

JISC



UNIVERSITY OF BIRMINGHAM



MIDESS Functional and Technical Requirements Specification

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 London (UCL).

As part of work-package 2 of the MIDESS Project, the team has undertaken an evaluation of a range of different digital repository software packages in order to investigate potential solutions to requirements raised in [Work-Package 3 \(User Needs Analysis\)](#). The full user needs analysis report is available on the MIDESS website (*URL in here once available*).

The results of the evaluation are presented here, and provide the basis for a functional and technical requirements specification for the implementation of digital repositories for images and multimedia at the partner institutions. The document also provides a set of recommendations for repository selection for each partner, and the justification for these recommendations.

MIDESS: WP 2 – Functional and Technical Requirements Specification

Table of Contents:

1	Introduction	3
2	Methodology.....	3
3	Functionality analysis.....	7
3.1	Evaluation criteria.....	7
3.2	Analysis against MIDESS requirements:	9
4	Summary analysis.....	15
5	Financial considerations	17
6	Overall conclusions.....	20
6.1	University of Leeds.....	20
6.2	University of Birmingham	20
6.3	LSE.....	21

1 Introduction

As part of work-package 2 of the MIDESS Project, the team has undertaken an evaluation of a range of different digital repository software packages in order to investigate potential solutions to requirements raised in [Work-Package 3 \(User Needs Analysis\)](#). The full user needs analysis report is available on the MIDESS website (*URL in here once available*).

2 Methodology

The Registry of Open Access Repositories (ROAR)¹ lists a wide variety of different repository installations, using a range of different software packages. A number of projects have attempted to provide comparisons of repository packages – for example, see Thom Hickey's blog which provides an up-to-the-minute ongoing evaluation of tools and packages². However, many of these evaluations have so far concentrated on the use of repositories for managing primarily textual document such as research outputs.

Jedy DeRidder in the paper *Choosing Software for an Institutional Repository*³ divided repository software into three broad categories:

- Software oriented towards grey literature and research.
- Software oriented towards learning objects
- Software oriented towards total content management

MIDESS is concentrating specifically on software oriented towards the management of multimedia materials and digitised images, so it leans closer toward the 'total content management' area, although does not fully fit within this category.

Given the particular focus of the MIDESS Project, the team used the repository survey carried out by the CURL Task Force on Digital Content Creation and Curation (Digital Resource Management Sub-Group) as its initial starting point⁴. Although this survey was primarily geared towards digital preservation functionality, it provide a useful starting point for identifying key systems as it focused on products which could be used to support the management of digital image collections.

This survey had identified 6 repository packages – including a mix of open source and commercial tools – and these packages provided the initial basis for the MIDESS project. However, the survey had a number of key gaps – in particular, DSpace was not represented. A number of new commercial repository packages had also entered the market since the survey was completed in 2004.

The team identified four further systems which should be considered for evaluation – DSpace, Symposia (a commercial product from Innovative Interfaces Inc.), DigiTool® (a commercial product from Ex Libris™) and EMC Documentum. Documentum is an enterprise document and content management system which provides a huge amount of functionality over and above the requirements for the project. However, it was felt that an initial analysis of Documentum should be conducted by the project in order to establish whether such tools provide a good 'fit' for institutional repository purposes. Documentum was selected as a single example of enterprise document management products.

¹ <http://archives.eprints.org/>

² http://outgoing.typepad.com/outgoing/2006/03/repository_comp.html

³ <http://aztec.lib.utk.edu/~deridder/comparison.doc>

⁴ <http://ahds.ac.uk/preservation/curl/DPSoftwareSurvey.pdf>

MIDESS: WP 2 – Functional and Technical Requirements Specification

An initial analysis of the products was undertaken at a high level, focusing on the core technical requirements for the project:

- Ability to handle a variety of file formats – this was considered essential for the management of a wide variety of different file types that might be typical to a multimedia and image repository.
- Interaction with third party systems - for example, support for integration with portals, VLE's, Library management systems etc. This was flagged as an area for investigation in the project bid.
- Authentication via LDAP or SSO (single-sign-on) – considered essential for integration with local authentication infrastructures.
- Support for OAI-PMH – an essential feature of an institutional repository, as it should be able to expose its metadata for external harvesting.
- Support for Z39.50 – considered important for integration with library services, such as federated search tools.
- Support for OpenURL – considered important for integration with library services.
- Support for OAIS – important for digital preservation purposes.
- Support for persistent URL's – important for digital preservation purposes.
- Support for compound object formats such as METS, MPEG 21 or IMS CP – this was flagged as an important area for investigation as part of the project bid.
- Support for rights management and content submission buffer – important for day-to-day management of the repository.

These core requirements had been established at the funding bid stage as key success factors for the project, and they were also backed up by the work undertaken by the CURL Task Force.

Digital repository suppliers and developers were contacted to provide information on the extent to which their products were able to support this initial list of requirements. A summary of the results of this exercise is shown below:

MIDESS: WP 2 – Functional and Technical Requirements Specification

Table comparing software against outline technical requirements

	ContentDM (DiMeMa, Inc)	Curator (Endeavor)	Symposia (Innovative)	Digitool (Ex Libris)	Documentum (EMC)	Eprints	Fedora	Greenstone	Dspace
Website	http://contentdm.com/	http://www.endinfosys.com/prods/curator.htm	http://www.iii.com/	http://www.exlibrisgroup.com/digitool.htm	http://www.documentum.com/	http://www.eprints.org/	http://www.fedora.info/	http://www.greenstone.org/cgi-bin/library	http://www.dspace.org/
Contact used by the MIDESS project	DiMeMa, Inc	Endeavor	Innovative	Ex Libris (UK) Ltd	EMC	University of Leeds and LSE has a E-Prints digital repository	Library of Wales	University of Oxford	University of Cambridge
Ability to handle variety of file formats	Yes	Yes	Yes	Yes	Yes	Limited – primarily intended for documents	Yes	Yes	Yes
Interaction with third party systems	No	Yes – Web Services	Yes – Metafind	Yes – Web Services	Yes – Web Services	None currently known although several proposed	Yes – Web Services	None found	Yes – Web Services
Authentication via LDAP or SSO	No	Yes	Yes	Yes	Yes	Yes	No – in development	No	Yes
Support OAI-PMH	Yes	Yes	Yes	Yes	No – but could be build	Yes	Yes	Yes	Yes
Support Z39.50 and/or SRU/SRW	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No

MIDESS: WP 2 – Functional and Technical Requirements Specification

Support OpenURL	No	Yes	Yes	Yes	No	No	No – in development	No	No
Support OAIS	No	No	No – in development	Yes	No	No	Yes	No	No – in development
Support persistent URLs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Support for compound object formats such as METS or MPEG 21 or IMS CP.	No	No	Yes - METS	Yes - METS	Yes – metadata structures flexible	No	Yes – METS	Yes - METS	No – in development
Rights management support	Yes	Yes	Yes	Yes	Yes	Partial	Yes	Yes	?
Content submission buffer	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

MIDESS: WP 2 – Functional and Technical Requirements Specification

The exercise resulted in the list of potential products being narrowed down quite substantially. Three products were ruled out at this stage for the following reasons;

- ContentDM – no support for LDAP authentication.
- Eprints – primarily focused on text document not multimedia.
- Documentum – no support for OAI-PMH.

It was then agreed to undertake a detailed evaluation of 6 products which were felt to best match the initial requirements.

The 6 products identified at this stage were:

Product	Status
Symposia (Innovative Interfaces Inc.)	Commercial
Curator (Endeavor)	Commercial
Digitool (Ex Libris)	Commercial
FEDORA	Open Source
DSpace	Open Source
Greenstone	Open Source

The 6 products were then evaluated against a more detailed set of requirements drawn from the results of the User Needs Analysis. Evaluation was undertaken using the following mechanisms:

- Repository evaluation day held at the University of Leeds on 19th Dec 2005. All 6 products provided a 45 minute presentation, plus written responses to a set of questions previously circulated to them, based on the user needs analysis. All project partners attended the event and scored the systems individually. Their scores and comments were then collated following the evaluation day. A number of interested participants from other institutions and other projects also attended.
- Detailed follow-up with individual suppliers and demonstrators with further questions following the presentations.
- Information received on costs, licensing and implementation issues. Whole-life costs were developed for each product.

Each project partner was then invited to make their own recommendations for the best fit package for their own institution.

3 Functionality analysis

3.1 Evaluation criteria

The criteria used for evaluation are derived from the recommendations resulting from [Work Package 3 - User Needs Analysis](#). Various groups of potential users and content creators were approached at each partner institution to provide their requirements for a digital image and multimedia repository.

The key requirements established by the UNA can be summarised as follows:

Essential:

- The repository packages will need to be capable of holding potentially very large collections, or of having the scalability to achieve this.

MIDESS: WP 2 – Functional and Technical Requirements Specification

- 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:

- The repository infrastructure will need to be flexible enough to enable the creation and management of a variety of different metadata schemas to meet different needs across the subject disciplines (for example, to support the use of SNOWMED and MESH for medical images).
- 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 through 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.
- Digital preservation facilities are considered very important; actual preservation activity may, however, take place outside of the repository.

In addition, the project team identified the need to establish staffing and support requirements, accessibility conformance, administrative features in the repository and hardware requirements as part of the evaluation.

MIDESS: WP 2 – Functional and Technical Requirements Specification

3.2 Analysis against MIDESS requirements:

The following table shows how each product meets the identified requirements:

	Open Source Products			Commercial Products		
Product issues	DSpace	FEDORA	Greenstone	Innovative Interfaces – Symposia	Endeavor - Curator	Ex Libris - DigiTool
<i>Staffing requirements for installation, support and ongoing management of the software for a large institution</i>	0.5 FTE Systems admin. 0.5 FTE developer. 1 FTE user training and support. Java programmer recommended.	Unix, Apache-Tomcat, XML, Java/PERL/PHP, SQL, XSL, CSS. Near out-of-the-box solutions are available with FEZ, ELATED and VITAL.	Programming skills not required. PERL is useful for creating filters to convert documents to Greenstone Archival Format (GSAF).	Programming and other specialist staff not required. Familiarity with metadata structures and repository concept is required.	Systems librarian for installation, technical and routine system administration. Programmer not required.	0.5 FTE Systems Librarian.
<i>Is there any physical limit on the amount of material the repository can hold?</i>	No theoretical limit.	Only disk space.	No limits noted.	None – beyond overall disk capacity.	No.	No.
<i>Support for file versioning</i>	No built-in file versioning support.	Keeps a version history. If you replace a datastream then the original is kept. The FEDORA admin client allows you to go back to previous versions.	Does not appear to have inbuilt functionality of this type.	Versions are handled with separate digital items.	Not supported.	Support will be in next versions of DigiTool
<i>Ability for institution to specify own metadata</i>	Metadata is extensible in v1.4, but only at the	Any metadata schema can be	GSAF has a metadata	Qualified Dublin Core is the primary metadata	Support for multiple DTD	Descriptive metadata can be

MIDESS: WP 2 – Functional and Technical Requirements Specification

<i>structures.</i>	system level (not at collection level).	added as an additional datastream. Metadata display will be limited by the front-end (eg: ELATED restricts metadata to a discrete set. Custom front-ends can be developed.	element which can incorporate any field from a schema.	format. METS records can be output. Future plans for increasing metadata format flexibility.	such as EAD, Qualified Dublin Core, EAD and TEILite is present. Locally defined formats are also possible. Support for XML schemes and MODS is under development for the next release, v4.0.	stored in Qualified Dublin Core and MARC21. Locally defined fields can be included in both schemas.
<i>Security features</i>	Authorisation is controllable at the Community, Collection and Item level.	Policies can be set up at repository-wide level, or object-specific level.	No information provided.	Authorisation controllable at community, collection and item level.	Rights lists can be applied at the repository, collection, container and object levels. Access can be controlled by domain name, IP address and user class.	Controllable at the individual object level.
<i>Statistics generation</i>	This is available in v1.4.	Not currently available.	No information provided.	Statistics available for access of items, communities and collections.	Rights management component can be used to track use of metadata and associated objects.	Standard reports on use of objects. Future versions will offer a Reporting Center component to provide more flexible reporting capabilities.

MIDESS: WP 2 – Functional and Technical Requirements Specification

<p><i>Linking to VLEs, portals and Library systems</i></p>	<p>No activity on linking with LMS. VLE integration taking place with a number of projects.</p>	<p>No information provided.</p>	<p>No information provided.</p>	<p>Developing integration with VLE/Portal via Web Services data transport layer.</p>	<p>This requires the Course Content Integrator module which works with Blackboard 6.0 using its Building Blocks API, and WebCT Vista 2.1 using its PowerLinks API. Functions include search of pre-selected collections, deep linking to individual objects or collections, access to full search interface within the course management system and single-sign-on. SRU/SRW are supported for federated searching.</p>	<p>Deep linking from VLE's is supported. Instructors can also link into their personal collections on Digitool. APIs are available to support resource discovery and for the repository. Z39.50 server available to support federated searching. X-server for harvesting using standard XML. SOAP can be used for loading objects directly from 3rd party systems such as a VLE.</p>
<p><i>Management of permissions for depositing material</i></p>	<p>Administrator can ensure that depositors can only add materials to their own collections.</p>	<p>This is supported through open source add-ons.</p>	<p>By default there is no concept of a depositor in Greenstone. You can circumvent Greenstone's</p>	<p>Administrator can ensure that depositors can only add materials to their own collections.</p>	<p>Administrator can ensure that depositors can only add materials to their own collections.</p>	<p>Administrator can ensure that depositors can only add materials to their own collections.</p>

MIDESS: WP 2 – Functional and Technical Requirements Specification

			own administrative system and use your own.			
<i>Support for hierarchical collections with unlimited nesting</i>	Supported.	Supported.	Supported.	Supported.	Supported.	Support for hierarchical collections is provided.
<i>Use of web interface for deposit and administration of collections</i>	Some administration tasks require system access.	Java client used for administration.	Web interfaces for submission have to be built locally.	Web-based submission interface. Administration managed through a Java client.	Staff and admin interfaces are Windows clients. There is also a web interface for data entry.	Resource Discovery, Management, Approver, Collection Management, Deposit and Web Ingest modules are web-based. Metadata editor is a client module.
<i>Availability of default metadata templates</i>	Templating mechanisms available.	Add-ons such as VALET provide these kinds of functions.	Not supported.	Templating not currently supported at the collection level. Metadata can be extracted from the file properties to reduce amount of user input needed.	Templating mechanisms are available and will be extended in v4.0.	Templating mechanisms available.
<i>Management of compound objects (where multiple file formats etc might be linked together).</i>	This is not handled well in Dspace.	Available in v 2.1.	Any relationship handled by XML can be handled by Greenstone.	Multiple files can be attached to any digital object.	Multiple media objects can be linked to metadata records.	Handles compound objects (such as multi-resolution images or multi formats of a

MIDESS: WP 2 – Functional and Technical Requirements Specification

						single text object).
<i>Handling of deleted records</i>	By default nothing is deleted. A deleted record leaves a tombstone which is removed from the indices but still present on the system.	Deletion is permanent. However, objects can be made inactive.	Would need to incorporate your own system for handling this.	Deleting is permanent. Items can also be 'removed' which hides them from public viewing.	Deletions are permanent. A roll-back facility is not provided.	Permanently deleted. Roll-back may be possible from system back-up facilities.
<i>Authentication</i>	LDAP supported. Shibboleth not yet supported. Work at LSE on Shibboleth support is currently underway.	LDAP supported. Shibboleth not yet supported but is on the development plan.	No information as to whether LDAP or Shibboleth supported.	LDAP supported. Shibboleth not yet supported.	LDAP supported. Shibboleth compliance is planned for a future release and is currently being developed in partnership with LSE.	LDAP supported. Shibboleth compliance in progress via a project at University of Newcastle.
<i>Range of acceptable file formats</i>	Will accept all file formats.	No limits.	There are plug-ins provided for a variety of file formats which convert them to GSAF.	Will accept all file formats.	File formats do not need to be defined within the system.	Pre-configured with numerous file formats. Others can be added.
<i>Migration of file formats for digital preservation purposes</i>	Not currently supported. The Infocon project is looking at automatable format migration.	Not currently supported.	Not supported.	Conversion of files is being investigated for future development.	Not currently supported but under consideration for development.	Not currently supported.
<i>Real-time indexing and updating</i>	Supported.	Supported.	Background batch processing.	Supported.	Supported.	Supported.
<i>E-commerce facilities</i>	No	No.	No.	Under consideration for	Yes – several	Planned for

MIDESS: WP 2 – Functional and Technical Requirements Specification

				future release.	customers have been using this facility for a number of years through third party products.	future release.
<i>Compliance with accessibility legislation</i>	V1.4 will be WAI compliant.	FEDORA ships with a basic front end which is not compliant. The recommendation is to build a custom-front end for user access.	Unfortunately it does not comply with current UK legislation.	Compliant.	Meets all ADA and Section 508 guidelines.	Complies with WAI priority 1 and most of priority 2.

4 Summary analysis

	Perceived strengths	Perceived weaknesses
<i>Greenstone</i>	<p>Open Source</p> <p>XML based</p> <p>Easy to extract data if required.</p> <p>Simple to get up and running to a basic level.</p> <p>Metadata specifications seemed highly flexible and configurable.</p> <p>High speed searching once indexed.</p> <p>Supported by UNESCO</p> <p>Browse and search features seemed good.</p>	<p>Very few users in the UK.</p> <p>Weak end-user interface</p> <p>Proprietary approach to metadata.</p> <p>Full metadata can only be retained by storing as a separate object.</p> <p>Unusual structure (gsaf) etc.</p> <p>Use of proprietary archive/metadata format (gsaf) which adds a requirement for an extra layer of conversions in/out.</p> <p>No real time indexing.</p>
<i>FEDORA</i>	<p>Open source.</p> <p>API's for searching and storing objects so Fedora should be easy to integrate with other systems.</p> <p>Rich Object Behaviour provides for mechanisms for describing the tools that act on objects within the repository.</p> <p>Good Scaleability.</p> <p>Batch upload support.</p> <p>Fully functioning versioning system.</p> <p>PDF conversion services.</p> <p>Image manipulation services.</p> <p>Could provide the backbone of a very powerful system.</p> <p>GUI/customisation software available from the OS community.</p> <p>Any metadata can be added as an additional datastream (METS, MARC, MODS).</p> <p>Referenced and managed content – referenced content stored outside of the system so as to balance storage requirements.</p> <p>Support for SSL</p>	<p>Considerable programming resources expected in order to tailor this product to requirements.</p> <p>MIDESS partners may require more "out of the box" functionality delivered from day 1.</p> <p>Fedora is more of a building block than a finished repository.</p> <p>Some questions as to the long term future of the product.</p> <p>Upload process looked complex.</p> <p>Need to get heavily involved with the community to ensure success.</p> <p>We are aware of only two current users in the UK.</p>
<i>DSpace</i>	<p>Open source.</p> <p>Commonly in use throughout the UK and USA especially at universities. There is therefore a large active academic developer community.</p> <p>Submission process seemed straightforward – simple web-based wizard.</p>	<p>Weak end-user interface – not clear how to access the digital objects.</p> <p>Java programmer needed in order to tailor the product to requirements.</p> <p>Poor support for relationship management between objects.</p> <p>Primarily focused on Dublin Core as a metadata schema. Poor support for other metadata.</p>

MIDESS: WP 2 – Functional and Technical Requirements Specification

		<p>Much of the important functionality appears to be available in the next release (1.4)</p> <p>Minimal tools for batch support (system administrator must do bulk import via command line).</p> <p>No Support for versioning</p> <p>Possible problems with scalability</p> <p>Lack of detailed user statistics on reporting</p>
<i>Digitool</i>	<p>Very mature product.</p> <p>Linking functionality which can be embedded into a VLE</p> <p>Web services developments very positive.</p> <p>Support for open standards.</p> <p>Good user interface.</p> <p>Good support for JPEG 2000</p> <p>An evaluation version could be obtained for the MIDESS project.</p> <p>Generally good metadata support but limited in range of DTD's provided.</p> <p>The functionality provided to load and create content seems easy to handle by non-library staff.</p> <p>Handling of multiple documents attached to single metadata record felt to be strong.</p> <p>Additional functionality built in via licensing other 3rd party products should provide best tools available</p>	<p>No direct access to external objects. They are stored in the repository and accessed from within the repository rather than as separate objects.</p> <p>Questions regarding flexibility of Metadata. Dublin Core and MARC seem to be the main expected options.</p> <p>DigiTool doesn't support versioning except as a compound object.</p> <p>Metadata editor is a client tool, although other modules are web-based. This assumes a model whereby only the administrator or other library staff are creating detailed metadata – this may be the appropriate model to use, but do we need/expect academics to be creating metadata as well? If so they will need to download and install the Metadata editor.</p> <p>Additional functionality built in via licensing other 3rd party products may have cost implication since Digitool may not be responsible for costs/upgrades of 3rd party products.</p> <p>No ability to sell digital material from the digital repository in the current version</p>
<i>Symposia</i>	<p>Links with University of Leeds existing infrastructure.</p> <p>Java admin client is already familiar to Leeds staff.</p> <p>Integration with local federated search infrastructure.</p> <p>Support for open standards.</p>	<p>.Appears to be focussed primarily on documents rather than images, video, sound or other multimedia material.</p> <p>Appears to be aimed primarily at librarians rather than end users.</p> <p>No evaluation copy available.</p> <p>Ability to import data in anything other than Dublin Core appeared to be limited.</p>
<i>Curator</i>	<p>Powerful navigation and search.</p> <p>Good handling of multi page JPG's</p> <p>Admin interface uses powerful drag and drop facility.</p>	<p>Some improvements in interface required (promised second quarter 06).</p> <p>Client based administration rather than web.</p>

MIDESS: WP 2 – Functional and Technical Requirements Specification

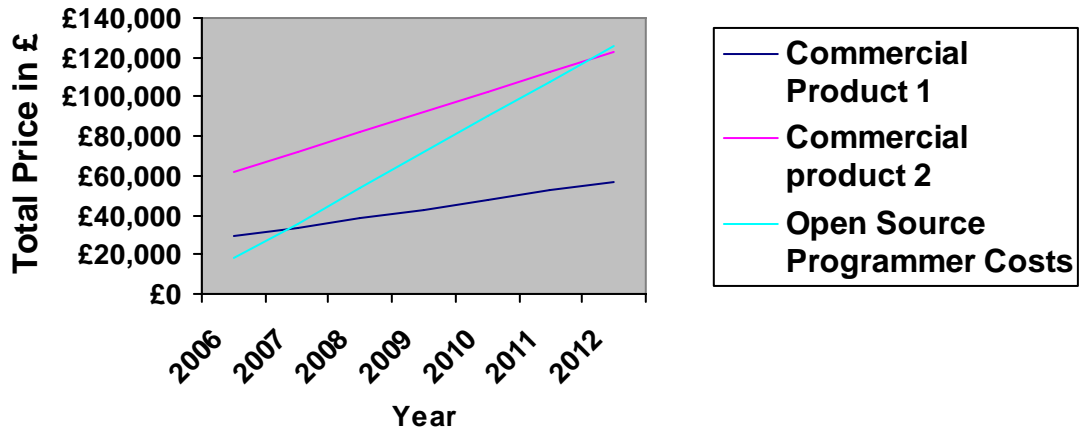
	<p>Impressive list of existing users (including the Getty Museum). Strong on archive/special collections type records (EAD). Appeared to have good support for metadata formats. Strong commitment to open standards. Ease of use by non-library staff. Can link to external objects outside of the system Support for SRW/SRU Ability to sell digital material if required (with third-party supplier shopping basket software). Good WebCT/Blackboard integration. Storage repository layer flexible and scalable.</p>	<p>No current built-in METS viewer. Drag and Drop functionality was felt to be confusing for a minority of people. Preservation functionality not currently supported (planned in a future release).</p>
--	--	--

5 Financial considerations

As part of the product evaluation, whole-life costings were developed for the products. Whole-life costings were developed for both commercial and open source products, using the following assumptions:

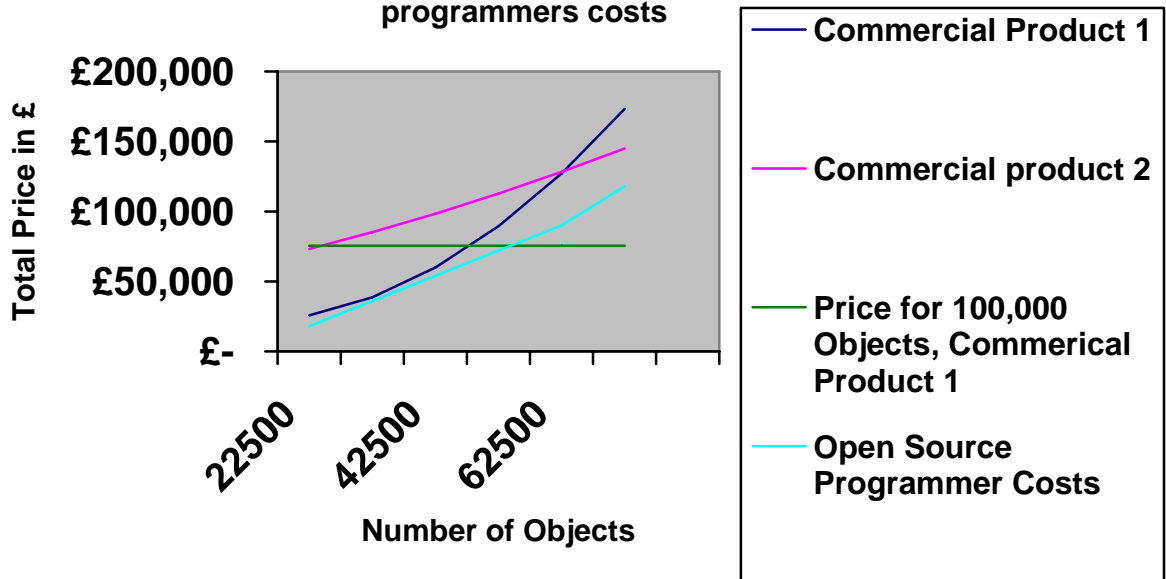
- 1) There are no software costs for the open source products. However, the ongoing staffing requirements have been shown to be greater than for commercial products, as significant development resource is required on an ongoing basis. We have assumed that 0.5 FTE systems staffing is required.
- 2) Commercial products are offered under a variety of licencing arrangements. These include the number of objects in the repository, the number of simultaneous end-users of the repository and the number of depositors to the repository. We have taken two different pricing models and compared these using a range of bands for number of objects and users.

Commercial Product 1 vs Commercial product 2 -Total prices assuming 20,000 Objects

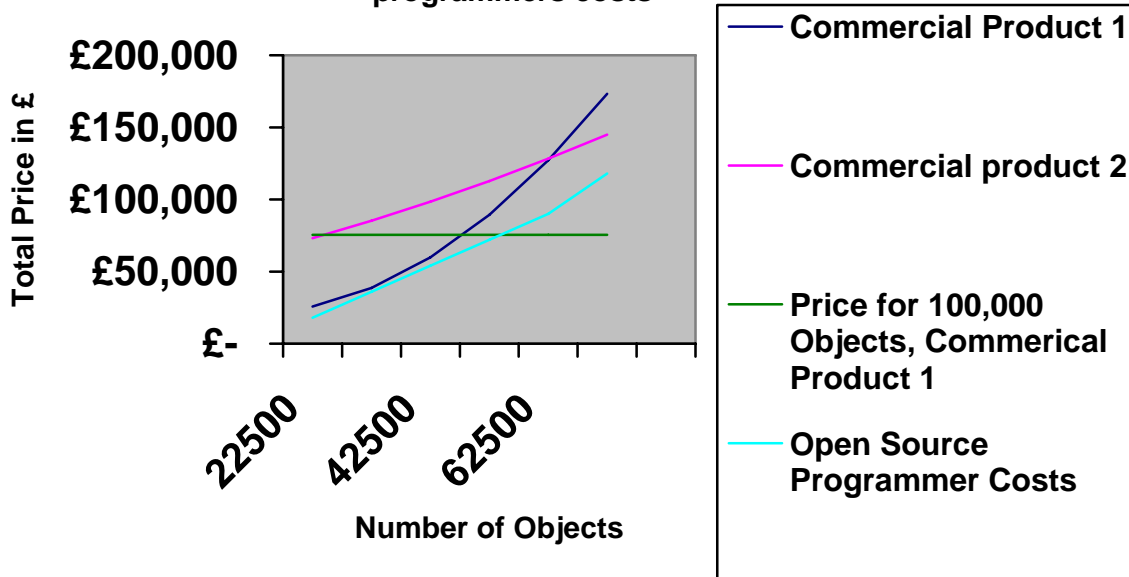


The graph above assumes a total of approximately 20,000 digital objects in the repository. This demonstrates that when the number of objects in the repository remains at around 20,000 then commercial product 1 is more cost-effective than commercial product 2.

Price vs Number Of Objects for Commercial Product 1, Product 2, Product 1 with 100,000 objects and open source programmers costs



Price vs Number Of Objects for Commercial Product 1, Product 2, Product 1 with 100,000 objects and open source programmers costs

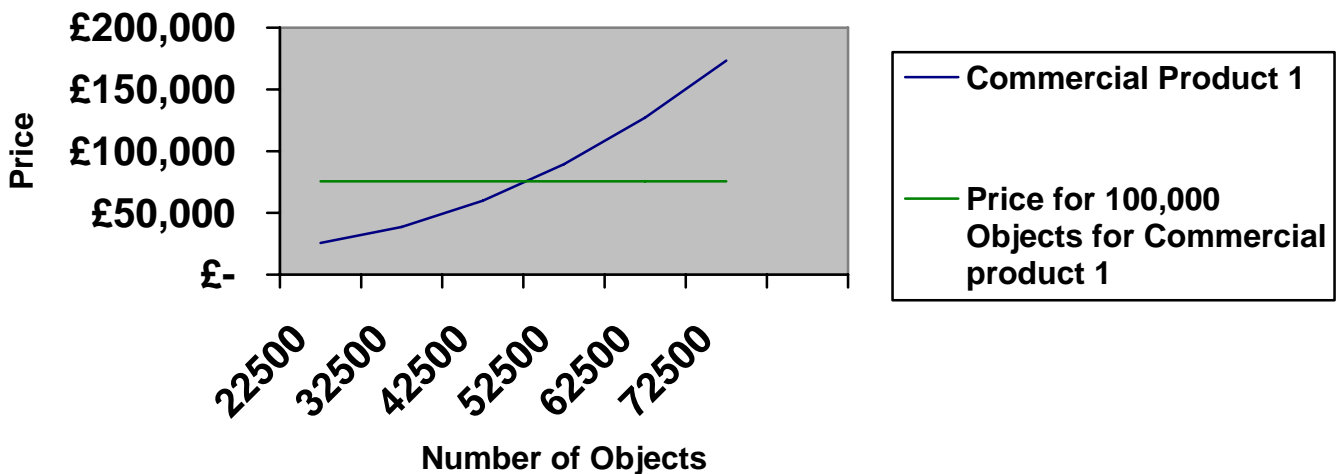


All figures exclude VAT.

Commercial product 1 increases in price with number of objects more rapidly than commercial product 2 however at a 100,000 objects there is a fixed price for commercial product 1. This is represented by the green line.

This is shown more clearly in the graph below.

**Chart showing intersection for Price vs Number of Objects
Incremental price (per 5,000 objects) vs fixed price for 100,000 objects**



All figures exclude VAT.

At 100,000 objects there is a reduced price for commercial product 1. The graph above demonstrates that when the number of objects exceeds around 52,500 objects (the

intersection of the two lines on the graph) then it is more cost effective to purchase a 100,000 object license for commercial product 1 than to continue purchasing object licenses in 5,000 segments above the 52,500 number of objects point.

6 Overall conclusions

Based upon the feedback from the presentations, Greenstone and Symposia were ruled out by all the MIDESS partners. Greenstone was ruled out because it was felt to be too limited in functionality to meet the needs which had emerged from the User Needs Analysis. Symposia was ruled out because it was felt that the product was in early beta release stage and lacked the functionality that the MIDESS project required.

6.1 University of Leeds

As stated in the User Needs Analysis, the University of Leeds' requirements are for an institution-wide repository rather than for a departmental or subject repository.

There is a huge variety of digital material at the University of Leeds. This consists of digital images, digital sound and digital video. The metadata for this material varies from non-existent to extremely complex over a wide range of subject areas. Therefore one of the major deciding factors in the choice of repository for Leeds was flexibility, both in the type of digital material being stored and in the metadata schemas supported. The system also needed to have the potential to integrate with other systems being used within the university such as the VLE and the Library Management System and the new University Portal System.

DSpace was ruled out at the University of Leeds because of the inflexibility found in the metadata structures. Metadata can only be defined at the system level, whereas Leeds has identified a need to define metadata structures at the collection level to match the needs of specific groups of users. DSpace v1.4 will provide support for output to METS; however, it cannot easily associate multiple digital objects with an individual metadata record. The batch upload features were also felt to be basic.

Whole-life costings were produced for Digitool, Curator and FEDORA. Costs were produced for a 5-year planning period for both commercial products. FEDORA was costed based on an estimate of the staffing required over a 5 year period to support and develop the system. A commercial front-end for FEDORA was also costed as an alternative to development of a front-end from scratch.

The analysis of costs and licensing issues led to a recommendation for the purchase of the Curator product, which provides the best model in terms of whole-life cost and licensing, and provides a close match with the functionality required.

6.2 University of Birmingham

The University of Birmingham plans to develop a repository primarily focused on digital library content produced from their digitisation programmes. The metadata for these collections exists in Qualified Dublin Core format, so there is not a strong requirement for flexibility in the metadata schemas available in the product.

The University feels that Digitool probably offers the most solidity and stability in terms of product, although the only UK HE institution using the product is Liverpool John Moores University. There is a close fit with other e-Library software at Birmingham and integration with the institutional portal has already been established with these sister products.

The main drawback of a commercial system is up-front costs. There were also some concerns about the ease of getting data out of a proprietary system, should a need to move to another package arise.

MIDESS: WP 2 – Functional and Technical Requirements Specification

The open source solutions were felt to offer less specifically for images, as they tend to focus on more generic solutions. They are easy to harvest from, and make extensive use of standards so would offer some safety net for any uncertainty ahead – but do need more (or at least different) allocation of technical resource. Open source also requires a commitment to take an active part in the community. Of those demonstrated, DSpace is recommended as it is sympathetic with existing Java application support available within the institution. Birmingham is also looking at Eprints as part of the MIDESS project since they have an existing Eprints digital repository.

6.3 LSE

LSE have already experience of working with DSpace through other initiatives. Consequently, their decision is to work with open source products for the purposes of the MIDESS project. Thus, LSE have chosen to install DSpace in conjunction with Fedora as their digital repository for MIDESS.

The FEDORA/DSpace combination has been chosen for the following reasons:

- FEDORA is being used by some respectable institutions (e.g. the whole of the Australian national repository, ARROW; the British Library; National Library of Wales) – it is felt that FEDORA is emerging as the open source repository of choice.
- The ability to support different versions (eg: multiple file formats can be included for a single item).
- Support for rich metadata and ability to use METS.
- Potential ability to link different databases in the institution to provide a single combined repository.
- Felt to have strengths in Web Services, security and compatibility with Shibboleth
- The ability to include relationships between objects and deal with complex digital objects. It can automatically assess parent/child relationships.
- Felt to have better provision for preservation metadata than Dspace.
- Handle system plug-in for Persistent ID generation.