Digital lifecycles and file types: final report

Rights and Rewards Project

Digital lifecycles and file types: final report

Melanie Bates, Sue Manuel, Steve Loddington, Charles Oppenheim

May 2006

List of Contents

List of Tables

List of Figures

1. Introduction

The Rights and Rewards in Blended Institutional Repositories Project is funded by the Joint Information Systems Committee (JISC) under the Digital Repositories Programme. This represents a cooperative venture between the Department of Information Science (DIS), the Engineering Centre for Excellence in Teaching and Learning (engCETL) and the University Library. The two year project aims to establish a single Blended repository to meet the teaching and research needs of this institution. It will address the motivational issues facing depositors of teaching materials with a focus on the associated Rights and Rewards.

This digital lifecycles study will identify the most appropriate materials for submission to the project's demonstrator repository. This takes into account factors like: granularity, persistence and multimedia types that can be supported for both teaching and research materials. It also documents the existing lifecycles of these items and the tools and specifications needed within a repository frameworks to support these lifecycles. For example, it will identify appropriate granularity of teaching resources and appropriate methods for content packaging.

The results of the study will help to identify which types of files are currently in use, which formats should be supported by the repository system ultimately selected for the demonstrator repository. This information is likely to be of benefit to other projects and institutions in the process of setting up an Institutional Repository (IR). These tasks and the methods by which they will be accomplished are outlined in Table 1.

Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Data types for deposit, standard lifecycle and most appropriate metadata	Which data type to be accepted for deposit and longevity of the file types?	Data from questionnaires and information on digital lifecycles	Data types identified, lifecycles recorded and metadata identified

Table 1: Evaluation plan

2. File formats

Repository managers have important decisions to make about what types of files their repository will accept. LeFurgy (2003) states that:

> Few file formats are currently suitable for long-term preservation. A format is often controlled as the intellectual property of a commercial entity, which typically has a vested interest in hiding the underlying code base.

Investigations to identify file formats in use by lecturers at universities and their expectations for the long-term access to these files have been carried out. This has provided clarification of current views held, and highlights the often contradictory information surrounding expectations and practicalities. The Rights and Rewards project undertook research into digital lifecycles to help inform the decision as to whether to support all of the file formats available, potentially a vast number, or to concentrate on the full support of a narrower range of file formats that are routinely used by the majority of teaching staff at Loughborough.

2.1. Methods

Several methods were employed to identify materials for submission to the test repository including: recording files already available in repositories, databases and Loughborough's Virtual Learning Environment (VLE), a workflow mapping exercise and by making comparisons with other studies in this area.

Research to identify the most popular digital file formats currently deposited in a range of repositories and databases of materials to support teaching has already been completed. The repositories and databases targeted were derived from the Rights and Rewards project survey (Bates *et al*., 2005) of the attitudes of academics and learning and teaching specialists towards repositories of materials to support teaching. Loughborough University's VLE (Learn) was also searched and file formats regularly used by teaching staff were noted.

Academic teaching staff at Loughborough were interviewed to gather information about their workflows in relation to the creation of their teaching material. This workflow mapping exercise identified specific files the repository will need to support by responses to the following question:

1. In the main what types of files do you create? And in what formats?

The information gathered here has enabled us to focus our further investigations on a narrower range of file formats. The exercise also uncovered academics' expectations for how long their digital material should be accessible. Their responses to the following questions provided this information:

2. How often do you refresh your material?
3. Have you had any problems in the past with accessing files you have created in the past?

Studies at other institutions have been consulted to provide corresponding information about file formats used in teaching and research activities.

2.2. Identification of files in use

A full list of files identified as part of investigations into repositories and databases used for locating items for use in teaching are listed in Table 2. Of these the most frequently found were Microsoft Word documents, PowerPoint presentations, web pages and images (JPEG and GIF). Audio, video, database and spreadsheet files were much less frequently apparent.

Text files	Images
Word Documents (DOC) Rich Text Format (RTF)	Joint Photographic Experts Group (JPEG) Graphics Interchange Format (GIF) Tagged Image File Format (TIFF)
Audio	**Video**
MPEG layer 3 (MP3) WAVE (WAV) Musical instrument digital interface (MIDI)	Movie (MOV) Windows Media Video (WMV) Audio Video Interleave (AVI) Flash (SWF) QuickTime Virtual Reality (QTVR) QuickTime Movie (MOV)
Spreadsheet	**Databases**
Excel (XLS)	Microsoft Database (MDB)
Presentation	**HTML**
PowerPoint (PPT, PPS)	Hypertext Mark up Language (HTML, HTM)
Compression	**Other**
Zone Information Protocol (ZIP)	Executable (EXE)

Table 2: Files located in repositories and databases

Recording materials available in Learn resulted in a similar list of files. The most regularly submitted files were: Word documents, PowerPoint presentations and images. Other files recorded were: movie, compressed and executable files.

These results can be compared with a survey carried out by the JISC funded RepoMMan project. This survey investigated a range of issues surrounding the effective use of a digital repository for research output. They asked respondents the following question:

To help us understand the type(s) of file that you store as parts of your 'works-in-progress''

	Docs	Images	Audio	Video	S/sheet	Stats	Diags
Freq	93.0%	20.1%	4.8%	0.0%	38.4%	25.3%	5.2%
Smtms	4.8%	35.4%	10.5%	7.4%	34.1%	22.3%	15.3%
Rarely	0.4%	30.1%	24.5%	30.6%	12.7%	18.3%	18.3%
Never	1.7%	14.4%	60.3%	62.0%	14.8%	34.1%	61.1%
	Databs	**Presentn**	**HTML**	**txt/XML**	**Archive**	**Special**	**Other**
Freq	10.9%	41.5%	27.5%	27.9%	8.7%	6.1%	3.5%
Smtms	24.9%	44.1%	32.8%	24.9%	24.5%	3.1%	1.7%
Rarely	21.8%	10.5%	18.8%	23.6%	29.3%	4.4%	9.6%
Never	42.4%	3.9%	21.0%	23.6%	37.6%	86.5%	84.7%

Table 3: File types for ‘work in progress’

As a guide the following examples of file types were suggested to participants:

Document files	(For example: .doc .rtf/rtfd .pdf .xsd .ps)
Image files	(For example: .jpg/jpeg .gif .png .psd .tif/tiff .eps)
Audio files	(For example: .wav .mp3 .aac)
Video files	(For example: .wmv .avi .rm .mpg (and its variants))
Spreadsheet files	(For example: .xls .xsc)
Statistics files	(For example from a package like SPSS)
Diagrams or CAD	(For example from packages such as Visio or AutoCAD)
Database files	(For example SQL, MySQL, Oracle or Access files)
Presentation files	(For example, PowerPoint files)
Web pages	
Simple text files	(This would include .txt and XML files, for example)
Archive formats	(For example Zip or Stuffit files)
Specialist text formats	(For example from LaTEX)

The resulting information, shown in Table 3 (Green, 2005), clearly illustrates that text documents, presentations, and spreadsheets were the most commonly used files. This corresponds with observations made of the contents of repositories, databases and the Loughborough Learn server. It was noted that the use of spreadsheets was not as frequently seen on Learn.

Another source of useful information was the JISC funded MIDESS project, which addresses issues with the management of digital content in an IR. One section of the survey conducted by this project was directed at creators of digital material, the responses from respondents based at Leeds University are detailed in Table 4[1]. The questionnaire asked:

What file format do you use?

Image	
JPG	69.3%
GIF	35.3%
TIFF	25.5%
BMP	24.8%
PSD	9.2%
PCT	0.7%
Sound	
MP3	32.0%
WAV	26.1%
AU	3.3%
Video	
WINDOWS MEDIA	34.6%
MPEG	32.7%
QUICKTIME	26.1%
AVI	25.5%
REAL PLAYER	22.2%

Table 4: Multimedia file formats

Unlike the previous survey the focus of the MIDESS survey was into teaching output, hence the emphasis on multimedia file formats. The results demonstrate a preference for image files, with sound and video being much less frequently used.

[1] Email contact with Stephen Charles, 12th December 2005.

An exploratory investigation to discover the range of file types currently in use in a number of repositories and databases was undertaken. The repositories investigated were those referred to by Harris (2006). A full table of the results of this exercise can be seen in *Appendix A*.

2.2.1. Workflow at Loughborough University

As part of the workflow exercise Loughborough's IR Manager was consulted about the text files academics used for recording their research outputs. At present the University Library operates a policy of mediated deposit with the repository manager taking receipt of all incoming files. To date only Microsoft Word documents have been offered. Where copyright agreements permit these are converted to PDF format before depositing into the repository.

A total of 10 Academic Teaching staff at Loughborough University were interviewed to get a idea of the file formats that are routinely created as part of their teaching activity. Table 5 lists the files Academics reported using.

Text files	**Web based files**
Microsoft Word (9) PDF (4)	HTML (8)
Other Microsoft Office files	**Image files**
Microsoft PowerPoint (6) Microsoft Excel (3) Microsoft Access (1)	GIFF (4) JPEG (2) BMP (1)
Video/Animation files	**Other files**
Flash (1) QuickTime (1)	CAD (1) FileMaker Database(1) Hot Potatoes (1) LaTeX (1)

Table 5: File formats used by Academics at Loughborough University

From these surveys we can conclude that the files that should be considered for support in the project's demonstrator repository are: PDF, DOC, RTF, PPT, XLS, JPEG, TIFF, GIF, HTML, XML, ZIP, MP3, WAV, AU, BMP, PSD. Initially it seems sensible to support, as fully as possible, the most commonly used files. However, developments in teaching and learning are such that in the future a greater range of file formats may have to be investigated and supported. These includes a variety of multimedia files such as sound and video. To this end we will focus our research on lifecycles and specifications for the more limited range of file formats that this initial scoping study has identified. But, examples of a few additional file formats will also be considered.

2.3. Specific file formats

2.3.1 Image file formats

Two of the main categories of image formats are raster (bit-mapped) and vector. Raster images are made up of individual units, ‘pixels’, in a grid formation. Vector images are based on mathematical instructions that are rendered by drawing programs into an image. Bit-mapped images are the most common and they include TIFF, JPEG and GIF formats. Vector images are typically displayed within drawing programs and consist of lines and curves. The drawing tools in Microsoft (MS) Word create vector images as do the graphs and charts in Excel and PowerPoint. High end drawing programs like Adobe Illustrator, CorelDraw and AutoCAD also create vector images. Many proprietary vector formats exist but there is an Open Standard, Scalable Vector Graphics (SVG). TASI (Technical Advisory Service for Images) provide a comparison of the properties of both of these formats, see Table 7 (TASI: An introduction to the vector image format 2005). These properties offer both advantages and disadvantages for users of these two formats.

Raster formats (bitmaps)	Vector formats (object-oriented)
Used for representation of continuous tones - suited to 'natural' photo-realistic imagery	Used for drawings and diagrams that can be described by mathematically defined shapes and attributes
Grid/matrix structure	Mathematical or textual description
Resolution dependant: scaling-up will diminish quality	Resolution independent: scaling-up is easy with no loss in quality
Most common capture type	Not normally used for capture, apart from some specialised mapping applications
Most common Web type	Less common Web type
Layering of images less common	Layering of images easy and common
Usually larger in file size	Usually smaller in file size
Usually quicker to display	Usually slower to display
Difficult to convert to a vector format	Easily and often converted to a raster format

Table 6: Raster and vector images compared

Additionally raster formats can be compressed while vector images do not support compression. Vector images are easy to re-edit without loss of quality, unlike raster formats.

Scalable Vector Graphics (SVG) is on open format that is supported by Adobe, Macromedia’s Flash (SWF) is a proprietary vector image format. Both can be delivered over the web although Flash is more widely supported (TASI: Image File Formats for Teaching and Learning Vector Images Part I: Overview 2004).

There is a third category of image, the Metafile. These can contain both raster and vector images as well as textual information and the instructions for displaying them. They are sometimes used as containers for vector images when native formats are not supported, web display is one example (TASI: Image File Formats for Teaching and Learning Vector Images Part I: Overview 2004). Encapsulated PostScript is one example of a metafile that is based on PostScript. Most imaging software can save in this format and EPS has the advantage of being platform independent, although more reliable on MACs than PCs. Because of its ability to lock images and content for editing EPS became an industry standard for delivery to commercial printers. It has now been superseded by the PDF (TASI: File formats and compression 2005).

2.3.2. RAW file format

Some digital capture devices, e.g. digital cameras, have their own native capture format or RAW format. This format holds all data, in uncompressed state, at the original capture stage before any colour interpolation. There are two advantages to using the RAW format:

Advantages

1. The file size is smaller than converting it to a TIFF file.
2. As an archive file this is the closest to the original source.

There are also disadvantages to RAW formats:

Disadvantages

1. No standard RAW format exists.
2. Because they are proprietary manufacturers software must be relied upon to read and convert them to a standard file format. (TASI: digital cameras 2003)

Adobe have introduced the Digital Negative (DNG) Specification in an attempt to overcome the problem of standardisation of camera file formats. Their DNG format is intended for use with all cameras and makes support by software vendors much simpler (Adobe: Introducing the digital negative specification 2005). DNG also reduces the risk of not being able to access a file in the future because software manufacturers are no longer supporting a particular proprietary RAW format.

2.3.3. Text file formats

2.3.3.1. MS Word documents

MS Word is a proprietary format that is commonly used by the education sector. Some of the advantages and disadvantages in the use of this format are listed below:

Advantages

- Widely available and commonly known software package.

- Backwards compatible.
- Can be read by most text reader technologies.

Disadvantages

- MAC and PC versions of documents are not always compatible.
- End-user has to have MS Word or free Word Viewer 2003 installed.
- Not a good software package for long documents.
- This is a proprietary format and so there are risks associated with long term preservation.

Microsoft provide free viewers for some of their most popular office software. Word Viewer 2003, for example, allows you to view Word 2003 documents as well as previous versions of Word. Files saved in the following additional formats can also be opened:

- Rich Text Format (.rtf)
- Text (.txt)
- Web Page formats (.htm, .html, .mht, .mhtml)
- WordPerfect 5.x (.wpd)
- WordPerfect 6.x (.doc, .wpd)
- Works 6.0 (.wps)
- Works 7.0 (.wps)
- XML (.xml) (Word viewer 2003 N.D.)

Viewers for Excel and PowerPoint are also free, no Access Viewer is available.

2.3.3.2 PDF

PDF files are intended for cross-platform delivery. They display text as the creator of the file intended, and for this reason are widely used. It is possible to apply a variety of restrictions to the use of a PDF file including disabling copy / paste functionality and the option to save a copy of the file. Digital signing of a PDF is also becoming more common.

MAC OS X has a built-in PDF creation utility. The PDF created has to be converted for reading on a PC. One disadvantage of using this option is that it is not option is that the PDF file is not optimized, it is likely to be larger than a PDF file created with Adobe Acrobat Distiller.

Advantages

- Good cross-platform compatibility.
- Appearance of a document can be controlled.
- PDF files tend to be smaller than other formats.
- Superior font support.
- Security features allow control over viewing, printing and editing.

- Watermarks can be added.

Disadvantages

- Not everyone has access to Adobe Acrobat Reader or Apple's Preview.
- PDF files can be inaccessible to screen readers – documents have to be tagged for this purpose.
- There are preservation risks as this is a proprietary format.

2.3.3.3. HTML

Advantages

- Most universal and accessible format as long as coding rules for HTML 4 or XHTML are followed.

Disadvantages

- HTML created by exporting from software like MS Word can include strange coding.

2.4. Embedding or linking diagrams, charts and images

It is now common practice to embed diagrams, charts or images in a file, such as a Microsoft Word document. The advantages and disadvantages are outlined below:

Advantages

- The embedded objects travel with the file and will be displayed and printed without the user having the software package used for creation of the objects.
- The embedded object can still be edited from within the document.
- The embedded document is stored separately from the original file so that it remains intact.

Disadvantages

- File sizes are greatly increased by saving additional information within the file.

Linking to diagrams or charts is also a possibility with programs like Microsoft Word. Again there are both advantages and disadvantages:

Advantages

- Any changes made to the original are reflected in the document.
- File size is not inflated by linking to a diagram or chart.

Disadvantages

- If a document has to be moved to a new folder moving, storage device or laptop the accompanying diagram has to moved as well or the link will fail.
- The link to the diagram must be broken if changes to the copy in the document must not be reflected in the original, or vice versa.

The most appropriate option is, to a great extent, determined by the intended use of the document.

2.5. Character encoding

Character encoding "is a method of converting a sequence of bytes into a sequence of characters" (Ragget 1999). Over the Web this conversion is handled by servers sending the HTML document, in byte format, to a user agent (e.g. a browser) that interprets them as characters. Characters are not always rendered correctly by user agents' if the agent lacks a suitable font. Using character set standard hexadecimal values for missing characters may alleviate the problem (Ragget 1999).

2.5.1. Unicode standard

Unicode addresses problems with: duplication of characters among different character sets, different characters sharing the same numeric value, and limitations to the number of characters that can be encoded. (Tull 2005. p2)

> Unicode™ is a standard for a universal character set for encoding the characters in the scripts of all the world's languages. (Tull 2005. p1)

Microsoft and Apple operating systems both support Unicode as do Internet Explorer, Netscape Navigator, and Opera. Unicode is included in numerous Internet standards, for example HTML uses Unicode as its character set, and it is the default character set for XML (Extensible Markup Language) (Tull 2005. p4). The World Wide Web Consortium (W3C) Recommend that Unicode encoding is used for HTML and XHTML documents; if other character sets are used these should be declared (Ishida 2005). However, the Unicode standard does not define the way characters are visually represented. This is handled by font or text rendering software and it includes formatting options for bold, italics, etc. (Tull 2005. p3-4) Thus, rendering of characters can vary from one platform to another.

2.6. Fonts

Fonts can be Operating System (OS) specific which means that they will not be available to users of other systems. There are a core group of fonts that are more common to Windows and Mac systems. This list includes: Arial, Book Antiqua, Comic Sans MS, Georgia, Courier New, Tahoma, Times New Roman, Trebuchet MS, and Verdana. But these are by no means universal and may not be installed on Unix and Linux computers (Fonts N.D.1).

2.6.1. Embedding fonts

Embedding fonts ensures that pages display in the format that the creator of the page intended. This is because the fonts used travel with the document rather than relying on the fact that the end user will have them installed on their system.

2.6.2. Embedding fonts in PDF files

Embedding fonts in a document, such as a PDF file, will increase the size of that file. If fonts are not embedded in a PDF file and are not available on the users system the PDF viewer will emulate the missing font. This ensures that the document can be read but it may not display as intended; text may not flow correctly.

Acrobat Reader includes the following fonts and therefore they do not need to be embedded:

- TimesRoman, TimesBold, TimesItalic, TimesBoldItalic.
- Helvetica, HelveticaBold, HelveticaOblique, HelveticaBoldOblique.
- Courier, CourierBold, CourierOblique, CourierBoldOblique.
- Symbol, ZapfDingbats.

(Improving the Performance of Adobe® LiveCycle™ Designer Forms 2005. p7)

However, “font matching, font substitution, and font synthesizing have not proven reliable at achieving true font fidelity” (TrueDoc: overview 2005). This practice also raises important issues of font ownership (TrueDoc: Overview 2005).

One of the advantages of embedding fonts is that the author of the document has some control over what the user of that document can do with it. For example TrueType fonts can have the following properties: installable, editable, print and preview or restricted license. If the print and preview option is selected when embedding fonts users will only be able to open the document as read-only. For users of the document this can be a disadvantage as the text can not be edited and the content repurposed.

2.6.3. Embedding fonts in web pages

It is possible to embed fonts into web pages. Three of the major font formats can be used: PostScript Type 1, TrueType and OpenType. These have first to be converted from their native format to either Embedded OpenType (.eot) from Microsoft, or TrueDoc (.prf) from Netscape and Bitstream. Neither of these formats work with all font formats or are supported by every web browser (Mulder N.D.1).

2.6.3.1. Embedded OpenType

Microsoft have released a free conversion tool, the Web Embedding Fonts Tool "WEFT", that allows designers to create ‘font objects’. These are linked to their web page and allow for grater control of typographic display. Earlier versions of IE and other browsers do not support these embedded fonts but the text will display using the browser default font instead (WEFT3 Overview 2000).

Advantages

- The onscreen quality of EOT fonts is superior to TrueDoc.

- Microsoft have released a free conversion tool, WEFT, to create EOT fonts. Only commercial tools are available to create TrueDoc.

Disadvantages

- A security flaw exists in EOT, developers have been able to copy the complete font to their system.
- The conversion tool is only available for use with Windows OS.
- OET fonts can only be viewed by IE.
- Type 1 fonts cannot be converted.
- Because there is no progressive rendering of a web page with EOT the user has to wait for the embedded font file to be downloaded before viewing the page.
- One file has to be created for each font, this increases the load on a server and leads to increased download times. (Mulder N.D.1)

2.6.3.2. TrueDoc

"TrueDoc is completely independent of platform, operating system, application, resolution, and device" (TrueDoc: overview 2005). Its fonts are fully scaleable at any resolution and it also has built-in 'psuedo-bold' and 'psuedo-italic' for use when memory is an issue. Both can be used to reduce the number of fonts that are required to be stored in memory. TrueDoc can be used for display of all world languages (TrueDoc in embedded devices, 2005). A TrueDoc viewer needs to be installed before these fonts will display, but a greater variety of fonts then become available, at no extra cost, without having to install each font on a users system.

Advantages

- TrueDoc does not copy or use hinting information from the original font. Fonts are generated in the form of character shapes and are not permanently installed are installed on a users system. They are only temporarily available for the purpose of rendering and viewing the document (TrueDoc: overview 2005). This ensures that these fonts are not used by third parties, thus making it more secure than EOT.
- The PRF can be embedded into a document, or it can travel with it (TrueDoc: Features and specifications 2005).
- TrueDoc is supported by both Netscape and Internet Explorer (IE).
- Type 1 and TrueType fonts can be used.
- Web pages initially display the default font until the PRF is loaded, at which point the screen redraws to display the embedded font. Therefore the viewer has faster access to the content of the page.
- Multiple cross-platform tools are available for creating TrueDoc fonts. (Mulder N.D.2)

Disadvantages

- Full cross browser support is not as complete as it could be. Older browsers do not display these fonts and support in IE is for Windows OS only.
- The official tool for the creation of TrueDoc fonts is not as robust as Microsoft's WEFT.
- The quality of onscreen font display is inferior to that of EOT. This is due to the fact that the outlines and hinting of the original font are not used. (Mulder N.D.2)

2.7. Web delivery

Standards and guidelines for mark-up languages for the web are provided by the W3C. HTML has now been around since 1997 and some newer browsers are unable to render HTML 1.0. HTML is being superseded by XHTML, and XHTML and CSS are becoming standard. The W3C recommendation for XHTML is currently in its second edition with a working draft of XHTML 2.0 available. The use of valid XHTML, automatic tools, and manual checks for accessibility problems should be undertaken. XHTML documents declare their media type to user agents, such as browsers, so that they know how to expect the text and how to render it. Older browsers do not recognise XHTML declarations but many legacy browsers will display XHTML if it is declared as media type text/html. Many modern browsers will accept the media type application/xhtml=xml, Microsoft Internet Explorer will display XHTML only if it is declared as application/xml. XHTML. The use of style sheets will eventually lead to more standard display of documents across browsers (Pemberton 2004). Repository systems do not all support the deposit of websites, DSpace is one example. However, it is possible to create a PDF file of a web site, with internal links intact. This file can provide a record of the site, its presentation and structure for archival purposes.

2.8. Accessibility

2.8.1. Overview

Standards compliance and accessibility issues are interlinked. The Special Educational Needs and Disability Act (SENDA) was introduced into law in the UK in 2002. This act requires that educational establishments make 'reasonable adjustments' to ensure that all students have full access to educational information. This includes electronic teaching materials and building accessible websites. The W3C's Web Accessibility Initiative (WAI) (2006) produced a number of guidelines for the creation of accessible websites. The W3C also make available automatic validation tools to assist developers with the task of checking pages for accessibility. The Quality Assurance Agency's (QAA) Code of practice for the assurance of academic quality and standards in higher education outlines 24 precepts in relation to students with disabilities; these must be followed by institutions wanting to gain QAA accreditation (Quality Assurance Agency 1999). The guidelines cover all aspects of student experience including learning and teaching provision. Wherever possible standards and guidelines should be adhered to, digital resources should be non browser or platform specific and assistive technologies should be catered for.

2.8.2. Website accessibility

2.8.2.1. W3C Technologies

Device-independent accessible content over the web is becoming an ever increasing possibility. The W3C is an important source information for developers and specific areas of interest to the community are those technologies for creating Web material which will more fully incorporate semantic information about that material. These technologies are all at the recommendations stage and include:

XML: Extensible Markup Language, plus associated technologies XSL (Extensible Style Language), XSLT (Extensible Style Language Transformation) and XML Schemas.
SMIL: Synchronised Multimedia Integration Language.
SVG. Scalable Vector Graphics.
MathML:. Mathematics Markup Language.
XForms: The next generation of Web forms. (French 2000, p40).

2.8.2.2. Separate Content from Presentation

Maximum accessibility can be gained by giving users choice. Keeping content separate from the styling of a webpage can be achieved by the use of Cascading Style Sheets (CSS) or database-driven pages (French 2000, p38-39). Web pages with CSS enable users to apply their own formatting preferences by setting their browser to display their preferred style sheet.

2.8.3. Accessibility issues with embedded fonts

In the process of embedding fonts into a document they are transformed into vector shapes. This means that the information in the text is no longer editable, giving a degree of protection to the content. Vector shapes do scale without loss of quality but there may be implications for screen readers. Fonts that are embedded into a PDF file can be unembedded and scanned images, that screen readers are incapable of reading, can be converted into readable text (Downie 2004). There is ongoing debate as to whether PDF files are accessible for all users. Some believe that they are as accessible as HTML as long as they are appropriately prepared, this includes structure and tagging for screen readers (Defining acrobat PDF accessibility N.D.).

2.9. Cross-platform issues

Cross-platform file types are preferred as they have the potential to be accessed by the greatest number of individuals. PDF may fit the bill for access to information, but for providers of these files there are costs associated with purchase of the software and this may be a barrier for some. There is also the issue as to whether PDF files comply with the Special Educational Needs and Disability Act 2001 (SENDA). Under this act institutions must make reasonable adjustments to ensure that students with a disability are not treated 'less favourably' than non-disabled students. This includes access to information. The W3C state that "content authors can no longer afford to develop content that is targeted for use via a single access mechanism" (Gimson 2001). Designing device independent content and applications ensures that content is not fragmented and authors can deliver information to users via their preferred access mechanism.

2.10. Open v proprietary products

2.10.1. Proprietary products

Proprietary products may not be available to all users as payment is required. In addition, files created by the use of commercial products can often only be opened by individuals having the same product on their system. Some commercial vendors, like Microsoft and Adobe, may be relied upon to be around in the future, but where companies cease trading there are potential problems for migration of their products. Another drawback with commercial vendors is the degree to which they make information about their code sequences available, and the accuracy of some of this information. Without this information tools to render these objects over time cannot be developed (Survey and assessment of sources of information on file formats and software Documentation N.D.).

2.10.2. Open standards

Open Standards may provide a viable alternative as information about their file formats is freely available. OpenOffice is one example that is an Open Source multiplatform product. It is compatible with every other major office suite (OpenOffice.org N.D.). Other software is available that supports the OpenDocument Standard, for example StarOffice, and use of these products may resolve some of the issue for certain file formats. These software are based on an XML file format and therefore support preservation as well as avoiding the issues of being hampered by lack of code information and being locked into single vendor licenses. However, many institutions have invested heavily in their current software base. A change of policy would be costly and time consuming for computing service departments and users of new products alike. The disadvantage of relying on Open Source products is that they may not be developed and maintained, or they may not develop in the way that you had planned for.

2.11. Appropriate use of formats

Open, portable, or de facto standards are preferred (XML, RTF, PDF) for preservation.

Text files - ODF (Open Document For office applications) is a standard that has been drafted by the Organisation for the Advancement of Structured Information Standard (OASIS). It is their attempt to provide an alternative option to Microsoft's Office suite of programmes.

Image creation – ideally images should be created in either TIFF or PNG format. These are non-compressed and non-lossy formats.

Archive images – capture device proprietary format, RAW file, or TIFF (Uncompressed Baseline TIFFv6). These are all uncompressed files and therefore have large file size.

Delivery (Web) – to reduce the file size for delivery compression is important. Use JPEG (JFIF), GIF and PNG (open source 'standard'). PNG is an appropriate alternative for GIF. Greater functionality can be had by using JPEG 2000 (JP2), the image can be viewed in looked at in detail using the zoom tool and it has integral rights management.

LZW compression - should not be used as it is a copyrighted technology and not suitable for long-term storage and retention (TASI Choosing a file format N.D.).

Efforts should be made to create digital resources that are accessible to all users, where this is not possible an accessible alternative format (HTML, PDF, TXT) should be provided. One way to achieve this is to provide appropriate support staff to assist with digital resource creation.

2.12. Repository support for file formats

The AHDS (Arts and Humanities Data Service) (2005) list 14 different digital resource types and defines their preferred deposit formats, acceptable deposit formats, problematic deposit formats and problematic aspects if there are any. Their definitions are driven by the possibility for preservation for particular file formats. Preferred formats are popular formats that have good software support. AHDS can successfully preserve the significant properties of items of these formats. Acceptable formats are those that they believe they can probably preserve given their current software and skills base. Problematic deposits are those they have identified as being "very *difficult* to ingest and preserve" because of lack of expertise of knowledge within ADHS, difficulties in obtaining the appropriate software or over reliance on software or hardware that can not be replicated elsewhere.

Digital Resource Type	Word Processor Document
Preferred Deposit Formats	Rich Text Format
Acceptable Deposit Formats	PDF, Word WordPerfect StarOffice / OpenOffice
Problematic Deposit Formats	Early versions of word processor packages. Word processor packages for platforms other than Windows, Mac, Unix, Linux
Problematic Aspects	-

Digital Resource Type	Raster Image
Preferred Deposit Formats	TIFF v6+, PNG

Acceptable Deposit Formats	GIF BMP PCX Photoshop* Paintshop Pro* CGM PhotoCD GeoTIFF
Problematic Deposit Formats	Any lossy compression (e.g. JPEG) Minority image formats (e.g. .bob) PDF
Problematic Aspects	-

Table 7: Example of preferred, accepted and problematic deposit format

TASI provide advice on file format selection for storage and preservation. The choice of a file format will depend on the type of image and how it is to be used. The choice is normally a compromise between quality (large file size) and usability (small file size). Different factors affect the choice of file format at each stage of the process:

- **Capture**: this is the most important step where it is important to retain as much data about the image, capture device and settings as possible.
- **Archival storage**: again as much information as possible should be retained, preferably in a standard lossless format that will be readable in the future. The ability to include metadata is useful.
- **Editing**: work in progress can be undertaken with proprietary formats. Maintaining an audit trail of image processing undertaken is also useful.
- **Delivery**: the output device and its capabilities should be considered along with the delivery method; and any quality requirements (compression, file size). A common, standard format is the preferred option. (TASI: File formats and compression 2005).

Levels of support for file formats deposited in a repository can be defined by DSpace repository managers. For example, the University of Rochester provides a list of file formats and the degree to which the repository can support them. The levels of support are defined under the following headings:

- **supported**: we fully support the format
- **known**: we can recognize the format, but cannot guarantee full support
- **unsupported**: we cannot recognize the format; these will be listed as "application/octet-stream," aka Unknown (UR research format support 2005).

Defining levels of support may be one means of accepting less commonly used file formats for deposit into the repository without attaching a commitment to preserve these files. Another possibility would be to create an 'informal' repository space where users can deposit a wide range of file formats and content types. This would be accompanied by a more strictly controlled 'formal' repository space where persistent access, re-use, discovery and delivery of accessible resources was maintained.

3. Digital lifecycles

3.1. Introduction

This section provides an overview of the subject of the lifecycle of a digital item for use in teaching. It outlines why repositories need to consider the lifecycle of these objects and touches on the subject of preservation of digital items. Examples of digital lifecycles are given; these are based on the experiences and practice of academic lecturers at Loughborough University. Hypothetical lifecycles are also introduced to demonstrate possible future digital item lifecycle. Factors that impact on digital lifecycles are also discussed.

3.2. Digital lifecycle and persistence

The lifecycle of a digital resource for use in teaching and learning depends on the nature of the resource, the length of time it remains useful and whether it can be re-used. Campbell *et al*., (2004, p6) state that:

> As digital learning resources of all types proliferate and evolve it is necessary to review the need for related information management functions such as access, preservation and disposal.

The persistence of such resources is a vital consideration within lifecycle management. Persistent resources may be either static or dynamic. The former remains the same throughout its lifecycle while the later can be modified, with content being added or removed from the original. Different management strategies are therefore required for each type of resource.

Campbell *et al*., (2004, p7) describe the National Library of Australia (NLA) policy that takes into account three factors to ensure persistence, these can be summarised as:

1. maintaining an original archival version of the resource;
2. maintaining an acceptable copy of a digital item and access to it;
3. maintaining reliable access to the location of an item.

Repositories can provide the functions that are required for the preservation of digital items over time. The original file can be stored, a PDF version can be provided for repository users to access and each item is allocated a unique URL. The can also support the long-term preservation of a range of file formats.

The Survey and assessment of sources of information on file formats and software Documentation (N.D.) states that a digital object may have a lifetime as short as 5-10 years. Expectations for access to digital items are at odds with the realities of digital preservation of certain file formats. These formats are proprietary ones where companies do not make details of their code sequences publicly available. The accuracy of some of the information that is made available is also questionable. Without this information tools to render these objects over time cannot be developed (Survey and assessment of sources of information on file formats and software Documentation N.D.). Other issues to consider are being locked into single vendor licenses and whether software vendors' products are backward compatible. Discussions with practitioners at Loughborough have shown that, up to now, access to digital files has not proved problematic. This may be due to the yearly cycle of access and updates to digital teaching items. One notable exception is access to some material available through the University's CAL Launcher (Computer Aided Learning). The current CAL Launcher does not support 16 bit software and so approximately 50% of these materials are no longer accessible.

Preservation of materials in an Institutional Repository is one aspect that several respondents to the Rights and Rewards project academic survey commented on as being a positive reason for contributing to such a repository (Bates *et al*., Rights and rewards project academic survey: final report, 2005). For this reason it is important to ensure that the projects demonstrator repository can support the lifecycles of digital objects it accepts for deposit. Contributors may not want to submit materials to a repository that they do not trust to store and preserve their digital items. A trusted digital repository manages and preserves files to ensure long-term access to those files (Research Libraries Group, 2002).

3.3. Versioning

Managing content and keeping track of different drafts of a document can be problematic. The difficulties academics at the University of Rochester have faced in this area have been highlighted by Gibbons (2004, p15). Information gathered at interviews revealed that some academics resort to sending themselves a copy of the current version of a document so that they have a date-stamp for the file. She adds, that this is further complicated by the practice of maintaining different copies in multiple locations, and where collaborative documents are concerned. Gibbons believes that for these reasons there is a need for an authoring system to help with version control. However, this may not be an issue with the Rights and Rewards Project. Interviews with academics have indicated that this may not be such an issue for them as their teaching materials tend to be placed in the Learn server. This organises their materials by module and year. Amendments to these files involve overwriting the previous version so that multiple different copies are not stored. That is not to say that the demonstrator repository the project will develop should not store more than one version of a file but that those lecturers interviewed did not have problems in tracing different versions of their files. It is possible that the sample did not represent a ‘typical lecturer’ at Loughborough and it is important to note that the projects demonstrator repository is not intended to replace or take on the functions of a VLE.

Policy needs to be defined to make it clear whether changes to items in the demonstrator repository will be permitted. If works in progress will be accepted, or if alterations and corrections are allowed, then a versioning system will be necessary. Gibbons (2004, p33) identifies two important issues here: should one version replace another or should multiple versions be supported and displayed in the repository. And, if versioning is supported which version will be associated with the persistent URL. Will it be the item that was originally deposited, the current version, or the whole collection of versions? There may also be case for different guidelines for different content depending on the type material, formal / informal materials, and the various stages of the content lifecycle.

3.4. Methods

An analysis of lecturers’ workflow in the creation of their teaching materials has provided an indication of how resources are created and amended over time. Interviews with 10 academic teaching staff at Loughborough University have provided detail of the composition of resources; a timescale for the life of an individual item; at what intervals (when and how often) amendments occur; and whether these changes are minor or major. Hypothetical lifecycles have also been created to facilitate investigations into possible future scenarios for digital lifecycles at Loughborough. The process of identifying and charting digital lifecycles will be used to define policy for a demonstrator repository at Loughborough University. It will help to identify whether items should remain static or will dynamic resources be permitted. If revisions are allowed which categories of item should be modified over time and for what reasons.

3.5. Digital lifecycles at Loughborough University

The examples that follow are based on current practice at Loughborough as well as hypothetical examples introduced to suggest possible future impacts on digital lifecycles. This analysis has raised some questions for consideration:

- What constitutes the beginning and end of the lifecycle of a digital item?
- Is a minor alteration sufficient for us to say that an item has come to the end of its lifecycle and a new item should be created?
- What constitutes a minor revision? – A change to the text perhaps. But what if the change was needed due to a factual error?
- What constitutes a major revision? – The swapping of an image or introduction of a significant amount of additional text?
- Do minor revisions force the creation of a new item?
- Do major revisions force the creation of a new item?
- Is a new lifecycle begun at the change of a file version, i.e. saving an MS Word document from MS 97 to MS 2000?

There are many factors to be considered when looking at digital lifecycles, not least of which is the concept of a ‘useful’ item. When a module is removed from the teaching programme it may remain on Learn for future reference. Does the fact that a module is not taught imply that the digital items for teaching it are no longer useful? It may be that elements can be reused in another similar module or the module may be reinstated at a later date. There may also be a case for their use as an example or guide for new lecturers.

Lecturers have reported using the Learn server as an archive for their teaching material. There may come a time when archived material is no longer accessible or files may not be compatible with new versions of software. There is a danger in relying on the Learn server to perform the functions of an archive store. Repositories, on the other hand, can provide archival services to ensure long-term access to valuable resources.

3.5.1. Categories of digital lifecycle

A range of potential digital lifecycles have been identified; these fall into three broad categories:

1. Resources created from scratch by the lecturer (Figure 1);
2. Resources created using some items from external sources (Figure 2); and
3. Resources adapted from an external source (Figure 3).

Each of these categories of digital lifecycle is represented by one of the three diagrams that follow.

Lifecycle / workflow A

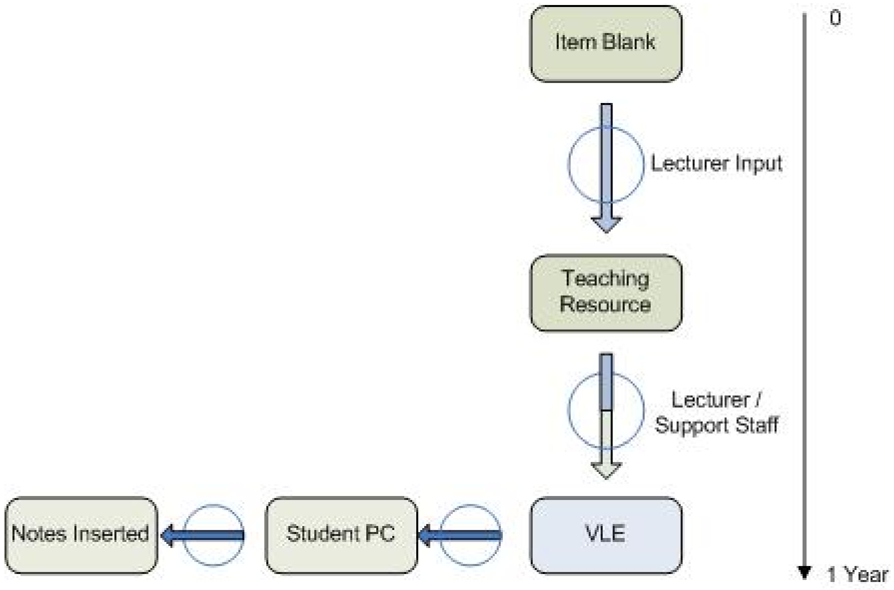

Figure 1: Digital item with a lifecycle of one year

In the example above a lecturer creates a digital item, from scratch, for use in their teaching. Having created the item a copy is sent to departmental support staff for deposit into the VLE. This digital item is held in the VLE for the current student cohort. A student copies the item to their PC and may add course / lecture notes, if it were a PowerPoint file for example. The lecturer's original remains in the VLE; the following year a copy is placed into the relevant directory in the VLE. Minor alterations are made to it. In this way the file remains current, so long as the institution updates software versions.

Lifecycle / workflow B

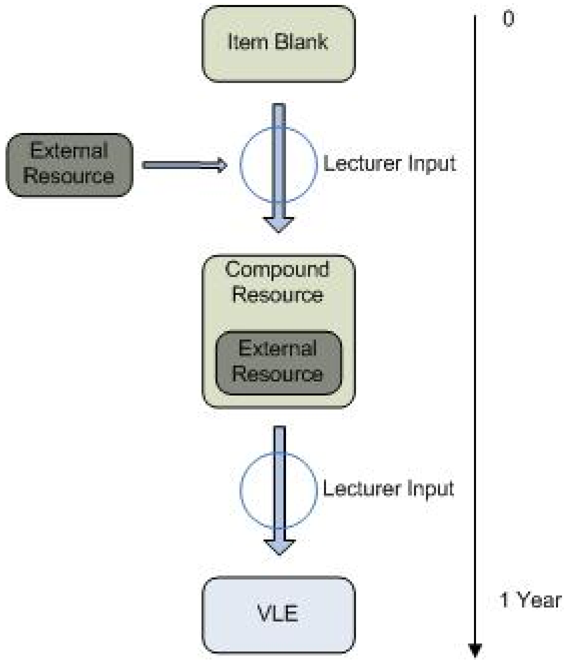

Figure 2: Lifecycle / workflow of a compound resource

This example illustrates the incorporation of an external item into the lifecycle / workflow process. This item could be an image that has been imported into a MS Word document, creating a compound (blended) resource. Its creator deposits the item into the VLE. Again this lifecycle follows a yearly cycle with the digital item is copied into the correct directory in the VLE. It is amended in readiness for use in the new academic year.

Lifecycle / workflow C

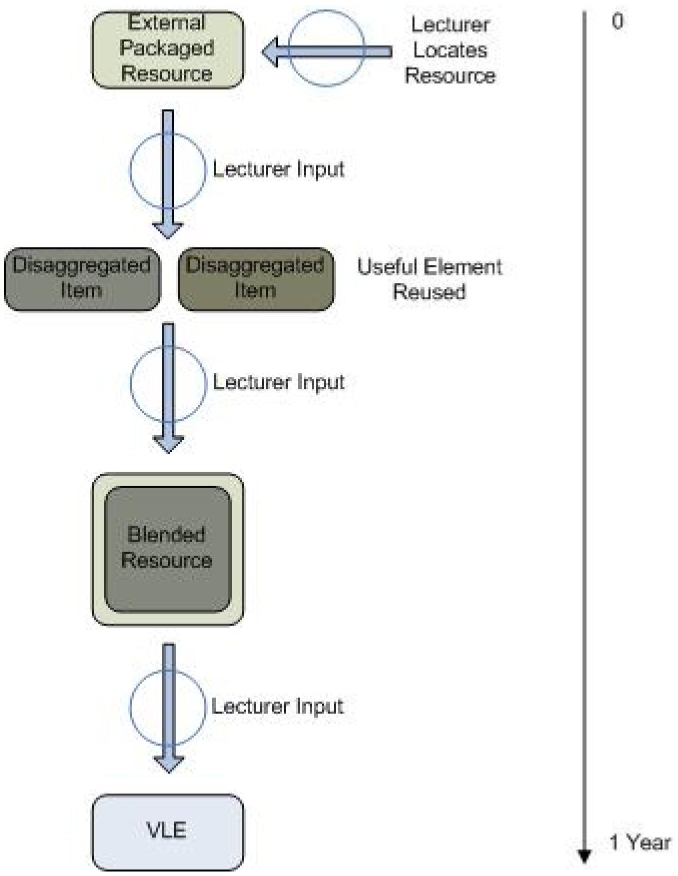

Figure 3: Lifecycle / workflow of a repurposed resource

This example shows how an external item initiates a new lifecycle / workflow process. This item could be a teaching resource that has been packaged to conform to the IMS packaging standard. Its creator had placed it in a repository for use by other teachers. It has been disaggregated and amended for use in a new teaching situation and the lecturer places it into the VLE. Once again this lifecycle follows a yearly cycle with a copy of the digital item being placed in the VLE in readiness for the new academic year.

3.5.2. Impact of the workflow on a digital lifecycle

Several examples of digital lifecycle have been gathered at the workflow analysis stage of the project. Many lecturing staff at Loughborough reported that digital items have a yearly cycle. Materials are created in the summer, prior to a new academic year, they are then placed in the VLE for students to access. In the summer of the second year items are copied to the appropriate VLE folders and amended in readiness for student access.

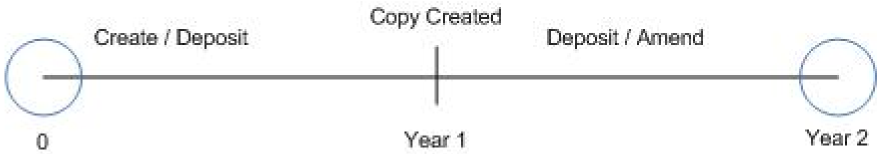

Figure 4: Digital lifecycle

In some cases a module may be removed or a new one created. This marks the end of one digital lifecycle and the start of another. This end of a lifecycle assumes that the digital file is no longer of use to its creator and is not accessed again.

3.5.3. Additional factors

To these base lifecycles illustrated in Diagrams 1-3 the impact of the following should also be considered:

- Refreshing and editing processes;
- Collaborative working methods; and
- Conversion of one file format into another e.g., MS Word to Adobe PDF.

The nature of some digital items means that they may require more frequent or major revision. Reading lists, exam questions and subjects where new ideas or data are regularly being discovered are examples lecturers gave of these. There are also times when factual information needs to be accurate and must be kept up-to-date.

Should an Institutional Repository mirror the working lifecycle of a digital resource? Digital items at Loughborough are ‘dynamic, they are edited on a yearly basis at the very least, but should a repository support this activity by permitting different versions to be deposited? If this is encouraged who will bear the responsibility of updating information? Lecturers are consistently expressing the view that they are overburdened with work. It is, therefore, difficult to imagine that they would be willing to take on this additional workload.

Collaborative working at Loughborough appears to involve a range of processes surrounding the creation of digital resources. For some collaborative modules lecturing and digital materials to support this activity are undertaken on an individual basis. For others the collaborative process impacts on the initial creation of a module, the division of lectures to colleagues, and suggestions for the creation of materials. Once this has been decided lecturers create their own digital items and make their own amendments.

4. Standards and specifications

4.1 Introduction

A range of standards and specifications are available for repository use. This section sets out why they are important, the benefits that can be gained by adhering to standards and specifications for individual repositories and for the wider community.

4.2. Standards and interoperability

Interoperability can be defined as: "The ability of systems and data to work seamlessly together" (CETIS Reference 2004). Adhering to standards is one way to ensure that maximum interoperability is achieved but it must also be noted that interoperability is hampered by the fact that different standards use different identifiers. For example, IMS Global Learning Consortium uses DOI (Digital Object Identifier) while OAI uses URI (Uniform Resource Indicator) (Looms, and Christensen 2002, p6-2). For storage, discovery and retrieval of learning objects from digital repositories the relevant standards are metadata, for describing the objects and content packaging for importing and exporting the objects (Duncan 2003. p6). Content packaging is discussed in Section 5. Although the Rights and Rewards Project is not setting up a Learning Object Repository, standards are equally important for a repository of teaching and research materials.

Dspace, one institutional repository system for investigation by this project, is a digital object asset management system. By default it supports Qualified Dublin Core (DCQ) and the interoperability protocols Z39.50 and Open URL. It also supports alternative metadata standards like Metadata Encoding Transition Standard (METS) and MAchine Readable Cataloguing (MARC). It provides persistent URLs. DC is widely used by repository systems as it represents the minimum level of metadata specified by the OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) for discovery interoperability (MacColl *et al.*, 2006, p22).

4.3. Metadata

Metadata, data about data, is used to describe objects so that they can be found. It facilitates useful and relevant search results, therefore aiding resource discovery. Complying with standards ensures that metadata can be harvested effectively by a range of discovery services across variety of repositories.

This is termed ‘federated’ searching and it attempts to maximise the search options so that the most appropriate resource is returned to the search query.

Day (2005) argues that metadata is now much more than just ‘data about data’, he believes “the importance of metadata is directly related to the roles they play in supporting the discovery, management and stewardship of digital resources”. Further enhancement to metadata can be achieved by the use of controlled vocabularies and formal classification schemes. Differing levels of metadata description can be utilised; a balance between describing an item for discovery and use and the cost that this entails must be drawn. The Learning for a Healthier Nation project was funded by the JISC in the X4L (eXchange for Learning) programme. A valuable lesson can be learnt from this project:

> …..good-quality metadata was essential if academic staff and library staff are to find relevant resources to support planned learning activities quickly and effectively. (Comrie 2005, p233).

Metadata can also be used to describe the relationships between individual digital items e.g., images in a collection or pages from a book. You can “mix and match” standards and you do not have to use all the elements in a standard if they are not appropriate for your needs. The role of standards is to provide guidance (VET Learning Object Repository Project 2003). One thing that is need is robust tools for the creation of metadata.

4.3.1. Dublin Core

DC has 15 elements but these are not sufficient to fully describe material for use in teaching. DC lacks descriptive elements for contextual information which, in many cases, is required for resources to be reused. Therefore, an extended schema might be required for richer additional metadata. To maintain interoperability additional metadata schemas can be stored internally, then mapped to DC. DC is not recommended for use in the museum community as it does not support the rich descriptions that museums need to share. This might also prove to be the case for items submitted to the repository by staff at the School of Art and Design at Loughborough University. The Dublin Core metadata initiative maintain a list of metadata tools (Tools and software N.D.). The list includes utilities, templates, DC metadata creation tools, and tools to automate the extraction and creation of metadata.

4.3.2. LOM

IEEE Learning Object Metadata Standard (LOM) was developed specifically for describing the educational attributes of a resource. It has nine categories of LOM and almost 80 elements. “LOM implements the IMS Content packaging standard, and so performs a similar task to METS in supporting the cataloguing of compound objects but within a pedagogical context.” (MacColle *et al*., 2006, p25).

LOM category	Description
General	Presents general information about the learning object such as title, language, and keywords
Life cycle	This category describes the history of the development of the learning object including changes that have occurred in features during its evolution, e.g., version and current status
Meta-metadata	Information about the metadata instance itself rather than the learning object that the metadata instance describes
Technical	This category describes the technical requirements and technical characteristics of the learning object
Educational	Outlines educational and pedagogic characteristics of the learning object
Rights	This category describes the intellectual property rights and conditions for using the learning object
Relation	Defines the relationships between the learning object and other related learning objects
Annotation	Provides comments on the education use of the learning object. Includes details about who created the comments and when
Classification	Describes the learning object in relation to a particular classification system

Table 8: Nine categories of learning object metadata (LOM). (Draft Standard for Learning Object Metadata 2002)

DC and IEEE LOM do not handle intellectual rights particularly well (Jones 2004, p.9). Digital Rights Expression Languages (DREL) are being developed so that this area may be covered in the metadata (Jones 2004, p.14). Standards in this area are not mature but it is clear from the responses to the Rights and Rewards survey that educational practitioners do have concerns about rights protection and this may be a barrier for sharing resources.

4.3.2.1. UK LOM Core

"The UK LOM Core is an application profile of the IEEE LOM that has been optimised for use within the context of UK education. The aim of the UK LOM Core is not to be prescriptive, but to identify common practice and provide guidelines for metadata implementers, creators and users." (UK Learning Object Metadata Core 2003).

There are currently two profiles in development:

- JORUM Application Profile draft 1.0 (Stevenson 2005)
- RDN LTSN application profile Version 1 (Powell 2005)

The RDN/LTSN profile has been developed to allow record sharing between these two organisations. It is also available for use by JISC X4L projects and it is suitable for teaching support material.

4.3.3. Exif data

When recording an image with a digital camera vital data about the camera and capture settings are created by the camera. Information is embedded in the image file as Exif (Exchangeable Image File) data, a specification for storing interchange information in image and audio files. The specification, Exchangeable image file format for digital still cameras: Exif Version 2.2, (2002), is published by JEITA (Japan Electronics and Information Technology Industries Association). The metadata stored in the image file can be read by web browsers, image editing and organising software and some printer drivers. Software tools are available which allow viewing and editing of Exif data.

4.3.4. Adobe XMP

Adobe's "eXtensible Metadata Platform (XMP) is a framework for adding machine-readable labels, or semantic content to application files, databases and content repositories" (Adobe: a manager's introduction to adobe eXtensible Metadata Platform, the adobe XML metadata framework, N.D.). XMP is being added to all Adobe software products and it gives creators of a digital items the opportunity to add their own metadata to files. Metadata can be recorded at all stages of the lifecycle of the file, from file capture onwards. The metadata provides information about the content of the file, its usage, author details, copyright data and other detailed information. More metadata can be added so that a record of changes can be made and retained. The metadata is saved within individual files; it is preserved even if a file is altered or converted to a different format. XMP is freely available via an open source license and it is standard compliant (Adobe: Controlling the data chaos by adding intelligence to media 2004).

4.4. Metadata Application Profiles

"Metadata Application Profiles are templates that enable organisations or communities to define the elements that are most useful to meet their particular needs." (Jones 2004, p.9-10). These profiles can include elements from a range of standards as well as custom fields designed to meet the needs of the institution.

4.5. Metadata Schemas

There are a number of metadata schemas available including: METS, MPEG-21 DIDL and SCORM. SCORM is discussed in Section 5, other schemas are detailed below.

4.5.1. Metadata Encoding Transmission Standard

Metadata Encoding Transmission Standard (METS) has been developed by OCLC and the Library of Congress. It is a standard for describing learning objects or learning materials. It is based on XML and includes four types of metadata:

- **Descriptive** – this includes the usual information contained in a library catalogue record.
- **Administrative** – information that is needed for handling, delivering, maintaining, and archiving an object. This includes access rights and copyright information.
- **Structural** – this describes the internal structure of an object.
- **Behavioural** – records details of software packages needed to render items.

METS has a hierarchical structure with a header, sections (with file groups and files) and a structural map. As with any XML file tags are used to describe, and point to, the file(s) in question. The metadata can be embedded in a METS file or held externally and a reference to it included in the METS file. Larger files are created by including the metadata in a METS file and this may be an issue in some circumstances.

The METS board endorses the use of certain schemas within METS. For descriptive metadata these include:

- Dublin Core
- MODS (Metadata Object Description Schema)
- MARCXML MARC 21 Schema (MARCXML).

For administrative metadata they recommend:

- Schema for Technical Metadata for Text (NYU)
- Library of Congress Audio-Visual Prototyping Project
- NISO Technical Metadata for Digital Still Images (MIX)
- METS Schema for Rights Declaration.

Institutions wanting to use METS may register their METS profile at the Library of Congress. These profiles are available for use of reference by anyone wanting to use METS.

4.5.2. MPEG-21

The MPEG-21 framework initiative is being driven by the need to share multimedia items across networks and devices (Bormans & Hill 2002). It is based on a user model where users are defined as an “entity that interacts with the MPEG-21 environment or makes use of a Digital Item” (Bormans & Hill 2002). MPEG-21 addresses one problem with digital items, that of dependencies. An example of a dependency is a JavaScript attached to a web page. When defining the extent of a digital item it is necessary to know the authors intent. The solution MPEG-21 provides is with the Digital Item Declaration (DID). This document specifies the markup, structure and organisation of a digital item, much like a content package contains an xml file to describe these associations. Figure 5 shows the hierarchical structure of the Digital Item Declaration Model.

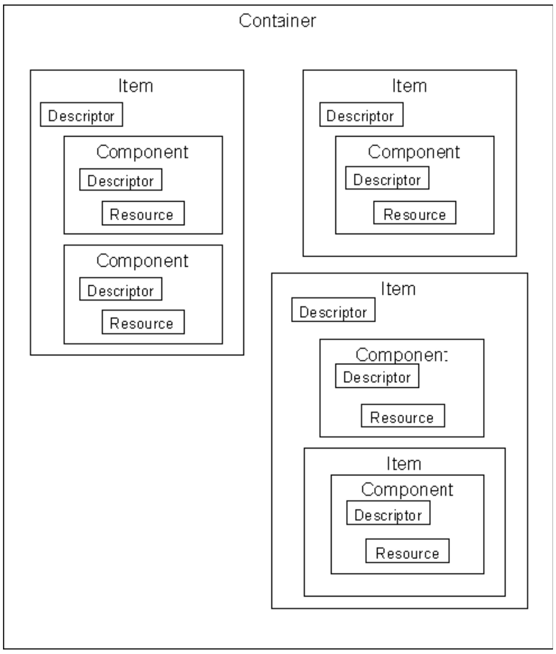

Figure 5: Relationship of the principle elements within the Digital Identification Declaration Model (Bormans & Hill 2002)

The MPEG-21 standard consists of a number of Parts defining digital item declaration, digital Item identification, Intellectual property management and protection, rights expression language, rights data dictionary, digital item adaptation, reference software and file format. Reference software packages are made freely available by MPEG (Moving Picture Experts Group).

MPEG-21 standard is nearly complete but the need for a new suite of standards has been identified. This is the Multimedia Application Format standard know as MPEG-A. This standard can be used by applications to build applications based on standards. The first application to be built is a music player. In the future it will be possible to develop a music album application that will include the album package, artwork, text and MP3 file (MPEG completes more elements of the MPEG-21 standard 2005).

4.5.3. MPEG-7

Is another standard from the Moving Picture Experts Group, it provides a set of tools for describing and encoding multimedia content so that audiovisual information can be exchanged between computer systems. MPEG-7 is designed to be application independent, supporting as broad a range of applications as possible. Its Description Schemas (DS) are metadata structures for describing and annotating audio-visual content like still pictures, graphics, 3D models, audio, speech and video. It provides tools for the description and management of a range of activities including:

- Content creation and production processes - director, title and so on.
- Content usage – for example copyright information and usage history.
- Content storage features - storage format and encoding.
- Content structural information - scene cuts, segmentation in regions and region motion tracking.
- Content low level features - colours, textures and sound timbres.
- Conceptual information - objects and events, interactions among objects.
- Content browsing - summaries, variations, spatial and frequency subbands.
- Collections of objects.
- User interaction with the content - user preferences and usage history.

The metadata is stored in XML format and descriptions at different levels of granularity are allowed. (Martinez 2004).

4.5.4. Metadata Object Description Schema (MODS)

MODS is a resource description format which is expressed using XML schema and language based tags rather than numeric ones. It was created by The Library of Congress who are the creators of MARC which "provides the mechanism by which computers exchange, use, and interpret bibliographic information". MODS can support MARC record metadata as well as original resource description. The Library of Congress also promote other standards that are relevant to this research, such as the Metadata Encoding & Transmission Standard (METS) and the Preservation Metadata (PREMIS) standard (MODS: uses and features 2006)

4.5.5. PREservation Metadata Implementation Strategies (PREMIS)

The PREMIS Standard is one which promotes the preservation of digital materials and using this standard allows digital materials to self document over time. Preservation is an issue that repositories need to investigate in order for materials within them to be sustainable. In 2005 the PREMIS working group released a data dictionary and created XML schema to support this. The data dictionary is suitable for a range of institutions and systems due varied nature of vocabularies, glossaries and special topics. Its international presence ensures that a breadth of support is available to those who chose to use PREMIS (PREMIS 2006).

4.6. Search protocols

4.6.1. Search and Retrieve Web services (SRW) 1.1

The SRW (Search & Retrieve Web Service) initiative is part of an international collaborative effort to develop a standard web-based text-searching interface.

The three main operations for SRW are:

- The **searchRetrieve** Operation: the basic operation by which queries and retrieval requests, and their responses, are passed between client and server.
- The **scan** Operation: enables the client to browse terms from indexes defined for at the server.
- The **explain** Operation: retrieves a document describing the capabilities of the server. (SRW service description 2004)

SRW is the operation used by DSPACE as a plug in for transmitting a search and getting back results. It uses Simple Object Access Protocol (SOAP) for the information exchange mechanism and a Web Service Description Language (WDSL) for record description (Search & Retrieve Web Service / Search & Retrieve URL Service N.D.).

4.7. Identifiers

4.7.1. Open URL

The OpenURL standard "is a protocol for interoperability between an information resource and a service component, referred to as a link server, which offers localized services" (OpenURL overview N.D.). However, in order to use this standard the information service being used must be OPEN-URL aware and furnish an Open URL for each object. Importantly users are directed to the most appropriate resource, this depends on their institutions subscription and access policies. Metadata is used to retrieve suitable objects such as journal articles or html web pages that comply with these policies.

OpenURL has a range of metadata formats for different types of materials, including, journal articles, books, dissertations and patents. Many information resources are Open URL enabled including: Biomed Central, Google Scholar, Oxford University Press and Swets Information Services (SFX Sources (OpenURL Enabled Resources) N.D.).

5. Granularity and Packaging Tools

5.1. Introduction

The issues associated with of the appropriate size of an object for reuse in learning are important ones that need to be addressed when setting up an Institutional Repository (IR). These issues include whether an item is a composite resource, comprising a number of individual items, and whether items need to be placed together to form a single unit that can be stored in a repository for future use and reference. Due consideration of the issues is needed so that appropriate solutions can be implemented. If resources are to be reused then they need to be of appropriate size and suitably described with information about the resource and a 'learning context' provided. If resources have to be packaged into a single file then a suitable packaging tool will be required. If other systems are able to share and understand these resources then standards must be adhered to. This report introduces these issues, sets out the reasons why they are so important to IR implementers, and describes some of the most popular content packaging tools in use.

An analysis of lecturers' workflow for the creation and deposit of items for use in teaching has been conducted at Loughborough University as part of the Rights and Rewards in Blended Institutional Repositories Project. The results of this analysis have provided information about the type of resources lecturers' routinely make available to their students. With this detailed data it is possible to suggest the nature of some of the items that a blended repository will be required to support. The Learning for a Healthier Nation project, funded by the JISC in the X4L (eXchange for Learning) programme, reported that in their experience:

> A key issue for academics was ensuring that technical support in their respective institutions was available when it came to incorporating resources into the VLE. (Comrie 2005, p233).

This is an important issue that needs to be taken into consideration when selecting an appropriate packaging tool for academics to use.

If an institution's resources for use in teaching and resources to support teaching activities are single items such as a PDF file or PowerPoint show than there may be no need to use packaging tools. But, if an institution wishes to make its resources available to a wider range of individuals then it becomes important to: describe these resources for discovery; ensure appropriate size or granularity of these resources; and to package individual items into a Learning Object (LO) with a tool that conforms to appropriate standards.

5.2. Granularity and learning objects

Learning materials may be an individual course or a collection of courses. LOs are digital items that can be reused to support the activity of learning. The possibility of sharing digital objects across institutions is being investigated by a number of JISC funded projects in the Digital Repositories Programme (2006). If we consider that reuse is a key factor then the granularity of an object becomes important. Duncan (2003, p12) states that “Granularization is a necessary condition for learning objects to be shared and reused.” McGill *et al*., have stated that it is “relatively easy to reuse small resources” but they “have limited context that could comprise their educational value” (2005, p637). Another vital element to be considered is the associated contextual information. This gives details of the purpose of a LO in an educational setting and as Duncan (2003, p15) states, small granules, such as an image file, may not contain this all-important information in its metadata, as the metadata does not describe the image file itself.

MacColl *et al*., (2006, p.16) explain that it is important to be able to describe resources at different levels of granularity. This means at the document level, which could be regarded the highest level of granularity, and the lowest level, its constituent parts.

LOs can be aggregated (put together) for reuse by others, but they then have to be disaggregated (taken apart) by the person wishing to access and make use of the resources in the LO. For this to happen specialist software is needed and for reuse across institutional boundaries interoperability is also a prerequisite. Therefore, adherence to the relevant standards and specifications is essential so that different systems are capable of reading these objects.

The Rights and Rewards survey (Bates *et al*., 2005) of academics and teaching and learning specialists highlighted a few issues that are of concern to these groups. Four specific comments respondents made were:

> keep content in as small as a size as possible. Keep content as modifiable as possible.
>
> I believe the level of granularity is either to high or to low. Producing something with a specific learning out come - would make reusability hard. Uploading just images or text also would make the amount of effort to compiling something worthwhile would also be time consuming. There has to be a happy medium.
>
> That the content be split up into bite size chunks so that I do not have to download a large file when browsing.
>
> What I don't want is ……….. to do anything to promote the illusion that learning/teaching can be divided up into little chunks and learnt by rote.

These comments indicate a range of concerns about the reuse of items for teaching and learning. The first comment stresses the need to get the size of the chunks of information in repositories right could be crucial to their success. There has to be a balance between overly large items with large file sizes that will be off putting to teachers, and items that are too small and would require too much work to repurpose for use in teaching. Jackson & Cooper (2003, p24-25) outline some of the problems associated with large LOs, these include some of the issues practitioners have identified, by their response to the Rights and Rewards survey, and also: problems with navigation being included in the packaged LO, this may cause broken links if the object is broken down into smaller items for reuse, and if LOs with their own internal navigation are brought into Learning Management Systems (LMS) then conflicts or duplication of navigation systems can occur. The use of small LOs can also create problems. Learning is enhanced by interactivity, LOs consisting of a single or limited number of items do not support this activity. Single items do not provide enough learning context and their value is therefore limited. Jackson & Cooper (2003, p25) comment that such approaches to the packaging of LOs "do not deliver 'good content' and can undermine a standards-based approach". Adopting a common sense approach to the issue of granularity and using metadata to describe resources as fully as is possible is advisable.

The latter comment, to the Rights and Rewards survey, would indicate that it is important for a clear definition of what a repository of teaching materials holds and what the intended use of those items is.

The repository that the Rights and Rewards Project will be setting up is not intended to house LOs, it is likely to contain single items that it will not be appropriate to package. It is not envisaged as a Learning Object Repository (LOR), neither is it a copy of the content of a VLE. A VLE may contain notes and working documents prepared with a student audience in mind. We propose a repository of good quality items designed to support the activities of teaching and learning. The audience for this material could be very different from the intended audience of a VLE. However developments at this institution may provide content packaging tools for staff and it may then be appropriate to accept these into the repository.

5.3. Content packaging tools

A content package is a structured file, usually a zip file, containing all relevant raw content files together with an XML manifest. The manifest defines all the content and their relationships to one another. The IMS Content Packaging specification (IMS CP) provides a common framework for the packaging and description of learning material. Metadata is a key component of an IMS CP. Compliance with this standard ensures that content can be distributed across a range of compliant LMS (VET 2003, p7).

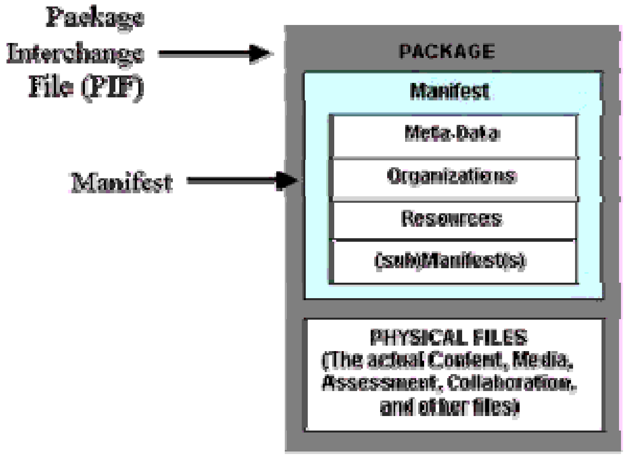

Figure 6: IMS Content Package (IMS Content Packaging Information Mode 2003)

Packaging files in a standard way facilitates interoperability and the distribution of packages in a digital environment. Packages describe the resources they contain, provide information about their structure and location (online). They also provide a definition of some content types. Packaging defines how learning content, of all types, can be exchanged between systems in a standardised way. They enable content created for use in one system to be used in another system, as long as both conform to standards.

The select an appropriate content packaging tool that will work with existing systems, has institutional backing, and with sufficient and appropriate support in place for users will assist take up. There are advantages and disadvantages in using content packaging tools and some of these are listed below.

Advantages

- Material can be uploaded to a repository in one zip file;
- Users can download material in one zip file;
- Content can be described with metadata;
- Conforms to specifications and is therefore sharable.

Disadvantages

- Learning curve for creators and users;
- An additional step is introduced to the process of making material available via a repository.
- Resources have to be disaggregated before they can be used.

Intrallect (N.D.) have identified a range of popular content packaging tools; a selection of these are listed below. Three of the most popular are RELOAD, CourseGenie and ScoBuilder. A brief description of these, and a few other noteworthy tools, is provided below.

- RELOAD

- Microsoft LRN
- CourseGenie
- ScoBuilder by Westcliff Data Services
- SCORM Manifest & Metadata Generator by JCA Solutions.

5.3.1. RELOAD

RELOAD rovides a range of tools including a Content Package and Metadata Editor, SCORM Player, and Learning Design editor and Player. RELOAD is rapidly becoming the de facto suite of tools for viewing, editing and creating IMS and SCORM Content Packages. Electronic learning resource creators can use the Reload tools to tag their materials with appropriate metadata, package it and implement Scorm runtime behaviours. (CETIS N.D.). The tools can be used with a variety of content such as web pages, images or text. A zip file is created and metadata is held within the content package.

RELOAD conforms to IMS specifications for:

- IMS content packaging (IMS CP), and
- IMS metadata specification.

RELOAD has been successfully used by several institutions and with popular VLE systems like Moodle and WebCT. This makes it a good choice for the Rights and Rewards Project and it certainly warrants further investigation.

5.3.2. CourseGenie

CourseGenie is a plug-in tool for Microsoft Word 2003. It can be used to create dynamic online courses. Once installed it can be accessed from the tools option in MS Word. When a course is generated CourseGenie tests for well formed XML; an error is generated if there are any problems in the code. The resulting HTML pages and java scripts are SENDA compliant. Multimedia elements can be linked to from the MS Word document. The associated media files can be automatically be copied into the output folder when the course is created. Supported movie types are: Windows Media, Flash and QuickTime. Streaming audio and video are also supported.

Metadata for a course can be entered by opening the metadata dialog box from the CourseGenie Menu within MS Word. Metadata is information about the course that can be used by IMS compliant learning management systems. Data for the following categories can be entered: Title, Description, Keywords, Objectives, Author, Organisation, Copyright, Version and Date. The same metadata can be used for several documents. This is achieved by selecting the option to export metadata from one document and importing it into another. The exported metadata is saved in a file called metadata.dat in the same folder as the Word document.

Content creators have the option to export in IMS or SCORM conformant content packages. The IMS package is generated in Microsoft's Learning Resource Interchange (LRN) Format. This is Microsoft's own implementation of the online learning content specifications of IMS (Banhazl 2000). Both of these content packages can be edited with the freely available Microsoft LRN (Learning Resource Interchange) Toolkit. The zipped content package can also be imported into a VLE.

CourseGenie has recently been introduced at Loughborough University to help academics make teaching material available to students via the VLE. Training sessions, and online material in its use have been created. Templates have also been created for a Module in Word and a web page. The Rights and Rewards Project could capitalise on the availability, intended use of and experience in this tool by supporting its use for the test repository. In this way it will not be too steep a learning curve for academics to create IMS or SCORM compliant material. It might also be possible to utilise existing support channels for its use. The Learning for a Healthier Nation project was funded by the JISC in the X4L (eXchange for Learning) programme have noted that:

> "A key issue for academics was ensuring that technical support in their respective institutions was available when it came to incorporating resources into the VLE." (Comrie 2005, p233).

This equally applies to the deposit of resources into an Institutional Repository. The table below lists some of the potential advantages and disadvantage to the adoption of CourseGenie as a content packaging tool at Loughborough University.

Advantages	Disadvantages
Readily available	Microsoft product – vendor lock in
Plugin to MS Word. staff at Loughborough are familiar with Word	Microsoft IMS may have restrictions / implications for interoperability
Support is available – training sessions, materials and templates	Can only be used with Microsoft products
Digital assets can be linked to	

Table 9: Advantages and disadvantages to the use of CourseGenie at Loughborough University

If LOs are to be included in the project's demonstrator repository then further consideration of the advantages and disadvantages of adopting CourseGenie needs to be undertaken before a decision to recommend it use is taken.

5.3.3. Microsoft LRN

Microsoft LRN (Learning Resource iNterchange) toolkit is free to download and it can be used for building content packages. It supports IMS content packaging and SCORM.

5.3.4. SCORM

SCORM conforms to IMS CP. It is able to create two types of content package, a resource package and a content aggregation package where metadata can be kept in an external file. SCORM provides support for learner management systems (LMS) through the use of five elements:

- Prerequisites
- Maxtimeallowed
- Timelimitation
- Datafromlms
- Masteryscore

These elements control the path a learner takes through the learning content. This additional functionality will not be required for the Rights and Rewards project repository as it will not be functioning in the way a VLE does. The materials in the repository will primarily be aimed at teachers and teaching and learning specialists. They will be for use in a teaching, and to assist teachers to create richer more interactive resources for student learning.

6. Frameworks and reference models

6.1. Introduction

The Research Libraries Group study (2002) states that “all trusted digital repositories must accept responsibility for the long term maintenance on behalf of its depositors”. The demonstrator repository that will be set up for this project is no exception and must consider preservation as one of its unique features. The Open Archival Information System (OAIS) reference model gives repositories a solution to preservation. The whole system or parts of it can be based around the applicable models to improve functionality and operation. Both conceptual and technical frameworks can inform many aspects of repositories. For example, research carried out by the E-Learning Framework (ELF) (2006) can be used to inform repositories on common services such as workflows, content packaging and metadata management.

6.2 The Open Archival Information System (OAIS) Reference Model

The reference model for an OAIS was created by the Consultative Committee for Space Data Systems (CCSDS) to “create a consensus on what is required for an archive to provide permanent, or indefinite long term, preservation of digital information.” OAIS is a de facto standard for digital archive architecture (Cervone 2004).

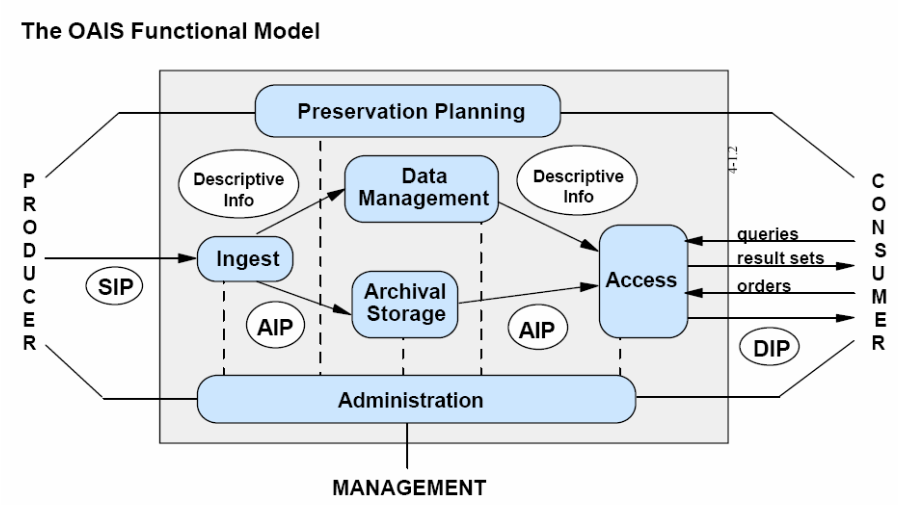

Figure 7: The OAIS Functional Model (CCSDS, 2002)

The preservation of materials within a repository is something that can be seen by contributors as a reason for contributing materials to it. Therefore the repository has to consider preservation strategies and techniques and promote preservation to users as being a unique benefit of the repository. In the Rights and Rewards UK Survey, many academics did not see any difference between a VLE and a teaching material repository, yet they are very different. The preservation of materials is just one of the major unique feature of a repository compared to other systems that material can be placed within, such as VLEs.

There are a number of OAIS-modelled repositories, including CEDARS, Harvard University's Digital Repository and the DSpace Digital Repository System. The Rights and Rewards Project aims to use DSpace as repository software for the demonstrator repository and therefore the OAIS model gives a good understanding of how DSpace works.

6.3 The eLearning Framework

The eLearning Framework is a joint initiative between the UK's JISC and Australia's Department Education Science and Training (DEST). Wilson *et al*., (2004a) state that:

> The ultimate aim of a Framework is, for each identified Service, to be able to reference one or more open specifications or standards that can be used in the implementation [of] the Service.

Figure 8 shows the three different levels of the framework covering a variety of services and subjects (The E-Learning Framework 2006)

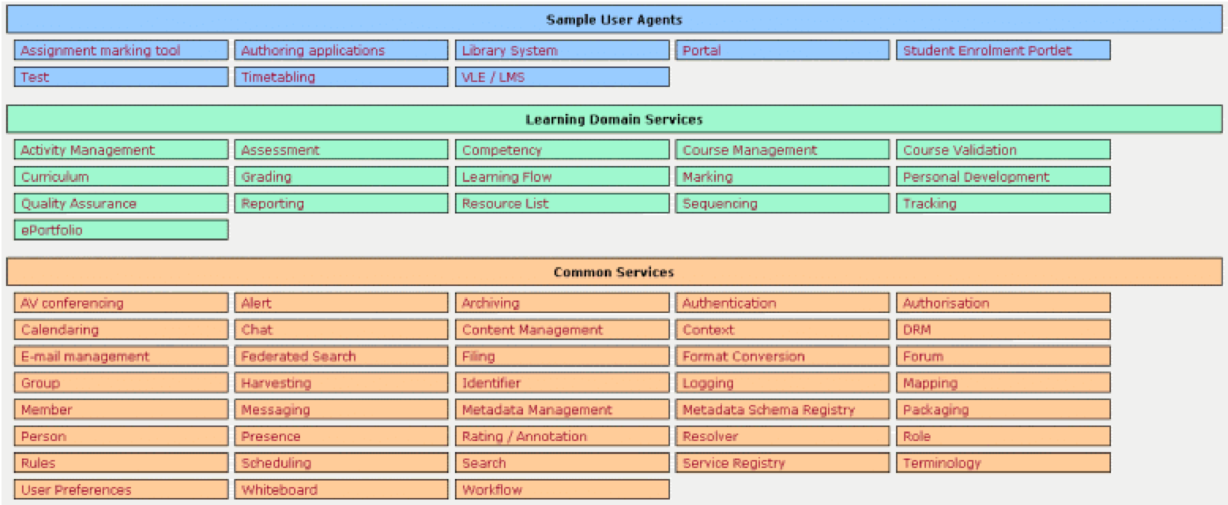

Figure 8: The eLearning Framework (The E-Learning Framework 2006)

Wilson *et al*., (2004b) stress that:

> The intention is not to provide a blueprint for an open-source solution, but rather to facilitate the integration of commercial, home-grown, and open source components and applications within institutions and regional federations, by agreeing upon common service definitions, behaviours, data models, and protocols.

This may turn out to be a useful framework for the Rights and Rewards project in the future when a number of the issues within the framework need further investigation. The range of open source and commercial solutions available means that there will be a variety of choice. With further work being carried out in many of these areas, means that information regarding any changes to existing specifications and standards will be available, as well as newly developed ones.

6.4. The e-Framework

The e-Framework for education and research is another JISC / DEST initiative. Oliver *et al*., (2005, p2) describe the main goal of the initiative as being "to produce an evolving and sustainable, open standards based service oriented technical framework to support the education and research communities". The e-Framework will allow service oriented approaches to the specification of network services. These can be documented in the form of reference models with input from domain experts and practitioners. The e-framework will benefit the community as it will allow interoperable services conforming to open standards to be derived.

6.5. The Open Knowledge Initiative

The Open Knowledge Initiative (O.K.I.) is an initiative between the Massachusetts Institute of Technology (MIT), partner universities and specification and standards agencies. Its aim is to develop and promote "specifications that describe how the components of a software environment communicate with each other and with other enterprise systems" (The open knowledge initiative 2006). The integration and interoperation of software is achieved by separating the ‘what’ from the ‘how’, or to put it another way the ‘consumer’ from the ‘provider’ (O.K.I.™ Architectural Concepts (Draft) 2003). Use of the O.K.I. architecture can benefit vendors, educators and developers alike and “at the core of the OKI are fully public software standards that define an open programming environment to foster pedagogical experimentation and sharing” (Siddall 2002). As the open programming environment’s intended method of delivery is via web browsers, software written to O.K.I. should be accessible to all users with a computer and Internet connection (Siddall 2002). The O.K.I. framework consists of four layers:

1. Campus infrastructure,
2. Common services,
3. Educational services, and
4. Educational applications.

The O.K.I. defines these layers, the functions of any services at each layer and how the services interact with one another (Siddall 2002).

The O.K.I. has published the Open Service Interface Definitions (OSIDs), these “define important components of a Service Oriented Architecture (SOA) as they provide general software contracts between service consumers and service providers” (The open knowledge initiative N.D.). Because OSIDs are only service contracts applications can be constructed independent of service environments. Additionally OSIDs are compatible with most other technologies and specifications and they can therefore be used with existing technologies and are equally at home with open source or vended solutions. The O.K.I. have developed the Repository Open Service Interface Definition (OSID) which is a specification for the storage and retrieval of digital information.

6.6. Ecology of a repository – a cosmic view of the repositories space

The model ‘the cosmic view of a repositories space’ highlights the fact that there are “no common methodologies as of yet to deal with multiple permutations of context” (Blinko & McLean 2004).

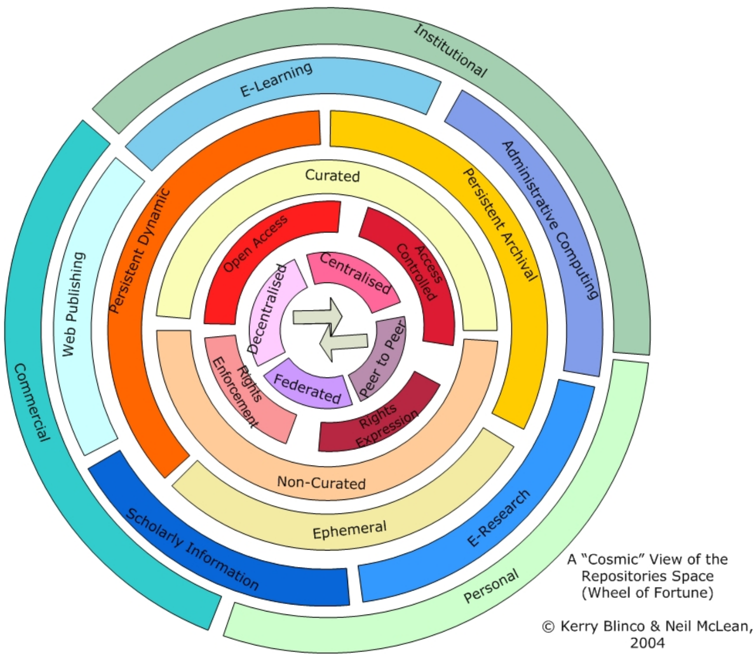

Figure 9: A cosmic view of the repositories space

It illustrates the many different contexts of repositories and the access, rights and dynamics that a repository can have. It is therefore important to make sure that these different contexts and features work together in a coherent fashion and that all different permutations are explored. Some of the associated issues were investigated in the Rights and Rewards survey (Bates *et al*., 2005) including access and copyright. The second outer layer shows the range of materials and strategies that a repository has, and for this project, it is important to investigate whether learning material and research output can be 'blended' into a single repository. Some factors such as the difference in metadata of these two types of materials, might mean that they cannot be put into the same repository.

7. Conclusion

Maintaining persistent access to resources, ensuring that different systems can locate and retrieve resources, updating files to preserve them for future use and making resources available at the appropriate level of granularity are all complex and important issues. This report has raised awareness within the project team of the complexity of the issues associated with digital file formats, lifecycles, metadata and standards relevant to repository services. Research conducted has shown that a range of file formats are currently being used by teaching staff at Loughborough and other UK Universities. However, the file formats in use by the majority of lecturers in their day-to-day activities of teaching are much more limited. They include: MS Office files, PDF, HTML, GIF, JPEG and BMP. The lifecycle of each of these objects may be a simple yearly cycle of: creation, use and deletion, or it may be much more complex with a range of individuals, processes, other digital items or file migration activities intervene. Some digital items are created and once lecturers are happy with the content only minor amendments are made; others are regularly refreshed and updated.

Standards and specifications are important for repository systems and future interoperability may be hampered if attempts are not made now to adhere to these. Recording file creation information, preservation and rights details in the metadata of each resource will aid content discovery and re-use.

It is likely that Learning Objects will not be included in the project's demonstrator repository for materials relating to teaching and learning. This is because the content the project team envisage for the repository will not require the use of content packaging tools. If academics and support staff are expected to use these tools then support channels should ideally be in place and widely advertised so that support is easy to locate. If a content packaging tool is required at Loughborough then there are obvious benefits to the adoption of CourseGenie as the preferred packaging tool. CourseGenie is already available at this institution and is being used by early adopters, for the creation of websites, and support is available.

As one of the repository systems that this project will consider is DSpace, the OAIS model may be used in future stages of the project. The successful preservation of materials was identified in the results of the Rights and Rewards survey as an incentive for contributing to a repository (Bates 2005), this feature is not something that all storage systems can implement. The size and scope of the e-framework means that the project will benefit by some of its activities. New developments in a range of common services including metadata, will be most useful for keeping current.

This report has highlighted a range of important issues associated with file formats, their lifecycle, current standards and specifications that are applicable to repository systems. The issues raised affect repository systems, their maintenance and development as well as users of repositories (deposit and retrieval of items). Careful consideration of relevant issues at an early stage in the process of setting up an Institutional Repository is advisable.

Bibliography

Adobe: Controlling the data chaos by adding intelligence to media, (2004). Available online at: http://www.adobe.com/products/xmp/pdfs/xmp_data_chaos.pdf.

Adobe: Introducing the digital negative specification, (2005). Available online at: http://www.adobe.com/products/dng/pdfs/dng_primer.pdf.

Arts and Humanities Data Service, (2005). *ADHS Deposit Format List*. Available online at: http://www.ahds.ac.uk/depositing/forms.htm.

Banhazl, P. (2000). *IntraLearn supports LRN 2.0 specification to accelerate midsize enterprise e-learning market*. Available online at: http://www.intralearn.com/4040_PR071400.asp.

Bates, M., Loddington, S., Manuel, S., & Oppenheim, C. (2005). *Rights and rewards project academic survey: final report*. Available online at: http://rightsandrewards.lboro.ac.uk/index.php?section=21.

Blinko, K., & McLean, N. 2004. The Ecology of repository services: a cosmic view. Available online: http://www.ecdl2004.org/presentations/mclean/n-mclean.pdf.

Bormans, J., & Hill, K. (2002). *Organisation Internationale de normalisation ISO/IEC JTC1/SC29/WG11 coding of moving pictures and audio*. Available online at: http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

Campbell, L. M., Blinko, K., & Mason, J. (2004). *Repository management and implementation: a white paper prepared for alt-I-lab 2004*. Available online at: http://www.educationau.edu.au/papers/repositories_Altilab.pdf.

CCSDS, (2002). Reference Model for an Open Archival Information System – Blue Book. Available online: http://public.ccsds.org/publications/archive/650x0b1.pdf.

Cervone, F. (2004). The repository adventure. *Library Journal*, 129(10), 44-46. Available online at: http://www.libraryjournal.com/article/CA421033.html.

CETIS Reference, (2004). Available online at: http://www.cetis.ac.uk/encyclopedia.

CETIS, (N.D.). *Tools for working with specifications*. Available online at: http://www.cetis.ac.uk/members/x4l/articles/tools.

Comrie, A. (2005). Learning and teaching: learning for a healthier nation. *Health Information and Libraries Journal*, 22(3), 232-234.

Day, M. (2005). *DCC Digital Curation Manual: Instalment on Metadata*. Available online at: http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata/metadata.pdf.

Defining acrobat PDF accessibility, (N.D.). Available online at: http://www.webaim.org/techniques/acrobat/.

Learning Objects from Cultural and Scientific Heritage Resources, (2003). *DigiCULT*, Thematic Issue 4. Available online at: http://www.digicult.info/downloads/digicult_thematicissue4_lres.pdf.

Downie, A. (2004). *The pros and cons of PDF files*. Paper presented at the Round Table Conference on Print Disability Sydney, 2004. Available online at: http://www.une.edu.au/eeo/writing/prosconspdf.pdf.

Doyle, C., & Robson, K. (2002). *Accessible Curricula: good practice for all*. Available online at: http://www.heacademy.ac.uk/embedded_object.asp?id=17323&filename+WDP012.

Dublin Core metadata initiative, (N.D.). *Tools and software*. Available online at: http://dublincore.org/tools/index.shtml.

Duncan, C. (2003). *Digital Repositories: e-Learning for Everyone*. Presented at eLearnInternational, Edinburgh 9-12 February 2003. Available online at: http://www.intrallect.com/products/intralibrary/papers/elearninternational_edin_feb2003.pdf.

Duncan, C. (2003). Granularization. In A. Littlejohn, (Ed.), *Reusing Online Resources: A Sustainable Approach to eLearning*, London: Kogan Page.

The E-Learning Framework, (2006). Available online at: http://www.elframework.org/framework/.

Exchangeable image file format, (2006). Available online at: http://en.wikipedia.org/wiki/EXIF.

Japan Electronics and Information Technology Industries Association (2002). *Exchangeable image file format for digital still cameras: Exif Version 2.2*. Available online at: http://www.exif.org/Exif2-2.PDF.

Final draft standard IEEE 1484.12.1-2002, (2002). Available online at: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.

Font embedding for the web, (2001). Available online at: http://www.microsoft.com/typography/web/embedding/default.htm.

Fonts, (N.D.). Avalable online at: http://www.webaim.org/techniques/fonts/?templatetype=3.

French, D., *et al*., (2000) Disability and Social Inclusion for the e-University. In P. Bacsich, & S. F. Bristow (Eds.) *The e-University compendium: Cases, Issues and Themes in Higher Education Distance e-Learning*, Vol 1. Available online at: http://www.heacademy.ac.uk/learningandteaching/eUniCompendium_chap23.DOC.

Green, R. (2005). Report *on research user requirements on-line survey*. Available online at: http://www.hull.ac.uk/esig/repomman/downloads/R-D3-research_survey_data_10.pdf.

Gibbons, S. (2004). Establishing and institutional repository. *Library Technology Reports*, 40(4).

Gimson, R. (2001). *Device independence principles*. Available online at: http://www.w3.org/TR/2001/WD-di-princ-20010918/.

Harris, J. (2006). *Institutional Repositories – Who is using what?* E-mail to: JISC-REPOSITORIES@JISCMAIL.AC.UK [24.01.06].

Information Environment Service Registry: Metadata, (2005). Available online at: http://iesr.ac.uk/metadata/.

Improving the Performance of Adobe® LiveCycle™ Designer Forms, (2005). p7. Available online at: http://partners.adobe.com/public/developer/en/pdf/designer_performance_1_0.pdf.

IMS Content Packaging Information Model, (2003). Available online at: http://www.imsglobal.org/content/packaging/cpv1p1p3/imscp_infov1p1p3.html.

Intrallect, (N.D.). *IntraLibrary v2.4: Support FAQ*. Available online at: http://www.intrallect.com/support/intralibrary/faq2.4/.

Ishida, R. (Ed.), (2005). *Authoring Techniques for XHTML & HTML Internationalization: Outline View*. Available online at: http://www.w3.org/International/geo/html-tech/outline/html-authoring-outline.html.

Jackson, C. & Cooper, A. (2003). *Learning object structure: a critical assessment*. In DigiCULT Learning Objects from Cultural and Scientific Heritage Resources, (2003)

Thematic Issue 4. Available online at:
http://www.digicult.info/downloads/digicult_thematicissue4_lres.pdf.

JISC Digital Repositories Programme, (2006). Available online at:
http://www.jisc.ac.uk/index.cfm?name=programme_digital_repositories.

Jones, S. (2004). *Learning Objects Quick Guide Developing concepts and technologies in the VET sector*. Available online at:
http://pre2005.flexiblelearning.net.au/projects/resources/learningobjectguide04.pdf.

LeFurgy, W. (2003). PDF/A: Developing a File Format for Long-Term Preservation. *RLG DigiNews* [online], 7(6). Available online at: http://www.rlg.org/legacy/preserv/diginews/diginews7-6.html.

Looms, T., & Christensen, C. (2002). *Advanced distributed learning emerging and enabling technologies for the design of learning object repositories*. Available online at:
http://xml.coverpages.org/ADLRepositoryTIR.pdf.

Martinez, J. M. (2004). *MPEG-7 overview (version 10)*. Available online at:
http://www.chiariglione.org/MPEG/standards/mpeg-7/mpeg-7.htm.

MacColl, J., Jones, R., & Jones, T. (2006). *The institutional repository in the digital age*. Oxford: Chandos Publishing.

McGill, L., Nicol, D., Littlejohn, A., Grierson, H.. Juster, N., & Ion, W. J. (2005). Creating an information rich learning environment to enhance design students learning. *British Journal of Educational Technology*, 36 (4), 629-642.

MPEG completes more elements of the MPEG-21 standard, (2005). Available online at:
http://isotc.iso.org/livelink/livelink/fetch/2000/2489/186491/187238/3165287/SC29WG11PressRelease22.pdf?nodeid=4076117&vernum=0.

MODS: uses and features, (2006). Available online at:
http://www.loc.gov/standards/mods/mods-overview.html.

Mulder, S. (N.D.1). *Embedding fonts tutorial: Lesson 1*. Available online at:
http://webmonkey.wired.com/webmonkey/99/45/index0a_page8.html?tw=design.

Mulder, S. (N.D.2). *Embedding fonts tutorial: Lesson 2*. Available online at:
http://webmonkey.wired.com/webmonkey/99/45/index1a.html.

O.K.I.™ Architectural Concepts (Draft). (2003). Available online at: http://okiproject.org/filemgmt-data/files/OkiArchitecturalConcepts.pdf.

Oliver, B., Roberts., & Blinco., K. (2005). The e-Framework for Education and Research: An Overview Available online at: http://www.e-framework.org/resources/eframeworkrV1.pdf

The open knowledge initiative, (N.D.) Available online at: http://www.okiproject.org/.

The open knowledge initiative, (2006) Available online at: http://okicommunity.mit.edu/staticpages/index.php?page=OverOKI.

OpenOffice.org, (N.D.). Available online at: http://www.openoffice.org/.

OpenURL overview, (N.D.). Available online at: http://www.exlibrisgroup.com/sfx_openurl.htm.

Pemberton, S. (2004). *HTML and XHTML Frequently Answered Questions*. Available online at: http://www.w3.org/MarkUp/2004/xhtml-faq.

Powell, A. (2005). *RDN LTSN application profile Version 1*. Available online at: http://www.rdn.ac.uk/publications/rdn-ltsn-ap/.

PREMIS, (2006). Available online at: http://www.loc.gov/standards/premis/.

Quality Assurance Agency (1999). *Code of practice for the assurance of academic quality and standards in higher education: Section 3 Students with disabilities*. Available online at: http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section3/COP_disab.pdf.

Raggett, D., Le Hors, A., & Jacobs, I. (Eds), (1999). *HTML 4.01 Specification*. Available online at: http://www.w3.org/TR/html401/.

Research Libraries Group (2002). *Trusted Digital Repositories: Attitudes and Responsibilities*. Available online at: http://www.rlg.org/en/pdfs/repositories.pdf.

Robson, R. (2004). *Digital Rights Management A White Paper for Alt-i-Lab 2004*. Available online at: http://www.jisc.ac.uk/uploaded_documents/Altilab04-DRM.pdf.

Search & Retrieve Web Service / Search & Retrieve URL Service, (N.D.). Available online at: http://www.oclc.org/research/projects/webservices/default.htm.

SFX Sources (OpenURL Enabled Resources), (N.D.). Available online at:

http://www.exlibrisgroup.com/sfx_sources.htm.

Siddall, S. (2002). Non-Technical Brief on the Open Knowledge Initiative. Available online at: http://siddall.info/cms/okitechbrief.html.

Stevenson, A. (2005). *JORUM Application Profile draft 1.0*. Available online at: http://www.jorum.ac.uk/docs/pdf/japv1p0.pdf.

SRW Service description, (2004). Available online at: http://www.loc.gov/z3950/agency/zing/srw/service.html.

Survey and assessment of sources of information on file formats and software Documentation, Final Report, (N.D.). The Representation and Rendering Project, University of Leeds. Funded by the JISC. Available online at: http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf.

TASI: An introduction to the vector image format, (2005). Available online at: http://www.tasi.ac.uk/advice/vector/pdf/vector-intro.pdf.

TASI: Choosing a file format, (N.D.). Available online at: http://www.tasi.ac.uk/advice/creating/format.html.

TASI: Digital cameras, (2003). Available online at: http://www.tasi.ac.uk/advice/creating/pdf/camera.pdf.

TASI: File formats and compression, (2005). Available online at: http://www.tasi.ac.uk/advice/creating/fileformat.html.

TASI: Image File Formats for Teaching and Learning Vector Images Part I: Overview, (2004). Available online at: http://www.tasi.ac.uk/advice/vector/pdf/tl_vector1.pdf.

TrueDoc: Features and specifications, (2005). Available online at: http://www.bitstream.com/font_rendering/products/truedoc/features.html.

TrueDoc in embedded devices, (2005). Available online at: http://www.bitstream.com/font_rendering/products/truedoc/embed.html.

TrueDoc: Overview, (2005). Available online at: http://www.bitstream.com/font_rendering/products/truedoc/overview.html.

Tull, L. (2005). Unicode. *Encyclopaedia of Library and Information Science*. 1(1). Available online at: http://www.dekker.com/sdek/abstract~db=enc~content=a713532118.

UK Learning Object Metadata Core, (2003). Available online at:
http://www.ukoln.ac.uk/metadata/education/uklomcore/.

UR research format support, (2005). Available online at:
http://www.lib.rochester.edu/index.cfm?PAGE=1360.

Web accessibility initiative (WAI), (Last updated date 2006). Available online at: http://www.w3.org/WAI/.

WEFT3 Overview, (2000). Available online at:
http://www.microsoft.com/typography/web/embedding/weft3/weft00.htm.

WestOne (2003). *Introduction to standards and specifications for learning object repositories*. Available online at:
http://pre2005.flexiblelearning.net.au/projects/resources/VLOR_introduction_to_standards_and_specifications.pdf.

WestOne (2003). *Using metadata for online learning resources*. Available online at:
http://pre2005.flexiblelearning.net.au/projects/resources/VLOR_using_metadata_online_learning_resources.pdf.

Wilson, S., Blinco, K., & Rehak, D. (2004a). *Service-Oriented Frameworks: Modelling the infrastructure for the next generation of e-Learning Systems*. Available online at:
http://www.jisc.ac.uk/uploaded_documents/AltilabServiceOrientedFrameworks.pdf.

Wilson, S., Blinco, K., & Rehak, D. (2004b). *An e-Learning Framework: a Summary*. Available online at:
http://www.jisc.ac.uk/uploaded_documents/Altilab04-ELF.pdf.

Appendix A

Key:

E-Prints	Dspace	Fedora	Biomed	Other

Institution	System	Repository Type	Accepted file formats
Aberdeen	Dspace (AURA)	Research	PDF
Bath	E-prints	Unknown – site inaccessible	
Birmingham	E-prints	Research	PDF, DOC, GIF, JPEG, ZIP
Brighton	Planned 07/05		
Bristol	Dspace (Bristol Repository of Scholarly Eprints (ROSE))	Research	PDF, TXT, ASCII
Bromley College	Dspace	Learning objects	JPEG, PDF, ZIP, DOC, PPT
Cambridge	Dspace	Research & Teaching	TXT, PDF, HTML, JPEG, TIFF, CML, RTF
Cambridge – Computer Lab Technical Reports	Web page with docs available as pdf	Not a repository	
CCLRC	Epubs (inhouse system)		
Cranfield – Cranfield QUeprints	Dspace	Research	PDF
De Montfort	unknown	Research	PDF
Durham	E-prints	Research	PDF, HTML, Postscript
Edinburgh – Edinburgh Research Archive (ERA)	Dspace	Research	PDF, Postscript, DVI
Glasgow	Dspace	Research & Teaching	PDF, PPT
Glasgow	E-prints	Research	PDF

Hull	Fedora	Repository aims to be up and running July 06	-
Imperial	E-prints	Research (7 Items)	PDF, Postscript
Lancaster	E-prints	Unknown – site inaccessible	
Leeds/Sheffield/York – White Rose Consortium	E-prints	Research	PDF, HTML, ASCII
Leicester	Dspace (just started project Jan 06)	Beginning of project	-
Lincoln – Faculty of Technology	E-prints	Research	PDF
Liverpool	Biomed Central Open Repository	Research	PDF, HTML
London – Birkbeck	E-prints	Research	PDF
London - Kings	E-prints	Research	PDF
London - SOAS	E-prints	Research	PDF, ASCII
London - UCL	E-prints	Research	PDF, HTML, RTF, ASCII
London School of Economics – Research Online	E-prints (Project team is using Dspace –not an IR as such)	Research	PDF, HTML, PPT
London Southbank	Dspace	Research (6 items)	PPT, DOC
Loughborough	Dspace	Research	PDF
Manchester Metropolitan	Biomed Central Open Repository	Research (5 items)	PDF, DOC
Manchester?	*E-prints??*	*N/A*	*n/a*
National Library of Wales	Fedora	No Information available	
Newcastle	E-prints	Research	ASCII, HTML , DOC, PDF Postscript, RTF
Nottingham	E-prints	Research	PDF
Open University	E-prints	Research	PDF

Oxford + Oxford Mathematical Institute	E-prints	Research	PDF
Portsmouth	E-prints	Research	PDF
Royal Holloway	E-prints	Research	PDF, Postscript
Sheffield Computer Science	E-prints	Research	PDF, DOC, HTML
Southampton + Southampton Electronics/Computer Science	E-prints	Research	PDF
St Andrews	E-prints	Research	PDF, HTML, Postscript
Stirling	Dspace	Research (Thesis/Dis)	PDF
Strathclyde	E-prints	Research	PDF, HTML
Surrey	ProQuest - Digital Commons	Research	PDF (system automatically turns uploaded RFT and Word files into PDF)
Wales Aberystwyth	Dspace	Nil Items	-
Warwick – Centre for Scientific Computing	E-prints	Research	PDF, HTML, Postscript, ZIP
Westminster	E-prints	No Information available	
Wolverhampton	Looking at systems	N/A	N/A