UNIVERSITY OF BIRMINGHAM

MIDESS Project User Requirements Analysis Executive Summary

Executive Summary

The MIDESS Project is a JISC project funded under the *Digital Repositories Programme*. MIDESS explores the management of digitised content in an institutional and cross-institutional context through the development of a digital repository infrastructure. 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 of London (UCL).

An on-line questionnaire was built and publicised throughout three of the MIDESS partner Universities: - Leeds, Birmingham and LSE. UCL did not take part in the survey since they were specifically assigned the copyright workpackage for the project. The goal of this questionnaire was to identify how respondents were creating, storing and using digital material. The results from the on-line questionnaire were then gathered and then analysed by summarising answers to each specific question within the questionnaire. The analysis was completed by Dr Stephen Charles, the MIDESS projects manager based at the University of Leeds.

In total there were 205 replies to the on-line questionnaire from within the MIDESS partner institutions. These were followed up with one to one interviews to develop a set of case studies on current and potential use of digital content.

Key Outcomes

1. There is a clear need for the development of digital repositories to assist with the management of digital content at both the departmental and institutional level. Key requirements include the ease of adding digital material to the system, the requirement for long-term storage, the ability to password-protect or restrict specific digital collections, the ability to search across collections, the effective management of consistent metadata schemas, and the provision of easy access to the digital material.

2. In the case of an Institutional repository, the variety of digital material and metadata schemas required necessitates a digital repository software system which is both powerful and flexible. Digital repository software that has less functionality is much more suited to subject matter digital repositories which can be satisfied by simpler metadata requirements.

3. Support needs to be provided throughout the entire digitisation, storage and location of material process.

   This includes advice on:-

   - Digitising material,
   - Adding the digital material to the digital repository,
   - Recommending appropriate metadata schemas,
   - Adding the metadata to the digital repository
   - Maintaining the collection.
   - Copyright issues
   - Technical issues
   - How to search the digital repository
   - How to use the resulting digital repository to best effect within the respective departments, and throughout the university.

4. Digital repositories are expected to connect to, and be accessible from, other systems within institutions such as the Library system, the Virtual Learning Environment (VLE) and university portals.

Table of Contents

MIDESS Project

MIDESS User Requirements Analysis Report (Work-Package 3)

The MIDESS Project is a JISC project funded under the *Digital Repositories Programme*. MIDESS is exploring the management of digitised content in an institutional and cross-institutional context through the development of a digital repository infrastructure. 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 of London (UCL).

1. Aims of User Requirements Analysis Work-Package:

WP 3 – the user requirements analysis work-package, is intended to meet the following aims:

- To develop a prioritised list of user requirements for digital content repositories, to include requirements and expectations for searching, display of images, management of content, storage, metadata requirements, long-term preservation requirements and re-use in other contexts (e.g.: in e-learning).
- To identify content that might be drawn on by the MIDESS project for repository population and testing.
- To establish links with potential users which may be used at a later stage (e.g.: for usability testing).

The user requirements analysis work-package concentrated on soliciting responses from participants at Leeds, Birmingham and LSE. UCL did not contribute to this work-package as their role in the project does not include the development of a repository, so it was felt to be less relevant to receive responses from them.

2. Work-Package Proposed Activity Breakdown:

The following activities and timescales were proposed for the Work-package, and approved by the Project Steering Group.

Activity	Dates	Owner	Notes
Produce stakeholder lists to identify stakeholders at each partner institution	End Aug 05	All partners	Stakeholders are staff currently involved in developing digital content collections, or who have expressed serious interest in doing so, or who have previous experience in this are
Semi-structured interviews with a cross-section of the listed stakeholders	19th Sept – 14th Oct	Project manager	Numbers interviewed will depend on availability and willingness to be interviewed, but should aim for approx 6 staff from each institution). Interviews could be face-to-face or by telephone.
Questionnaire	30th Sept – 21st Oct	Project manager	To be published and promoted via internal mailing lists and intranet; aim for target set of responses of 50

			staff from each institution.
User Requirements analysis report	24th Oct – 18th Nov	Project manager	Circulated to steering group and working groups, dissemination via project website

3. Interviews:

Semi-structured interviews were carried out with 9 participants from The University of Leeds, 2 participants from the University of Birmingham and 6 participants from LSE.

Each MIDESS partner has a local project working group. Selected members of the working groups at each partner site were interviewed on a one-to-one basis to ascertain some initial requirements, review project expectations and identify institutional expectations.

Other potential participants for the interviews were drawn from the responses to the on-line questionnaire. Where a respondent indicated that they had a large amount of material or needed more information about potential digital repository developments, the project manager contacted them personally and arranged a follow-up interview.

3.1 Case studies

Case Studies from Leeds

Case Study 1

Virtual Museum of Old Physics Equipment Currently Stored in the Department of Physics at The University of Leeds.

Contact names Graeme Gooday, Denis Greig. University of Leeds

There is a considerable amount of historical physics equipment currently stored at the University of Leeds that is expected would be of considerable interest to both science historians and people in general who are interested in physics. This including some of Bragg's equipment (Bragg later won the Nobel Prize for Physics for his X-Ray diffraction work).

It is felt that this equipment should be photographed, digitised and made available to as wide an audience as possible via the Internet. It is suggested that an Institutional digital repository would be a suitable storage location for this material.

Thus the proposal is to construct a virtual museum where images of this equipment (and possibly video clips) can be viewed.

The description of the physics equipment should be both detailed enough for science historians and also suitably understandable for less technical people. It is suggested that the Dublin Core metadata schema be used.

Case Study 2

Images and Video stored within the Medical School and Medical departments at The University of Leeds.

Contact name Pat Harkin, University of Leeds.

The Medical Faculty and the various medical departments within the University of Leeds use a wide variety of digital images and video both for teaching and learning as well as for reference material. There is currently no centralised digital repository where this material is held. Existing digital material is either stored on local PC's and/or departmental servers. Although digitised, the material is not usually shared and rarely has descriptive metadata attached.

Two of the most popular medical metadata schemas are SNOMED and MESH. These schemas are specific to the medical profession and enable medical specialists to describe components of the body, diseases, causes of disease etc in great detail. Broadly speaking for each image there are multiple numbers of descriptors. These descriptions are obtained via a hierarchical schema system with increasingly detail at each hierarchical level.

Patient confidentiality prevents medical staff from making medical digital material widely available via the internet. For this reason common metadata schemas such as Dublin Core are of relatively limited relevance as metadata standards in medicine.

Because of the sensitivity of the material (much of which may be of medical images of patients) if medical digital material is to be stored in a digital repository, the material must be restricted primarily to medical staff and medical students and therefore collections need to password protected and limited to access by authorised staff. Particularly sensitive material may also be required to be protected at an individual file level and further restricted by IP address, thereby ensuring that the material can only be accessed at specific computers within the medical community at Leeds.

Case Study 3.

Medieval Illuminated Manuscripts Digitisation and Storage. Special Collections. The University of Leeds.

Contact names Oliver Pickering and Katja Airaksinen The University of Leeds.

The University of Leeds Library holds over 200,000 books and thousands of manuscripts in its special collections department. 688 illuminated pages. Illustrations from illuminated manuscripts dating from the 12th to the 16th century have been digitised.

High resolution TIFF images have been provided as the default format. These TIFF files varied in size between 10Mb and 253Mb in size. These TIFF images will be included in the repository along with corresponding lower resolution JPEG images.

The high resolution TIFF images will be protected by a password or other such protection technique to ensure that once the collection has been made available via the Internet, that the TIFF files are not downloaded by online visitors viewing the images. The lower resolution JPEG images will have no such restrictions on viewing.

The metadata schema to be used is VRA Core.

A Microsoft Access database has been created as a short term solution to catalogue and store the digitised images. The digitised images are hyperlinked to appropriate records in the Microsoft Access database. The Microsoft Access file structure and data will be transferred into the digital repository once the appropriate system has been chosen.

Case Study 4

Computer based learning material in MP3 format for medical students.

Contact name John Sanders

It has been proposed that medical students are provided with learning material in the form of MP3 files sound files. Sound clips from medical consultants, lectures and commercial suppliers could be collected, digitised and converted into MP3 sound files. The resulting MP3 files could then potentially be stored in a digital repository.

Medical students could then download the MP3 files on their IPod's or other similar MP3 playback devices, and listened to at their convenience. The content would primarily consist of medical revision notes. The medical school at the University of Leeds has a commercial supplier of sound based training material who is potentially willing to play an active part in the project.

Password protection of the MP3 medical files would be required in order to restrict the material to staff and students in the medical school and suitable metadata would need to be specified to enable students to locate the appropriate material.

Case Study 5

Godfrey Bingley Photographic Collection

Contact name Special Collections University of Leeds

This collection of around 10,000 photographs was donated to the University in 1913 by Godfrey Bingley himself and is an archive of his life's work.

Spanning the years 1884-1913 (when he gave up photography due to failing eyesight) the photographs cover a wide range of subjects. There are many Yorkshire scenes, including places such Kirkstall Abbey, Fountains Abbey, Headingley, Bolton Abbey and Scarborough. However, the collection also covers the rest of the UK and some of Europe, America and the West Indies.

Professor P.F. Kendall of The University of Leeds , in accepting the gift in 1913, described the archive as 'probably the most magnificent collection ever made of lantern slides, illustrating architecture, archaeology, geology and scenery in all parts of England, but especially Yorkshire'.

The condition of the photographs varies. Given how old the photographs are it is felt that unless some form of preservation is undertaken relatively swiftly then the photographs will continue to decompose.

Case Study 6

School of Music

Contact Name Kia Ng

Music scores, information from motion capture equipment used to capture dance movements from dancers. The files produced are very large and in a variety of formats. Variety of formats (some files are very large). This material would be very useful however the viewer will need a variety of applications to view the files. This material is to be made publicly available via the web.

Case Study 7

Media Services, Film and Video Collection. University of Leeds
Contact Media Services

The Media Services archive at the University of Leeds contains several hundred films and videotapes produced over the last 70 years for academic departments within the University of Leeds. Mechanical deterioration of the material and radical changes in recording and playback standards are making it increasingly difficult to access the material. Included among this material are numerous old 16mm film which includes unique footage during the periods of both the Spanish Civil War and the Second World War.

Clips from these films and videotape could potentially be digitised and stored in a digital repository enabling much wider access via the Internet. Also required would be a suitable video capture device and large quantities of computer storage space to accommodate this material.

Case Study 8

Fine Art School's Slide Collection, University of Leeds.
Contact - Benedict Read, Department of Fine Arts

The Fine Art School's Slide Collection at the University of Leeds currently holds in the region of 150,000 Fine Art 35 mm slides. These are currently catalogued in an online database with detailed metadata for each slide. Metadata fields for this slide collection includes Name, Forename, Country, Title, Date, Medium, Dimensions, Location, Period, Keywords, Slide Source, Status, Country, Type, Accession Date and Accession Number. However currently this online database does not contain the corresponding digitised image of the slides.

Potentially a large quantity of the slides could be digitised and the resulting digitised images stored in an Institutional digital repository with either open or password protected access to the Internet. There are however particular problems regarding copyright and these would need to be addressed.

Case Study 9

Streaming media in/from the Language Centre
Contact Name Peter Howarth, Language Centre

The Language Centre at the University of Leeds currently receives 20 satellite TV channels including numerous foreign language channels. Recordings of individual TV programmes are made on request from tutors. These have been supplied on VHS and more recently on DVD. Similar requests are made for audio recordings from radio stations.

It has been proposed by the Language Centre to transfer as much classroom video and audio material as possible onto central servers to set up round-the-clock recording on hard disk of core satellite TV channels, retained for a period during which time tutors can request programmes to be archived or copied onto DVD, the remainder of the programmes to be automatically wiped.

Potentially a digital repository could be used to store the resulting digital material. This would ensure that students could access the material irrespective of their location provided they have access to the Internet.

Case Studies from Birmingham

Case Study 1

Barber Institute of Fine Arts Coin Collection

Contact Name Dr Eurydice Georganteli

The Barber Institute of Fine Arts possesses one of Europe's finest collections of Byzantine coins. In 1967 the Trustees supplemented the Institute's works of art by acquiring over 15,000 coins from Philip Whitting and Geoffrey Haines, two distinguished collectors and numismatists.

The majority of these coins have been digitised but there is currently no means of delivery.

Case Study 2

Department of History of Art Slide Collection

Contact Name Dr. Jutta Vinzent

The Department of History of Art has a substantial slide library. The collection holds more than 70.000 catalogued slides. Although the vast majority of the slides are of Fine Art images, these are complemented by a substantial collection of Architectural images and include collections of applied arts and miscellaneous subjects.

The slide library has a searchable on-line catalogue which is accessible to students and staff via the intranet however as yet there is no visual catalogue of the slides.

Case Studies from LSE

Case study 1

Digitised lectures from LSE

Contacts: Sarah Leach, Centre for Learning Technology, LSE

LSE has been video and audio recording lectures for specific courses and programs at the LSE. The Video lectures are made available to students, through WebCT, for revision purposes and further understanding. The video recordings come in varying formats such as Windows Media Player, Quick Time and Real Player. The material is currently being catalogued in the CLT media database which uses UKLOM as its metadata standard. There are approximately 116 items recorded.

Case study 2

Digitised photograph collections from LSE Archives

Contact: Sue Donnelly, Archivist, LSE Library

Archives have digitised an archive of 1,100 photographs taken by the anthropologist Bronislaw Malinowski (1884-1942) during his field work in the Trobriand Islands during 1915-1918. Malinowski later taught at LSE. The photographs are available as both high resolution TIFFs and JPEGS. The JPEGs are currently available via Archives Catalogues (http://archives.lse.ac.uk/) on the Library website.

A further 500 photographs were digitised from the archives of another anthropologist, Siegfreid Nadel (1903-1956), taken during his field work in Nigeria. The same specifications were used. The photographs are currently available through the archives database but will be accessible via the Archives Catalogue later in 2006. The metadata schema used is ISAD(G).

Case study 3

Digitised off air broadcasts held by CLT

Contacts: Kris Roger, Centre for Learning Technology, LSE

LSE currently digitise off-air broadcasts, including television and radio output, for use in WebCT courses by specific lecturers. This is done under the Schools Educational Recording Agency (ERA) Licence. There are approximately 100 items in the collection.

The material is currently catalogued in the CLT media database which uses UKLOM as its metadata standard.

Case study 4

Charles Booth's Enquiry into London Life and Labour notebooks

Contact: Sue Donnelly, Archivist, LSE Library

The RSLP funded *Charles Booth Online Archive* (http://booth.lse.ac.uk/) project digitised 38 notebooks describing walks around London with members of the Metropolitan police force made during the late 19th century. The notebooks were scanned as high resolution TIFFs and JPEGS and are accessible via the website as both JPEGS and DjVU files. The metadata is ISAD(G).

Case study 5

Court of Directors' minutes, Ionian Bank

Contact: Sue Donnelly, Archivist, LSE Library

The Ionian Bank project, funded by Alpha Bank, Greece, digitised the Court of Directors' minute books, 1845-1917 (12 volumes). The volumes were scanned as high resolution TIFFs and JPEGS and are made available as PDF files via the Archives Catalogue (http://archives.lse.ac.uk/) and also via the Ionian Bank project webpages. The metadata is ISAD(G).

Case study 6

Fathom public lectures and interviews

Contact: Sarah Leach, Centre for Learning Technology, LSE.

The CLT hold the LSE produced content from the now defunct website Fathom.com, a consortium of universities, libraries and other educational institutions. Video content includes Audio and video in the form of lectures and interviews with academics and other experts. The format is Real Player. All clips have been catalogued in the CLT media database which uses UKLOM as its metadata standard. There are currently 336 Fathom items listed in the database.

4. Questionnaire:

An on-line questionnaire was developed in order to determine the following:

- What digital content is being created by academic and support staff at the MIDESS partner Institutions?
- The level of expertise of respondents at the partner institutions in creating and using digital material?
- The requirements of respondents creating and using digital material?

The on-line questionnaire can be viewed at:-
http://www.leeds.ac.uk/library/midess/questionnaire.htm

The questionnaire was developed using Active Server Pages (ASP). Once the questionnaire had been completed and submitted by the respondent the data was sent in the form of an email to the project manager, and to the lead contact at each of the partner institutions.

4.1 Response issues

There was no easy way to contact everybody who might be potentially interested in digital repositories. Contacting potentially interested parties consisted primarily of three techniques.

1) Recommendations from members of the MIDESS working groups.
2) Sending out requests to fill in the on-line questionnaire via the various University mailing lists. This was surprisingly effective, and resulted in a large number of responses.
3) Use of Intranet facilities for promoting the project and the questionnaire.
4) Offer of a £50 prize for one respondent completing the questionnaire.
5) Use of partner university websites to identify potential interested parties by examining the quantity and quality of the digital material that was being placed on pages in the university website.

4.2 Questionnaire structure

The on-line questionnaire was split into two major parts thereby addressing the needs of two potentially very different types of respondents.

Section A of the questionnaire was designed for university staff such as academics, researchers and administrative staff who were actively involved in creating and using digital multimedia material to support either their learning and teaching or research.

Section B of the questionnaire was designed for university staff involved in supporting the creation and use of digital media, although these staff would not normally themselves be directly involved in creating the digital material. Typical staff here might include library, IT staff, media services staff and e-learning support staff.

Some people in the partner universities have a dual role of both creating digital material (relevant to section A of the questionnaire) while also assisting and supporting other colleagues (relevant to section B of the questionnaire), these respondents usually completed both sections of the questionnaire.

4.3 Response rate

Total response rate (submissions after the closing date were excluded)

205 replies

University of Birmingham	31 replies
University of Leeds	157 replies
London School of Economics (LSE)	17 replies

181 respondents replied to section A only, 60 Respondents replied to section B only. 35 respondents replied to both section A and section B.

4.4 Variation in questionnaire responses across MIDESS partners

The main variation in responses across the MIDESS partners was in the number of responses, with by far the largest number of responses coming from the University of Leeds.

5.Conclusions drawn from User Needs Analysis:

5.1 General issues

- There is a large amount of activity within the partner institutions – either in creating digital content, or in planning to develop digital content. A large number of respondents to the questionnaire had either used or developed digital material to support a variety of activities.
- The technology used to create digital material is sufficiently mature that many university staff in the partner sites now seem to feel confident in creating their own digital content without requiring a large amount of support. The questionnaire served as a reality check for a number of respondents who felt it had assisted them in identifying potential issues for consideration in relation to creating digital content.
- Digital content was primarily created in order to support learning and teaching activities. In a number of cases, academic staff have created their own personal digital images and digital content, which often replaced a previous reliance on personal slide collections in order to support teaching.
- It was found that most important goals when creating or using digital material were to make teaching and research more interesting and to ensure that the material was available long term.
- Having the material in a digital format also made the material easier to distribute around the university.
- It was considered that saving time and sharing material with other institutions was less important overall.
- Digital material was also created to support research activities. A significantly smaller number of respondents use digital material to provide reference material.

5.2 Issues with creating digital content

- Technical problems are the biggest reported headache for respondents looking to digitise material and add it to a repository. These are generally down to a lack of detailed understanding of technical issues at the project outset.
- Respondents were sometimes unfamiliar with the software, had a poor understanding of file formats, struggled with provision of access through the web (access and security issues), did not have time to learn new software in detail, did not plan for storage spare requirements, or adequately consider long-term digital preservation issues.
- Copyright is also regarded as a big issue for respondents. There was a very low awareness of copyright legislation amongst respondents.
- Respondents often found it difficult to find suitable external material, and there was a generally low awareness of the availability of external copyright-cleared collections such as EMOL. Respondents tended to rely on search engines such as Google for tracking down external digital images and multimedia materials, but reported information overload when they do so.
- The majority of respondents had no metadata associated with their digital material. Awareness of the requirement for metadata, or of current metadata formats and schemas was low.
- The majority of respondents had not adapted digital material for users with special needs. With the introduction of SENDA (the special Educational Needs and Disability Act (2001), this is now a legal requirement and should be addressed as soon as possible.

5.3 Funding issues

- The vast majority of respondents had no external funding support and did not make use of specific websites developed to assist them with their teaching or research. Clearly additional support and more effective search engines/search techniques would be of assistance here.

Where external or internal funding had been available this was via a range of sources, including:

Funding Sources for Respondents
Research Councils – e.g.; EPSRC, ESRC, BBSRC, AHRC etc
HEFCE and JISC
Research and consultancy funding, e.g. local authorities, local government (ODPM)
Internal funding programmes such as a University Learning Development Unit
TQEF
White Rose Fund
LTSN
Heritage Lottery Fund
Clothworkers' Foundation
NHS, Wellcome, MRC
Charitable Foundations – e.g.: Gatsby
Internal School funds
RSLP

5.4 Storage, security and back-up of material

- Respondent's digital material was primarily stored on a local PC although the method used to backup their files varied, with the majority indicating that they were backing up to a departmental server rather than an institutional server.

  At the University of Leeds the difference between a departmental server and an Institutional server is somewhat blurred. In practise respondents are backing up files on their network drive however the exact physical location and nature of the network drive is unclear to them. For example at the University of Leeds many network drives are in fact partitions of the large centralised Storage Area Network (SAN) although this fact is largely hidden from the person backing up the files to the network drive. Thus in many cases respondents are backing up to what they think is a departmental server via the network drive but which is in fact a storage area set aside for their department on the Institutional SAN.
- CDs and DVDs were also used as long term backup devices despite the questionable long term storage reliability of these mediums.
- The vast majority of respondents expected to continue to use their digital material for either years or indefinitely rather than just days or months. This demonstrates the importance of backwardly compatible software and hardware in order to ensure the material is accessible long term.
- Most respondents considered that digital material could often be stored offline, rather than requiring instant access. It is likely that increased speed of access to off-line storage media such as CD-ROMs has played a part here.

5.5 File formats

- The most commonly used file format for images was JPG, due largely to the increasing use of this format on the web.
- Both sound and video had no clear winners when it came to the most popular format. The MP3 sound format was slightly more popular than WAV format, while for video AVI, MPEG, QUICKTIME, REAL PLAYER, and Windows Media were all equally popular with both developers and users.

5.6 Sharing digital content

- A large amount of the digital content that has been created was not freely available. (90 respondents specified that their digital content was not freely available.).
- Respondents primarily cited shortage of time as a reason for not making their content freely available.
- Respondents were also concerned about copyright issues (uncertainty about whether their content is 'legal' rather than protecting their own copyright) as a reason for not making content freely available.
- Some material needed to be closely restricted for data protection reasons (e.g.: medical images).
- Some content was available via password-protected websites, and a small minority of respondents requested the implementation of a payment system in order to provide access to their digital material.
- The majority of respondents considered that their material would be of value to others undertaking teaching or research in similar areas (though not to other users in general) and they also considered that the most popular ways of finding their material would be either via a search engine (such as Google) or via an Institutional website.

5.7 Responses from technical and support staff

Section B of the questionnaire was designed primarily for technical staff and others supporting the creation of digital material rather than those respondents creating or using digital material. It was therefore expected that the responses would show a greater

understanding of the issues involved with digital media and this in fact proved to be the case although perhaps not to the extent we would have hoped. Section B of the MIDESS questionnaire was designed primarily to be completed by staff supporting the creation of digital media however their lack of knowledge of digital media leads the MIDESS project to question how much support for the creation of digital media they can actually give in practise.

- The majority of respondents to section B were aware of interest in using and /or creating digital content within their university departments/schools.
- There was an equal split between interest in creating digital images and digital video with a significantly lower interest by respondents in creating digitised sound files.
- There was also an equal split between those respondents who were aware of specific projects within their university with those who were not aware of any specific projects within their university.
- The majority of respondents to section B of the questionnaire had at some time supported users in developing digital material. This somewhat contrasts with the user's responses that one of the major issues with digital material was the perceived lack of technical support.
- The majority of respondents completing section B of the questionnaire reported providing assistance with technical support. They also assisted with adding digital material to databases and making disk space available for digital material.
- Respondents were equally split between users requests for the material to be made available via an unrestricted website or via a password-protected website.
- A greater proportion of support staff responding to section B of the questionnaire were aware of external sources of digital material than respondents to Section A (usually users of digital material), however the majority of technical staff still remained largely unaware of external sources of digital material.
- Where the support staff were aware of the external sources of digital material they made efforts to inform users within their departments. Given the responses from the users was that they were largely unaware of other sources, efforts by the technical staff to inform users of potential sources of digital material would appear to be having limited effect.
- There was a much greater understanding of the copyright issues by support staff than by users; however one third of all support staff were still unaware of the copyright issues and legal restrictions related to digital media. This highlights the importance of ensuring that all staff are made aware of copyright issues when depositing digital material into an institutional digital repository.
- The majority of support staff thought that users of digital material would be interested in sharing their digital material and also expected users to be interested in viewing and using other people's material.
- There was widespread support for a central repository amongst staff and they expected that this would encourage users to share digital material.
- The major barriers to sharing and re-use of this digital material were considered to be the complexity involved and the fact that much of the material was primarily intended for use within the university rather than for wider distribution. Security of the digital material (for example in the case of medical images) was also an important factor.
- There was a strong agreement amongst support staff that support would be required for the sharing and use of digital material.
- Support staff thought that the most important features of a digital repository would be easy access to the material, a large amount of material to choose from and no copyright issues.
- The majority of respondents to Section B thought that the digital repository should interface to other systems, including the VLE, Portal, Library System, Intranet etc.
- There was an equal split between those support staff who were familiar with metadata standards and those who were not familiar with these. Dublin Core, IMS

and MARC were the most widely recognised standards. Emerging metadata standards such as METS were considerably less well known.

- More than three-quarters of the respondents to Section B thought that metadata was relevant for a digital Repository even if in some cases they didn't use it themselves.

6. Implications of Conclusions

An institutional repository can be seen as a means of displaying an institution's digital assets and the term can encompass a wide range of meanings. These include:-

- Subject repositories (repositories containing digital material specific to a subject such as biology),
- Teaching and Learning repositories where teaching material such as computer based learning courses can be located,
- Preservation repositories which are specifically designed for the long term preservation of digital material
- Research repositories which contain both papers (published and unpublished) and thesis.

Thus it is important to decide what the functionality of an institutional repository should be and to determine whether the digital content is of value to the institution and worth preserving. An Institutional repository should not simply be seen as an additional storage space for digital material or a means of making raw digital material more widely available. For potential submitters there are often issues as to whether digital materials should be stored at a departmental level, at an institutional level or at a national level. Until quite recently institutional repositories were quite rare and many academics submitted their digital material to national repositories simply because these were the only places where such material could be submitted.

However with the rapid increase in the number of institutional repositories there is now the opportunity to store material in an institutional repository rather than a national repository. National repositories are increasingly 'harvesting' this institutionally based digital material for storage in their national repositories.

Institutional Repositories are primarily designed to hold collections of static objects, whether they are digital images, digital video or digital documents. Along with the objects should be metadata information relating to those objects.

It is therefore important to define the scope of the repository, determining what will be included in the repository and what will not, this allows the repository administrator to make an estimate, however imperfect, on the quantity of data, its likely growth and thus its recurrent costs.

There are also institutional barriers to obtaining content for digital repositories. Potential contributors may wish to store their digital material locally where they feel they have total control of its content. There is often the perception that in the progression from local computer – departmental server – institutional repository – national repository, that there is a resulting loss of control of both the digital material and how the digital material will be used by people viewing it.

Free disk space may also be provided at the local computer or departmental level while additional funding may be required for more centralised storage. The cost of the initial digitising may also be prohibitive for more potential submitters of digital material.

Once the repository is actually running there are also questions regarding the funding for submitted digital material. Most commercial products base their prices on the number of objects in the repository. Under these circumstances should potential depositors be

charged for both the number of objects they wish to deposit and for the storage space they use in depositing the digital objects, or should this all be funded centrally by the institution?

Once the digital objects have been added to the repository, who is responsible for entering the metadata corresponding to the digital object? Should this be the depositor? Should this be Library staff or should this be specialised personnel, who have a detailed understanding of the digital subject matter being deposited? Who owns the collection once the digital material is added to the repository? Who has the rights to add, modify and delete material from the collection? The administrator? The depositor? Others?

There are a large number of respondents who are developing, or considering developing, digital material. The costs of developing digital content appear to be relatively low to the end user, who might simply purchase a cheap scanner or digital camera and set about creating content without any serious thought as regards wider technical requirements or the metadata issues involved.

Users do appear to quickly run into difficulties when considering making their material more widely available; however, they are often faced with a bewildering range of software packages, hardware platforms and other digitisation facilities. Technical support at this more advanced level is available at each of the MIDESS partner institutions, but it is often sparsely provided, and users can often struggle to gain access to this support and expertise. Support was found to be very welcome by users where available.

There does not appear to be a joined-up strategic approach in place at any of the partner institutions, as regards the creation and long-term management of digital content. This results in fragmented user support and a lack of availability of clear, consistent advice on a range of issues from storage to copyright to metadata schemas.

In order to make the digital repositories at the various institutions a success there clearly needs to be a great deal of user support at all stages of the process from initial consultation and advice through to active management of the deposited material. We suggest that this should consist of a co-ordinated response across the support services who hold the required expertise (e.g.: library, central IT staff, e-learning support staff and media services staff).

Project teams need to be formed to co-ordinate support, and a key aspect of this is taking time to understand requirements and explaining how the digital repository can provide a key component for support. There is a clear need for the development of digital repositories to assist with the management of digital content at an institutional level; and key benefits include the potential for long-term storage of the material, the ability to password-protect or restrict specific digital collections, the ability to search across collections, the effective management of consistent metadata schemas, and the provision of access to the digital material.

This should be followed by advising them on digitisation (if the material has yet to be digitised) and even possibly practically demonstrating how the digitisation is undertaken. It is assumed that the administrator of the digital repository will not be digitising material for anything other than a simple demonstration and where possible the administrator would expect the potential depositor to self-deposit material themselves rather than be required to deposit material for them.

The lack of understanding and appreciation of copyright issues shows a clear need of well defined guidelines in this area. The MIDESS project has a specific work-package addressing these issues.

Before submitting digital material to a digital repository both the potential depositor of digital material and the administrator of the digital repository need to be clear that the material submitted does not breach copyright and both also need to be clear what the potential use of the material will be once it has been deposited in the digital repository. For example while depositors may have no problem with viewers of the digital repository downloading digital material and using it in part of their lectures, depositors do not normally wish to allow their digital material to be then sold for profit by the viewer.

A minority of respondents had metadata associated with their digital material and most had little or no understanding of how metadata might assist in helping people to find the digital material. On interviewing several respondents on a one-to-one basis, all respondents were familiar with search engines such as Google, Yahoo etc in which often a single phrase is entered, but fewer respondents were familiar with the potential advantages of more detailed metadata schemas such as Dublin Core, LOM and the advantages such metadata provided in locating information.

7. Recommendations

7.1 For development of the digital repositories

Essential:

- The repository packages will need to be capable of holding potentially very large collections, or of having the scalability to achieve this.
- Permissions management is crucial, as it will be essential that some collections can be limited to specific target audiences.
- The repository will need to be capable of handling a variety of file formats, including still images, sound and video.

Important:

Technical requirements

- The repository infrastructure will need to be flexible enough the enable the creation and management of a variety of different metadata schemas to meet different needs across the subject disciplines.
- It is unlikely that metadata will be bulk-uploaded into the repositories, as metadata is not currently available for many of the collections identified. Instead, the repository will need to have tools for creating metadata on the system which are easy to use.
- Bulk upload facilities for images and other content should be provided.
- Integration with VLE's, portals and other corporate systems is not essential at this stage of the project, but should certainly be considered as part of our longer term requirements.
- Single-sign on for authentication purposes (for data upload or access to restricted content) is important – and should be achieved mechanisms most appropriate to the local institutional architecture (e.g.: LDAP or Shibboleth).
- The repository should be capable of referencing externally held content on other servers.
- The repository should be capable of providing an interface to off-line storage facilities if possible.
- The repository should have flexible and powerful search facilities.

- The repository should if possible provide e-commerce facilities so that individual images can be sold.

- Licensing conditions for individual images and collections should be clear and easy to identify on the system.

Functional requirements

- There should be support for the digital repository from the highest level within the institutions.
- Digital preservation facilities are considered very important; actual preservation activity may, however, take place outside of the repository.
- Partner institutions will need to develop a co-ordinated approach to support for the repository and for digital content creation activities in particular. This will certainly require a cross-service approach.
- Institutions should develop models for digital preservation of content held in the repository, including costing of long-term storage.
- Repository launch should be accompanied by launch of a set of best-practice guidelines for content creation – to include advice on project planning, use of software and tools, copyright, metadata creation, long-term preservation and storage.
- The digital repository administrator should identify an owner of the deposited collection who the administrator can contact with issues relating to the collection.
- Potential depositors to a digital repository should be contacted as soon as possible to discuss how they wish their material be used in the digital repository. Ideally this should be before they have digitised their material rather than after digitisation.
- Assistance should be available throughout the entire digitisation and deposit process with easy access to help via phone, email and personal contact for the depositor.
- The benefits of the digital repository should be promoted to the individual/department etc. These include the long term preservation of the material, the ability to easily locate the material and the expected prestige for the individual/department gained by sharing the material.
- Guidelines should be backed-up by a co-ordinated approach to ongoing training events and workshops at the institutional level.
- The digital repository should be publicised and actively promoted throughout the institution. This can be achieved through techniques such as presentations, workshops, tutorials, user groups, personal contacts etc.
- Institutions should promote external copyright-cleared materials such as EMOL more extensively.
- There should be funding for adequate hardware to ensure that the necessary storage space for the digital material is available.
- The barriers to potential deposit of files for non-digitised material should he lowered by identifying possible sources of digitisation equipment for those potential depositors who have no direct access to the necessary hardware.
- The most appropriate metadata schemas for the digital material should be identified and these should be discussed with the person depositing the digital material. Agreement should also be made as to who enters the appropriate metadata for each collection/digital object.
- Wider consideration needs to be given to how the repository will ‘fit’ into the wider institutional architecture – e.g.: integration with VLE’s and portals, ongoing bandwidth requirements, long-term funding etc.
- The input formats and the output formats that the digital repository supports should be identified to ensure that digital material can be exported and imported into other digital repository systems to ensure portability of data.

Statement of Deliverables for Repository Implementation

The University of Leeds will be implementing an institutional repository primarily for teaching and learning. The focus of the University of Leeds Institutional repository will be on digital images, digital video and digital sound rather than text based research papers or thesis. Leeds currently has an E-Prints system which handles research papers. The institutional repository at Leeds must have the capability of storing both a variety of digital material and different metadata schemas in order to accommodate the wide variety of requirements throughout the university. The intention is for an institution – wide repository for deposit of materials both by academics and central service within the University of Leeds. Long term it is expected that this repository will also be closely integrated with the new virtual learning environment and portal that is to be installed at the University of Leeds.

The University of Birmingham's repository is initially to be focussed on digital material available within their library collection which consists of text based material and digital images and includes medieval manuscripts, films scripts and correspondence. It is hoped that the University of Birmingham's repository will be expanded at a later date into an institutional repository and at that stage accept additional digital material from other departments around the university.

LSE's digital repository will hold both library collections and research output. The research output is currently being held in an e-prints repository. The library collections at the LSE consist of prints, maps, card catalogues, notes and pamphlets.