added). The latest version of DigiTool (3.5) supports integration with other external persistent identifier methods such as DOI (<http://www.doi.org/>) and URN (<http://www.cni.org/tfms/1998a.spring/handout/LDaigle98Stf.html>)

In order for digital repositories such as DigiTool, DSpace and Fedora to support GUID, there needs to be acceptance of this identifier method and its widespread adoption by the repository community. Mandatory adoption of the GUID identifier method by influential bodies within the medical community (such as the General Medical Council or Medical Research Council) and insistence of its use in repositories containing medical information would encourage repository suppliers to implement GUID as standard functionality.

In the meanwhile, the work of CHERRI can certainly inform the processes associated with the ingest of clinical recordings into repositories in the MIDESS institutions and, in particular, care can be taken to associate appropriate rights metadata where this is available.

Conclusions

Discussions among the project partners highlighted the shared concerns that each institution has to establish an information architecture which will allow 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 have been identified, but these should be considered within the context of broader information provision. Various protocols such as Z39.50, SRU/SRW and OAI-PMH can be used to enable cross-platform resource discovery. Even so, connecting services can at present be difficult or 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 necessary functionality for such integration. Such developments will need to encompass a broad range of issues, including authentication and authorisation.

Resource discovery is however only the first stage of the process: providing access – including cross-institutional access - to the digital objects stored in repositories is the real objective. Here the experience of the MIDESS project has been less satisfactory. Initially, METS was seen by the project partners as a standard which would permit the transmission of digital objects between systems, allowing a Delivery Information Package to be exported from one repository in order to be ingested into a second repository as a Submission Information Package. However in reality, the METS schema is so permissive in its approach that it cannot enable such interoperability at the present time. The MIDESS project had access to 4

MIDESS Workpackage 9: Resource Discovery and Shared Services

different repository platforms, but it proved very difficult to achieve the smooth transfer of a METS object without significant manual intervention and editing, even after the Repository Bridge software had been installed.

In contrast, LSE were able to work very successfully with METS to achieve bulk ingest of objects into their Fedora system. In this instance, the METS structure could be tailored to the needs of Fedora without any concerns for interoperability. The conclusion to be drawn is that METS simply provides an XML envelope within which a digital object, its metadata and its structure can be wrapped: interoperability depends on some degree of commonality in the contents as well as in the wrapper. Implementation of appropriate application profiles could provide a solution to this issue.

Within the project, issues around commonality of metadata structures particularly came to the fore and these seemed to focus on two key questions:

- Can the repository system ingesting material cope with the specified metadata schema? This encompasses diverse issues including:
 - support for the schema itself (Digitool did not initially support MODS)
 - version (Digitool still only supports MODS 2.0)
 - local implementations (the Birmingham coin collection created local extensions to qualified Dublin Core which were not exported by DSpace and would not have been easily ingested into another repository)
- Can the repository system correctly interpret the contents of the specified metadata schema and exploit it for resource discovery, etc.? An example of this is from Fedora, where it was possible to ingest and store metadata in qDC or EAD but a parallel DC metadata stream was required for indexing purposes.

The project worked on many different collections, identifying appropriate metadata schemas for each. The limitations of Dublin Core – even qualified DC - were very apparent, and yet DC proved the most reliable vehicle in resource sharing.

What is true for descriptive metadata is even more true for administrative and rights metadata, where standards are generally less developed. There is a clear need to find ways of embedding such metadata within the repository structures, not only for storage and explicit retrieval but also to control access to resources and/or to force display of pertinent warnings. A simple example of this is the restriction of ERA recordings to on-campus users. LSE is able to block access to these recordings from off-campus but there is no easy and flexible way for this to be explained to users of the Fedora repository. If the information were to be included explicitly in the dc:rights element, and LSE subsequently signed up to the ERA Plus licence which does permit off-campus access, then there would be no easy way of updating this data in several hundred metadata records. Similar concerns apply to privacy and data protection information for recordings made within a clinical context.

While some of the issues identified within the MIDESS Project are likely to be resolved as more experience is gained in implementing and integrating repository systems and as the software itself evolves in capability and sophistication, there is a definite need for interoperability standards to be developed and agreed in this area. METS, in particular, provides little more than a broad structure for packaging digital objects, complete with their metadata. Significant investment is required in developing a range of application profiles which will permit 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.

References

1. OAI-ORE Project <http://www.openarchives.org/ore/>
2. Copeland, S., Penman, A and Milne, R (2005), "Electronic theses: the turning point" , Program, Vol 39 No.3 pp 185-197
3. CLT Media Database to Fedora Ingest. Marie Lagerwall 2007
http://www.leeds.ac.uk/library/midess/CLT_media_database_to_Fedora_ingest.pdf
4. http://www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories/project_repository_bridge.aspx
5. <http://cadair.aber.ac.uk/dspace/bitstream/2160/203/1/Using+OAI-PMH+and+METS.pdf>
6. <http://www.cherri.mvm.ed.ac.uk/cherri.pdf>

Appendix 1

Appendix 1a. UKETD_METS

```

<?xml version="1.0" encoding="UTF-8" ?>
: <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
  http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2007-06-21T13:57:03Z</responseDate>
  <request metadataPrefix="uketd_mets" verb="GetRecord"
    identifier="oai:midess.bham.ac.uk:123456789/171">http://midess.bham.ac
    .uk:8080/dspace-oai/request</request>
: <GetRecord>
: <record>
: <header>
  <identifier>oai:midess.bham.ac.uk:123456789/171</identifier>
  <datestamp>2007-01-18T11:40:33Z</datestamp>
  <setSpec>hdl_123456789_167</setSpec>
  </header>
: <metadata>
: <mets OBJID="hdl:123456789/171" LABEL="DSpace Item"
  xmlns="http://www.loc.gov/METS/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3"
  xsi:schemaLocation="http://www.loc.gov/METS/
  http://www.loc.gov/standards/mets/mets.xsd
  http://www.loc.gov/mods/v3
  http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
: <metsHdr CREATEDATE="2007-06-21T14:57:03">
: <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
  <name>DSpace at the University of Birmingham</name>
  </agent>
  </metsHdr>
: <dmdSec ID="DMD_hdl_123456789_171_mods">
: <mdWrap MDTYPE="MODS">
: <xmlData>
: <mods:name>
: <mods:role>
  <mods:roleTerm type="text">author</mods:roleTerm>
  </mods:role>
  <mods:namePart>Slater, Terry R.</mods:namePart>

```

```

</mods:name>
= <mods:extension>
  <mods:dateAccessioned encoding="iso8601">2007-01-
    18T11:40:33Z</mods:dateAccessioned>
  </mods:extension>
= <mods:extension>
  <mods:dateAvailable encoding="iso8601">2007-01-
    18T11:40:33Z</mods:dateAvailable>
  </mods:extension>
= <mods:originInfo>
  <mods:dateIssued encoding="iso8601">2007-01-
    18T11:40:33Z</mods:dateIssued>
  </mods:originInfo>
  <mods:identifier
    type="uri">http://midess.bham.ac.uk:8080/dspace/handle/123456789/171
  </mods:identifier>
  <mods:note>image copyright 2003 Terry R. Slater</mods:note>
  <mods:abstract>This house is dated to 1425 by dendro-chronology. It stands
    in the Begijnhof (the nun's square - hence the modern statue) on the
    edge of the medieval city. It is listed for conservation and is now used as
    a restaurant. Photograph, May 2003.</mods:abstract>
  <mods:note type="provenance">Submitted by Adrian Dover
    (a.i.dover@bham.ac.uk) on 2007-01-18T11:40:13Z No. of bitstreams: 1
    amsterdam_15.jpg: 867654 bytes, checksum:
    f35b7504297aa263c819680532eccfd9 (MD5)</mods:note>
  <mods:note type="provenance">Approved for entry into archive by Adrian
    Dover(a.i.dover@bham.ac.uk) on 2007-01-18T11:40:33Z (GMT) No. of
    bitstreams: 1 amsterdam_15.jpg: 867654 bytes, checksum:
    f35b7504297aa263c819680532eccfd9 (MD5)</mods:note>
  <mods:note type="provenance">Made available in DSpace on 2007-01-
    18T11:40:33Z (GMT). No. of bitstreams: 1 amsterdam_15.jpg: 867654
    bytes, checksum: f35b7504297aa263c819680532eccfd9
    (MD5)</mods:note>
= <mods:physicalDescription>
  <mods:extent>867654 bytes</mods:extent>
  </mods:physicalDescription>
= <mods:physicalDescription>
  <mods:internetMediaType>image/jpeg</mods:internetMediaType>
  </mods:physicalDescription>
= <mods:subject>
  <mods:topic>Begijnhof</mods:topic>
  </mods:subject>
= <mods:subject>
  <mods:topic>restaurant</mods:topic>

```

```

    </mods:subject>
  : <mods:subject>
    <mods:topic>1425</mods:topic>
    </mods:subject>
    <mods:titleInfo>Het Houten Huys, Amsterdam</mods:titleInfo>
    <mods:genre>Image</mods:genre>
    </xmlData>
  </mdWrap>
  </dmdSec>
  : <amdSec ID="TMD_hdl_123456789_171">
  : <rightsMD ID="rights_123456789_171_bin">
  : <mdWrap MIMETYPE="text/plain" MDTYPE="OTHER"
    OTHERMDTYPE="TEXT">
    <binData>TGijZW5zZSBncmFudGVklGJ5IEFkcmlhbiBEb3ZlciAoYS5sLmRvd
      mVyQGJoYW0uYWMudWspYWx0ZXJhdGlvbWw3RoZXlkdGhhbiBhcyB
      hbGxvd2VklGJ5IHRoaXMkbGljZW5zZSwgdG8geW91
      ciBzdWJtaXNzaW9uLgo
    </binData>
    </mdWrap>
  </rightsMD>
  </amdSec>
  : <fileSec>
  : <fileGrp USE="ORIGINAL">
  : <file ID="f123456789_171_1" MIMETYPE="image/jpeg" SIZE="867654"
    CHECKSUM="f35b7504297aa263c819680532eccfd9"
    CHECKSUMTYPE="MD5"
    OWNERID="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/
    171/1/amsterdam_15.jpg" GROUPIP="GROUP_f123456789_171_1">
    <FLocat LOCTYPE="URL" xlink:type="simple"
      xlink:href="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/1
      71/1/amsterdam_15.jpg" />
    </file>
  </fileGrp>
  </fileSec>
  : <structMap>
  : <div DMDID="DMD_hdl_123456789_171_mods"
    ADMID="TMD_hdl_123456789_171 rights_123456789_171_bin">
  <fptr FILEID="f123456789_171_1" />
  </div>
  </structMap>
  </mets>
  </metadata>
  </record>

```

MIDESS Workpackage 9: Resource Discovery and Shared Services

</GetRecord>

</OAI-PMH>

Appendix 1b. METS

```

<?xml version="1.0" encoding="UTF-8" ?>
: <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2007-06-21T14:00:21Z</responseDate>
  <request metadataPrefix="mets" verb="GetRecord"
    identifier="oai:midess.bham.ac.uk:123456789/171">http://midess.bham.ac
    .uk:8080/dspace-oai/request</request>
: <GetRecord>
: <record>
: <header>
  <identifier>oai:midess.bham.ac.uk:123456789/171</identifier>
  <datestamp>2007-01-18T11:40:33Z</datestamp>
  <setSpec>hdl_123456789_167</setSpec>
  </header>
: <metadata>
: <mets OBJID="hdl:123456789/171" LABEL="DSpace Item"
  xmlns="http://www.loc.gov/METS/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3"
  xsi:schemaLocation="http://www.loc.gov/METS/
    http://www.loc.gov/standards/mets/mets.xsd
    http://www.loc.gov/mods/v3
    http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
: <metsHdr CREATEDATE="2007-06-21T15:00:21">
: <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
  <name>DSpace at the University of Birmingham</name>
  </agent>
  </metsHdr>
: <dmdSec ID="DMD_hdl_123456789/171">
: <mdWrap MDTYPE="MODS">
: <xmlData>
: <mods:name>
: <mods:role>
  <mods:roleTerm type="text">author</mods:roleTerm>
  </mods:role>
  <mods:namePart>Slater, Terry R.</mods:namePart>
  </mods:name>
: <mods:extension>

```

```

<mods:dateAccessioned encoding="iso8601">2007-01-
  18T11:40:33Z</mods:dateAccessioned>
  </mods:extension>
= <mods:extension>
  <mods:dateAvailable encoding="iso8601">2007-01-
    18T11:40:33Z</mods:dateAvailable>
    </mods:extension>
= <mods:originInfo>
  <mods:dateIssued encoding="iso8601">2007-01-
    18T11:40:33Z</mods:dateIssued>
    </mods:originInfo>
<mods:identifier
  type="uri">http://midess.bham.ac.uk:8080/dspace/handle/123456789/171
  </mods:identifier>
<mods:note>image copyright 2003 Terry R. Slater</mods:note>
<mods:abstract>This house is dated to 1425 by dendro-chronology. It stands
  in the Begijnhof (the nun's square - hence the modern statue) on the
  edge of the medieval city. It is listed for conservation and is now used as
  a restaurant. Photograph, May 2003.</mods:abstract>
<mods:note type="provenance">Submitted by Adrian Dover
  (a.l.dover@bham.ac.uk) on 2007-01-18T11:40:13Z No. of bitstreams: 1
  amsterdam_15.jpg: 867654 bytes, checksum:
  f35b7504297aa263c819680532eccfd9 (MD5)</mods:note>
<mods:note type="provenance">Approved for entry into archive by Adrian
  Dover(a.l.dover@bham.ac.uk) on 2007-01-18T11:40:33Z (GMT) No. of
  bitstreams: 1 amsterdam_15.jpg: 867654 bytes, checksum:
  f35b7504297aa263c819680532eccfd9 (MD5)</mods:note>
<mods:note type="provenance">Made available in DSpace on 2007-01-
  18T11:40:33Z (GMT). No. of bitstreams: 1 amsterdam_15.jpg: 867654
  bytes, checksum: f35b7504297aa263c819680532eccfd9
  (MD5)</mods:note>
= <mods:physicalDescription>
  <mods:extent>867654 bytes</mods:extent>
  </mods:physicalDescription>
= <mods:physicalDescription>
  <mods:internetMediaType>image/jpeg</mods:internetMediaType>
  </mods:physicalDescription>
= <mods:subject>
  <mods:topic>Begijnhof</mods:topic>
  </mods:subject>
= <mods:subject>
  <mods:topic>restaurant</mods:topic>
  </mods:subject>
= <mods:subject>

```

```

<mods:topic>1425</mods:topic>
  </mods:subject>
<mods:titleInfo>Het Houten Huys, Amsterdam</mods:titleInfo>
<mods:genre>Image</mods:genre>
  </xmlData>
  </mdWrap>
  </dmdSec>
= <amdSec ID="TMD_hdl_123456789/171">
= <rightsMD>
= <mdWrap MIMETYPE="text/plain" MDTYPE="OTHER"
  OTHERMDTYPE="TEXT">
<binData>
  TGIjZW5zZSBncmFudGVkIGJ5IEFkcmIhbiBEb3ZlciAoYS5sLmRvdmVyQ
  GJoYW0uYWMudWspYWx0ZXJhdGlvbWw3RoZlIgdGhhbiBhcyBhbGxv
  d2VklGJ5IHRoaXMkbGljZW5zZSwgdG8geW91 ciBzdWJtaXNzaW9uLgo
  </binData>
  </mdWrap>
  </rightsMD>
  </amdSec>
= <fileSec>
= <fileGrp USE="ORIGINAL">
= <file ID="123456789_171_1" MIMETYPE="image/jpeg" SIZE="867654"
  CHECKSUM="f35b7504297aa263c819680532eccfd9"
  CHECKSUMTYPE="MD5"
  OWNERID="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/
  171/1/amsterdam_15.jpg" GROUPIP="GROUP_123456789_171_1">
<FLocat LOCTYPE="URL" xlink:type="simple"
  xlink:href="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/1
  71/1/amsterdam_15.jpg" />
  </file>
  </fileGrp>
  </fileSec>
= <structMap>
<div />
  </structMap>
  </mets>
  </metadata>
  </record>
  </GetRecord>
  </OAI-PMH>

```

Appendix 2

2a. XML data from Repository Bridge in MODS format

```

<xmlData>
<mods:name>
<mods:role>
  <mods:roleTerm type="text">author</mods:roleTerm>
  </mods:role>
  <mods:namePart>Slater, Terry R.</mods:namePart>
  </mods:name>
- <mods:extension>
  <mods:dateAccessioned encoding="iso8601">2007-01-18T11:40:33Z</mods:dateAccessioned>
  </mods:extension>
- <mods:extension>
  <mods:dateAvailable encoding="iso8601">2007-01-18T11:40:33Z</mods:dateAvailable>
  </mods:extension>
- <mods:originInfo>
  <mods:dateIssued encoding="iso8601">2007-01-18T11:40:33Z</mods:dateIssued>
  </mods:originInfo>
<mods:identifier
  type="uri">http://midess.bham.ac.uk:8080/dspace/handle/123456789/171</mods:identifier>
<mods:note>image copyright 2003 Terry R. Slater</mods:note>
<mods:abstract>This house is dated to 1425 by dendro-chronology. It stands in the Begijnhof (the
  nun's square - hence the modern statue) on the edge of the medieval city. It is listed for
  conservation and is now used as a restaurant. Photograph, May 2003.</mods:abstract>
<mods:note type="provenance">Submitted by Adrian Dover (a.l.dover@bham.ac.uk) on 2007-01-
  18T11:40:13Z No. of bitstreams: 1 amsterdam_15.jpg: 867654 bytes, checksum:
  f35b7504297aa263c819680532eccfd9 (MD5)</mods:note>
<mods:note type="provenance">Approved for entry into archive by Adrian
  Dover(a.l.dover@bham.ac.uk) on 2007-01-18T11:40:33Z (GMT) No. of bitstreams: 1
  amsterdam_15.jpg: 867654 bytes, checksum: f35b7504297aa263c819680532eccfd9
  (MD5)</mods:note>
<mods:note type="provenance">Made available in DSpace on 2007-01-18T11:40:33Z (GMT). No.
  of bitstreams: 1 amsterdam_15.jpg: 867654 bytes, checksum:
  f35b7504297aa263c819680532eccfd9 (MD5)</mods:note>
- <mods:physicalDescription>
  <mods:extent>867654 bytes</mods:extent>
  </mods:physicalDescription>
- <mods:physicalDescription>
  <mods:internetMediaType>image/jpeg</mods:internetMediaType>
  </mods:physicalDescription>
- <mods:subject>
  <mods:topic>Begijnhof</mods:topic>
  </mods:subject>
- <mods:subject>
  <mods:topic>restaurant</mods:topic>
  </mods:subject>
- <mods:subject>
  <mods:topic>1425</mods:topic>
  </mods:subject>
<mods:titleInfo>Het Houten Huys, Amsterdam</mods:titleInfo>
<mods:genre>Image</mods:genre>
</xmlData>

```

2b. XML data from Repository Bridge changed to Dublin Core format

```
<xmlData>
  <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
    <dc:creator>Slater, Terry R.</dc:creator>
    <dc:date>2007-01-18T11:40:33Z</dc:date>
    <dc:date>2007-01-18T11:40:33Z</dc:date>
    <dc:date>2007-01-18T11:40:33Z</dc:date>
    <dc:identifier>
      http://midess.bham.ac.uk:8080/dspace/handle/123456789/171
    </dc:identifier>
    <dc:description>image copyright 2003 Terry R. Slater</dc:description>
    <dc:description>
      This house is dated to 1425 by dendro-chronology. It stands in the Begijnhof (the nun's square -
        hence the modern statue) on the edge of the medieval city. It is listed for conservation and is
        now used as a restaurant.
      Photograph, May 2003.
    </dc:description>
    <dc:format>867654 bytes</dc:format>
    <dc:format>image/jpeg</dc:format>
    <dc:subject>Begijnhof</dc:subject>
    <dc:subject>restaurant</dc:subject>
    <dc:subject>1425</dc:subject>
    <dc:title>Het Houten Huys, Amsterdam</dc:title>
    <dc:type>Image</dc:type>
  </oai_dc:dc>
</xmlData>
```

Appendix 3

3a. dc.xml file supplied by Birmingham exported as METS from DSpace.

```
<?xml version="1.0" encoding="UTF-8" ?>
<mets OBJID="hdl:123456789/169" LABEL="DSpace Item"
xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:mods="http://www.loc.gov/mods/v3" xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
  <metsHdr CREATEDATE="2007-04-13T16:12:49">
    <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
      <name>DSpace at the University of Birmingham</name>
    </agent>
  </metsHdr>

  <dmdSec ID="DMD_hdl_123456789_169_mods">
    <mdWrap MDTYPE="MODS">
      <xmlData><mods:name><mods:role><mods:roleTerm
type="text">author</mods:roleTerm></mods:role><mods:namePart>Greenwood, C. and
J.</mods:namePart></mods:name>
<mods:extension><mods:dateAccessioned encoding="iso8601">2007-01-
18T11:27:34Z</mods:dateAccessioned></mods:extension>
<mods:extension><mods:dateAvailable encoding="iso8601">2007-01-
18T11:27:34Z</mods:dateAvailable></mods:extension>
<mods:originInfo><mods:dateIssued encoding="iso8601">2007-01-
18T11:27:34Z</mods:dateIssued></mods:originInfo>
<mods:identifier
type="uri">http://midess.bham.ac.uk:8080/dspace/handle/123456789/169</mods:identifier>
<mods:abstract>Part of C. and J. Greenwoods Map of Warwickshire, surveyed 1820-22,
published 1822 at a scale of one inch to one mile. This was the first county survey to show
parish boundaries. This section of the map, south of Stratford-upon-Avon is especially
colourful since Hundreds and County divisions are shown in different
shades.</mods:abstract>

<mods:note type="provenance">Submitted by Adrian Dover (a.l.dover@bham.ac.uk) on
2007-01-18T10:53:18Z
No. of bitstreams: 1
greenwood_stratford_boundaries.jpg: 1503047 bytes, checksum:
2e0806b7a300dc3b405968b9be145562 (MD5)</mods:note>
<mods:note type="provenance">Approved for entry into archive by Adrian
Dover(a.l.dover@bham.ac.uk) on 2007-01-18T11:27:34Z (GMT) No. of bitstreams: 1
greenwood_stratford_boundaries.jpg: 1503047 bytes, checksum:
2e0806b7a300dc3b405968b9be145562 (MD5)</mods:note>
<mods:note type="provenance">Made available in DSpace on 2007-01-18T11:27:34Z (GMT).
No. of bitstreams: 1
greenwood_stratford_boundaries.jpg: 1503047 bytes, checksum:
2e0806b7a300dc3b405968b9be145562 (MD5)</mods:note>
<mods:note type="sponsorship">University of Birmingham. School of Geography, Earth
& amp; Environmental Sciences.</mods:note>
```


MIDESS Workpackage 9: Resource Discovery and Shared Services

```
</amdSec>
<fileSec>
  <fileGrp USE="ORIGINAL">
    <file ID="f123456789_169_1" MIMETYPE="image/jpeg" SIZE="1503047"
CHECKSUM="2e0806b7a300dc3b405968b9be145562" CHECKSUMTYPE="MD5"
OWNERID="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/169/1/greenwood_
stratford_boundaries.jpg" GROUPID="GROUP_f123456789_169_1">

      <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/169/1/greenwood_str
atford_boundaries.jpg"/>
    </file>
  </fileGrp>
</fileSec>
<structMap>
  <div DMDID="DMD_hdl_123456789_169_mods" ADMID="TMD_hdl_123456789_169
rights_123456789_169_bin">
    <fptr FILEID="f123456789_169_1"/>
  </div>
</structMap>

</mets>
```

3b. Modified dc.xml file ingested into DigiTool

```
<?xml version="1.0" encoding="UTF-8" ?><mets OBJID="hdl:123456789/169"
LABEL="DSpace Item" xmlns="http://www.loc.gov/METS/"
xmlns:xlink="http://www.w3.org/TR/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:mods="http://www.loc.gov/mods/v3"
xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd
http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
<metsHdr CREATEDATE="2007-04-13T16:12:49">
  <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
    <name>DSpace at the University of Birmingham</name>
  </agent>
</metsHdr>
<dmdSec ID="DMD_hdl_123456789_169_mods">
  <mdWrap MDTYPE="MODS">
    <xmlData><mods:name><mods:role><mods:roleTerm
type="text">author</mods:roleTerm></mods:role><mods:namePart>Greenwood, C. and
J.</mods:namePart></mods:name>
<mods:extension><mods:dateAccessioned encoding="iso8601">2007-01-
18T11:27:34Z</mods:dateAccessioned></mods:extension>
<mods:extension><mods:dateAvailable encoding="iso8601">2007-01-
18T11:27:34Z</mods:dateAvailable></mods:extension>
<mods:originInfo><mods:dateIssued encoding="iso8601">2007-01-
18T11:27:34Z</mods:dateIssued></mods:originInfo>
<mods:identifier
type="uri">http://midess.bham.ac.uk:8080/dspace/handle/123456789/169</mods:identifier>
<mods:abstract>Part of C. and J. Greenwood's Map of Warwickshire, surveyed 1820-22,
published 1822 at a scale of one inch to one mile. This was the first county survey to show
parish boundaries. This section of the map, south of Stratford-upon-Avon is especially
colourful since Hundreds and County divisions are shown in different
shades.</mods:abstract>
<mods:note type="provenance">Submitted by Adrian Dover (a.l.dover@bham.ac.uk) on
2007-01-18T10:53:18Z
No. of bitstreams: 1
greenwood_stratford_boundaries.jpg: 1503047 bytes, checksum:
2e0806b7a300dc3b405968b9be145562 (MD5)</mods:note>
<mods:note type="provenance">Approved for entry into archive by Adrian
Dover(a.l.dover@bham.ac.uk) on 2007-01-18T11:27:34Z (GMT) No. of bitstreams: 1
greenwood_stratford_boundaries.jpg: 1503047 bytes, checksum:
2e0806b7a300dc3b405968b9be145562 (MD5)</mods:note>
<mods:note type="provenance">Made available in DSpace on 2007-01-18T11:27:34Z (GMT).
No. of bitstreams: 1
greenwood_stratford_boundaries.jpg: 1503047 bytes, checksum:
2e0806b7a300dc3b405968b9be145562 (MD5)</mods:note>
<mods:note type="sponsorship">University of Birmingham. School of Geography, Earth
&amp; Environmental Sciences.</mods:note>
<mods:physicalDescription><mods:extent>1503047
bytes</mods:extent></mods:physicalDescription>
<mods:physicalDescription><mods:internetMediaType>image/jpeg</mods:internetMediaType
></mods:physicalDescription>
<mods:subject><mods:topic>england</mods:topic></mods:subject>
```

MIDESS Workpackage 9: Resource Discovery and Shared Services

```
<mods:subject><mods:topic>county</mods:topic></mods:subject>
<mods:subject><mods:topic>hundreds</mods:topic></mods:subject>
<mods:subject><mods:topic>parishes</mods:topic></mods:subject>
<mods:titleInfo>Map of Warwickshire, surveyed 1820-22</mods:titleInfo>
<mods:genre>Map</mods:genre>
</xmlData>
</mdWrap>
</dmdSec>
<fileSec>
  <fileGrp USE="REFERENCE">
    <file ID="f123456789_169_1" MIMETYPE="image/jpeg" SIZE="1503047"
CHECKSUM="2e0806b7a300dc3b405968b9be145562" CHECKSUMTYPE="MD5"
OWNERID="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/169/1/greenwood_
stratford_boundaries.jpg" GROUPID="GROUP_f123456789_169_1">
      <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/169/1/greenwood_str
atford_boundaries.jpg"/>
    </file>
  </fileGrp>
</fileSec>
<structMap>
  <div DMDID="DMD_hdl_123456789_169_mods" ADMID="TMD_hdl_123456789_169
rights_123456789_169_bin">
    <fptr FILEID="f123456789_169_1"/>
  </div>
</structMap>
</mets>
```

However while the JPG graphic imported

MIDESS Workpackage 9: Resource Discovery and Shared Services



the METS files produced a Servlet error below

Servlet Error

Error

```
java.lang.NullPointerException
```

Stack Trace

```
java.lang.NullPointerException at  
com.exlibris.digitool.viewers.metsviewer.archrepos.ArchObjGeneratorMetsJd.processMDWrap(ArchObjGeneratorMet  
sJd.java:430) at  
com.exlibris.digitool.viewers.metsviewer.archrepos.ArchObjGeneratorMetsJd.processDmdSec(ArchObjGeneratorMet  
sJd.java:266) at  
com.exlibris.digitool.viewers.metsviewer.archrepos.ArchObjGeneratorMetsJd.processArchObj(ArchObjGeneratorMet  
sJd.java:187) at  
com.exlibris.digitool.viewers.metsviewer.archrepos.ArchObjGeneratorMetsJd.generateTree(ArchObjGeneratorMetsJ  
d.java:157) at  
com.exlibris.digitool.viewers.metsviewer.archrepos.ArchObjGeneratorMetsJd.generateDoc(ArchObjGeneratorMetsJd  
.java:1316) at  
com.exlibris.digitool.viewers.metsviewer.archrepos.ArchRepositoryImpl.getArchObject(ArchRepositoryImpl.java:219)  
at com.exlibris.digitool.viewers.metsviewer.archservlet.ArchObjServlet.doGet(ArchObjServlet.java:360) at  
javax.servlet.http.HttpServlet.service(HttpServlet.java:697) at  
javax.servlet.http.HttpServlet.service(HttpServlet.java:810) at  
org.apache.catalina.core.ApplicationFilterChain.internalDoFilter(ApplicationFilterChain.java:237) at  
org.apache.catalina.core.ApplicationFilterChain.doFilter(ApplicationFilterChain.java:157) at  
com.exlibris.digitool.common.SetCharacterEncodingFilter.doFilter(SetCharacterEncodingFilter.java:50) at  
org.apache.catalina.core.ApplicationFilterChain.internalDoFilter(ApplicationFilterChain.java:186) at
```

MIDESS Workpackage 9: Resource Discovery and Shared Services

org.apache.catalina.core.ApplicationFilterChain.doFilter(ApplicationFilterChain.java:157) at
org.apache.catalina.core.StandardWrapperValve.invoke(StandardWrapperValve.java:214) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:104) at
org.apache.catalina.core.StandardPipeline.invoke(StandardPipeline.java:520) at
org.apache.catalina.core.StandardContextValve.invokeInternal(StandardContextValve.java:198) at
org.apache.catalina.core.StandardContextValve.invoke(StandardContextValve.java:152) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:104) at
org.jboss.web.tomcat.security.SecurityAssociationValve.invoke(SecurityAssociationValve.java:72) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:102) at
org.jboss.web.tomcat.security.JBossSecurityMgrRealm.invoke(JBossSecurityMgrRealm.java:275) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:102) at
org.apache.catalina.authenticator.AuthenticatorBase.invoke(AuthenticatorBase.java:540) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:102) at
org.apache.catalina.core.StandardPipeline.invoke(StandardPipeline.java:520) at
org.apache.catalina.core.StandardHostValve.invoke(StandardHostValve.java:137) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:104) at
org.apache.catalina.valves.ErrorReportValve.invoke(ErrorReportValve.java:117) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:102) at
org.apache.catalina.authenticator.SingleSignOn.invoke(SingleSignOn.java:417) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:102) at
org.apache.catalina.valves.AccessLogValve.invoke(AccessLogValve.java:535) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:102) at
org.apache.catalina.core.StandardPipeline.invoke(StandardPipeline.java:520) at
org.apache.catalina.core.StandardEngineValve.invoke(StandardEngineValve.java:109) at
org.apache.catalina.core.StandardValveContext.invokeNext(StandardValveContext.java:104) at
org.apache.catalina.core.StandardPipeline.invoke(StandardPipeline.java:520) at
org.apache.catalina.core.ContainerBase.invoke(ContainerBase.java:929) at
org.apache.coyote.tomcat5.CoyoteAdapter.service(CoyoteAdapter.java:160) at
org.apache.coyote.http11.Http11Processor.process(Http11Processor.java:799) at
org.apache.coyote.http11.Http11Protocol\$Http11ConnectionHandler.processConnection(Http11Protocol.java:705) at
org.apache.tomcat.util.net.TcpWorkerThread.runIt(PoolTcpEndpoint.java:577) at
org.apache.tomcat.util.threads.ThreadPool\$ControlRunnable.run(ThreadPool.java:683) at
java.lang.Thread.run(Thread.java:534)

Appendix 4

4a dc.xml METS file from Birmingham before modification

```

<?xml version="1.0" encoding="UTF-8" ?>
<mets OBJID="hdl:123456789/171" LABEL="DSpace Item"
xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd">
  <metsHdr CREATEDATE="2007-04-04T09:42:16">
    <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
      <name>DSpace at the University of Birmingham</name>
    </agent>
  </metsHdr>

  <dmdSec ID="DMD_hdl_123456789_171_oai_dc">
    <mdWrap LABEL="Default Dublin Core Record" MDTYPE="OTHER" MIMETYPE="text/xml"
OTHERMDTYPE="UNSPECIFIED">
      <xmlData>
        <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
          <dc:creator>Slater, Terry R.</dc:creator>
          <dc:date>2007-01-18T11:40:33Z</dc:date>
          <dc:date>2007-01-18T11:40:33Z</dc:date>
          <dc:date>2007-01-18T11:40:33Z</dc:date>
          <dc:identifier>
            http://midess.bham.ac.uk:8080/dspace/handle/123456789/171
          </dc:identifier>
          <dc:description>image copyright 2003 Terry R. Slater</dc:description>
          <dc:description>
            This house is dated to 1425 by dendro-chronology. It stands in the Begijnhof (the nun's
            square - hence the modern statue) on the edge of the medieval city. It is listed for
            conservation and is now used as a restaurant.
            Photograph, May 2003.
          </dc:description>
          <dc:format>867654 bytes</dc:format>
          <dc:format>image/jpeg</dc:format>
          <dc:subject>Begijnhof</dc:subject>
          <dc:subject>restaurant</dc:subject>
          <dc:subject>1425</dc:subject>
          <dc:title>Het Houten Huys, Amsterdam</dc:title>
          <dc:type>Image</dc:type>
        </oai_dc:dc>
      </xmlData>
    </mdWrap>
  </dmdSec>
  <amdSec ID="TMD_hdl_123456789_171">

```

MIDESS Workpackage 9: Resource Discovery and Shared Services

```
<rightsMD ID="rights_123456789_171_bin">
  <mdWrap MIMETYPE="text/plain" MDTYPE="OTHER" OTHERMDTYPE="TEXT">

<binData>

TGljZW5zZSBncmFudGVkIGJ5IEFkcmllb3ZlciAoYS5sLmRvdmVYQGJoYW0uYWMudWsp
IG9uIDlwMDctMDEtMThUMTE6NDA6MTNalChHTVQpOgoKTk9URTogUExBQ0UgWU9VUjBpV04gTEID
RU5TRSBIRVJFCiRoaXMgc2FtcGxllGxpY2Vuc2UgaXMgcHJvdmlkZWQgZm9yIGluZm9ybWF0aW9u
YWwgchVycG9zZXMGb25seS4KCK5PTi1FWENMVVNJVkUgREITVFJJQIVUSU9OIEExJQ0VOU0UKCKJ5
IHNpZ25pbmcgYW5kiHN1Ym1pdHRpbmcgdGhpcyBsaWNlbnNlCB5b3UgKHROZSBhdXR0b3l0cykg
b3Iy29weXJpZ2h0cm93bmVYKSBncmFudHMgdG8gRfNwYWNlIFVuaXZlcnNpdHkgKERTVSkgdGhl
IG5vbi1leGNsdXNpdmUgcmlnaHQgdG8gcmVwcm9kdWNlLAp0cmFuc2xhdGUgKGFzIGRlZmluZWQg
YmVsb3cpLmVsb3p3IjZlZldHJpYnV0ZSB5b3VyIHN1Ym1pc3Npb24gKGlUy2x1ZGlUZwp0aGUg
YWJzdHJhY3QpIHdvcmxkd2lkZSBpbWwcm9kdWNlLAp0cmFuc2xhdGUgKGFzIGRlZmluZWQg
IGFueSBtZWZlZm9sCmluY2x1ZGlUZyBidXQgZm90IGxpbiWl0ZWQgdG8gYXVkaW8gb3IgdmlkZW8u
CgpZb3UgYWdyZWUgdGhhdCBEU1UgbWF5LmVsb3p3IjZlZldHJpYnV0ZSB5b3VyIHN1Ym1pc3Npb24g
cmFuc2xhdGUgdGhlcN1Ym1pc3Npb24gdG8gYW55IG1ZG11bSBvciBmb3JtYXQgZm9yIHROZSBw
dXJwb3NlIG9mIHByZXNlcnZhdGlvbi4KClldSBhbHNvIGFncmVlIHROYXQgRfNlVIG1heSBzZWVw
IG1vcuUgdGhbiBvbmUgY29weSBvZiB0aGlzIHN1Ym1pc3Npb24gZm9yCnB1cnBvc2VzIG9mIHNI
Y3VyaXR5LmVsb3p3IjZlZldHJpYnV0ZSB5b3VyIHN1Ym1pc3Npb24gZm9yCnB1cnBvc2VzIG9mIHNI
IHN1Ym1pc3Npb24gaXMgeW91ciBvcmlnaW5hbCB3b3JrLmVsb3p3IjZlZldHJpYnV0ZSB5b3VyIHN1
cmVsb3p3IjZlZldHJpYnV0ZSB5b3VyIHN1Ym1pc3Npb24gZm9yCnB1cnBvc2VzIG9mIHNI
YWxzbyByZXByZXNlbnQkdGhhdCB5b3VyIHN1Ym1pc3Npb24gZm9yCnB1cnBvc2VzIG9mIHNI
IG9mIHlvdXlga25vd2xIZGdILCBpbmZyaW5nZSB1cG9uZmFueW9uZSdzIGNvcHlyaWdodC4KCKlm
IHRoZSBzdWJtaXNzaW9uIGNvbnRhaW5zIG1hdGVyaWFslGZvciB3aGljaCB5b3UgZG8gZm90IGhv
bGQgY29weXJpZ2h0LAp5b3UgcmVwcmVzZW50IHRoYXQgZm91IGhhdWUgZm90YXQgZm90IGhhdWU
cmVzdHJpY3RlZCBwZXJtaXNzaW9uIG9mIHROZQpjb3B5cmVsb3p3IjZlZldHJpYnV0ZSB5b3VyIHN1
IHRoZSBzaW9uZm90IGhhdWUgZm90YXQgZm90IGhhdWUgZm90YXQgZm90IGhhdWUgZm90YXQgZm90
IHRoZSBzaW9uZm90IGhhdWUgZm90YXQgZm90IGhhdWUgZm90YXQgZm90IGhhdWUgZm90YXQgZm90
IFNVQk1JU1NJT04gSVMGkFTRUQgVVBPTiBXT1JlIFRlQVQgSEFTIEJFRU4gU1BPTINPukVEIE9S
IFNVUFBUiFRAPcWSBtBBR0VOQ1kgT1gT1JHQU5JWkFUSU9OIE9USEVSIFRIQU4gRfNVLmVz
T1UgUkVQUkVTRU5UIFRlQVQgWU9VIEhBVkUKRlVVRkIMTEVEIEFOWSBSUdIVCBPBIbSRVZJRvcg
T1gT1RIRVlGTOJMSUdBEIPTIMgUkVRVUISRUQgQlkgU1VDSApDT05UUKFDVCPUIBBR1JFRU1F
TIQuCgpEU1Ugd2lscCBjbG9mcmx5IGlkZW50aWZ5IHlvdXlgbmFtZShzKSbhcYB0aGUgYXV0aG9y
KHMplIG9yIG93bmVYKHMplIG9mIHROZQpzdWJtaXNzaW9uLmVsb3p3IjZlZldHJpYnV0ZSB5b3VyIHN1
YWx0ZXJhdGlvbiwgb3R0ZXlkdGhbiBhcyBhbGxvd2VklGJ5IHRoXMKbGijZW5zZSwgdG8geW91
ciBzdWJtaXNzaW9uLgo=

</binData>

  </mdWrap>
</rightsMD>
</amdSec>
<fileSec>
  <fileGrp USE="ORIGINAL">
    <file ID="f123456789_171_1" MIMETYPE="image/jpeg" SIZE="867654"
CHECKSUM="f35b7504297aa263c819680532eccfd9" CHECKSUMTYPE="MD5"
OWNERID="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/171/1/amsterdam_
15.jpg" GROUPID="GROUP_f123456789_171_1">
      <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/171/1/amsterdam_15
.jpg"/>
    </file>
  </fileGrp>
</fileSec>
</amdSec>
</rightsMD>
</mdWrap>
</rightsMD>
```

MIDESS Workpackage 9: Resource Discovery and Shared Services

```
</fileGrp>
</fileSec>
<structMap>
  <div DMDID="DMD_hdl_123456789_171_oai_dc" ADMID="TMD_hdl_123456789_171
rights_123456789_171_bin">
    <fptr FILEID="f123456789_171_1"/>
  </div>
</structMap>
</mets>
```

4b modified METS file from Birmingham for ingest into DigiTool.

```
<?xml version="1.0" encoding="UTF-8" ?>
<mets OBJID="hdl:123456789/171" LABEL="DSpace Item"
xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd">
  <metsHdr CREATEDATE="2007-04-04T09:42:16">
    <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
      <name>DSpace at the University of Birmingham</name>
    </agent>
  </metsHdr>

  <dmdSec ID="DMD_hdl_123456789_171_oai_dc">
    <mdWrap LABEL="Default Dublin Core Record" MDTYPE="OTHER" MIMETYPE="text/xml"
OTHERMDTYPE="UNSPECIFIED">
      <xmlData>
        <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
          <dc:creator>Slater, Terry R.</dc:creator>
          <dc:date>2007-01-18T11:40:33Z</dc:date>
          <dc:date>2007-01-18T11:40:33Z</dc:date>
          <dc:date>2007-01-18T11:40:33Z</dc:date>
          <dc:identifier>
            http://midess.bham.ac.uk:8080/dspace/handle/123456789/171
          </dc:identifier>
          <dc:description>image copyright 2003 Terry R. Slater</dc:description>
          <dc:description>
            This house is dated to 1425 by dendro-chronology. It stands in the Begijnhof (the nun's
square - hence the modern statue) on the edge of the medieval city. It is listed for
conservation and is now used as a restaurant.
            Photograph, May 2003.
          </dc:description>
          <dc:format>867654 bytes</dc:format>
          <dc:format>image/jpeg</dc:format>
          <dc:subject>Begijnhof</dc:subject>
          <dc:subject>restaurant</dc:subject>
```

MIDESS Workpackage 9: Resource Discovery and Shared Services

```
<dc:subject>1425</dc:subject>
<dc:title>Het Houten Huys, Amsterdam</dc:title>
<dc:type>Image</dc:type>
</oai_dc:dc>
</xmlData>
</mdWrap>
</dmdSec>
<fileSec>
  <fileGrp USE="REFERENCE">
    <file ID="f123456789_171_1" MIMETYPE="image/jpeg" SIZE="867654"
CHECKSUM="f35b7504297aa263c819680532eccfd9" CHECKSUMTYPE="MD5"
OWNERID="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/171/1/amsterdam_
15.jpg" GROUPLD="GROUP_f123456789_171_1">
      <FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="http://midess.bham.ac.uk:8080/dspace/bitstream/123456789/171/1/amsterdam_15
.jpg"/>
    </file>
  </fileGrp>
</fileSec>
<structMap>
  <div DMDID="DMD_hdl_123456789_171_oai_dc" ADMID="TMD_hdl_123456789_171
rights_123456789_171_bin">
    <fptr FILEID="f123456789_171_1"/>
  </div>
</structMap>
</mets>
```

Appendix 4b-1

METS file contained within first part of the successful ingest of the above file (appendix 4)

```
<?xml version="1.0" encoding="UTF-8" ?>
<_ oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:creator>Slater, Terry R.</dc:creator>
  <dc:date>2007-01-18T11:40:33Z</dc:date>
  <dc:date>2007-01-18T11:40:33Z</dc:date>
  <dc:date>2007-01-18T11:40:33Z</dc:date>
  <dc:identifier>http://midess.bham.ac.uk:8080/dspace/handle/123456789/
171</dc:identifier>
  <dc:description>image copyright 2003 Terry R. Slater</dc:description>
  <dc:description>This house is dated to 1425 by dendro-chronology. It stands
in the Begijnhof (the nun's square - hence the modern statue) on the
edge of the medieval city. It is listed for conservation and is now used as
a restaurant. Photograph, May 2003.</dc:description>
  <dc:format>867654 bytes</dc:format>
  <dc:format>image/jpeg</dc:format>
  <dc:subject>Begijnhof</dc:subject>
```

```
<dc:subject>restaurant</dc:subject>  
<dc:subject>1425</dc:subject>  
<dc:title>Het Houten Huys, Amsterdam</dc:title>  
<dc:type>Image</dc:type>  
</oai_dc:dc>
```

Appendix 4b-2

The 4 METS files contained within the second part of the successful ingest from appendix 4.

File 1 mets_section structMap

```
= <structMap xmlns="http://www.loc.gov/METS/"  
  xmlns:xlink="http://www.w3.org/TR/xlink"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" id="DTL1">  
= <div DMDID="DMD_hdl_123456789_171_oai_dc"  
  ADMID="TMD_hdl_123456789_171_rights_123456789_171_bin"  
  id="DTL2">  
<fptr FILEID="f123456789_171_1" id="DTL3" />  
</div>  
</structMap>
```

File 2 mets_section metsHdr

```
= <metsHdr CREATEDATE="2007-04-04T09:42:16"  
  xmlns="http://www.loc.gov/METS/"  
  xmlns:xlink="http://www.w3.org/TR/xlink"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">  
= <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">  
<name>DSpace at the University of Birmingham</name>  
</agent>  
</metsHdr>
```

File 3 mets_section fileSec

```
= <fileSec xmlns="http://www.loc.gov/METS/"  
  xmlns:xlink="http://www.w3.org/TR/xlink"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">  
= <fileGrp USE="REFERENCE">  
= <file ID="f123456789_171_1" MIMETYPE="image/jpeg" SIZE="867654"  
  CHECKSUM="f35b7504297aa263c819680532eccfd9"  
  CHECKSUMTYPE="MD5"  
  OWNERID="http://midess.bham.ac.uk:8080/dspace/bitstream/1234567  
  89/171/1/amsterdam_15.jpg" GROUPID="GROUP_f123456789_171_1">  
<FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="METSID-1" />
```

```
</file>  
</fileGrp>  
</fileSec>
```

File 4 mets_custom_dmd

```
<?xml version="1.0" encoding="UTF-8" ?>  
= <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"  
  xmlns:dc="http://purl.org/dc/elements/1.1/">  
  <dc:creator>Slater, Terry R.</dc:creator>  
  <dc:date>2007-01-18T11:40:33Z</dc:date>  
  <dc:date>2007-01-18T11:40:33Z</dc:date>  
  <dc:date>2007-01-18T11:40:33Z</dc:date>  
  <dc:identifier>http://midess.bham.ac.uk:8080/dspace/handle/123456789/  
    171</dc:identifier>  
  <dc:description>image copyright 2003 Terry R. Slater</dc:description>  
  <dc:description>This house is dated to 1425 by dendro-chronology. It stands  
    in the Begijnhof (the nun's square - hence the modern statue) on the  
    edge of the medieval city. It is listed for conservation and is now used as  
    a restaurant. Photograph, May 2003.</dc:description>  
  <dc:format>867654 bytes</dc:format>  
  <dc:format>image/jpeg</dc:format>  
  <dc:subject>Begijnhof</dc:subject>  
  <dc:subject>restaurant</dc:subject>  
  <dc:subject>1425</dc:subject>  
  <dc:title>Het Houten Huys, Amsterdam</dc:title>  
  <dc:type>Image</dc:type>  
</oai_dc:dc>
```