Preservation Watch Report

Author: Jane Stevenson for the JORUM Team

Date: July 2005

Draft Version: 1.0

Status: Final

Authorised: Signed off by JISC

Version Changes to this File

Revision	Date	Change	Details
	01.08.2005		Checked by Jackie Carter and sent to JISC
	01.11.2005		Signed off by JISC

Related Documents

RELATED DOCUMENT	Details

Contents:

*****

****

“Where stone tablets could be expected to survive for tens of thousands of years, a floppy disk or magnetic tape may only last 10 years. The hardware and software required to perceive or experience the information will be lucky to survive even that long.” [1]

"Historians will look back on this era and see a period of very little information. A 'digital gap' will span from the beginning of the wide-spread use of the computer until the time we eventually solve this problem. What we're all trying to do is to shorten that gap." (Danny Hillis, Disney Chief of Research and Development)[2]

“It is estimated that we have created and stored since 1945 one hundred times as much information as we did in all of human history up until that time!” [3]

*****

NB: all Web references provided in the footnotes were accessed in July 2005.

[1] *Time&Bits: Managing Digital Continuity*.Available at: www.wired.com/news/culture/0,1284,10301,00.html
[2] Meloan, S. *No Way to Run a Culture*. Wired News, 13 February 1998. Available at: http://wired-vig.wired.com/news/print/0,1294,10301,00.html
[3] http://www.longnow.org

1. Executive Summary

This report aims to be both a reflection of current activities, standards and recommendations relating to digital preservation, both nationally and internationally, and a starting point for further work on the possibilities for implementing a long term preservation strategy for learning objects within the JORUM learning object repository.

Successful digital preservation requires the necessary organisational structures, resources, skills and commitment to ensure that the technical solutions are correctly implemented so that access to materials is maintained after the software and hardware originally required have become obsolete. This involves maintaining adequate metadata and ensuring the security and integrity of materials over time. The decision to undertake long term preservation must therefore rest upon strong arguments in favour of providing medium and long term access to materials. It is important to remember that preservation is essentially about maintaining access and that this is likely to involve some kind of preservation process becoming necessary sooner rather than later with today's fast changing technologies.

JORUM is currently not an archive repository, in the sense that it does not undertake preservation of learning objects. If an archive facility is created it could be a facility that exists within JORUM, or it could be separate. It may be that JORUM continues to provide access, with preservation being carried out separately, so that objects pass between JORUM and an archiving facility, which undertakes preservation but does not provide public access. Alternatively, JORUM may act as both the point of access and the preservation repository. Whatever is decided about how the preservation process is carried out, it would seem sensible for JORUM to continue to be the single access point for all learning materials. As well as the creation of preservation copies, an archive will store master archive copies, and it is advisable to provide a separate store for these masters, or indeed separate stores for multiple master copies.

Many e-learning initiatives offer a whole range of services and promote the concepts of reusing and repurposing materials. However, it would appear that most of them have not directly addressed the issues of long term preservation. Projects often aim to create and disseminate learning materials, and promote repurposing, but they do not specifically address, or do not prioritise, accessing the learning objects once they require preservation. A common scenario within e-learning is for the metadata to be held and searched centrally whilst the objects are still held locally, so that the local repositories are, in principle, responsible for digital preservation, rather than a central facility. But it seems unlikely that preservation practices will be widely adopted by relatively small scale local learning repositories. Digital archiving will be a technically and organisationally challenging task, and whilst many institutions may maintain digital repositories, it seems unlikely that a large number of institutions will have the motivation, skill, or resources to undertake the long term archiving of e-learning objects. This, therefore, provides a compelling argument for JORUM, as a centralised and national repository, to investigate the possibilities for maintaining long term access to learning mateirals. In looking at this challenging topic, JORUM is keen to engage with the wider community in the discussion of issues and the development of relevant standards. Whilst there are few digital preservation initiatives within e-learning, it is undoubtedly the case that a great deal can be learnt from initiatives within the heritage, cultural and archival communities, where the materials to be preserved are often diverse, and where the intention is both to maintain effective access and preserve materials over time. A JORUM team could work closely with the Digital Curation Centre and other centres of expertise, such as the National Data Centres, UKOLN and the Arts and Humanities Data Service, as well as benefiting from the findings of the many JISC-funded digital preservation projects. As the JISC Digital Preservation Strategy states, "digital preservation requires new workflows, scarce skills, and close co-operation across different professions from traditional preservation management to computing science."[4] A central archive employing a collaborative approach to development may provide the best option.

Clearly, before embarking on any sort of preservation programme, it is necessary to be convinced that learning objects should be preserved. There are compelling arguments in favour of long term preservation, resting on the principles of reuse, potential use and historic value, and recognising the high cost of creating complex materials. These arguments need to be set against the time and resource intensive processes that will be required to maintain access. If it can be seen that learning objects provide a stream of benefits over an extended period of time, then preservation should eventually become an accepted part of the life-cycle of the resources, and as the number of digital assets produced by HE and FE institutions grows, it should become increasingly accepted that they are a vital part of the knowledge base of the learning and research

[4] Beagrie, N. *A Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) 2002-2005*. Available at: http://www.jisc.ac.uk/index.cfm?name=pres_continuing

community. Their overall value should increasingly be recognised and the need to take measures to ensure their long term preservation should become more of a priority.

JISC continues to have a commitment to developing the digital preservation agenda of the UK, but a successful approach to preservation clearly rests most importantly upon the establishment of a demand, or at least a potential demand, for the use of learning materials over the longer-term by the community of users. Preservation is only worthwhile if the stakeholders have a vested interest. People will be willing to deposit materials in a learning object repository if they have a compelling interest to do so. Depositors should certainly benefit as users, and it may also enhance their profile as an author of learning materials. It may also be that potential contributors will appreciate the benefits of depositing materials in JORUM because the preservation that is required to ensure access over the long term is problematic for individual institutions to carry out and therefore a central archive can provide a unique service.

If materials are not preserved then new materials will be created that may duplicate what has previously been available, when it might be more appropriate and cost effective, especially for more complex and interactive objects, to reuse and modify those that already exist. A key question to consider is whether the cost of constantly recreating learning materials will be higher than carrying out preservation processes so that materials can be reused over time. This may prove difficult to establish, as there are a lack of business models or cost models available for digital preservation activities, and there is little empirical data available, as the notion of preserving digital materials is still relatively new. Even without this handicap, it is notoriously difficult to predict the demise of current hardware and software and therefore the frequency and complexity of interventions necessary to avoid obsolescence. Indeed, it may be that future processes will change accepted methods of digital preservation, which introduces a further level of uncertainty.

The concept of a trusted digital repository, providing reliable long term access, may be important in developing a good reputation that in turn encouraging deposit. An archive repository will need to make sure that it has a clear and comprehensive policy as well as effective, clearly documented procedures for maintaining access and it will need to provide secure storage for the digital materials. The policy should set out the aims and objectives of the archive and the responsibilities that it undertakes to meet, ensuring that there is full accountability. If the archive is following the Open Archival Information System (OAIS) model,[5] then the policy should reflect the mandatory responsibilities that are set down in the OAIS (*see* section 15).

This report set out some of the key issues that need to be addressed within a digital preservation policy. These include quality of materials, appraisal processes, rights issues, validation to ensure authenticity, completeness and version control. The the level of granularity at which resources are described, stored and accessed has major implications for their preservation and reuse over the longer-term (*see* section 10.9). This is a particularly difficult issue in relation to learning objects, which can be complex hierarchical resources containing a whole range of formats. The existence of virtual learning objects, where metadata is provided within JORUM but the resource is maintained externally, creates a further complication that will need to be addressed. JORUM cannot be held responsible for preserving materials that are not held within the repository, but any one learning object may include both stored objects and virtual objects, and the archival policy will need to address this scenario.

Preservation metadata is essential for the long term management and preservation of resources, and it is critical that the right type of metadata is recorded and maintained to enable access to and reuse of digital data in the future. The JORUM application profile[6], defining the metadata elements that are used for JORUM learning objects, is based on the LOM (Learning Object Metadata)[7], and it includes many elements that are necessary and useful for preservation. However, it is likely that additional metadata will be required, and that metadata will need to be provided for individual components within aggregate learning objects. The digital preservation community have begun to coalesce around the core preservation metadata that is required, but the levels of detail that are described in metadata element sets differ considerably. A decision on the metadata to use will need to consider how necessary or useful each element is and, crucially, whether there are the resources are available to create that metadata. This study considers some of the recent initiatives and metadata schemas available, and looks in particular at the PREMIS core set of preservation elements. This core set builds upon the OAIS reference model, and is intended to reflect elements that are actually implemented in practice, so it should provide a useful basis for selecting the key preservation metadata elements for learning objects

[5] Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System OAIS.* January 2002. Available at: http://www.ccsds.org/documents/650x0b1.pdf
[6] *JORUM Application Profile*, 2005. Available at: http://www.jorum.ac.uk/contributors/chelp/index.html, July 2005
[7] *UK LOM Core* (v0.2). May 2004. Available at: http://www.cetis.ac.uk/profiles/uklomcore

An archive repository is likely to define and support a limited number of file formats and those that are maintained are likely to be standards based and preferably open source. Documentation should be provided on all formats that are supported and how each format is to be treated. The availability of file format information is crucial to the successful design and implementation of rendering tools. This report highlights some initiatives relating to preservation formats and maintaining information on formats, and refers in particular to XML and PDF, which are often cited as possible formats for long term preservation.

Materials cannot remain accessible without undergoing preservation processes of some kind. Whilst an archive may undertake to retain the bit stream of all the objects, it also needs to provide a renderable and meaningful object and this involves preservation processes. There are two main options for preservation, described in this report, namely emulation and migration. When considering the process to adopt, the look and feel characteristics that are considered essential must be identified, and it is preferable to chose a method that minimises the number of preservation processes, as these will inevitably alter the record, as well as having cost implications. All processes applied should endeavour to maximise future options for access rather than imposing increasing limitations. The priority should be for the user to be able to access the record as easily as possible. A successful solution to the problem of archiving would need to find a comprehensive, systematic, and dynamic means for preserving virtually any kind of electronic record, free from dependence on any specific hardware or software. Many commentators believe a preservation process should preserve the 'essence' or 'significant properties' of the material. The argument is that where resources are limited and choices must be made, it is crucial to determine and document the essential characteristics as soon as possible in order to concentrate on preserving what is important.

This report makes substantial reference to the Open Archival Information System model, an ISO standard developed by the Consultative Committee for Space Data Systems (CCSDS). It provides a framework for the whole range of functions associated with digital preservation, including ingest, storage, management, administration, preservation planning and access. An OAIS archive has to conform to the model and responsibilities set down in the standard. This report covers in particular the Information Model set down by the OAIS, which provides a model for associating the data object with the necessary information to render it meaningfully, as well as associating it with the requisite preservation metadata and descriptive metadata. The OAIS is a conceptual model and not a system for implementation. Whilst it may serve as a foundation for an e-learning archive, defining the basic functional components of the archive and listing the minimum requirements that an archival system should meet, it is not a guide for implementation but a starting point for setting out what an implemented system should incorporate.

The report ends with suggestions for the next steps that could be taken in working towards implementation of a digital preservation programme for the learning objects that are stored in JORUM.

2. Key Definitions

2.1 Digital Preservation

For the purposes of this report, digital preservation is defined as "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary." [8]
Fundamentally, to experience digital records, a person must have the right combination of hardware and software, and it is this which most obviously sets digital records apart from paper records. The concept of the 'original' record is arguably not applicable to digital records, as many users of digital records can experience equivalent copies. A digital record can more accurately be thought of as a performance, which is reliant on the data, hardware and software working together: "digital records are fundamentally performances and not objects".[9] Digital preservation is about maintaining the performance by ensuring that a user has the ability to view the data via suitable software and hardware.

Digital preservation involves the application of technologies and methods to ensure that information remains accessible and meaningful in the long term. The International Council on Archives state that the hardware

[8] Jones, M. and Beagrie, N. *Preservation Management of Digital Materials: A Handbook*, 2003, Available at: http://www.dpconline.org/graphics/handbook/
[9] Heslop, H (National Archives of Australia). *An Approach to the Preservation of Digital Records. 2002. Available at:* http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf

and software that create a record will typically last five years at current rates of technological change[10]. If this is the case, it is likely that maintaining access for even ten years will require preservation processes to be undertaken.

2.2 Digital Curation

Digital curation is about digital preservation within the wider context. The Digital Curation Centre expresses it thus:

"The term digital curation is used … for the actions needed to maintain digital research data and other digital materials over their entire life-cycle and over time for current and future generations of users. Implicit in this definition are the processes of digital archiving and preservation but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge."[11]

Digital curation is about active management over the whole life-cycle of the materials, from creation to destruction (or indefinite preservation), and about adding value to data. It may also encompass promotion and support for research activities, financial and legal issues. Whilst the JORUM Research and Development workpackage, under which this report was produced, refers specifically to preservation, it may be prudent to take a broader perspective that considers all of the issues relating to the curation of learning materials.

2.3 Learning Object

JORUM has defined a learning object as "any resource that can be used to facilitate learning and teaching that has been described using metadata".[12]

2.4 Non-aggregate Learning Object

One item, such as an image or PDF file, that exists as a single learning object within JORUM and therefore has metadata attached.

2.5 Aggregate Learning Object

Any number of resources, which may include different formats, organised into a hierarchical structure, where metadata will be attached at the top level, but is not necessarily attached to each individual resource. An example might be a learning object for a course that includes several audio files, a number of TIFF images and a PDF file, organised into a defined structure.

2.6 Content Package

An IMS Content Package provides a standard for packaging learning objects. It consists of the manifest file and the physical files (i.e. the actual content). The manifest file is a special XML file describing the content organisation and resources in a Package. This comprises metadata, organisations (organisations of the content within the manifest), resources (references to all of the actual resources and media elements needed for a manifest) and sub-manifest files (nested manifests). See appendix (1) and also http://www.imsglobal.org/content/packaging/cpv1p1p2/imscp_infov1p1p2.html for further information.

---

[10] International Council on Archives. *Electronic Records: A Workbook for Archivists.* April 2005. Available at: http://www.ica.org/biblio/Study16ENG_4.pdf

[11] http://www.dcc.ac.uk/about/what/

[12] JORUM *Scoping and Technical Appraisal Study,* 2004. Available at: http://www.jorum.ac.uk/research/archive/research/publications.html

An OAIS Content Package is similarly a conceptual container that packages the physical files with associated metadata (*see* section 15 on the OAIS).

2.7 Virtual Object

Also known as a resource stub, this is a link to other websites hosting content that users may download and integrate into their own VLEs.

2.8 Archive Repository / Digital Repository

This report uses the term archive repository (or digital archive repository) in order to distinguish a repository that provides reliable long-term access to resources, thus requiring some form of preservation process to be undertaken, as opposed to a repository that provides access to resources in the short term, without altering them in any way.

The term *digital repository* is not used consistently, in so far as something that refers to itself as a digital repository may or may not undertake long term preservation.

3. Scope of the Report

The intention of this report is to set down the issues involved in the creation of an archive repository for e-learning objects, provide some indication of current thinking on these issues and discuss possible options for a long term repository. It draws from a wide variety of documentary sources, not only from initiatives within e-learning but also from many other sources relevant to digital preservation, both nationally and internationally.

Digital preservation is a large and complex area, and given the time constraints within which this report was created, some issues are covered more comprehensively than others. The technical issues are clearly complex and it is the intention only to briefly describe current practice and define the possible technical solutions.

This report is the outcome of a preservation watch, which is part of the JORUM Research and Development programme, and the scope is to a large extent defined by the current and recent projects, initiatives and implementations of digital preservation systems. This is a constantly shifting arena, where new technologies are rapidly developing, and so this report can only try to reflect the situation at the time that is was written.

The Digital Preservation Coalition (DPC) is intending to commission a consultancy to complete a UK Digital Preservation Needs Assessment. The study will analyse and synthesize existing sources of data on digital preservation activity in the UK and gather additional information to present a detailed analysis of the status quo regarding digital preservation in the UK. This DPC report could help to inform and influence the development of a preservation strategy for learning materials.[13]

The long term preservation of learning materials is a very new area of research and development, and there are no real practical implementations to use as a model. In this report, an attempt has been made to look at initiatives within digital preservation and extrapolate practices and recommendations into the domain of e-learning. The challenges cannot be underestimated, but in the same way that JORUM itself is a ground-breaking service, so the development of a long term repository could also be an opportunity to be leading the way in this field.

4. JORUM and its relationship to an Archive

The JORUM repository will be providing access to a wide range of digital materials, with a diversity of format, from simple text based files to sophisticated multimedia resources. It is also inevitable that JORUM will take in new, as yet unknown formats.

[13] http://www.dpconline.org/graphics/reports/index.html#ukneeds

A JORUM learning object may be one item, or it may be an *aggregate object*, consisting of more than one item and possibly consisting of a large number of items with a complex hierarchical relationship between them. In addition, JORUM provides access to *virtual objects*, also known as 'resource stubs', which are not actually hosted by the repository. In the case of virtual objects, JORUM stores the metadata and provides links to the learning materials, which are stored elsewhere.

All JORUM learning objects are required to use a minimum set of required metadata elements, as defined by the JORUM application profile, of which some are compulsory and some are optional. Metadata is attached to each learning object at the top level, and it may or may not also be attached to any number of the individual items within an aggregate object.

The metadata for JORUM learning objects will be added by both the contributor and, at least initially, by expert cataloguers, as well as some automated entry. The use of cataloguers provides an opportunity to ensure that the metadata attached to each learning object is comprehensive.

JORUM learning objects can be exported by users, integrated into their own learning materials and put into Virtual Learning Environments. Ideally, objects that have been repurposed would then be put back into JORUM, but licensing issues currently preclude the reuse of resources in new JORUM learning objects. The central principle is that objects can be shared, reused and repurposed.

Currently JORUM does not indicate a lifespan for the materials deposited, and the issues of appraisal and possible deletion of materials have not been actively addressed in the service set up. However, if JORUM is to be seen as a secure environment for learning materials, then it could be argued that the archival responsibility is inextricably linked to the service – even if learning objects are only kept for a short number of years, they need to remain accessible from the outset, and there may be instances where this requires positive action for preservation early on in the life of a resource.

There are a number of options for JORUM to consider for the possible relationship between the JORUM and an archival function:

1. JORUM does not take on the responsibility of providing long term access

    1a) JORUM has the authority to delete materials once they are no longer accessible using current hardware and software, or no longer of significant pedagogical value
    1b) JORUM may transfer materials to other bodies once they require preservation, possibly to a separate archive repository, and take no further responsibility for providing access to these materials

2. JORUM does not take on responsibility for preservation, but does provides long term access

    Materials could be transferred to an archive when they require preservation, and transferred back to JORUM so that the archive would provide the preservation function and JORUM continue to provide the access function.

3. JORUM provides a storage facility, keeping materials for the long term but not carrying out preservation activities. This would be an interim measure, possibly storing the bit stream and functional version (*see* section 14.5 on bit level preservation) in order to maintain the potential for continued access.

4. JORUM takes on responsibility for preservation and for providing long term access

    JORUM takes on the responsibility for undertaking preservation processes and provides access to all materials.

Comments on the above scenarios

It would seem inefficient and potentially confusing to set up two means of accessing materials, with JORUM providing access to 'current' materials and an archive providing access to 'preserved' materials. It is important to remember that non-preserved and preserved materials are both being kept for the same purpose. Long term preservation is really about providing long term access, and from a user's point of view it is preferable to have one place to go to access learning materials. It would also seem to be a duplication of effort to create two infrastructures for the storage and access of materials. If the intention is to ensure that archived materials (i.e. those that have undergone some kind of preservation process) are interoperable

with current materials as far as the users are concerned, then the pragmatic option would seem to be to view JORUM as a national repository for learning materials, with responsibilities for maintaining access over time. One alternative would be to maintain the JORUM repository as the point of access for all learning objects, but also maintain an archive where preservation activities are carried out.

An interesting example of the separation of the two functions of access and preservation is the DAITSS (Dark Archive in the Sunshine State),[14] covering the state of Florida, where the preservation repository is separate from the institutional repositories. DAITSS has no public interface and allows no public access. The institutional repository sends items that they want to preserve, and the preservation repository can ship items back to the institutional repository, maybe after they have been migrated. It is then up to the institutional repositories to provide access. The preservation functions of the DAITSS are normalization (reformatting to a more 'archivable' format) and migration, and it writes three independent master copies of all Archival Information Packages (AIPs). It implements the functional model of the OAIS reference model and uses METS documents as descriptors for Submission Information Packages (SIPs), Archival Information Packages (AIPs) and Dissemination Information Packages (DIPs). It is intended that the DAITSS software will be released as open source software.

The separation of the preservation repository from the access repository in DAITSS makes sense because the archive is serving a whole number of institutional repositories. Following this model, a preservation repository for learning objects could be set up and used not only by JORUM but also by institutional repositories.

The JORUM Scoping and Technical Appraisal Study describes the JORUM service as a key part of the JISC's Information Environment, "providing the means by which access to, deposit of, and sharing and re-use of, freely available learning and teaching content can be undertaken by teaching staff and their colleagues." It goes on to refer to the likelihood of JORUM having "both transient and archival instances, to satisfy both the need for day-to-day use of learning and teaching materials, and the long term retention of some of them".[15] This implies separate access and preservation functions, both part of the JORUM responsibilities. It is advisable to also maintain a separate store for the master archive copies, or indeed separate stores for multiple master copies in a separate location (i.e. on separate hardware).

5. Current Initiatives in Digital Preservation and Learning Object Repositories

In many countries there are a whole host of initiatives in digital preservation, and some national programs have emerged such as the US National Digital Information Infrastructure and Preservation Program (NDIIPP)[16]. Increasingly, there are international initiatives that pool expertise within digital preservation, such as the Electronic Resource Preservation and Access Network (ERPANET)[17] and Preserving Access to Digital Information (PADI)[18].

There are a range of initiatives to provide access to e-learning materials, and many of these offer a range of services and promote the concepts of reusing and repurposing materials. However, it would appear that most of them have not directly addressed the issues of long term preservation. For the purposes of this report, the main source of information consulted is the web, and so it may be that other work is going on behind the scenes that has not been described on the websites of these projects.

This section looks at some digital preservation initiatives in the UK that may have relevance for the creation of a learning object archive, and also at e-learning initiatives in other countries. Some of the projects that are referred to below provide similar functionality to JORUM, many provide a central store for metadata and do not actually store the learning materials, others are initiatives providing networked access to materials and others are consortiums looking to pool resources and expertise and to further develop standards and best practice. All projects emphasise the need to ensure that learning objects can be repurposed, which is often linked to the need to provide sustainable and durable resources, but there is little mention of the need for preservation processes. Many of these projects are in the early stages, and it may be that long term issues will be addressed in due course. In addition, many projects are keen to engage with the wider community in the discussion of issues and the development of standards, which are likely to address preservation issues.

[14] DAITSS overview available at: http://www.fcla.edu/digitalArchive/pdfs/DAITSS.pdf
[15] JORUM Scoping and Technical Appraisal Study, 2004.
[16] http://www.digitalpreservation.gov/
[17] www.erpanet.org/
[18] http://www.nla.gov.au/padi/

The following initiatives have been considered:

5.1 UK

There are many projects relating to digital preservation, particularly those funded by the JISC as part of their Continuing Access and Digital Preservation Strategy[19]. A number of projects have been running as part of the Supporting Digital Preservation and Asset Management in Institutions programme.[20] It is not possible to look in detail at their findings and recommendations here, but three of these that might be of particular interest are:

The Assessment of UK Data Archive and The National Archives compliance with OAIS/METS project: Assessing how the operational structures of The National Archives and UK Data Archive can be informed by OAIS and exploring the potential for interaction between existing metadata standards utilised within the two institutions and METS.[21] *See* section 15 on the OAIS (Open Archival Information System and appendix (5) on METS, the Metadata Encoding and Transmission Standard.

The Digital Asset Assessment Tool – DAAT[22] - is being developed by the University of London Computer Centre, with a number of partners, as an aid to help to formulate an approach to preservation, prioritising those materials where the risk of loss and cost of loss is greatest.

The SHERPA Digital Preservation project: using the OAIS reference model to develop a persistent preservation environment for the SHERPA consortium, assigning rights and responsibilities and establishing protocols and work flow processes that will ensure the long term preservation of the repository content.[23]

Personal Archives Accessible in Digital Media - paradigm[24] - is a collaboration between the libraries of the Universities of Oxford and Manchester to explore the issues involved in preserving digital private papers through gaining practical experience in accessioning and ingesting digital private papers into digital repositories, and processing these in line with archival and digital preservation requirements. It is looking to produce best practice guidelines and it will be investigating the DSpace and Fedora digital repository software. Although this project is focussing on private political papers, many of its conclusions will no doubt prove valuable for all digital preservation initiatives.

An Effective Strategic model for the Preservation and Disposal of Institutional Assets – espida[25] - is a project based at the University of Glasgow that is developing a sustainable business-focused model for digital preservation at FE/HE institutions. It has as its focus the creation of a model of the relationships, roles and responsibilities, costs, benefits and risks inherent in institutional digital preservation.

Mandate is a JISC-funded project running from December 2004 to November 2005, which is being undertaken to develop a toolkit for digital asset management and preservation. This is being run by the John Wheatley College (JWC), in partnership with the Centre for Digital Library Research (CDLR) at Strathclyde University and the Scottish Library and Information Council.[26] In this case, the material to be preserved is that created in the development of a new programme of study, and includes learning/teaching materials, assessment materials and procedural documents relating to a comprehensive approval process. The toolkit will support a college-wide strategy for development and storage of this material. It will take into account issues of workflow within an OAIS (Open Archival Information System) framework, and the creation, storage, retention, retrieval, and preservation of material. The project will format, index, and store existing unstructured digital materials created in currently disparate processes and from the research practice create and test

[19] http://www.jisc.ac.uk/index.cfm?name=pres_continuing
[20] http://www.jisc.ac.uk/index.cfm?name=programme_404
[21] http://www.jisc.ac.uk/index.cfm?name=project_oais
[22] http://www.jisc.ac.uk/index.cfm?name=project_daat
[23] http://www.jisc.ac.uk/index.cfm?name=project_sherpa2
[24] http://www.paradigm.ac.uk/
[25] http://www.gla.ac.uk/espida/index.shtml
[26] http://www.jwheatley.ac.uk/mandate/

templates and workflow models for routine use in John Wheatley College and for application across the sector. The project will develop appropriate structures for web-based storage of information that will contribute to the end goal of improved digital preservation.

As well as the digital preservation projects, it is worth mentioning the Digital Curation Centre, as this provides a national focus for research into curation issues and promotes expertise and good practice within digital curation. Collaboration with the DCC would be central to the development of an archive for learning materials.[27]

5.2 Canada

eduSource Canada– http://www.edusource.ca/
Not a repository itself, eduSource Canada is a project to create a test bed of linked and interoperable learning object repositories and to provide leadership in the development of the necessary infrastructure.

Belle – http://belle.netera.ca
A lifelong learning environment to develop a prototype educational learning object repository

CAREO – http://www.careo.org
A project that aims to create a searchable, Web-based collection of multidisciplinary teaching materials for educators

Pool - http://www.edusplash.net
The Portal for Online Objects in Learning, developing an infrastructure for learning object repositories. Centralised repositories keep all of their objects in one site, but with this initiative, every user can have their own mini-repository, which can be linked together so users can search all the POOL sites from an individual application.

5.3 The USA

DSpace - www.dspace.org
DSpace is a digital repository software system, used for a variety of digital archiving needs. It is freely available as open source software, so it can be modified and extended as required. DSpace identifies two levels of preservation: bit preservation, and functional preservation. Bit preservation involves the preservation of the digital file exactly as it was deposited without any change or future alteration, ensuring that it remains exactly the same over time. Functional preservation is the preservation of a usable file, which will change over time. The principle is to preserve the functionality of a number of supported formats.

The PREMIS Survey on implementing preservation repositories[28] found that DSpace was the most popular application for preservation, but most of those surveyed used a combination of types of applications.

The DSpace website points to an article by Paul Wheatley that looks into the options for developing the digital preservation function of DSpace - http://dspace.org/news/articles/DpAndDSpace.pdf

Fedora – http://www.fedora.info/
Fedora is described as 'a general purpose repository service', developed jointly by The University of Virginia Library and Cornell University. The Fedora project is devoted to the goal of providing open-source repository software that can serve as the foundation for many types of information management systems. Fedora employs XML and METS (Metadata Encoding Transmission Standard) to create digital objects by encapsulating content, along with metadata about the content, and actions that can be performed on the content. This linkage of the content to the applications that are used to search, render, and display it distinguishes Fedora from other digital repository systems.

MERLOT – www.merlot.org

---

[27] http://www.dcc.ac.uk/
[28] PREMIS Working Group (OCLC/RLG). *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community.* Sept 2004. Available at: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf

Multimedia Educational Resource for Learning and Online Teaching. MERLOT does not store the learning objects, only the locations. It aims to establish a 'community of use', providing the opportunity for cross-sectoral communication and development.

Advanced Distributed Learning - http://www.adlnet.org
The Advanced Distributed Learning (ADL) Initiative, sponsored by the Office of the Secretary of Defense (OSD), is a collaborative effort between government, industry and academia to establish a new distributed learning environment that permits the interoperability of learning tools and course content on a global scale. It aims to develop tools, specifications and guidelines that ensure content is accessible, adaptable, interoperable and durable across revisions of hardware and software. It may be worth tracking developments within the ADL initiative, especially those that may help to support preservation and long term reuse of learning objects. It is currently involved in updating the IMS Content Packaging specification, along with the Learning Technology Standards Committee of the IEEE[29] and the IMS Global Learning Consortium.[30]

US Research Awards for Digital Preservation - http://www.digitalpreservation.gov/about/pr_050405.html
The Library of Congress has taken the lead in the preservation and archiving of digital materials in the US. The digital strategy is based on a 2002 document "Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program" (www.digitalpreservation.gov).
In May 2005, The Library of Congress National Digital Information Infrastructure and Preservation Program (NDIIPP) and the National Science Foundation awarded eleven university teams a total of $3 million to undertake pioneering research to support the long term management of digital information. These awards are the outcome of a partnership between the two agencies to develop the first digital-preservation research grants program. They include projects looking at scalable digital archives, technologies for automated ingestion and long term preservation and the automation of metadata capture.

5.4 Australia

Colis – www.colis.mq.edu.au
Collaborative Online Learning and Information Services, developing a test bed for the development of collaborative online learning and information services.

The Learning Resource Catalogue - http://www.learnet.hku.hk:8052/
The Learning Resource Catalogue (LRC) was developed at the University of New South Wales and includes 25 institutions worldwide. The project promotes collaboration and the sharing and reuse of high quality learning materials.

The LRC holds metadata pertaining to the learning objects rather than the objects themselves. The web-application that provides access to tools for collaboration, cataloguing and sharing materials is the LRC3. The LRC website provides a great deal of informative information about creation, searching, repurposing and the general pursuit of excellence in the development of high quality learning materials, but it does not appear to specifically address the long term issues.

Arrow - http://www.arrow.edu.au/
Australian Research Repositories Online to the World – a project to identify and test software or solutions to support best practice institutional digital repositories comprising e-prints, digital theses and electronic publishing. It is intending to use the Fedora repository software.

5.5 The Netherlands

Nedlib - http://www.kb.nl/coop/nedlib/index.html
Aims to construct the basic infrastructure upon which a networked European deposit library can be built. The website refers to long term preservation as a key issue that is being investigated and includes reports on emulation and on metadata for long term preservation.

[29] http://www.ieeeltsc.org/
[30] http://www.imsglobal.org/aboutims.html

5.6 Conclusion

In general, it seems that international projects are putting considerable effort and resources into the creation, dissemination and reuse of e-learning materials, but they do not specifically address, or do not prioritise, accessing the learning objects once they require preservation. In many cases the metadata is held and can be searched centrally but the objects are still held locally, and therefore it may be that many local repositories will be responsible for digital preservation rather than one central facility.

It is necessary to look to other domains, in particular the archival domain, for work in the area of digital preservation. Most digital preservation policies that are currently publicly available are by cultural heritage institutions. Although e-learning materials may have their own particular issues, it is undoubtedly the case that a great deal can be learnt from archival initiatives, where the materials to be preserved are often diverse, and where the intention is both to maintain effective access and preserve materials over time. There are currently many conferences and workshops within the UK being organised by bodies such as the Digital Curation Centre and Digital Preservation Coalition that provide a useful forum for exchanging information and finding out about current initiatives within this fast-changing domain.

It may be that as digital preservation moves beyond the realm of smaller scale projects, it will become a routine component of a digital asset's lifecycle, and the indications are that this will happen particularly in areas such as national libraries and archives, but it appears far from the case with learning objects at present.

6. Purpose of long term preservation

It is important when considering whether to undertake preservation activities, to clearly define the reasons for keeping learning objects over the long term. One of the primary reasons for preserving many types of digital records, such as those of an institution or company, is their evidential value, in other words, to demonstrate actions that were taken, decisions that were made, why they were made, etc. However, this is not really pertinent to learning objects.

It is important to stress that preservation is not the aim in itself but rather the means to the end, which is to provide continuing access. This should increase the value of the materials created, and it may offer the benefit of avoiding the constant recreation of materials that have been deleted.

A 'transient' repository, which does not undertake preservation activities, can only provide access to digital resources in file formats for which the hardware and software are still readily available and accessible. If the intention is to enable users to take learning objects and reuse them in different ways and for different pedagogical purposes, then it seems likely that many learning objects will remain relevant over a reasonably long time period. For example, a learning object that demonstrates a law of physics may be used by tutors in different contexts, so that it might be used as one part of many different aggregate learning objects. If such objects are to continue to be reused in this way over time, they are likely to require preservation, in order to ensure that they can be accessed by the user community. An object created in Microsoft Word or Macromedia Flash in 2005 will almost certainly require digital preservation in order that it can be re-used in any new aggregate object created in 2010. If not, then users in 2010 are unlikely to be able to access, view or use the resource. In fact, without preservation the resource would become worthless, no matter what its pedagogical value. In other words, reuse is likely to occur where access is provided over time (i.e. over a number of years) and access over time requires preservation.

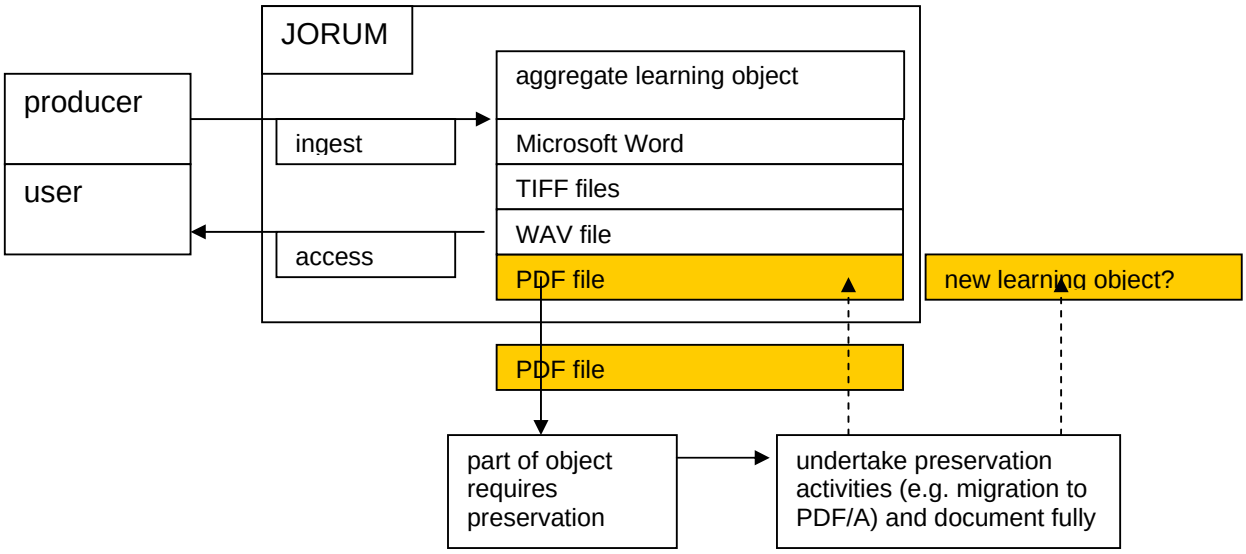


fig 1: illustration showing preservation activities

The illustration above show the way in which an aggregate learning object may be ingested into the JORUM repository, and accessed by users. When a part of the learning object requires preservation, it may, for example, be migrated and then continue to be part of the learning object, so that the whole object continues to be accessible. An alternative scenario would be to create a new learning object whenever preservation takes place, thus maintaining the original and a preserved copy.

The appropriate life span for a learning resources has been quoted as typically being around five years.[31] However, it is difficult to predict whether users of JORUM will continue to use materials over a longer time period, and whether repurposing materials will encourage this continued use. A system of appraisal will help to ensure that the materials that are maintained are the most relevant (*see* section 10.2 on appraisal). Materials could be removed if they are no longer considered worth retaining, or they could simply be stored in their original format, thus leaving the opportunity to make them accessible in the future.

E-Learning materials may be considered to have value:-

1. **Because they have continuing value in their original capacity**

Some learning materials will remain pedagogically germane for many years and therefore may continue to be used in the same context, for the same learning aims and objectives as they were originally conceived. An example might be something like a learning object explaining Newton's first law of motion – if this was clearly and effectively described within a learning object, it could continue to be integrated within aggregate learning objects.

2. **Because they can remain pedagogically valuable by being be re-purposed**

Ideally, learning objects can be modified, broken down and combined with other content to yield unique learning that meet individual learners' needs. Content within JORUM is available for exporting and integrating into other learning objects (though there are rights issues surrounding the re-deposit of learning objects into JORUM that reuse JORUM content). This approach means that materials can be used in a variety of ways, adding new meaning and value over time. Maintaining materials may encourage the community to build on what has already been created rather than constantly recreating them.

[31] See for example Barker, E., et all, *Long Term Retention and Reuse of E-Learning Objects and Materials*. June 2004.

The potential for repurposing is likely to be bound up with granularity (*see* section 10.9 on granularity).

3. Because they are an historic record

Traditionally, archives have had a different user base from current records – i.e. current records are used in the course of business or day to day activities, archives are used by researchers and give an historic perspective. This purpose may be applicable to some extent with learning objects, as a repository of learning materials can have value as an historic record. As more and more learning uses digital materials, they may become a key historical asset both for researching teaching and learning methods and for evidence of the evolvement of concepts, theories and methods over time. Historical value may be important, but it may not be seen as a sufficient reason by itself to justify the costs of preservation of materials such as learning objects. It is, however, a valuable argument when put together with pedagogical value, and it may be that a certain proportion of learning materials could be preserved because they are seen as particularly important in an historical perspective. This is an interesting issue for learning objects, because it may be that learning becomes increasingly based upon the use of such technologies, and therefore we need to think carefully about preserving a record, just as we now see important historic value in records of Victorian teaching methods.

4. Because they may have future potential

It is worth considering that the use of knowledge contained in learning materials may well be expanded by future technical developments. If materials are preserved, we may be able to use them in the future in ways that are not currently envisaged. Future users may have access to tools that are not currently available and may want to use the data in ways that were not predicted.

In summary, it is vital to prove the argument that "digital preservation offer a means to extend the productive life of valuable, often expensive, digital resources." [32]. If it can be seen that a learning object or learning resource provides a stream of benefits over an extended period of time then digital preservation should eventually become an accepted part of the life-cycle of the resources.

As the number of digital assets produced by HE and FE institutions grows, it should become increasingly accepted that they are a vital part of the knowledge base of the learning and research community. Their overall value should increasingly be recognised and the need to take measures to ensure their long term preservation should become more of a priority.

6.1 Incentives to Contributors to Preserve

The argument for long term preservation rests upon there being sufficient incentives to preserve. This issue is discussed at length in a report by the OCLC[33]. Whilst the report sets out the issues very much in terms of the economic arguments rather than educational and research benefits, it still provides a useful discussion of incentives, setting out roles and models under which digital preservation might take place.

The models outlined by the OCLC report set out three roles: rights holder, archive and beneficiary. In theory, both the current JORUM repository and an archive repository of learning materials should come under the 'demand-side model' defined in the report, in which the rights holder and beneficiary are the same and the archive is a separate entity. It is important to dissuade potential contributors from the view that the rights holder and beneficiary are separate, creating a 'centrifugal model'. Under a centrifugal model it is argued that the incentive to preserve is diminished, because potential contributors do not see themselves as beneficiaries.

It may be that potential contributors will appreciate the benefits of depositing materials in an archive because the preservation that is required to ensure access over the long term is problematic for individual institutions to carry out and therefore a central archive can provide a unique service. This should encourage depositors to see themselves as beneficiaries, because they can continue to access their materials, which otherwise would have become inaccessible over time. It will be important to convince the community that reusing and repurposing materials over the long term rather than constantly recreating them is a worthwhile activity.

---

[32] Lavoie, B.F. (OCLC) *The Incentives to Preserve Digital Materials: Roles, Scenarios and Economic Decision-Making*. April 2003. Available at: http://www.oclc.org/research/projects/digipres/incentives-dp.pdf

[33] *ibid*

The materials in JORUM will not necessarily be unique, as they may also be deposited in institutional repositories and people may choose to access them via this route. However, materials kept in an archive repository are more likely to be unique over the long term. A properly preserved copy becomes the only accessible copy over time, unless institutional repositories are also carrying out preservation processes.

In the end, people will be willing to deposit materials in a learning object repository if they have a compelling interest to do so. They should certainly benefit as a users, and it may also enhance their profile as authors of learning materials. Whilst JISC does have a commitment to developing the digital preservation agenda of the UK, as outlined in its *Strategy 2002-2005,* and does have a desire to ensure the continuing availability of digital materials for research, learning and teaching, there is clearly still a need to establish a clear demand, or at least a potential demand, for the use of learning materials over the longer-term by the community of users.

7. Required Resources

Preservation activities are part of the ongoing JORUM research and development programme, but not currently part of the JORUM service-in-development. In order to carry digital preservation forward there would need to be a significant investment to carry out management, administration, organisational and technical activities. The expertise and resources required would be considerable.

It could be argued that without investment in the long term, the cost of a repository like JORUM will remain high because materials will only be available within it for the short-term:

"The initial technical infrastructure costs required for creating and/or acquiring digital materials and providing access to them are substantial. It makes sense to consider means of protecting this investment from the outset though this has rarely been a consideration to date."[34]

If materials are not preserved then new ones will be created that largely duplicate what has previously been available when it may be appropriate to reuse and modify those that already exist. This duplication of effort may involve higher costs, especially for more complex and interactive objects, though a key question to consider is whether the cost is higher to constantly recreate learning materials or to carry out preservation processes so that materials can be reused over time.

7.1 Costs

There are a lack of business models or cost models available for digital preservation activities. There is little empirical data available, as the notion of preserving digital materials is still relatively new. Even without this handicap, it is notoriously difficult to predict the demise of current hardware and software and therefore the frequency and complexity of interventions necessary to avoid obsolescence. Indeed, it may be that future processes will change accepted methods of digital preservation, which introduces a further level of uncertainty.

Factors that would need to be taken into account include: the initial procurement of a digital storage system and systems utilising the storage system; technical staffing; the storage media; the length of time the materials are to be preserved; the level of access; the preservation strategy or combination of strategies adopted; the formats used; the degree of automation; the changes deemed necessary to the metadata. In addition to this, there would need to be a clear understanding of where the costs for digital preservation fall.

The development costs for the OCLC Digital Archive were $3.1 million dollars (approximately £2 million pounds). The staff and administrative costs were $2.4, and the hardware and software cost around $675,000.[35].

A cost benefit analysis would set out the costs and benefits of preservation, covering personnel, routine activities, storage, management, access, use, etc:

---

[34] Beagrie, N. *Preservation Management of Digital Materials, A Handbook*. 2003.
[35] Bellinger, M. (OCLC). *Cost and Business Models for Digital Preservation,* October 2002. PowerPoint presentation available at: http://www.dpconline.org/text/events/elearning.html.

"Because costs for both technical and organisational infrastructure are still not well defined organisations are confronted with the requirement to commit to the principle to safeguard significant digital assets, without a clear idea of the associated costs over time."[36]

For an archive to operate effectively there would need to be constant awareness of developments within the field of digital preservation, as it is inevitable that new solutions will be developed over time and new standards will be introduced nationally and internationally. Resources will be needed to assess the relevance and meet these developments where appropriate. It may be that new developments in technologies and formats may result in greater efficiencies and thus promote cost-effectiveness.

Projects within higher education often have short-term or medium-term funding, and this is clearly an issue for long term preservation, which really requires ongoing investment. Digital preservation in general has been poorly funded over the past decade, especially at an institutional level. This is now changing, and JISC have funded many projects covering digital preservation and asset management.[37] However, many of these projects run for one or two years, a level of funding that would prove problematic for an ongoing preservation programme.

The academic community may feel that putting resources into the long term preservation of digital learning materials is not a reasonable use of resources. It may be difficult to persuade communities of the benefits of an archive of e-learning materials, though this may change as more people become accustomed to using digital learning materials.

The JISC-funded espida project, running from January 2005 to December 2006, seeks to develop and implement a sustainable business focussed model for digital preservation, as part of a knowledge management agenda in higher education institutions. It will 'develop a model of the relationships, roles and responsibilities, costs, benefits and risks inherent in institutional digital preservation and implement this model by selling it to all the stakeholder groups, including senior management, administrative and clerical staff and academic teachers and researchers.'[38] It seeks to identify the costs and benefits of developing a sustainable approach to preservation, and should provide a useful model for the development of an archive of learning materials.

The report on Trusted Digital Repositories (2002) contains a recommendation to develop models for cooperative repository networks and services, as collaboration may provide cost advantages. Cedars recommended the development of distributed systems, as a means to specialise and streamline processes[39] and PANDORA, the Australian system, provides an example of this sort of distributed system in operation.[40] It would be worth considering designing an archive that could be federated at some point in the future.

8. Stakeholder Rights and Responsibilities

It is often said that digital objects need to be managed over their life cycle, preferably from the point of creation, if they are to remain accessible as long as is required. A strategy for digital preservation should ideally embrace all stakeholders involved in the process of the creation and maintenance of digital materials and the responsibility for management at all points in the life-cycle of the records needs to be clearly defined. However, the management of learning materials deposited in an archive may differ somewhat from a more generic model, where an archive is receiving digital objects from a whole range of sources and with a whole range of purposes.

---

[36] *Preservation Management of Digital Materials, A Handbook*. 2003.
[37] see the JISC *Supporting Digital Preservation and Asset Management in Institutions programme*, available at: http://www.jisc.ac.uk/index.cfm?name=programme_404
[38] http://www.jisc.ac.uk/index.cfm?name=project_espida
[39] *Cedars Guide to the Distributed Digital Archiving Prototype, section 6. March 2002. Available at:* http://www.leeds.ac.uk/Cedars/guideto/cdap/guidetocdap.pdf
[40] http://pandora.nla.gov.au/

8.1 Creator Rights

The archive's policy should clearly define the rights of those who deposit content in the archive. It may be that the archive seeks to become a trusted digital repository and intends to conform to the OAIS responsibilities, in which case depositors have the right to expect that their materials will be preserved in line with these standards. It may take on a lower level of responsibility and may not undertake to ensure that access is maintained to all materials that are deposited. The depositor should clearly understand what can be expected from the archive.

8.2 Creator Responsibilities

Many commentators take the view that the creator of an electronic record must take on some responsibility for preservation and access: "the role of preservation and access has to be recognised as the initial responsibility of the creator of an electronic publication."[41] This does not imply that the creator has primary responsibility, but that some issues need to be thought about at the creation stage. Responsibilities might include choice of format, addition of suitable metadata, maintaining access and ensuring the integrity of the record in its early stages. Many of these issues have been addressed by JISC as part of the X4L programme, which has looked at the reusability of learning materials and encouraging the use of formats that are more interoperable and therefore generally better for long term access.

In terms of digital preservation, a creator is very often part of an organisation, and the creator's responsibilities are usually considered within the general context of the activities of an organisation, with digital records forming part of the organisation's activities. For example, the Government creates electronic records in the course of its activities, with individual departments often creating records relating to specialist areas, and only a small proportion of these records will be kept long term. The organisation and context of the records may be vital to their meaning, and these records will be created for diverse purposes, from memos and emails to reports. It may be argued that only the creators can ensure that their integrity is maintained and that they are adequately described.

E-learning materials are created for a specific purpose – to provide materials for teaching and learning that can be used online. It is not the specific context that is important so much as the pedagogical value. An individual item may be given a context by placing it within a learning object, but a learning object is not tied down to the context of the individual, department or institution in which it was created. Indeed, the intention is to be able to take resources out of that context and repurpose them.

In order to encourage good practice in the creation of digital materials, depositors should be encouraged to create "rich digital resources worthy of long term preservation" and materials that are "more amenable to digital preservation practices and processes."[42] An archive repository would need to carefully consider how to achieve good practice and how much responsibility creators and providers should have.

The creators of learning objects for JORUM are not mandated to use particular formats or to consider preservation and wider access issues. At the outset, they only have to provide the most basic metadata, such as a title, brief description and one subject classification. Some of the information will be populated automatically, such as file size and format, and specialist cataloguers will complete the metadata in order to ensure that the resource is properly described. Increasing the responsibilities of contributors may deter them from depositing content. JORUM exists to provide a repository for contributors to be able to deposit their learning materials if they wish to (and some materials from JISC-funded projects that are required to be deposited). The core aim is to build up a repository of quality learning objects, in whatever format or whatever structure suits the pedagogical purpose. This being the case, it may not be practical to impose additional responsibilities on depositors, as this may deter them from contributing. However, learning objects may have particularly specialist content, and the metadata required to describe them accurately may require very specialist knowledge. In this case, the creator is likely to be in the best position to add the descriptive metadata. The question is whether the creator is willing to do this and whether it is feasible to distinguish

[41] Parrott, K. (PADI), *Towards Federation 2001 Preservation of Australian Digital Information Working Party*. Available at: www.nla.gov.au/padi/avccrep.html

[42] RLG/OCLC. *Trusted Digital Repositories: Attributes and Responsibilities*. May 2002. Available at: http://www.rlg.org/longterm/repositories.pdf

more specialist learning objects that require particularly specialist knowledge for effective description and classification.

8.3 User Rights

A trusted digital repository will ensure that it provides a good service with reliable access to its materials via an effective search interface. The level of access that will be maintained should be clearly defined by the repository, in terms of the functionality and look and feel of the resources, as well as the purposes to which downloaded materials may be put. The commitments made will depend on the legal situation, the resources available to the repository and any practical considerations. Preservation processes will affect the materials and will therefore affect the users' experience of the materials, and this should be made clear at the outset. In other words, a user should not necessarily expect that the repository can continue to maintain exactly the same functionality and look and feel that was provided by the original software after the materials have been migrated to a more stable format.

8.4 User Responsibilities

The user should have a minimal amount of responsibilities, and these should be set out in the terms and conditions under which they agree to access and use the materials. This has already been done for JORUM[43], and is not likely to change significantly for learning objects that are preserved and accessed in the long term.

9. Centralised and decentralised models

The concepts of *centralised* and *decentralised* (or distributed) can be used in a number of ways. In terms of computing, distributed usually refers to a situation where "computer programming and data that computers work on are spread out over more than one computer, usually over a network."[44] Learning object repositories within UK academia can be said to follow a decentralised model, as JORUM is only one repository amongst many, but for the purposes of this report, JORUM is defined as a centralised service, because it has a central structure and a central storage facility (though it does reference materials that are not stored within the repository, *see* section 10.10 on interdependence).

In a decentralised service, the responsibility for storing materials may be distributed, and there may also be a number of access points, possibly providing different interfaces and search capabilities, with a central interface that may employ harvesting technologies. In this scenario, long term preservation would either have to be undertaken in a distributed context, or the materials would have to be moved to a central facility for the purposes of preservation. Possibly a master copy and a preservation copy of a learning object could be stored centrally and an access copy could be provided locally.

An archive repository could be set up as a distributed enterprise between a number of institutions, possibly with a central archive providing preservation capabilities as well as institutions undertaking to preserve their own learning materials. This approach may offer the advantages of more flexibility and resilience and the costs could be spread more widely. It would require close communication and cooperation between a central repository and all of the institutions, in order to ensure that materials remained accessible, metadata was being kept up-do-date, migration processes carried out according to consistent standards, etc. It would necessitate individual institutions having the necessary expertise to carry out preservation processes and it is likely that there would be high investment in training staff and a duplication of expertise and resources. At present, many institutional repositories could be classified as "lightweight digital repositories"[45] which means that they are not considering resources to be in any way permanent. A distributed system would require a change to this way of thinking.

A distributed model is often felt to be the best option for digital repositories (e.g. Cedars, Advanced Distributed Learning Initiative, PANDORA). The present e-learning environment in the UK (and elsewhere) is made up of distributed systems, often maintained by individual institutions. However, it is less certain

[43] http://www.jorum.ac.uk/contributors/ctermsofuse/index.html
[44] http://whatis.com
[45] Long, P. *Learning Object Repositories, Digital Repositories and the Reusable Life of Course Content, Campus Technology Magazine* 2004. Available at: http://www.campus-technology.com/article.asp?id=9258

whether this would be the right option for an archive repository. Digital archiving will be a technically and organisationally challenging task, and whilst many institutions may maintain digital repositories, it seems unlikely that a large number of institutions will have the motivation, skill, or resources to undertake the long term archiving of e-learning objects.

A consistent, standards-based approach to digital preservation is important, and there may be some doubt as to whether this could be maintained in a distributed environment. Such an approach may also prove problematic due to funding sources, which are not always designed to facilitate collaborative approaches. Parties would need to undertake to commit a certain level of resources and work together to develop and implement methods and standards. A shared vision might be better achieved by a collaborative approach to develop a central, standards based archiving service. There are clear advantages in being able to draw upon a wide pool of expertise, shared services and tools when setting up a potentially large-scale archive that seeks to be national in its scope. It is generally acknowledged that digital preservation projects benefit from a substantial degree of collaboration, and many of the projects currently in existence seek to create a collaborative infrastructure. A central team could work closely with the Digital Curation Centre, which is being developed as a national centre of expertise and advice. The DCC is intending to develop and maintain central services, standards and tools and it seems appropriate for it to be involved in developing policies and procedures for an e-learning repository. It would also be useful to take advantage of other centres of expertise, such as the Digital Preservation Coalition, National Data Centres, UKOLN and the Arts and Humanities Data Service.

As the JISC Digital Preservation Strategy states "Digital preservation requires new workflows, scarce skills, and close co-operation across different professions from traditional preservation management to computing science."[46] A central archive employing a collaborative approach to management and development may provide the best option.

10. Preservation policy

The policy should be about more than just preservation in its narrower sense; it should ideally be looking at the broader concept of digital curation. It is vital to have a policy that sets out the aims and objectives of the archive and the responsibilities that it undertakes to fulfil, ensuring that there is full accountability. This would need to be monitored over time, in response to reviews of the system and general developments within the digital preservation and e-learning communities.

If the archive is an OAIS (*see* section 15 on the OAIS model), this will inform the policy that is formulated. A policy should cover the following areas:

- the commitments that the archive undertakes to abide by in terms of deposit, preservation, management and access
- the distribution of responsibilities
- the issues of ownership and rights – licences used and the implications for activities undertaken by the archive
- a clear definition of the relationship between the JORUM service and a long term store for learning materials
- recognised sources for learning objects and means of deposit
- the types of information that the archive is willing to accept, including limitations on format (if any)
- the treatment of virtual learning objects that are not stored in the JORUM
- the standards and specifications that are used, such as adherence to the OAIS and international standards for the physical environment
- preservation processes adopted to ensure continued access
- the level of responsibility for preservation
- the establishment of priorities for preservation
- the formats used for preservation

[46] Beagrie, N. *A Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) 2002-2005*. Available at: http://www.jisc.ac.uk/index.cfm?name=pres_continuing

- the addition of metadata for preservation and updating metadata for both access and preservation
- the appraisal process, including when materials are appraised and the criteria used for appraisal
- any restrictions on access
- responsibilities towards depositors, e.g. undertaking to ensure continued authenticity and accuracy of materials over time
- responsibilities towards users, e.g. ensuring continued access to all items within learning object packages or documenting where access is not necessarily guaranteed
- for an OAIS, a definition of the *Designated Community* (*see* section 15 on OAIS)
- commitments to research and development and to collaboration with other bodies undertaking similar work

If the archive becomes compliant with the OAIS then the policy should reflect the mandatory responsibilities that are set down in the OAIS recommendation. The OAIS responsibilities establish a Designated Community, which should be defined and should be able to understand the information provided without needing the assistance of experts.

The OCLC/RLG report on trusted digital repositories[47] establishes the attributes and responsibilities of a repository and also builds upon the OAIS model. It provides detailed information on what makes a repository reliable and trusted by depositors and users. A central recommendation of this report is the need for certification of repositories, either by audit or by adhering to standards established by appropriate agencies. The RLG and National Archives and Records Administration of the USA (NARA) are currently working on a certification process for digital archives.[48] The Digital Curation Centre (DCC) are intending to offer training on designing repositories with audit and certification in mind.

Creators may be reluctant to hand over their learning materials to an archive, where preservation processes will necessarily change the record. The concept of a trusted digital repository, providing reliable long term access, is important in encouraging deposit. Traditional archives and libraries are generally recognised as trusted because they have fulfilled their purpose of providing secure storage and effective access over time. If a learning object repository is to build the same level of trust, it will need to make sure that it has effective, clearly documented procedures for maintaining access and it will need to provide secure storage for the digital materials. By establishing itself as a trusted repository, an archive of e-learning materials would seek to "provide reliable long term access to managed digital resources to its designated community, now and in the future." The OCLC/RLG report recommends a program for certification, providing the basis for trustworthiness, which would dictate the criteria that must be met and employ mechanisms for assessment.

The following sections of this report set out some of the key issues that need to be addressed within a digital preservation policy.

10.1. Quality of Materials

Contributions can be made to JORUM from any member of staff in an institution that has signed the deposit licensing agreement and whom the institution nominates to deposit on its behalf. The JORUM is not taking on the role of pedagogic quality control, operating instead on the principle of allowing any content from bone fide contributors and therefore effectively placing the responsibility for pedagogic quality control with the contributors. JORUM is, however, taking on the role of providing quality control of metadata, working in partnership with the Resource Discovery Network (RDN).

When considering long term retention, it could be argued that the approach to pedagogic quality control should be revised. An undertaking to maintain access to resources over the long term requires a significant investment of time, effort and money. The quality of materials may be considered to be synonymous with the perceived pedagogical value, and this may influence the decision to preserve materials.

[47] RLG/OCLC. *Trusted Digital Repositories: Attributes and Responsibilities.* May 2002. Available at: http://www.rlg.org/longterm/repositories.pdf
[48] http://www.rlg.org/en/page.php?Page_ID=580

"It may be justifiable to spend more money to produce higher quality materials if they will last longer and be used by more people."[49]

This is a complex issue, as there are many factors impacting on the quality of resources. The definition of quality itself is open to question, and it may be that some materials deemed to be high quality do not turn out to be those most highly used.

If there is no appraisal process at any point, then it will be impossible to preserve materials on the basis of their quality, and the decision would have to be to preserve all or none. If it is decided to preserve all learning objects within JORUM then one option would be to measure quality at the point of deposit into JORUM and reject materials felt to be of low quality. This would, however, change the current ethos of JORUM, possibly discourage contributions, slow down the delivery to users and require additional resources.

The JORUM Scoping Study proposes a number of approaches to adding 'secondary metadata', namely peer review, informal user evaluations, star ratings and use of a forum.[50] These approaches could be considered for appraisal of learning objects for long term preservation.

The most likely scenario would seem to be a quality control process at the point of preservation. However, attempting to judge the quality of materials is complicated by the many uses to which they might be put and the fact that a particular learning object may be worthless to one user but be exactly suited to the purposes of another user. If it is decided that the significant level of resources involved in preservation preclude the idea of preserving all materials held in JORUM, then it is inevitable that the perceived quality of resources will play a part in the appraisal process.

10.2. Appraisal

Within the conventional archival context, materials are archived when they are no longer current or semi-current, that is, they are no longer used for the day-to-day activities of an individual or organisation. An archive generally works on the principle of preserving records for as long as they are judged to have value and significance. Whilst some standard criteria can be used for measuring this, it also comes down to the expertise and experience of the appraisers, balancing costs and benefits and anticipated future needs. This scenario would only apply to learning materials if they were to be kept solely for historic purposes. The reality is that long term preservation is about continuing access for the same purposes that the materials were developed – for teaching and learning within higher and further education.

Appraisal of the learning materials would need to be undertaken unless the decision is made to archive all materials. One option would be to store all materials ingested into JORUM in their original format, but only carry out preservation processes on the basis of appraisal criteria. Certainly, if only a selection of materials are to be preserved then appraisal becomes inevitable. The criteria for appraisal would need to be clearly established, though in the end appraisal always introduces a certain level of subjectivity.

Appraisal may be based on four areas:

1. Appraisal based on content

As learning objects are generally developed for specific subject areas, it may be advisable to take advice from individuals with expertise in the relevant subject. This could be combined with a broader view, looking at the potential for reusing across a broader range of subjects. Subject specialists could use their expertise and experience to judge the pedagogical value of the materials. They may consider the value of materials over the long term and look at the level of use of materials, though this will not necessarily provide an indication of future use. Clearly this form of appraisal introduces a level of subjectivity, and as materials are intended to be repurposed, it becomes very difficult to ascertain their value.

2. Appraisal based on metadata

[49] *Barker, E, et al (* JISC). *Long Term Retention and Reuse of E-Learning Objects and Materials.* Nov 2004. Available at: http://www.jisc.ac.uk/index.cfm?name=project_elo (Word doc, v1.4)
[50] *JORUM Scoping and Technical Appraisal Study, 2004, Vol V Metadata*

Along with quality of content, the quality of the metadata may have an impact upon appraisal decisions. Clearly a resource with sub-standard or inadequate metadata may severely diminish the effectiveness of the resource. In addition, aggregate objects where metadata only exists at the 'top level' would be more problematic to preserve than those where metadata is attached to each single resource (*see* section 10.9 on granularity).

3. Appraisal based on risk

An appraisal process may also consider prioritising those materials that are under the greatest threat in terms of the risk of loss of data and the possible cost of retrieval. The JISC-funded project, Digital Asset Assessment Tool (DAAT) may provide helpful guidance, as it aims to develop a tool for identifying the preservation needs of digital holdings, allowing scarce resources to be focussed on those assets where the risk of loss and cost of loss is greatest.[51]

4. Appraisal based on cost and technical implications

The cost of maintaining learning materials may have an impact on preservation – the more simple text based materials will be much cheaper to maintain than interactive, multi-media objects. It may also be that difficulties arise when dealing with obscure or complex proprietary formats. This may not be an ideal basis on which to select materials, but it is inevitable that cost will play a role and it is entirely feasible that learning materials not considered to be of high quality but easy to preserve could be selected over those that may be of higher quality but have major cost implications for preservation. The relationship between preservation and appraisal makes it sensible for decisions about both to be taken jointly.

10.3. Rights

Long term access requires the use of preservation processes and therefore the rights to do this must be obtained. The rights holders must be prepared to grant the repository the right to preserve, with the understanding that materials will inevitably undergo some changes during the process of preservation.

"Simply copying (refreshing) digital materials onto another medium, encapsulating content and software for emulation, or migrating content to new hardware and software, all involve activities which can infringe IPR unless statutory exemptions exist or specific permissions have been obtained from rights holders." [52]

The report on Trusted Digital Repositories states that "at best, a repository will need to arrange separate rights clearance for long term maintenance; at worst, preservation will be compromised because rights clearances for access cannot be obtained."[53] The report recommends further investigation of the complex relationship between digital preservation and intellectual property rights. It may be that long term maintenance and access requires separate agreements from short term access. However, the existence of two distinct agreements may complicated the situation, and only the materials from those contributors prepared to sign the agreement for long term access could be preserved.

The PREMIS Survey[54] found that many repositories have agreements with depositors that describe the uses to which the content can be put rather than detailing the mechanisms of preservation. An agreement should refer to the repository's right to preserve the material, allowing for the changes in preservation techniques and technology that will occur over time. It could refer in broad terms to the actions that the repository may perform in order to preserve material. Warranty clauses could further clarify the extent of the repository's preservation responsibilities. If the archive is an OAIS it should ensure that there is a legally valid transfer agreement that either transfers the intellectual property rights to the archive or states the rights that have been granted to the archive and any limitations that have been imposed. However the IPR issues are dealt with, an OAIS must maintain sufficient control over the materials to be able to preserve them for the long term, and that means that it must be able to migrate information when necessary.

---

[51] http://www.jisc.ac.uk/index.cfm?name=project_daat
[52] *Preservation Management of Digital Materials: A Handbook. 2003.*
[53] *Trusted Digital Repositories: Attributes and Responsibilities.* May 2002
[54] PREMIS Working Group (OCLC/RLG). *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community.* Sept 2004. Available at: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf

The MIT DSpace licence states that contributors must agree that MIT may, without changing the content, translate the submission to any medium or format for the purpose of preservation. They must also agree that MIT may keep more than one copy for the purposes of security, back-up and preservation.[55]

SCRAN, the Scottish Cultural Resources Access Network, is a learning image service which uses a licence-based business model, whereby the contributor retains commercial rights and SCRAN receives perpetual non-exclusive worldwide right to non-profit educational use. However, this may not be applicable for an e-learning repository, as SCRAN effectively offers grant aid for the digitisation of cultural treasures, in exchange for a non-exclusive licence for their educational use.[56]

The JORUM Deposit Licence does cover long term preservation quite comprehensively, and may not require any alterations in order to cover preservation activities. It states that HEFCE, as the licencee, "agrees… to ensure the future preservation and accessibility of the Licensed Material". This undertaking means that it is important for JORUM to continue to investigate preservation options and how future accessibility can be maintained. The institution giving content to JORUM (the licensor) must agree to give a perpetual right to ensure future preservation of the material. HEFCE sub-licenses material to institutions who have signed up as JORUM Users. The JORUM team and indeed users are able to "re-format the Licensed Material in any way to ensure its future preservation and accessibility".[57]

At present the legal complexities mean that learning materials cannot be redeposited in JORUM within new learning objects if they have originally been taken out of JORUM. They can, of course, be incorporated into learning objects within institutional virtual learning environments. The arguments for repurposing become stronger over time, and the ability to download learning objects and incorporate materials into new objects that could then be deposited into JORUM would seem to become increasingly important, adding value to the materials and therefore strengthening the arguments for preservation. This is therefore an area that would be worth further consideration.

10.4. Access

Access needs in terms of successful searching and identification of relevant materials should largely be met by the descriptive metadata already attached to learning materials, and used within the JORUM repository. However, descriptive metadata will need to change over time if it is to remain useful.

Providing users with access to learning objects that are relevant to them requires the metadata to reflect as far as possible the pedagogical purposes to which the objects might be put. Whilst some learning objects may, for example, describe laws or processes where the context is largely irrelevant, e.g. scientific laws, others may be created that describe current process or events where over time the context of the process or event may be seen differently. It is important to consider how the metadata might reflect this.

10.5. Authenticity

Authenticity refers to the need to ensure that a record is what it purports to be. An e-learning resource might be a description of a key experiment or a mathematical formula. For this to be useful in the future, tutors will want to be sure that the description or the formula have not been altered. So whilst these resources may be used in different ways, for different teaching purposes, they must be seen as authentic records.

Digital materials can be changed with alarming ease. Coupled with this, there may also be a need to make changes in order to preserve, and this has major implications for authenticity and integrity over time.

Various mechanisms for verifying digital objects have been developed. A recommendation of the Trusted Digital Repositories report is that metadata recording authentication information should be mandatory. Stored materials must be checked at submission and also at intervals to ensure that the byte streams are maintained, and other authenticity checks should be regularly carried out

---

[55] MacColl, J. (University of Edinburgh). *DSpace Institutional Repositories and Digital Preservation*. Oct 2004. Available at: http://www.dpconline.org/docs/events/041019maccoll.pdf

[56] Royan, B. *eLearning and the Business Case for Digital Libraries*. Oct 2002. Available at: http://www.dpconline.org/graphics/events/presentations/pdf/elearning1a.pdf.

[57] *JORUM Deposit Licence Agreement*. Available at: http://www.jorum.ac.uk/contributors/cregister/index.html,

One simple means of authentication is a checksum, which is a simple measure for protecting the integrity of data by detecting errors in data that is moved or stored. It works by adding up the basic components of a message, typically the bytes, and storing the resulting value. Later, anyone can perform the same operation on the data, compare the result to the authentic checksum, and (assuming that the sums match) conclude that the data was probably not corrupted. This simple process cannot detect the reordering of bytes or multiple errors that effectively cancel each other out, but there are some more sophisticated redundancy checks that consider both the value and position of the bytes.[58]

Digital signatures, which are a method for authenticating digital information, may also be employed to help maintain authenticity.

10.6. Completeness

For a set of records to be complete there needs to be confidence that no items have been removed or added other than in accordance with documented procedures. This requirement needs to be considered for JORUM learning objects – a learning object may consist of a number of items, as well as a number of links to virtual objects. If the links are no longer valid, the object may be said to be incomplete, but the question is whether this is unacceptable or whether the important consideration is that such things should be properly documented. This comes down to the responsibilities that the archive has for preserving learning objects. It would be an option for an archive to make learning objects available but inform users that it is up to them to check the objects are what they purport to be.

10.7. Version control

Long term preservation will need to take into account the possibility of maintaining several versions of an object, which may happen where the original learning object has been supplanted by a more up-to-date version. If all versions are to be kept, the need to classify the versions correctly is of paramount importance.

A repository may store a master archive copy, a preservation copy where the format has been changed or the original has been migrated, and an access copy. There would need to be a clear system for identifying these manifestations of the same object, and therefore persistent, unique identifiers would be essential.

10.8. Unique Identifiers

It is essential to adopt unique, persistent identifiers that conform to a widely accepted identifier scheme. The IEEE LOM (Learning Object Metadata standard) specifies that a globally unique label must identify the learning object. There are a number of identifier schemes available, and JORUM has adopted URIs (Uniform Resource Identifiers).

In an OAIS, a unique identifier or name needs to be provided for accessing the archived object (known as an Archival Information Package). Within the Preservation Description Information the Reference type should include identifiers for uniquely identifying the Content Information. These are also provided to enable external systems to refer to specific Content Information. *See* section 15 on OAIS for more information.

When establishing an archive, it may be advisable to carry out more work to compare different identifier schemes against the preservation and access requirements of the archive.

10.9. Granularity

The level of granularity at which resources are stored, accessed and described has major implications for their preservation and reuse over the longer-term.

"Granularity is the relative size, scale, level of detail, or depth of penetration that characterizes an object or activity."[59]

[58] http://en.wikipedia.org/wiki/Checksum
[59] www.whatis.com

In terms of JORUM learning objects, there are two ways of looking at granularity. In order to consider these it is necessary to think in terms of two types of objects: a single item, which may be described as a non-aggregate learning object (e.g. one Flash movie, one MPEG, one Word document), or an aggregate object, which is a number of items that may be in a number of formats, linked by a structure that is defined in the IMS Manifest File. A non-aggregate learning object, which is not structurally linked to other learning items, may be quite simple and have one focussed purpose (such as to describe a chemical experiment), or it may be very large and contain a whole range of information (including a whole range of experiments and formulae). Conversely, a learning object may be made of up smaller items, which may be simple in themselves (such as text files), but the process of aggregation creates a more complex object. Such a complex object may contain many items, with metadata attached at the top level. If metadata is also attached to the item level then these single items can be viewed as learning objects in their own right. If metadata is not attached at the item level, the items are not strictly speaking learning objects in their own right – they are items that only exist within the aggregate learning object.

An individual item that is complex in terms of its length, learning design and subject matter may cover a topic very thoroughly but its complexity may inhibit future reuse. There may also be implications for preservation. However, it should have comprehensive metadata attached, and this is key for the effective preservation and reuse of the object.

A learning object that has a complex hierarchy and is composed of many parts may provide a comprehensive learning tool that tutors can use off-the-peg, saving them the need to build an object from individual resources. However, taken as a complete object it may be less flexible in terms of future use. It may contain a whole range of formats and a number of links to virtual objects, making preservation a more challenging activity It could be broken down into individual items, but then there are two issues that need to be addressed – whether to also maintain it as a complete object, and whether there is sufficient metadata at the item level to make each part meaningful. (*See* section 11.3.1 on granularity and metadata).

Smaller items and smaller objects may be more flexible and potentially reach a much wider user base, as they can more easily be recontextualised. Many definitions of learning objects emphasise reusability as a common attribute, and if learning objects can be seen as recontextualisable, their value will surely increase. If a learning object is seen to exist within a very defined context, then it may be difficult to reuse it in a different context. As Koppi and Bogle state in their article on learning objects:

"…the maximum reusability depends on how readily the object can be removed from any particular context."[60]

In fact, a learning object at its most basic may not have any inherently pedagogical attributes – it may for example be a simple image. However, it may be distinguished as a learning object because it has metadata attached that facilitates its use within a learning and teaching environment.

If this argument is maintained, then it would suggest that the value of learning materials can be best maintained over the long term by ensuring that aggregate objects can be broken down into individual learning objects, which can then be used in a whole host of ways. Tutors tend to use material from a wide variety of sources and put material together in different ways. The content packaging specifications supported by the JORUM provide a way to structure items into a meaningful whole are designed to facilitate the recombining of learning materials in new ways. A content package may contain any number of items and a user of JORUM can export the content package and extract the item(s) that they wish to use.

In terms of reuse, it would seem that the important thing to consider is what is the most meaningful unit pedagogically, and this is likely to be a unit that is reasonably flexible, able to be used in any number of future learning objects. JORUM is encouraging contributors to create resources in fairly small chunks rather than as large, complex objects. Although there is no compulsion on contributors to create objects in any particular way, X4L contributors are encouraged to prioritise reuse and this is likely to become an increasingly accepted principle to follow when creating learning objects.

In an ideal scenario, if items are to be preserved within an IMS content package there would be adequate metadata attached and no necessary hyperlinks to other units, thus ensuring that there are no necessary dependencies. The aim would be to preserve a free-standing reusable item that is pedagogically useful.

[60] Koppi, T., Bogle, L. and Bogle, M. 'Learning objects, repositories, sharing and reusability' Feb 2005. *Open Learning, vol 20, no.1*

This is feasible for learning objects that consist of one item, but far more complicated for a learning object that consists of more than one item. It would be possible to create aggregate objects from non-aggregate objects. This would have the benefit that each item within an aggregate object, having originally been created as a single object, has meaningful metadata attached. It would make it easier to extract file objects and continue to repurpose them as well as facilitating preservation. However, the more common scenario is likely to be that packaged learning materials are created at the outset with minimal metadata attached to the individual parts.

10.10. Interdependence

Issues of interdependence occur where a learning object is dependent upon other resources (e.g. server side scripts) or external resources, such as links to web pages. This creates problems for long term preservation because the learning objects lose some of their value if preserved in isolation.

Although JORUM is a store for learning materials, not all of the materials will actually be hosted by the repository. In some cases, JORUM will have the metadata, so that the user can find the materials, and then the user will be directed to an alternative location for the actual materials, such as a link to a website. These virtual objects are known in JORUM as resource stubs, and they create additional complexity when considering preservation, with the danger of maintaining dead links and the adverse implications if the archive seeks to become a 'trusted digital repository'. It is not practicable for an archive repository to have responsibility for maintaining external resources, and whilst it would be reasonably straightforward to remove virtual objects that exist as single item learning objects when the resources are no longer available, it is a more problematic situation where they form part of an aggregate learning object. The preservation process may encompass virtual objects, possibly by taking a snap-shot of a web resource, or the decision may be taken not to take responsibility for materials held outside of the repository. One alternative would be to break down aggregate objects into single items, only preserving the items that exist within the repository as actual resources. It will be essential to have a clear policy on the responsibilities that the repository has, if any, for maintaining the ability to find independent linked-to objects referenced from learning objects.

10.11. Reusability

Reusability is probably a more exacting requirement that those of accessibility and understandability. The potential for reusability will depend upon a number of factors. The technical format will have major implications (*see* section 13 on formats). If the object references external material this will have implications for reusability, the more so if the links are a crucial part of the resource. In addition, any dependencies on other resources will be a problem, for example, HTML pages that require server side scripts or Java applets.

A resource that is highly reusable is one that does not reference external content, is not technically dependent on other external resources and can be delivered via a wide variety of platforms or technologies. Reusability is still potentially high if the resource is restricted to one delivery technology, but the technology is relatively ubiquitous.

A 2004 evaluation of learning resources for reusability looked at a number of e-learning projects in the UK and how these projects had approached reusability and interoperability.[61] It is interesting to note that at this time, of the 27 projects under scrutiny, nine said that they were hoping or intending to implement IMS metadata and a number had no plans to use metadata. Some projects intentionally blocked users from reusing resources, though the majority showed a willingness to consider this in principle but were hampered by a lack of time, money and expertise. A few of the projects were making significant efforts to make resources reusable – INHALE[62], which was a JISC funded project, allowed users to repackage the modules, edit them and upload the changed objects back to the database. INHALE was a two year project and long term issues were not covered in the aims and objectives. The subsequent INFORMS Project[63] that has superseded this is looking at the process of disseminating and scaling up the INHALE model for delivery in a wider range of institutions. One of its aims is to investigate the reuse of content, but this does not involve

[61] Currier,S, Campbell, L (University of Strathclyde). *Evaluating Learning Resources for Reusability: The "DNER & Learning Objects" Study*. July 2004. Available at: http://www.unitec.ac.nz/ascilite/proceedings/papers/059.pdf
[62] www.inhale.hud.ac.uk
[63] http://informs.hud.ac.uk/cgi-bin/informs.pl

looking at possible problems of access over time. Another JISC funded project, the Lifesign project,[64] aimed to evaluate the use of networked moving picture and sound material for learning and teaching in the life sciences. This was a project that was successful in many of its aims and objectives, and the evaluation report refers to it as creating "the most highly reusable resources", but the JISC funding period was for less than three years, and the final report to JISC stated that "evaluation should be conducted over a longer period in order to ascertain more fully the issues involved in streaming video in teaching and learning over the long term and the evolving use of the video streams." This is often an issue with projects funded over the short term. Investigation of long term issues becomes difficult.

It will not be practicable, and arguably not desirable, to direct contributors to JORUM to create content designed primarily for reuse. The essential purpose of JORUM is to provide a repository for learning objects that have a pedagogical purpose. It will contain any number of complex learning objects that may be designed with reuse in mind, or they may simply be created to fulfil their purpose as part of an online learning course, with no conscious thought of use beyond that remit. It will be necessary to take content such as this and make a decision about its value for reuse.

11. Metadata

All JORUM learning objects are required to use a set of elements, described in an application profile[65] based upon the UK LOM Core. A number of these are compulsory, and need to be completed before publication, and others are optional. Contributors will be required to add only a small amount of metadata, namely title, description, information on intellectual property rights and a classification. A number of metadata elements will be added automatically, such as date of contribution, role of contributor and file size. JORUM are currently employing a number of cataloguers from the Resource Discovery Network (RDN) to fill in most of the metadata fields.

The UK LOM Core was designed specifically to enable access rather than as preservation metadata, but a number of metadata elements required for preservation will already be included in the administrative, descriptive and technical metadata provided for learning objects. This is certainly the case for the JORUM application profile, which includes many elements that are necessary and useful for preservation. However, it is likely that additional metadata will be required. Digital preservation projects may largely agree on the most important metadata elements, but beyond that, the levels of detail that are described in metadata element sets differ considerably. A decision on the metadata to use will need to consider how necessary or useful it is, whether there are the resources are available to create that metadata, and what the implications are of not creating the metadata. Issues that will impact upon the metadata also include the preservation processes that will be used and the existence and nature of an appraisal process.

11.1 Preservation Metadata

11.1.1 What it is for

"Metadata is key to the management of digital information, however, for long term management and preservation it is even more critical that the right type of metadata is recorded and maintained to enable access and reuse of digital data in the future."[66]

Metadata is required for a number of purposes. It is necessary for resource discovery, and also for describing the technical nature of resources and associated systems, for supporting the management of resources and for preserving resources over time. If a collection is set up with the intended aim of preserving it over time, then metadata is crucial to support this function.

Five key functions of preservation metadata are:

- To store technical information that supports preservation decisions and action
- To document preservation action taken (e.g. migration or emulation)

[64] http://www.lifesign.ac.uk/
[65] http://www.jorum.ac.uk/contributors/chelp/index.html
[66] Patel, M.(UKOLN), *Scoping Study for a DCC Metadata Standards Registry* (draft). July 2004. Available at: www.dcc.ac.uk/docs/Scoping_Study_for_a_DCC_Metadata_Standards_Registry.pdf

- To record effects of preservation strategies
- To ensure authenticity of digital resources over time (e.g. by using digital signatures)
- To note information on collection and rights management

11.1.2 Why preservation metadata is essential for the long term

Within a very few years, hardware and software on which valuable data has been stored can be rendered obsolete. Solutions are constantly being developed and updated, but it is never enough simply to have the resource to hand. It is essential to have the technical and administrative information which provides the background to the creation, delivery, operation and administration of the digital material. Preservation metadata is "the infrastructure that supports the processes associated with digital preservation."[67] It is metadata that provides information on the technical context for the resource (its file format, file size, associated software, version etc), and other information (e.g. copyright information) that is crucial to its ongoing management. It is as integral a part of its preservation as its physical security. Without this information, some or all of the following vital questions, and others, will remain unanswered:

- what is the resource?
- how can it be used?
- how has it been changed?
- who has been involved in its creation/alteration?

Efforts have been made to develop standard specifications for digital preservation metadata, on an international scale as well as nationally and within specific domains. It would be prudent for an e-learning archive repository to take careful account of the work that has been done in this area and utilise broadly agreed and implemented standards. We cannot know for sure what will be required in order to manage digital preservation in the future, but the best option is to collaborate on and adopt national and international standards.

11.1.3 Trends in Preservation Metadata

There are many metadata standards and schemes, and so choosing which to adopt is not easy. The PREMIS report on implementing preservation repositories[68], indicates that the digital preservation community is beginning to coalesce around several emerging trends in the use and management of preservation metadata. This report was based on a survey of 48 repositories, mostly in the US and UK. Most were informed to some extent by the OAIS model and most drew metadata elements from various schemes – 33 different metadata element sets were referred to in the survey. METS (*see* appendix 5) was frequently used for structural metadata and as a container for descriptive and administrative data. Metadata is often stored redundantly in an XML or relational database and with the data objects. Storing metadata in a database allows for fast access for use whilst storing it with the data object makes the object self-defining outside the context of the preservation repository. Multiple versions of objects are usually maintained (originals and some normalized or migrated versions, *see* section 14 for definitions).

The IMS Content Package is the standard way that learning objects are imported into JORUM and exported from JORUM. This consists of a Zip file, containing an IMS manifest file and the learning resources. The manifest file contains the metadata, which includes descriptive, technical and structural metadata.

The intention of the Digital Curation Centre to develop a metadata standards registry will facilitate making informed decisions about a preservation strategy. A registry will provide appropriate up to date information in an area where there are many standards and approaches for curating and preserving digital resources. The DCC proposes adopting the OAIS for managing the collected information and is looking at METS as a suitable means to pull together the metadata required for differing purposes to act as an OAIS Information Package.[69]

---

[67] OCLC/RLG Working Group on Preservation Metadata. *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects.* June 2002. Available at: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

[68] *Implementing Preservation Repositories For Digital Materials: Current Practice and Emerging Trends.* Sept 2004.

[69] *Scoping Study for a DCC Metadata Standards Registry*. July 2004.

11.1.4 Creating Metadata

The survey conducted by the PREMIS Working Group showed that for many projects metadata is added by a combination of depositor, repository staff and automated processes. Seventy five percent of the repositories in the survey expected some metadata to be supplied by submitters, but many of these did not require it, and there was a wide variety in expectations in terms of both quantity and quality of metadata supplied.

JORUM is initially intending to employ specialist cataloguers, and it may be that a long term preservation process would also require staff to be employed to modify and add metadata over time. The modification of descriptive metadata in order to maximise the pedagogical value of resources would seem to require skilled staff.

Automation is clearly preferable where it is practical and accurate, and it is often used for some technical metadata, though it can only be used for limited types of information.

"An argument that is heard very often is that adding metadata is very costly, in terms of creation, maintenance and effort in general. To be able to assess costs it is necessary, however, to take into account also what benefits are derived form it. Neither of these two questions has been answered properly and hardly any research has been done on them."[70]

11.2 Preservation Metadata Frameworks and Standards

11.2.1 OCLC/RLG Metadata Framework

This framework is described in a report by the OCLC/RLG Working Group on Preservation Metadata in June 2002, 'Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects'.

The OCLC/RLG metadata framework specifies three areas that need to be addressed:

- Viability: the bit stream (i.e. the basic sequence of bits) is intact and readable
- Renderability: translation of the bit stream into a form that can be viewed by human users or processed by computers
- Understandability: the content can be interpreted by users

This framework may be useful when considering the preservation of e-learning materials because it combines the consistency of a standard with the flexibility of a framework that can be applied in many different circumstances:

- the Working Group has an international membership of experts and the use of the framework should help to foster co-operation and collaboration with different domains
- it uses the OAIS framework, which is widely recognised as the de facto standard for digital preservation
- it makes no assumptions about the type of digital materials being preserved or the technical solutions adopted
- it is applicable to any type or structure of digital resource
- it combines a number of key preservation metadata schemas already in existence: Cedars[71], the National Library of Australia[72], the Networked European Deposit Library (NEDLIB)[73] and the Online Computer Library Centre (OCLC)[74]
- it can be used at various levels of specificity/granularity

[70] ERPANET. *Metadata in Digital Preservation.* Training Seminar Briefing Document. May 2003. Available at: http://www.erpanet.org/www/products/marburg/documents/erpaTraining-Marburg_BriefingPaper.pdf
[71] http://www.leeds.ac.uk/Cedars/
[72] www.nla.gov.au/preserve/pmeta.html
[73] http://www.kb.nl/coop/nedlib/
[74] http://www.oclc.org/

- elements and sub-elements described are optional
- elements can generally be repeated as appropriate

The OCLC/RLG framework proposes an implementation of two of the components of the OAIS model – *Content Information* and *Preservation Description Information (see* section 15).

The framework represents an expanded conceptual structure of the OAIS information model, defining precise components that capture specific types of information. Further investigation could compare the OCLC/RLG element set against the application profile employed for JORUM learning objects.

See appendix (2) for a list of the OCLC/RLG Metadata Framework elements.

11.2.2 PREMIS: Core Preservation Metadata Elements

Following on from the OCLC/RLG metadata framework, the OCLC Office of Research, OCLC Member Services, and RLG formed a working group aimed at developing a core set of preservation metadata elements'. This also provides guidance on implementation, though it is implementation independent, and suggestions for best practice for populating the elements.

The final PREMIS report was published in May 2005, entitled 'Data Dictionary for Preservation Metadata'[75] and including a data model and examples of implementation. This core element set, which builds upon the OAIS reference model, is intended to reflect elements that are actually implemented in practice, and should provide a useful basis for selecting the key preservation metadata elements for learning objects:

"The Data Dictionary defines and describes an implementable set of core preservation metadata with broad applicability to digital preservation repositories."

PREMIS defines 'core' elements as "things that most working preservation repositories are likely to need to know in order to support digital preservation." It includes minimal descriptive metadata or agent metadata, as these are already being addressed by others, and it covers rights only as far as they pertain to preservation. In addition, there is no format specific metadata and no media or hardware details. The next phase of this project will concentrate on issues of implementation and interoperability, and the intention is to base the future evolvement of the Data Dictionary on experience and feedback.

The PREMIS data model refers to metadata elements as *semantic units* because the working group wanted to emphasise that these elements define the information that a repository needs to know but did not want to specify ways of recording or representing the information. It defines an *Object* as a discrete unit of information in digital form. It defines an *Intellectual Entity* as a coherent set of content, such as a database or Web site. One Intellectual Entity may contain many others, such as a Web site containing web pages and images. An Object entity has three subtypes: representation, file and bit stream. A *representation* is the set of files and structural metadata needed to render the intellectual entity and a *file* is a named and ordered sequence of bytes that is known by an operating system. For example, a journal article may be in PDF format, so that this file equates to the representation. Another journal article may consist of an XML file and two image files, so that three files constitute the representation. The *bit stream* is data within a file that has meaningful common properties for preservation purposes.

JORUM content includes complex aggregate learning objects, and each one of these learning objects can be seen as a coherent set of content, or an Intellectual Entity, that contains a number of other Intellectual Entities or Objects. A learning object might consist, for example, of a PDF file and an MP3 audio file, and this, together with the structural metadata required, would constitute the original representation. JORUM would create an object description for this original representation and then it could create an object description for the for the PDF and an object description for the MP3 file. Thus, there are three object descriptions for this learning object. If the repository then creates a WAVE file from the MP3, an object description would need to be created for this, as well as an object description for the new representation of the original PDF and the WAVE file. There would therefore be more than one representation for the same intellectual entity.

[75] PREMIS Working Group. *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*. May 2005. Available at: http://www.oclc.org/research/projects/pmwg/premis-final.pdf

The structural relationship between objects within JORUM can be complex, and the Data Dictionary acknowledges that the structural relationships between the files that constitute a representation of an Intellectual Entity are essential preservation metadata.

The model also deals with events, agents and rights. Events can include validation and integrity checks as well as activities such as migration. An Object is associated with Events via elements that record the relationship between them. Agents are seen to influence Objects via Events, so the role of the Agent is a property of the Event entity. Rights are associated with Agents, by stating where an Agent grants permission for an action to take place relating to an Object.

Clearly the metadata required in order to maintain usable versions of intellectual entities is considerable, and preservation processes will inevitably create additional representations. The emphasis in the Data Dictionary is on metadata being automatically created, but there will always be a proportion of metadata that needs to be manually created. It will be interesting to monitor how the digital preservation community responds to the Data Dictionary as a practical model. There is a PREMIS implementors' group forum open to members of the PREMIS implementor community.[76]

Appendix (3) provides a summary of the PREMIS Data Dictionary.

11.2.3 United States Government Printing Office

In June 2004 a meeting of experts on digital preservation organised by the US Government Printing Office (the Federal Government's primary centralised resource for gathering, cataloguing and preserving published information) planned to determine a set of specifications for descriptive and preservation metadata. A discussion of the definition of preservation metadata showed differing opinions – for some it refers to administrative metadata and for others to technical metadata. The work of the PREMIS working group was seen as a key development. The group acknowledged the importance of provenance, which refers not only to the creation and background of a resource but also reformatting information and details of who has been responsible for what. METS was put forward as a possible solution for wrapping metadata around an object whilst allowing for a diversity of metadata schemas to be used. It was agreed that the best practice is to create metadata at the information creation stage as metadata collected at this point would be relatively consistent, reliable and automatic. A list of metadata elements was developed in response to the expert advice given to the GPO.[77]

11.2.4 The National Library of Australia

The preservation metadata set that has been developed by the National Library of Australia (NLA) is a data output model that does not specify a means of delivery, but sets down what has to be recorded so that preservation can be managed. The set is based upon the need to manage and describe different levels of resource. It defines collections, objects and files (sub-objects),with the object being the primary focus for management and description. The set distinguishes between a 'work' and its physical or virtual 'manifestations'.[78]

The metadata set is, like all others, based upon assumptions about future requirements. The aim is to support both migration and emulation which allows for some flexibility in terms of the technical solutions that might be adopted.

There are 25 elements in the metadata set. A summary of these elements is given in appendix (4).

11.2.5 The National Library of New Zealand

[76] http://www.loc.gov/standards/premis/

[77] U.S. Government Printing Office. *Report on the Meeting of Experts on Digital Preservation: Metadata Specifications*. June 2004. Available at: http://www.gpoaccess.gov/about/reports/metadata.html

[78] National Library of Australia. *Preservation Metadata for Digital Collections*. October 1999. Available at: http://www.nla.gov.au/preserve/pmeta.html

The NLNZ has a preservation metadata standards framework[79], which has been influenced by the National Library of Australia, the Cedars initiative, OCLC/RLG activities and the emerging consensus around the OAIS Model. The schema is predicated on the idea of a preservation master held in the digital archive retaining the preservation metadata. The preservation master is a rendition of the original into an object that can be preserved, managed and disseminated over time, and it may undergo migration to different formats over time.

The NLNZ has also been developing a software tool to automatically extract preservation metadata from digital files and output it in XML for uploading into a preservation metadata repository.[80] It is expected to be freely available, possibly as open source software. It consists of a generic application and a number of adaptors for specific file formats (adaptors available for Word, Excel, PowerPoint, TIFF, JPEG, WAV, MP3, HTML, GIF and BMP).

11.3 Factors impacting the metadata

This section looks at a number of factors that should be considered when defining the preservation metadata. There are also other considerations that may influence these:

- Types of material to be deposited
- Technical strategies adopted to preserve the digital objects
- The composition of the objects
- The level of granularity to which metadata is applied
- The number of copies, possibly a master copy and access copy, or additional manifestations
- The metadata elements that should be compulsory and those that should be optional or dependent on other factors
- how the metadata may change over time

Another key factor will be the time that it takes to capture this information, and whether there are the necessary resources for this. The level of detail implied by a standard such as the PREMIS Data Dictionary is significant. However, the question is whether effective preservation could be carried out without providing this level of detail.

11.3.1 Granularity

Metadata may be assigned at various levels of an object. In the case of JORUM learning objects, metadata will be assigned to the top level, but it may or may not be assigned to items within aggregate learning objects.

The content packaging specification that JORUM supports and the packaging tools commonly used, such as RELOAD and Microsoft LRN 3.0, allow metadata to be attached at any level, so items within aggregate objects may or may not have metadata attached and the quality of the metadata may vary widely. It is certain that the bulk of the descriptive and technical metadata will relate to the 'top level' object, whether it be an aggregate object or single item object.

Rights management may also be an issue here, as different parts of a packaged learning object may be owned by different parties. The metadata would need to make clear the IPR situation.

11.3.2 Manifestations

It may be that there are a number of manifestations of one object within a repository. In many digital repositories, an archive master is stored separately and the need migrate to new formats will create more versions. The need to preserve a number of manifestations would be a policy decision on the part of the repository.

---

[79] National Library of New Zealand, *Metadata Standards Framework – Preservation Metadata (Rervised)*. June 2003. Available at: http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf

[80] http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction

11.3.3 Compulsory Elements

Whilst it is common to include both compulsory metadata elements and optional elements, there may be some variation in the elements required, depending upon the type of object being described. It may be preferable to have an indication of the significance of each element and make decisions on the basis of what is appropriate for each object in turn. It may be considered that some core elements should always be present.

11.3.4 Rights Management

Some rights management metadata has been incorporated into the application profile for the JORUM repository, and some will be captured in Open Digital Rights Language and exported as extensions to the IMS manifest file. This should also be considered for long term preservation of content. Preservation processes will change resources over time, and this needs to be taken into account in the depositor's agreement. Although the essence of the object may be preserved, it will not be preserved as an identical manifestation of the original. An agreement would need to recognise that access will be provided over the long term. In addition, the arguments for re-incorporating materials into future learning objects within JORUM would become more compelling over time.

12. Preservation Process

Digital format is inherently fragile and unstable in the medium to long term. Materials are dependent upon specific hardware and software and so the longevity of the medium on which information is stored and of the hardware systems that permit users to perceive the stored information are key to maintaining accessibility. We will always be dependent upon machines to interpret the '1's and '0's on the storage medium. It is increasingly easy and cheap to copy and store information, which means that we have largely lost the process of 'natural selection' that existed when information was paper-based. Some commentators believe that it may well be that long term preservation can only really be achieved by storing on paper!

The task of sorting, collating, renewing and preserving a rapidly expanding collection of materials is going to prove increasingly difficult, compounded by the fact that technological change is rapid, and therefore obsolescence may be a threat within a short time frame, maybe only 3-5 years.

The technological solution that is adopted needs to solve the problem of keeping objects long term without degradation to their legibility or comprehensibility. For e-learning materials, reuse is of paramount importance. When considering the preservation methods to adopt, it is essential not to put any barriers in the way of reuse.

Preservation up until the present has typically been about maintaining access to information that has become essentially of historical value rather than continuing to be of current value. This is not the case for e-learning materials. They may need to undergo preservation processes whilst continuing to be relevant within their original pedagogical context, so they may be said to be current but requiring preservation processes to remain accessible:

"Because of the pace at which information technology develops, it is quite possible that the content held in learning objects may become technologically obsolete while it is still pedagogically valuable."[81].

Digital repositories generally have more than one copy of an object. Many store an archival copy, and maybe also an archival back-up copy, and then provide an access copy for users. It is preferable to keep master copies of materials in a separate location to ensure that the archives are secure and safe from unauthorised access.

The access copy may be maintained separately from the master copy, or it may be created on the fly, and a repository may undertake preservation of the access copy or create new access copies from the master as appropriate. Under this sort of system, it is important to maintain the relationship between the access and preservation copy.

[81] *Long Term Retention and Reuse of E-Learning Objects and Materials.* Nov 2004

The DELOS project to establish a framework for a digital preservation testbed environment and produce metrics for testing and validating digital preservation strategies may provide useful feedback on understanding the issues and finding suitable solutions. It concentrates on the planning process that is required before running a testbed – issues that need to be addressed include deciding how the evaluation should work, what the aim of the testbed is, how to evaluate the results, what is important and how to measure the importance of different elements.[82]

12.1 Preserving the Essence

Many commentators believe a preservation process should preserve the 'essence' or 'significant properties' of the material, and within the e-learning community, it may be useful to debate the general principles of significant properties of learning materials. The idea is that the essential characteristics may be defined for each type of record, so that, for example, the essence of a word-processed document or a PowerPoint is determined and documented. In other words, documenting the essence of each genre of record to ensure that the essential characteristics are preserved. It is arguable whether the significant properties of materials remain essentially the same, and therefore whether there can be a standard model for each format. However, it would seem that where resources are limited and choices must be made, it is crucial to determine and document the essential characteristics as soon as possible in order to concentrate on preserving what is important. The Trusted Digital Repositories report makes the recommendation that tools to identify the significant properties of digital objects must be developed and that for practical reasons, significant properties must be identified for formats, rather than individual digital objects. This will effect the technical solutions and required metadata.

"Determining the essence of records is not a science and is open to subjectivities…but it is essential to an efficient, effective and accountable preservation programme. Focusing on the essence of a record allows us to clearly state our archival requirements for the preservation of that record and to be held accountable against those requirements."[83]

A decision needs to be made about which aspects of a record are essential to the correct rendering and meaning of the record and which are non-essential, and therefore resources do not need to be employed in preserving them (e.g. toolbars, button functionality, colour of the interface). This must be well-documented so that future researchers are aware of the archival decisions that have been made.

12.2 Options for what to preserve

The preservation strategy for the archive should clearly set out what is to be preserved and how it is to be preserved.

There are a number of options that may be considered:

- To preserve all formats that come into the archive as comprehensively as possible - a high level of preservation that maintains the exact look and feel of the materials
- To preserve the essential characteristics of all formats
- To preserve the essential characteristics of a selection of formats (a choice based on documented criteria)
- To store all objects in their original format and migrate objects according to a set of criteria, such as level of demand or ease of preservation, or risk of loss.
- To prioritise preserving non-proprietary formats – these formats are available indefinitely and are not dependent upon commercial companies and intellectual property rights. Approaches to preservation may be shared with other repositories. Where non-proprietary formats are not an option, it is best to select proprietary formats that are offered by a number of vendors
- To use the minimal amount of preservation processes required, as migration from one format to another will inevitably cause some loss to the meaning of the record

[82] http://www.dpc.delos.info

[83] Heslop, H et al (National Archives of Australia Green Paper) An Approach to the Preservation of Digital Records. 2002. Available at: http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf

It has been argued that the level of preservation may impact on the demand for materials kept in an archive repository. Some users may require a higher level of preservation, for example, preservation of the exact look and feel of the materials. It would be worth considering what the demand might be within the user community, though user demands need to be measured against resources available.

13. Formats

A long term repository for e-learning materials would need to define the range of formats that will be supported, and have detailed information about how each format will be treated. The policy adopted by the National Archives of Australia, for example, is to take in any format. This is seen as essential to a comprehensive preservation programme. It could be argued that as the JORUM repository will take any format, from the more basic text files and images to video and Macromedia Flash movies, the archive should take any format, from common open source formats to obscure proprietary formats. An archive could agree to take in all formats whilst implementing preservation processes on the basis of documented criteria that might prioritise some formats over others. It may be easier to preserve content stored using file formats based on open standards, but contributors are not compelled to create learning materials using open standards or to use prescribed software. It is certainly important to try to keep options open for the future, rather than limit future preservation options by present processes.

Although the use of standard formats may facilitate preservation, the speed of change in this area means that even where standards do exist, documents created on older versions of the software may not be accessible using newer versions, and many learning objects combine a number of formats, adding to the difficulty of reducing dependence on hardware and software. However, it is still valid to consider the principle of standard formats for preservation:

"It does still make sense to encourage the use of standards where they exist and also to raise awareness of the inherent difficulties…of maintaining access to digital objects originating in non standard formats."[84]

Formats used must meet the needs of producers, users, and the archive. Producers wish to create documents easily and have a high level of flexibility. Restrictions on file formats would impinge on this. Users require objects that are reliable and appropriately functional and the archive itself will want to employ stable technologies and ensure that suitable metadata is attached to the resources. The system in which materials are preserved and the formats used are not necessarily those that will be used to provide access. Separation of access from preservation allows the choice of file formats and software systems for preservation to be largely independent of the requirements of the user community. Even if the best, open and widely accepted file formats are selected for preservation, it is inevitable that migration to newer formats will need to happen over time.

It will be a huge challenge to maintain the information contained within the more complex formats in a way that will allow it to be rendered correctly. It will simply not be possible to maintain all of the sophisticated software required to run dynamic and interactive content. An archive repository is likely to define and support a limited number of file formats and those that are maintained are likely to be standards based and preferably open source. It may also undertake to maintain the bit stream of materials even where the format is not currently supported. There may be further options that could be adopted, where the archive cannot absolutely guarantee to support the format.

In general, commentators agree that proprietary formats are unsuitable for long term preservation. The issues surrounding the intellectual property rights mitigate against this. It is preferable to have control over the technologies used to access records, but formats are often controlled as the intellectual property of a commercial entity, and competition drives frequent changes in individual file formats. Non-proprietary formats ensure that the archive has control over access to the records and may also have the advantage of spreading costs between like-minded institutions for the financial costs of development. In addition, versions of proprietary file formats tend to be short lived and application software does not generally provide import facilities for every previous file format version.

[84] PADI. *Digital Archives: Who keeps them and who pays?: Response on behalf of PADI Working Group to the study commissioned by the United Kingdom's National Preservation Office, August 1997*. Available at: http://www.nla.gov.au/padi/response.html

If there is no non-proprietary alternative, then it would be advisable to favour proprietary technologies that can be sourced from multiple vendors.

There is the question of whether the display format should be archived or the underlying data. Formats used for ready rendering on the Web frequently differ from the format of content in the underlying publishing system. HTML or PDF formats delivered on the Web may use data stored as XML. HTML or PDF may well maintain the look and feel of the original display, but many believe they will present archiving problems because the rendering software will certainly be superseded over time. The SGML or XML marked-up text will be less sensitive to technological change, but ensuring the ability to re-render it as it was originally displayed will be technically complex.

13.1 Current Initiatives and Developments

This section of the report highlights some initiatives and refers to two key formats. The JISC report, 'Long Term Retention and Reuse of E-Learning Objects and Materials'[85] has a useful section on formats.

The International Council on Archives 'Electronic Records Workbook'[86] has a useful list of features considered important when selecting a preservation format. The format chosen should ideally be:

- able to represent all information and relationships between information that are considered to be significant
- defined by an international, national or publicly available standard
- proven in terms of longevity
- usable for access purposes or capable of being transformed into formats which are usable for access purposes
- independent of any particular hardware or software environment
- capable of automated conversion from original formats to preservation formats

The availability of file format information is crucial to the successful design and implementation of rendering tools. This is discussed in the final report by the JISC funded Representation and Rendering Project based at the University of Leeds.[87] This provides an assessment of file format information available for various categories, from web sites to funded collections to open source developments. It also provides advice on file format identification and some discussion of user documentation. The report highlights PRONOM (*see* section 13.1.3) and the File Format Registry at Harvard University as leading initiatives to provide file format information. It also highlights the problems of obtaining information about the composition of their file formats from some commercial providers, which will have major implications for the preservation processes adopted. The key recommendations are to obtain and secure file format information for use by the whole digital preservation community to aid the development of rendering solutions and to collect and preserve technical hardware and software documentation, which is crucial, in the absence of file format information, as a means to inform the emulation development process.

The INFORM methodology (INvestigation of FOrmats based on Risk Management) has been developed by the OCLC as a means to measure the risk factors of digital formats and to provide guidelines for preservation action plans. An article by Andreas Stanescu of the OCLC provides a brief explanation of this methodology.[88]

13.1.1 Global Digital Format Registry

The idea of a Global Digital Format Registry (GDFR)[89] has been around for some time. An international ad hoc committee led by Harvard LTI and MIT DSpace was set up in 2002 with the aim of setting up a registry

[85] *Long Term Retention and Reuse of E-Learning Objects and Materials*. November 2004.
[86] International Council on Archives, *Electronic Records: A Workbook for Archivists*. 2005. Available at: http://www.ica.org/biblio.php?pdocid=285
[87] Representation and Rendering Project, University of Leeds. *Survey and Assessment of Sources of Information on File Formats and Software Documentation*. 2003. Available at: http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf
[88] Stanescu , A. (OCLC). *Assessing the durability of formats in a digital preservation environment: The INFORM Methodology.* D-Lib Magazine, November 2004. Available at: http://www.dlib.org/
[89] http://hul.harvard.edu/gdfr/

to “maintain persistent, unambiguous bindings between public *identifiers* for digital formats and *Representation Information* for those formats.”[90] The GDFR model defines OAIS Representation Information specifically as the significant syntactic and semantic properties of formats.

A significant amount of work has been done on developing a GDFR, and there is a trail implementation (http://tom.library.upenn.edu/) but it is not yet a reality.

13.1.2 DCC Representation Information Registry

The Digital Curation Centre is now involved in developing a Representation Information Registry, to be an OAIS, which is a broader concept than a data format registry, and there is a hope that this will develop into a distributed, global, collaboration and may converge with PRONOM.[91]

13.1.3 PRONOM

PRONOM[92] was developed by the Digital Preservation Department of the UK National Archives (TNA). It is a resource for technical information about the file formats used to store electronic records, and the software products that are required to create, render, or migrate these formats. It is intended to provide authoritative information about software products, their support lifecycles and technical requirements, and about the file formats which they support. It records information on the support life cycles of software products and the intention is to allow the automated reporting of products which are due to become obsolete. PRONOM will be including a tool to allow users to determine all the possible routes for migrating between two file formats, and the TNA Digital Preservation Department is also looking at developing precise methods for describing and quantifying the changes that may occur during migration.

13.1 4 Digital Formats for Library of Congress Collections

A website[93] providing information about digital content formats, providing an inventory of information about current and emerging formats.

13.1.5 JHOVE

JSTOR and the Harvard University Library are collaborating on a project to develop an extensible framework for format validation, called JHOVE (JSTOR/Harvard Object validation environment).[94] This will provide functions to identify formats and validate them in order to ascertain that they comply to the specification for the purported format. It will also provide functions for the characterisation of digital objects, determining the format-specific significant properties of an object of a given format. These characteristics all form part of an object’s representation information, as defined by the OAIS. The actions are performed by modules, and the JHOVE website provides a list of current standard modules (including HTML, JPG, GIF, TIFF, PDF, WAVE and XML).

13.2 Formats

13.2.1 PDF

PDF has been proposed as a possible archival format, and it does have the advantage of preserving the look and feel of a record. PDF as it stands would certainly not satisfy all requirements for digital records, and it is a proprietary format. However, a current initiative is looking to develop PDF as a file format for long term preservation. Known as PDF/A, the standard is intended to specify the use of PDF in a manner specifically

---

[90] Abrams, SL, Seaman, D. *Towards a Global Digital Format Registry. World Library and Information Congress*, August 2003. Available at: http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf

[91] http://dev.dcc.ac.uk/dccrrt/

[92] http://www.nationalarchives.gov.uk/pronom/default.htm

[93] http://www.digitalpreservation.gov/formats/index.shtml

[94] http://hul.harvard.edu/jhove/

geared to long term use and seeks to make this into an ISO standard. PDF does offer advantages over some other proprietary standards because it is based on a published specification and the most recent version offers a rich metadata capability, the Extensible Metadata Platform (XMP). PDF/A would define the set of PDF components that could be used and specify how they could be used. So, for example, javascript and executable files would not be accepted, encryption would be forbidden and audio and video content would be forbidden. The standard is described in a *DigiNews* article[95] and there is also a white paper by Adobe which argues the case for PDF as a suitable format for the long term:

"PDF represents not only the data contained in the document but also the exact form the document took. The file can be viewed without the originating application. In fact, ten years from now, and into the future, users will still be able to view the file exactly as it was created." [96]

The paper argues for PDF on the basis that it is being used by the Dutch National Archive, the Australian Victorian Electronic Record Strategy and the UK Public Record Office (now The National Archives).

13.2.2 XML

XML is a possible long term solution, and one which offers independence from specific applications. However, the cost is likely to be the loss of a significant amount of the formatting and layout. In addition XML is not particularly suited for non-textual documents, such as images, and may not, therefore, suit the needs of e-learning materials.

"[XML] offers some important advantages, chiefly that textual content will achieve a degree of independence from specific information technology configurations…Textual content may be reasonably well represented in the XML version, but much of the original document's formatting and layout will likely be lost…A further problem is that XML is sharply limited in its support for nontextual data such as photographs and other graphics." [97].

In Australia a tool known as XENA (XML Electronic Normalising of Archives) has been developed which converts a range of file formats to XML representations (*see* section 16.1).

It seems that XML is considered to be one of the most promising tools for preservation, but work on developing XML tools is still in the early stages.

14. Technological Options for Preservation

14.1 Technology Preservation

Preserving a digital object together with all of the software and hardware needed to interpret it. This approach would be resource intensive and lead to a huge amount of ageing hardware, so it would not really be viable for the long term.

14.2 Bit Level Preservation

It is generally standard practice to preserve an object in its original form, and a bit stream can be preserved indefinitely. This allows for future strategies, doesn't restrict the type of learning objects that are taken in by the archive and minimises the risk of data loss or corruption. Emulation consists of preserving the bit stream and using hardware and software emulators.

The archive repository may decide to differentiate between formats where usability is maintained and those for which it only undertakes to ensures that the bits are maintained as deposited and that the associated documentation is kept usable to support possible future activities.

---

[95] LeFurgy, W.G. *PDF/A: Developing a File Format for Long term Preservation RLG DigiNews Vol.7 No.6,* Dec 2003
[96] *PDF as a Standard for Archiving* (Adobe White Paper). 2002. Available at http://www.adobe.com/products/acrobat/pdfs/pdfarchiving.pdf,
[97] *PDF/A: Developing a file format for long term preservation.* Dec 2003.

The Cedars project on digital preservation referred to preservation of the byte stream rather than the bit stream, as it is current practice for digital data to be stored in bytes and the packing of 8 individual bits into a byte is effectively hidden from the system level user.

14.3 Emulation

This generally refers to running the original software on emulators that mimic the behaviour of obsolete hardware and operating systems. Most emulators just emulate a hardware architecture. If a specific operating system is required for the desired software, it must be provided as well (and may itself be emulated). Both the OS and the software will then be interpreted by the emulator, rather than being run by native hardware.

Emulation has the advantage of keeping the resource in its original state and also retaining the 'look and feel' of the original, though users would need to have access to the appropriate emulation software and be able to operate within the original computing environment. It clearly requires the skill to write the necessary programmes and issues are likely to arise to do with intellectual property when emulating proprietary software.

Users would need to learn the original computing environment in order to access the information and it may be, for example, that by 2080 the idea of mouse clicks on icons is completely foreign to most people. It does not seem practicable to ask users to install and use new software applications.

This method seems to have been less popular than migration in practice, but it may be argued that "emulation of the original computational environment, gives the very best hope for recreation of the experience of a preserved digital object".[98] If it is a requirement that the look and feel of the application is maintained, then emulation may be the appropriate approach to adopt, though this may be problematic for proprietary software in particular.

A recent article in RLG DigiNews compares the cost of migration and emulation:

"…migration applies to the entire collection repetitively: each and every single object in the digital archive has to be converted, again and again. This means that the bigger the archive gets, the more expensive migration will be. This is in contrast with emulation: emulation tools apply to the collection as a whole and do not affect the format of individual digital objects…Emulation and migration are inherently different in terms of life cycle management, which causes a serious difference in costs. While migration applies to all objects in the collection repetitively, emulation applies to the entire collection as a whole. This makes emulation most cost-effective in cases of large collections, despite the relatively high initial costs for developing an emulation device."[99]

14.4 Migration

Migration is the technique for long term retention of records that has been most widely practised. It involves both copying records periodically from one generation of technology to another and moving information from one file format to another.

Migration is intended to move information from obsolete or potentially obsolete formats to current or more stable formats. This may be combined with some standardisation of formats. However, it may lead to some loss in functionality or accuracy, so that the original 'look and feel' will be lost to some extent and the 'performance' of the object may not be exactly the same as previously. Metadata should record the successive migration processes and help to preserve the authenticity of the object.

There is likely to be some loss of data as a result of migration This will depend upon a number of factors: the new data format, the type of process, the number of treatments applied and the level of human intervention.

---

[98] Holdsworth, D, Sargeant, D.M (University of Leeds). *A Blueprint for Representation Information in the OAIS Model.* 1999. Available at: http://www.personal.leeds.ac.uk/~ecldh/Cedars/ieee00.html

[99] Oltmans, E., Kol, N. *A Comparison between Migration and Emulation in Terms of Costs.* RLG DigiNews 9(2), 15 April 2005. Available at: http://www.rlg.org/en/page.php?Page_ID=20571#article0

Migration could in theory involve using the application software, so that data is migrated from one version to another. This is likely to prove costly and impractical because file formats are generally short lived and therefore frequent migrations would be necessary. It also means relying on commercial software providers, which is a risky policy. This procedure is likely to result in loss of information, especially for more complex objects, and it does not provide for detailed documentation of migration processes. Migration to a standard, open format, such as XML, is generally considered a better option. This may encourage optimization of the material for future preservation actions and for enhanced delivery (*see* section 14.6).

14.5 Refreshing

Copying content from one storage media to the same storage media. This may be a simple method of preservation, but it is not likely to be a practical option in an environment where formats become obsolete so rapidly.

14.6 Normalisation

This is a method of migration that involves restrictions on formats, choosing only stable, normalised formats. It may involve a one-off conversion to a data format that suits the purposes of the archive, such as XML, and the archive would undertake to support the archival data formats indefinitely. The emphasis with normalisation is not on preserving the original process, but on preserving the 'essence' of the record. The variety of formats of digital objects in an archive will affect the cost and complexity of operation. To control such complexity and cost, an archive may want to normalise deposited objects into a set of preferred formats whenever possible. For example, all raster images are stored in TIFF, and JPEG or GIF images are converted into the TIFF format. This conversion process would create a 'normalised master' version of the record. If a record cannot be normalised (i.e. there is no conversion tool to change the data format of the original record into an equivalent archival data format), a record needs to be made to indicate that the original record is un-normalised.

This is thought to be a cost effective option and minimises the risk of data loss or corruption, though the look and feel of the materials may be compromised and there is a risk of information loss.

14.7 Preservation Processes: Conclusions

Materials cannot remain accessible without undergoing preservation processes of some kind. However, it is preferable to minimise the number of preservation processes. Not only are there the cost implications to consider, which are likely to be substantial, but there is also the fact that each process will inevitably alter the record. The look and feel characteristics that are considered essential must be identified and systematic methods used to preserve them. All preservation processes must be well documented.

All processes applied should endeavour to maximise future options for access rather than impose increasing limitations. The priority should be for the user to be able to access the record as easily as possible. This should partly be achieved by using non-proprietary solutions. Users should not have to learn new systems and commands or install new software.

As yet, there is little empirical data that can be used to recommend a particular method of preservation. A successful solution to the problem of archiving would ideally find a comprehensive, systematic, and dynamic means for preserving virtually any kind of electronic record, free from dependence on any specific hardware or software.

15. Framework: OAIS Model

The Open Archival Information System model is an ISO standard developed by the Consultative Committee for Space Data Systems (CCSDS).[100] It is probably the most influential model for how a digital archive

[100] Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System OAIS)*. January 2002. Available at: http://www.ccsds.org/documents/650x0b1.pdf

repository should work, providing a framework for the whole range of functions associated with digital preservation:

- ingest
- archival storage
- data management
- administration
- preservation planning
- access

An OAIS archive is one that conforms to the model and responsibilities detailed in the recommendation:

An OAIS is "an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community."[101]

The OAIS model is widely recommended. JISC has undertaken to "promote use of the broad understandings and concepts embodied in the Reference Model for Open Archival Information Systems (OAIS) as a conceptual model for construction and management of digital archives."[102]. The Digital Curation Centre has stated that the OAIS model will be used as a basis for their approach to curation and the RLG/OCLC report, 'Trusted Digital Repositories', recommends the OAIS Model as a useful framework. The demonstrator system developed by Cedars to recommend techniques for long term storage of digital data conformed to the spirit of the OAIS model. The National Library of Australia's preservation metadata element set was informed by the OAIS amongst other models.

Overall, there is general agreement that the OAIS model provides a useful and scaleable framework, and it also encourages the adoption of common terms to represent actions. However, in practice there are differences in the way that the model informs implementations of preservation repositories. According to the PREMIS Survey[103] many repositories find it useful as a guide for showing the requirements for a digital archive and for helping with the planning stages, but not all use it comprehensively. For the purposes of considering an archive of learning objects, it is useful to understand the model and appreciate how it might help with considering all of the functions of the archive, even if the model is not actually implemented to the extent of the archive becoming an OAIS.

It is important to remember that the OAIS is a conceptual model and not a system for implementation. Whilst it may serve as a foundation for an e-learning archive, defining the basic functional components of the archive and listing the minimum requirements that an archival system should meet, it is not a guide for implementation but a starting point, setting out what an implemented system should incorporate.

15.1 OAIS Responsibilities

To be OAIS compliant an archive needs to support the model and fulfil the responsibilities that the Recommendation sets down. These include obtaining sufficient control of information in order to ensure long term preservation, determining the *Designated Community* that the archive serves, following documented policies and procedures and ensuring that the Designated Community has access to the preserved information. The Designated Community is defined as the community which the archive serves. In the case of learning objects in JORUM, the Designated Community is the tutors, and indirectly the students, within higher and further education.

---

[101] *Reference Model for an Open Archival Information System*, section 1.1

[102] Beagrie, N. *A Continuing Access and Digital Preservation Strategy for the JISC 2002-2005*. Available at: http://www.jisc.ac.uk/index.cfm?name=pres_continuing

[103] PREMIS Working Group. *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community* Sept 2004. Available at: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf

15.2 OAIS Model and Preservation Metadata

The OAIS effectively consists of a number of models, and the *Information Model* is the part of the framework dealing with the metadata requirements of digital objects.

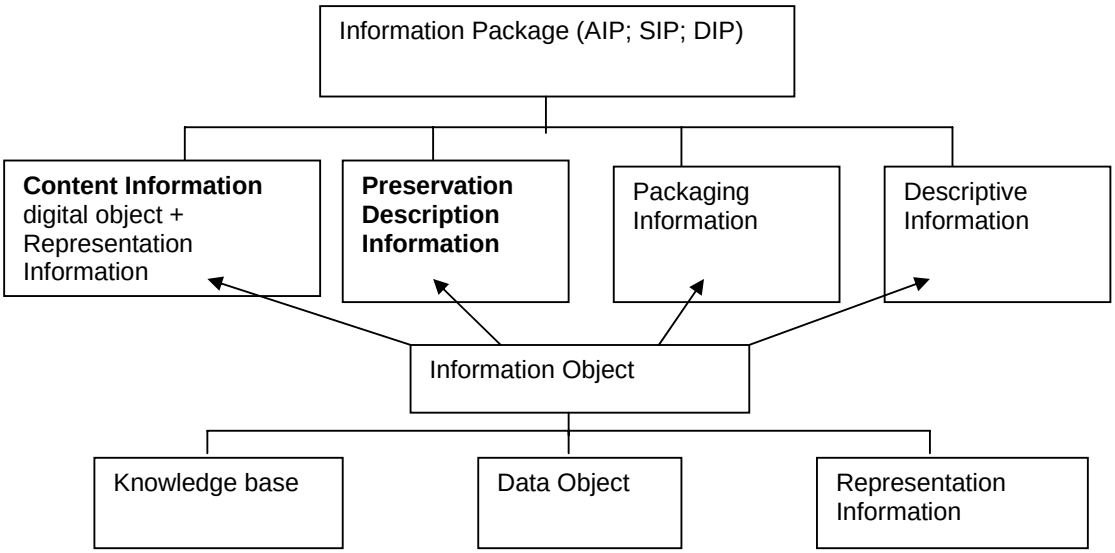


fig 2: OAIS Information Model

A Data Object may be physical or digital. In the case of e-learning, we are only looking at digital objects. Interpretation of the Data Object is achieved through the users' *Knowledge Base* and the *Representation Information*, which imparts meaning to the data. An Information Object combines the Data Object with the necessary Representation Information.

An *Information Package* contains *Content Information* and *Preservation Description Information* (PDI), in other words it contains the Data Object itself and the Representation Information (RI) needed to make the data meaningful, together with the appropriate preservation metadata. In addition to this, the *Packaging Information* binds, identifies and relates the Content Information and the PDI. The *Descriptive Information* supports discovery and retrieval. In the case of JORUM learning objects, the Descriptive Information largely equates to the metadata already provided and defined in the application profile, though certain parts of this metadata, such as the technical requirements for access, might be categorised as Representation Information.

15.3 OAIS Information Packages

Information Packages may be *Submission Information Packages (SIP)*, *Archival Information Packages (AIP)* or *Dissemination Information Packages (DIP)*. This reflects the fact that the submission (SIP) and dissemination (DIP) may not include all of the metadata required within an AIP. The Submission Information Package is transformed into one or more Archival Information Packages. A user will receive a Dissemination Information Package, which may or may not include all of the PDI and may include several AIP's.

The Archival Information Package is at the heart of an OAIS. It is a container that has all of the necessary information to enable long term preservation and access. The AIP is an Information Object containing the Content Information, which comprises the Data Object and Representation Information. In addition, the AIP contains the Preservation Description Information (which is effectively an Information Object in its own right). The AIP is delimited by and identified by the Packaging Information.

Packaging Information Object

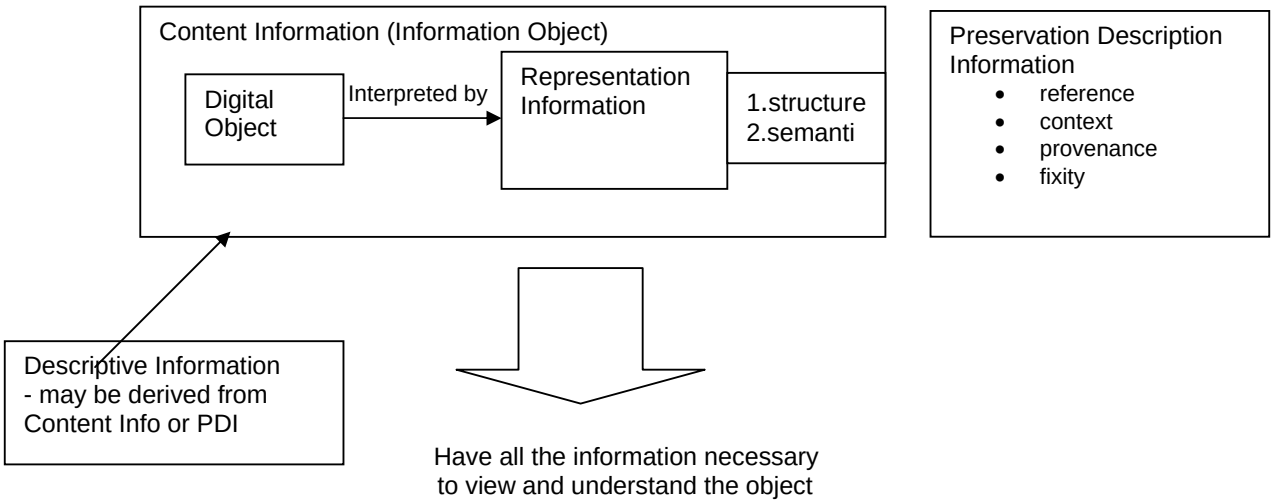


fig 3: OAIS Packaging Information Object

Content Information consists of the digital object and Representation Information (RI) metadata, which provides information required for meaningful access to the object. Representation Information supplements the Knowledge Base that the user already has (e.g. knowledge of English, knowledge of programming language).

15.4 Representation Information

The Representation Information is used to provide additional meaning and is necessary to render and understand the bit sequences constituting the Content Data Object. For example, a Microsoft Word document, which will have a complex format that requires Representation Information describing the format, together with Access Software that understands the Representation Information. Another example would be the ASCII definition that describes how a sequence of bits is mapped into a symbol or, for an ASCII table of numbers giving the coordinates of a location on the Earth, the Representation Information will describe the format of the numbers, their definitions as latitude and longitude and the definition of their units as degrees. The Representation Information is likely to include both Structure Information and Semantic Information, which provides additional meaning.

Representation Information will often reference other Representation Information, such as a reference to the ASCII standard, which is then additional RI that is needed for a full understanding of the Digital Object. The RI may also include information deemed necessary to adequately understand the meaning of the document. If a Word document is saved as an ASCII file, then the Representation Information required would change. There may be a *Representation Network*, with successive layers of Representation Information. This could be tied up by a physical document. It is therefore possible to have a huge amount of Representation Information, with learning materials referencing other information sources, which reference further information sources, and so on. It may be that practical boundaries need to be set, probably based upon the knowledge base of the Designated Community. It may be that many Information Objects require the same Representation Information, in which case it can just be provided once to the OAIS.

If the archive is aiming to make the learning materials available for the Designated Community, then the Representation Information provided may not need to be as comprehensive as that which would be required for a broader community. One of the problems with Representation Information is that it can in theory become recursive to a substantial degree, to the point, for example, where an English dictionary is required in order to ensure accurate understanding of the terms used in the Data Object. The assumption is that tutors wishing to use the materials have an understanding of the subject area and therefore an understanding of the content and purpose of the materials. In other words, they have a Knowledge Base that enables them to understand the materials that they download from the repository. A Data Object will always require Representation Information, to enable the user to interpret the data object correctly, but the RI provided need be sufficient only to meet the needs of the Designated Community. A learning object that refers to a particular scientific experiment may require RI in order to interpret the data, but it may not require further RI in order to explain the data further, as the Designated Community are familiar with the data.

However, when considering the future reuse and repurposing of the data, more comprehensive RI may help to promote wider use. The Designated Community may be defined not simply as scientists and science students, as a Data Object may be relevant to other disciplines which have a different Knowledge Base and its relevance may change over time. If a learning object is to be relevant to a wider community, it might involve adding explanations in support of the Representation Information. Within JORUM, learning objects are intended to be used by tutors and incorporated into a whole range of courses, aimed at students of differing expertise, many of whom will not yet be subject specialists, and so if the Representation Information enables the object to be accessible to students with a fairly minimal understanding of the subject, it means that it can be more readily incorporated into different courses, aimed at different levels of experience.

15.5 Preservation Description Information

The Preservation Description Information (PDI) provides descriptive metadata so that the Content Information can be understood indefinitely. It is categorised into four types of preserving information:

1. Reference Information

This identifies the mechanisms used to provide assigned identifiers for the Content Information. It allows outside systems to refer to the particular CI unambiguously. It therefore identifies and describes the Content and is therefore key to resource discovery. It may be worth considering ‘holding places’ for metadata schemes such as Dublin Core, to allow these to be reused where appropriate.

2. Context Information

This provides information on why the Content Information was created, i.e. its purpose, which is often crucial to understanding its meaning, and also how it relates to other Objects.

3. Provenance Information

This documents the origin or source and the history of the Content Information, recording any changes that have been made and the custodial history. It gives future users some assurance as to the reliability of the Content Information.

4. Fixity Information

This is used to ensure that the Content Object has not been altered in an undocumented manner and provides a means to document authentication mechanisms.

15.6 The application of the OAIS Model

15.6.1 Submission of Learning Objects to the Archive

In a scenario where learning objects are passed from JORUM to an archive, the materials received would equate to Submission Information Packages (SIPs), which would be in the form of IMS Content Packages. These SIP’s are likely to contain all of the necessary Descriptive Information and a certain amount of the PDI. These would then need to be prepared for storage and management within the archive. An AIP would need to be generated, complying to the archive’s data formatting and documentation standards.

JORUM may maintain learning objects in a whole variety of formats. However, typically, once the Data Object is transferred via a Submission Information Package, the form and content may change, as the archive may limit the number of preservation formats.

One SIP could equate to one AIP, though it would be possible to break down SIP’s into a number of AIP in order to end up with smaller chunks of materials that can be more easily repurposed. Each AIP would include full PDI necessary for preservation. One AIP may contain a collection of other AIP’s, and this may be a model suitable for aggregate learning objects.

The interaction between the OAIS and the contributor of the Information Objects would depend partly upon the relationship between JORUM and a learning archive (i.e.. whether JORUM is seen as the contributor, or whether JORUM is itself the OAIS). It might be that JORUM would submit SIP's to the archive. Whatever the situation, the OAIS data model provides a means to define the submission of data that is provided via SIP's.

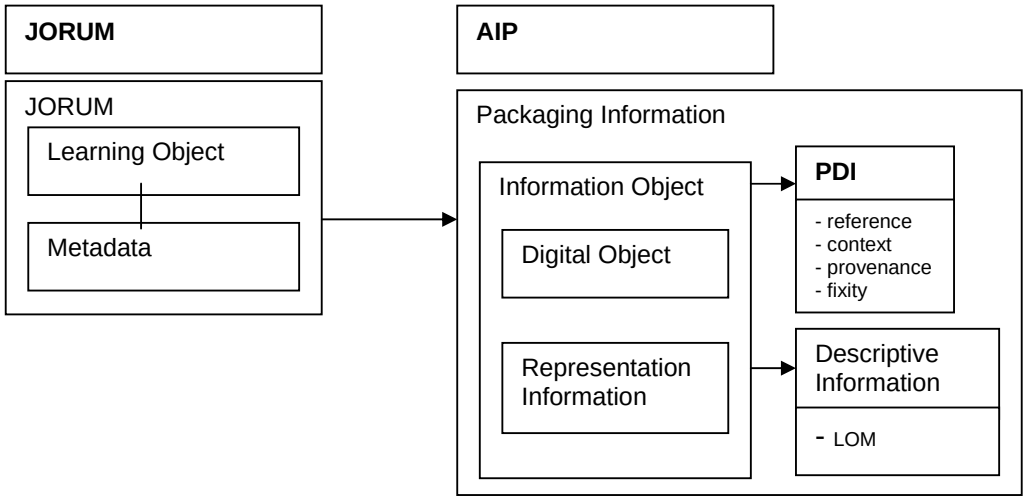


fig 4: Application of the OAIS Model to the JORUM

15.6.2 Dissemination of Learning Objects to Users

An OAIS would follow the model set out in the recommendation, for providing Dissemination Information Packages via Data Dissemination Sessions. Essentially, the OAIS provides all or part of an AIP to a user, which may include all or part of the PDI. The Descriptive Information may be provided with the Dissemination Information Package, or at any time before or after the transfer.

15.6.3 Current Work on Implementations of OAIS

The OCLC/RLG Metadata Framework proposes an implementation of Content Information and Preservation Description Information metadata and creates additional structure in order to set out a detailed metadata element set (*see* appendix 2).

A JISC funded project is currently underway as part of the 'Supporting Digital Preservation and Asset Management in Institutions' programme to explore implementations of the OAIS (Oct 2004 – Sept 2006). Assessment of UK Data Archive and The National Archives compliance with Open Archival Information System and Metadata Encoding and Transmission Standard[104] is being carried out by the UK Data Archive (UKDA) at the University of Essex in partnership with The National Archives (TNA). Both organisations have mapped their systems and metadata to the OAIS Reference Model and METS to test the assumption that both broadly comply with the standards.

15.7 Functional Entities

The OAIS describes *Functional Entities* for the full range of functions that an archive repository should undertake. These are ingest, archival storage, data management, administration, preservation planning and access. There is not the time to look at these in detail within this report, but they could provide a very useful model for ensuring that all of the requisite functions are carried out.

[104] http://www.data-archive.ac.uk/home/oaismets.asp

15.8 Conclusions

Five years from the issue of the OAIS (i.e. 2002) it will be reviewed to take into account new requirements and technologies, which should mean that it will remain relevant. It has already become an ISO standard, and if widely adopted it will encourage the standardisation of terms and concepts that are used within the digital preservation community, allowing solutions to be more easily compared and archival requirements to be more widely understood. If other digital repositories, and in particular e-learning repositories, choose to implement OAIS archives, this will provide a major benefit for JORUM to exchange ideas, compare techniques, and work together to develop solutions and broad consensus. One of the principles of the OAIS model is that different archives may take on the role of producer or consumer to other archives, thus promoting the interoperability of OAIS archives. The OAIS Recommendation has a section exploring the degree of interaction and cooperation among OAIS archives, both technical and managerial.

Use of the OAIS model may help to formulate a policy for an e-learning archive. If intending to become an OAIS, the minimum responsibilities of the archive are set out in the Recommendation and conformance to these would need to be made explicit in the policy.

16. Implementation Examples

This section provides information on some initiatives that may provide useful information and advise. It covers the National Archives of Australia, the Cedars Project and the US Electronic Records Archive. Clearly there are additional projects and services that would be of interest, and may be worth further investigation, such as the work of The National Archives of the UK, the National Library of New Zealand, and the Florida DAITSS project, which has briefly been mentioned earlier in this report.

16.1 National Archives of Australia

There is a great deal of investment in solutions to long term preservation in Australia, and the National Archives of Australia (NAA) has carried out some interesting work. The basic conceptual approach of the Digital Preservation Project, which began in 2001, is provided in a National Archives Green Paper, 'An Approach to the Preservation of Digital Records'.[105]

The NAA has developed what they call the 'performance model', which helps to clarify the fundamental nature of a digital record. This principle recognises the fact that the user does not actually access the source directly. The data file contains the meaning, but by itself it is meaningless. It requires the technology to render it correctly – the combination of data and process to create a performance. This is similar to the OAIS idea of Representation Information, which gives meaning to the Data Object.

The performance relies on the technology, so it cannot be said that each experience of a digital record is the same. The aim is to preserve the correct rendering (even though this is not identical for each user) rather than the original data source and the accompanying technology. As long as the digital record can be rendered over time, with as little change to the original rendering as possible, then the data source and the technology can be replaced.

The NAA preservation process is based upon preserving both the original digital record (in its original format) and a version converted into an archival data format. The NAA uses XML-based archive data formats, as XML is non-proprietary and a good solution for long term access across different platforms:

"The idea of creating our own data formats to meet the preservation needs of many record types is not as daunting as it at first seems. Mark-up language technology, and specifically XML, allows us to quickly and easily create our own non-proprietary archival formats that can preserve a record's essence."[106].

XML can therefore be used indefinitely, even if it is effectively replaced with another data format technology, so that there should be no need for migration to a new format, thus significantly reducing costs and the damaging effects of cyclical migrations. XML may not be a suitable format to maintain the essential meaning of a digital object, however, it is possible to use XML to create data formats that can preserve the essence of the record. This approach requires XML expertise and XML applications to be constantly maintained.

During the preservation process, all records are copied onto a new medium and then converted into an original bit stream object structured according to the NAA long term package format. Part of each original bit stream object is the original record in its original data format. Every record on the new medium is also converted into an archival data format that the Archive plans to support indefinitely. If this normalisation process cannot be carried out (i.e. there is no conversion tool) then the record is logged as being un-normalised. Normalisation converts the original source object into XML. The idea is that records only need to be migrated once into archival formats.

The National Archives of Australia have developed an XML normaliser, called Xena (XML Electronic Normalising of Archives)[107] a Java tool that converts the source object into XML. Xena is open source software with a plugin architecture, and the file formats that can currently be converted to standard XML

[105] Heslop, H. et al (National Archives of Australia) *An Approach to the Preservation of Digital Records*. Green Paper, 2002. Available at: http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf
[106] *An Approach to the Preservation of Digital Records. 2002.*
[107] http://xena.sourceforge.net/

include MS Word, Excel, Powerpoint, JPG, GIF, TIFF, HTML, RTF, PNG, BMP, plan text, relational databases and web sites.

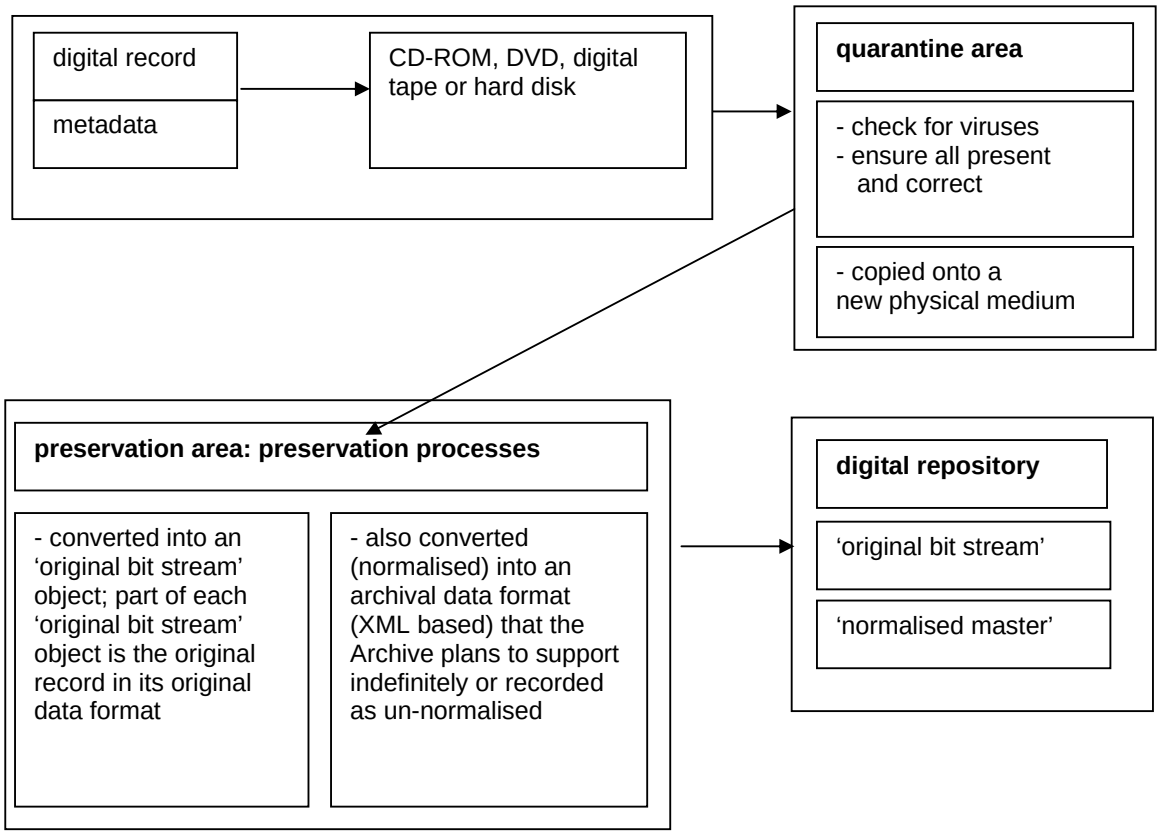


fig 5: Model based on the National Archives of Australia approach to digital preservation (see www.naa.gov.au/recordkeeping/preservation/digital/digital_repository.html)

16.2. Cedars (Curl Exemplars in Digital Archives)

The Cedars Project,[108] which ran from April 1998 to March 2002 was intended to "address strategic, methodological and practical issues and provide guidance in best practice for digital preservation".[109]

It worked on the creation of a Distributed Digital Archive Prototype[110] based on the Open Archival Information System model, and employing a metadata specification, which was subsequently integrated into the OCLC/RLG Metadata Framework. The Prototype was based on a federated archive, demonstrating the feasibility of linking between archives with different technical systems. This may not necessary be entirely relevant for an e-learning repository, although this model could in theory be implemented, with distributed archiving stores used.

Cedars advocated an overall management approach to the technical strategy and recommend facilitating interoperation with a distributed federation of archives. The Prototype report makes a number of recommendations including the promotion of media independence, the identification of significant properties of resources prior to ingest and the separation of the access and archival storage, with the AIP as a self-standing unit. The report also describes an approach to the management of the OAIS Representation Networks, which is the set of Representation Information that fully describes the meaning of an OAIS Data

[108] http://www.leeds.ac.uk/Cedars/
[109] http://www.leeds.ac.uk/Cedars/overview/overview.html
[110] Cedars Guide to the Distributed Digital Archiving Prototype. March 2002. Available at: http://www.leeds.ac.uk/Cedars/guideto/cdap/guidetocdap.pdf

Object. The Cedars Project sees the maintenance and development of these Networks over time as essential in order to manage technical obsolescence.

The Cedars Demonstrator Archiving Prototype worked on the principle of separating the content from its carrier as a means of challenging technological obsolescence. At ingest (i.e. deposit), an object is removed from its medium of acquisition and preserved as a stream of bytes: 'There is good reason to be confident that data held as a stream of bytes can be preserved indefinitely.'[111]

16.3. US Electronic Records Archive

The Electronic Records Archive[112] is being developed as an OAIS by the US National Archives and Records Administration (NARA) to be "a comprehensive, systematic, and dynamic means for preserving virtually any kind of electronic record, free from dependence on any specific hardware or software."[113]

In August 2004, NARA awarded a $20m contract to two companies to develop "a revolutionary system that will capture electronic information, regardless of its format, save it permanently, and make it accessible on whatever hardware or software is currently in use" .[114] The goal is to have a functional subset of the system by 2007.

16.3.1 Persistent Archives Method (PAM)

The Electronic Records Archive refers to the Persistent Archives Method (PAM). This is being developed within the context of an implementation of the OAIS as a means of keeping, and if necessary converting, records to a format that respects all its essential properties, but reduces or eliminates dependencies on specific hardware or software, enabling faithful reproduction over varieties and generations of technology. It aims to deal with a wide range of records and formats. This is a similar approach to the NAA, where the archive process involves converting to an archive format. The Persistent Archives Method is the same as the method referred to as Persistent Object Preservation (POP). It relies on the idea of capturing the essence of a record, which refers to content, structure, context, and presentation.

Electronic documents, which have both content and form or structure, are represented as complex real-world objects having attributes and behaviours that must be maintained. In more complex cases, such as a Geographic Information System (GIS), the electronic records would be represented as complex objects that include descriptions of behaviours that explain how their subcomponents interact with other objects to form the records. For example, a cartographic record displayed in a GIS may include elements drawn from dozens of different layers in the system. The behaviours of such a record include all of the operations needed to retrieve the required data from the appropriate layers, defined in the logical data model, and cause them to be displayed correctly in a map or chart.

Within a learning object repository, the 'original order' of the records would refer to the logical order of the learning materials forming an object. These connections would need to be preserved in order to preserve the structure of the records.

"Representing and encapsulating complex records from contemporary information systems as objects and collections requires a flexible, open, standardized, self-describing and extensible modelling language that will preserve the content, context, and structure of records. Persistently preserving records would entail encoding of record attributes, behaviours, content, and structure into an encapsulated object. This object would then represent all information required to re-instantiate the record at a future time. A parallel approach would be implemented to re-instantiate aggregates of records, such as files and series."[115]

---

[111] *Cedars Guide to: Digital Preservation Strategies*. April 2002. Available at: http://www.leeds.ac.uk/Cedars/guideto/dpstrategies/dpstrategies.html

[112] http://www.archives.gov/electronic_records_archives/

[113] NARA Press Release, August 3 2004. Available at: http://www.archives.gov/media_desk/press_releases/nr04-74.html

[114] ibid.

[115] NARA Electronic Records Archives. *Introduction to Preservation and Access Levels Concepts*. Dec 2003. Available at: http://www.archives.gov/electronic_records_archives/pdf/preservation_and_access_levels.pdf

17. The Next Steps

Rather than drawing conclusions from this report, it would seem more appropriate to consider how best to proceed from this initial analysis of current activities. This report is intended as a starting point, setting out many of the issues that are involved in long term preservation. In order to take this forward towards implementation, further research would need to be carried out it many of these areas, and it would be prudent to take advantage of the conclusions of the digital preservation initiatives that are currently underway, many of which have been referenced in this report.

The main issues that require further investigation are set out below, in a logical order (though none of these areas are discreet, and each will impact upon the others). This is *not* an exhaustive list of issues that need to be addressed, but provides a basic structure that can be used for the next phase of work, or modified in accordance with the feedback provided by JISC and the JORUM Steering Committee.

17.1 Is it worth preserving learning materials long term?

The first question to ask is whether it is worth preserving learning materials at all. It is important to put forward compelling arguments for long term preservation. However, this is not an issue that can be fully considered without taking into account the costs involved, which are bound up with the preservation processes used, the quantity of metadata required, and whether it can be automatically generated or requires manual creation, and the approach taken to appraisal, which relates to the quantity and quality of materials that might be preserved. It is therefore very difficult to put forward clear, comprehensive and well supported arguments for preservation without investigating many of these issues.

Recommendation: to set out a case for preserving learning materials as far as is possible within the constraints of available information and to present the case for preservation to the user community, to try to gauge whether there is a potential demand for learning materials over time and whether long term preservation will encourage deposit.

17.2 Costs and Value

This is an area of digital preservation that is receiving significant attention at the moment, and the espida project in particular may provide some useful tools and recommendations. The value of intangible assets, such as digital resources, is clearly difficult to measure, but it is this value that is at the heart of a reasoned cost/benefit analysis which is necessary in order to make a considered and persuasive argument in favour of long term preservation.

Recommendation: Keep a watching brief on current research projects and examine recent findings and case studies in order to create a cost model that is relevant for learning objects and articulate the value of learning objects, attempting to examine the different dimensions of that value and changes over time.

17.3 Functions, framework and responsibilities

The functions that form part of a repository system need to be clearly identified and defined. A framework can then provide the structure to link the different functional components together. (*Framework* is used here to mean a support structure to hold together all of the conceptual components and functional requirements of the system in a logical relationship).

The OAIS is now an international standard, and it should provide many advantages if it is used in terms of the adoption of common terms and concepts and aiding with the development of the strategy and techniques for preservation. It also sets out the responsibilities that need to be met by an archive, which may help to encourage the development of a learning object archive as a trusted digital repository. If the decision is taken to conform to the OAIS framework, then this will inform the functional requirements, though the model will need to be modified and extended in order to work within the particular environment of the JORUM/learning object archive.

The framework that is developed should show how JORUM will relate to an archiving facility, and there are a number of issues that need to be considered in order to decide whether JORUM should have

responsibility for access and for preservation, or whether the activities of access and preservation are separate. In addition, there is the need to consider the storage of master archival copies of learning objects.

Once functions and relationships have been established, it will be useful to set out proposals, as far as possible, for where responsibilities should lie for carrying out all of the required functions.

Recommendation: Investigation and further discussion of OAIS model, and whether this is appropriate to implement either in full or in part for an archive of learning materials. Identification of functional requirements and creation of a framework, or creation of a number of options for a framework, setting out the implications of each option. Identification of responsibilities for carrying out functions where possible.

17.4 Metadata requirements

One of the core issues in terms of commitment of resources is the addition of suitable metadata for long term preservation. A great deal of work has been carried out internationally on metadata schemas for preservation and JORUM has adopted its own metadata application profile, based on the LOM, which comprehensively covers descriptive metadata and does also cover technical and other metadata that is required for preservation. It will be important to carefully consider the implications of attaching metadata to each individual object within an aggregate object, something that will be resource-intensive, but may be felt necessary in order to successfully preserve and reuse resources.

Recommendation: make a detailed comparison of the JORUM application profile and preservation metadata schemas, in particular the PREMIS core elements, in order to identify the metadata that will be required for preservation, in addition to that which is already being created for JORUM learning objects. Make a recommendation about the levels at which metadata should be attached and the implications of this.

17.5 What to preserve

The decision about what to preserve involves a range of issues. Unless all objects are to be preserved, it will be necessary to adopt an appraisal process and therefore to set out appraisal criteria. A decision will also need to be made about what is essential to preserve in terms of the look and feel and functionality of the learning materials. This work will need to build upon the current report, which sets out the concepts of ‘essence’ and ‘performance’, in order to think more specifically about learning objects and about the user experience of the objects. It will be useful to take advantage of the work that has been carried out by the digital preservation community on the essence of different formats, and to look at any particular implications for learning objects.

The granularity of learning objects is clearly a key issue - whether to preserve aggregate learning objects as a whole and also preserve each individual part separately.

Recommendation: Set out the options for appraisal and how an appraisal process might work in practice. Look more closely at the work that has been carried out on preserving the essence of digital materials, and whether this is applicable to learning objects, or whether there are particular issues that need to be addressed. Consider the implications of preserving aggregate learning objects.

17.6 How to preserve

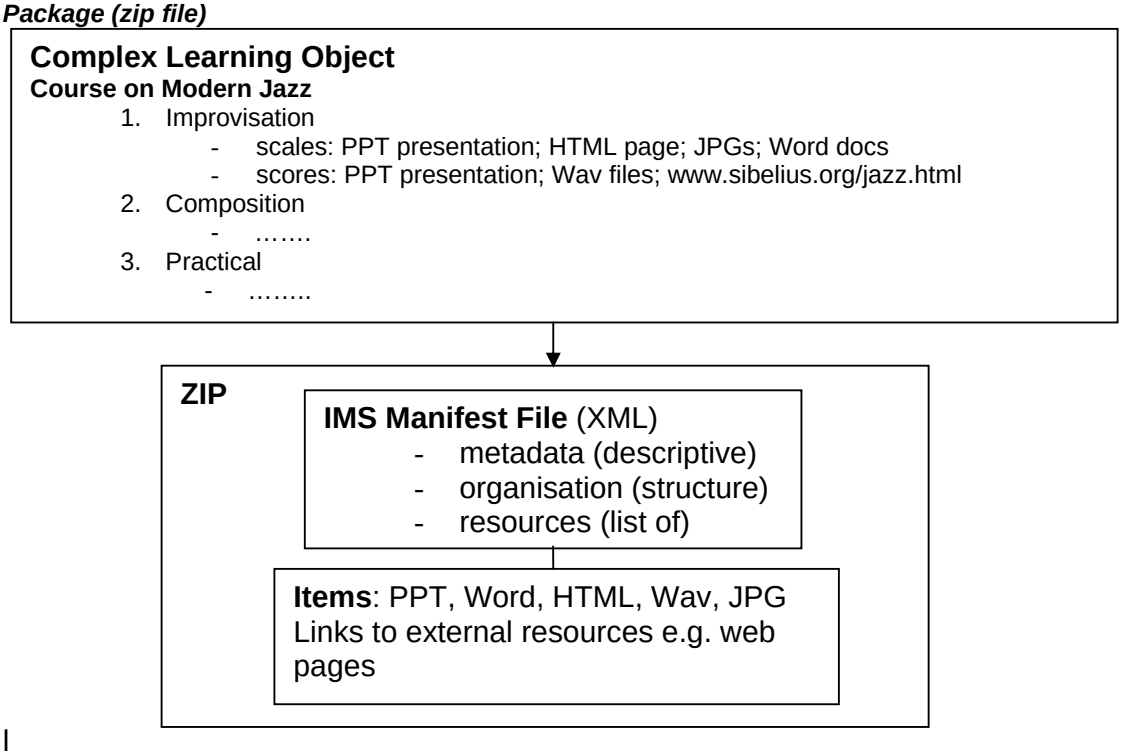
This report has not looked in detail at the practical preservation processes that may be used. It will be necessary to investigate these options further and make some assessment of the most appropriate option for preserving learning objects. However, a detailed analysis of preservation processes would require the input of individuals with a high level of technical skills.

Recommendation: Seek to benefit from current work being undertaken in practical implementation of digital preservation in order to help inform a strategy for learning objects. Look more closely at each option for preservation, seeking to clarify the advantages and disadvantages of each approach and whether there are particular issues relating to learning objects that might influence the decision.

18. Appendices

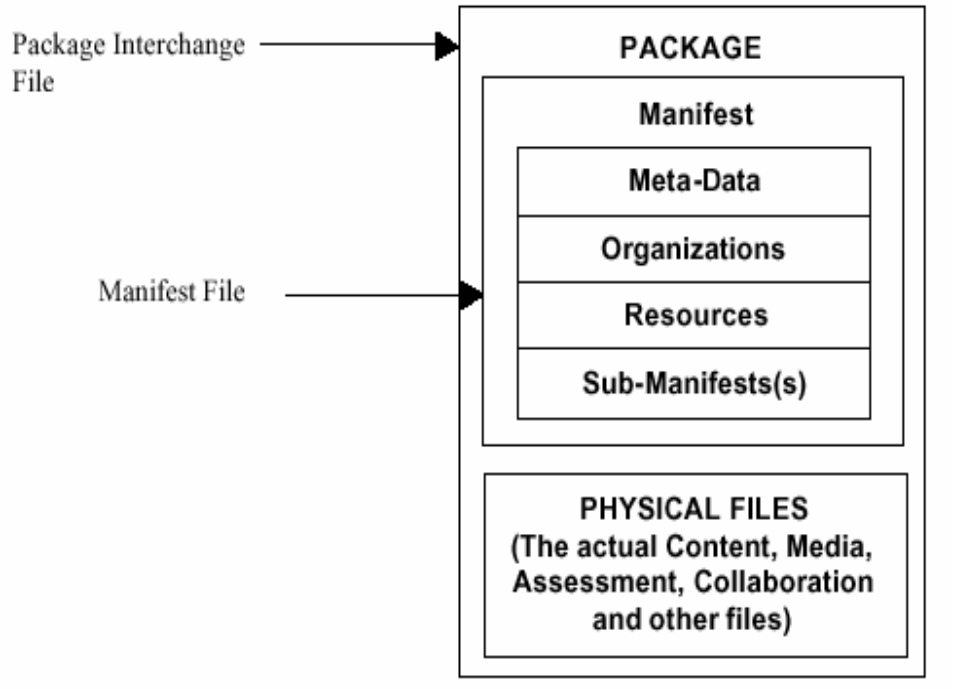
Appendix (1) JORUM Learning Objects

Example of an aggregate learning object indicating the way that this becomes an IMS Content Package (zip file)



I

IMS Content Package



Appendix (2) OCLC/RLG Metadata Elements

OCLC/RLG Metadata Elements For Content Information

A Content Information Package consists of the '**Content Data Object**' and its associated **Representation Information**.

Components	Elements	Elements		Description / Sub-Elements	
Content Data Object Description - characteristics and features of the content data object necessary to understand and render its content	**Underlying abstract form description**			human readable description e.g. of a file system or database components, to facilitate converting the archived byte stream into the correct components to render the object	
	Structural type			class of object e.g. sound, text, database	
	Technical infrastructure of complex object			internal structure and interrelationships, e.g. web page is one HTML file, 3 embedded static GIFs and one audio WAV file	
	File description			technical spec of the file(s), applying to formats used to access content, e.g. GIF:dimensions, resolution, colour palette, compression algorithms	
	Installation requirements			procedures required to install object e.g. ZIP must be unpacked and stored on local hard drive in specified directory prior to use	
	Size			in bytes	
	Access inhibitors			inhibitors such as encryption, watermarks, passwords	
	Access facilitators			facilitators such as time markers in audio files, navigational links in hypertext docs	
	Significant properties			properties to be preserved – maybe content rather than 'look and feel'	
	Functionality			functional attributes e.g. web page contains interactive javascript application	
	Description of rendered content			how content should be viewed and interpreted e.g. numbers may be clarified as temperature readings from specific experiment	
	Quirks			losses due to preservation process e.g. broken hyperlinks, javascript no longer functional	
	Documentation			supporting documentation such as manuals	
Environmental Description - hardware/software environment (could describe only what is supported/archived or minimum requirements or recommended environment)	**Software environment**	**Rendering Programs**	**Transformation process**	implementation/mechanism to transform byte stream into correct representation of components, e.g. unzip/untar a file	
				Transformer engine	identifies software engine for transformation process
					Parameters
					Input format
					Output format
					Location
					Documentation

			Display/Access Application	software programme for human-readable display
				input format
				output format
				location
				documentation
		Operating system	**OS name**	name of software platform e.g. windows, linux
			OS version	version of OS system
			Location	location of working copy of OS system
			Documentation	supporting documentation
	Hardware environment	**Computational resources**	**Microprocessor requirements**	necessary processing power – general e.g. 333Mz or specific e.g. intel pentium 333Mz
			Memory requirements	sufficient memory resources
			Documentation	supporting documentation
		Storage	**Storage information**	storage resources for operation of software environment and rendering of content, eg. 33 MB of hard disk space
			Documentation	supporting documentation
		Peripherals	**Peripheral requirements**	additional requirements such as sound cards, speakers, monitor resolution, CD-Rom drive
			Documentation	supporting documentation
		Hardware environment as a whole	**Location**	location of physical devices

OCLC/RLG Metadata Elements For Preservation Description Information

Components	Elements	Description / Sub-Elements	Description / Sub-Elements	
Reference Information - identification	**Archival system identification**	to uniquely identify the data object within the system		
		Value	actual value	
		Construction method	means by which id created and assigned	
		Responsible agency	entity responsible	
	Global identification	to uniquely identify the data object to external systems		
		Value	actual value, e.g. persistent URL	
		Construction method	means by which created and assigned	
		Responsible agency	entity responsible	
	Resource description	information for resource discovery extracted from metadata sources or created by the archive		
		Existing metadata	any metadata scheme used, eg. DC, MARC	
			Existing records	a single instance of a particular metadata scheme
Context Information - relationship of the content information with its environment	**Reason for creation**	why a data object was created		
	Relationships	relationships with other data objects		
		Manifestation	relationships with other manifestations of the same object e.g. versions in HTML and PDF	
			Relationship type	
			Identification	
		Intellectual content	relationships between the intellectual content of this object and other objects, eg. set of objects form a web page, part of digitized images of art collection	
			Relationship type	such as web page, collection, serial
			Identification	enables link to related object
Provenance Information - the history of the content information; changes since origination and custody	**Origin**	process by which object was created	**Event** (generic event associated with provenance) - designation (name of event) - procedure (details of procedure) - date - responsible agency (entity responsible) - outcome (description of) - note (any additional info)	
	Pre-ingest	history of object (maintenance, changes in custody) up until submission to archive		
	Ingest	process by which object is ingested e.g. migrated to standard format; object broken down to component parts		
	Archival retention	maintenance, management, etc of object whilst in archive		
	Rights management	legal uses of object (access permissions, legal responsibilities)		
Fixity Information	**Object authentication**	info to meet minimum requirements for authentication (eg digital signature)	**Authentication type**	technique used e.g. digital sig.

- data integrity checks; not altered in an undocumented manner			**Authentication procedure**	info on steps to implement technique
			Authentication date	most recent use of authentication type
			Authentication result	result of most recent use of authentication type

Notes:

There are often a number of structural layers defined for the metadata, so the upper layers may represent organisational information rather than a metadata element where information is recorded, and the lower layers may not necessarily be used – this is flexible, as metadata can occur at varying levels of specificity.

The result of an Event could be the generation of a new Content Data Object
Events may impact on other metadata
Event element is repeatable and can occur at multiple stages of the life cycle of the object
Ingest metadata might include archival retention decision – how long kept, who responsible for decision

Appendix (3) PREMIS Data Dictionary

Providing structure of semantic units with brief descriptions

Semantic Unit (1)	Semantic Unit (2)	Semantic Unit (3)	Semantic Unit (4)	Description
objectIdentifier				unique identifier designation
	objectIdentifierType			domain within which the identifier is unique
	objectIdentifierValue			value of identifier
preservationLevel				a value indicating the set of preservation functions
objectCategory				category of object to which metadata applies
objectCharateristics				technical properties of file or bitstream
	compositionLevel			whether the object is subject to processes of decoding or unbundling
	fixity			whether altered in an undocumented/unauthorized way
		messageDigestAlgorithm		algorithm used to construct the message digest
		messageDigest		output of the message digest algorithm
		messageDigestOriginator		agent that created original message digest
	size			size in bytes of file
	format			
		formatDesignation		identification of the format of the object
			formatName	designation of the format of file or bitstream
			formatVersion	version given in formatName
		formatRegistry		information about the format by reference to an entry in a format registry
			formatRegistryName	identifying format registry
			format RegistryKey	key to reference an entry for this format in the registry
			formatRegistryRole	purpose/use of registry
	significantProperties			important characteristics
	inhibitors			features intended to inhibit access/use/migration

		inhibitorType		inhibitor method
		inhibitorTarget		content/function protected by inhibitor
		inhibitorKey		decryption key or password
creatingApplication				information about the application that created the object
	creatingApplicationName			designation for name of software that created the object
	creatingApplicationVersion			version of software program that created the object
	dateCreatedByApplication			date and time object created
originalName				name of object before any renaming by repository
storage				how and where file stored
	contentLocation			info needed to retrieve file from storage system or access bitstream within file
		contentLocationType		means of referencing location of content
		contentLocationValue		reference to location of content
	storageMedium			physical medium
environment				hardware/software supporting use of object
	environmentCharacteristic			extent to which described environment supports its purpose
	environmentPurpose			uses supported by specified environment
	environmentNote			additional information
	dependency			non-software component or associated file needed to use the file e.g. a schema
		dependencyName		designation for component
		dependencyIdentifier		unique designation used to identify dependent reource
			dependencyIdentifierType	domain in which identifier is unique
			dependencyIdentifierValue	value of dependency identifier
	software			required to render/use the object
		swName		manufacturer and title of

				application
		swVersion		version(s) referenced in name
		swType		class/category of software
		swOtherInformation		additional requirements/instructions related to software
		swDependency		name/version of any software component needed by software for this object
	hardware			hardware components needed by software or user of software
		hwName		manufacturer,model,version
		hwType		class/category
		hwOtherInformation		additional requirements or instructions
signatureInformation				needed to use a digital signature to authenticate signer and/or information in object
	signatureInformationEncoding			encoding for values of sigValue, keyInfo, certificateInfo
	signer			responsible for generating signature
	signatureMethod			designation for encryption and has algorithms
	signatureValue			digital signature
	signatureValidationRules			operation to be performed as part of signature validation
	signatureProperties			additional information about generation of signature
	keyInformation			info about signer's pubic key needed to validate digital signature
		keyType		denoted by algorithm used to generate the key
		keyValue		value of signer's public key
		keyVerificationInformation		additional info needed to verify signer's public key
relationship				relationship between this and other objects
	relationshipType			high-level, nature of

				relationship
	relationshipSubType			specific, nature of relationship
	relatedObjectIdentification			identifier and sequential context of related resource
		relatedObjectIdentifierType		designation of domain within which identifier is unique
		relatedObjectIdentifierValue		value of related object identifier
		relatedObjectSequence		order of related object relative to other objects with same type of relationship
	relatedEventIdentification			identifier and contextual sequency of event associated with relationship
		relatedEventIdentifierType		eventIdentifierType of related event
		relatedEventIdentifierValue		eventIdentifierValue of related event
		relatedEventSequence		order of related event
linkingEventIdentifier				eventIdentifier of event associated with the object
	linkingEventIdentifierType			eventIdentifierType value of related event
	linkingEventIdentifierValue			eventIdentifierValue value of related event
linkingIntellectualEntityIdentifier				for an intellectual entity associated with the object
	linkingIntellectualEntityIdentifierType			designation of domain
	linkingIntellectualEntityIdentifierValue			value of identifier
linkingPermissionStatementIdentifier				for a permission statement associated with the object
	linkingPermissionStatementIdentifierType			designation of domain in which identifier is unique
	linkingPermissionStatementIdentifierValue			value of identifier

Appendix (4) National Library of Australia Metadata element set

NB: Elements in bold are essential (at object level) or essential if applicable

1	**Persistent Identifier**
2.	**Date of Creation**
3	**Structural Type** (sound, video, text)
4	**Technical Infrastructure of Complex Object** (format, no. of files, type of files)
5	**File Description** (image, audio, video, text, database, executables – including format, version, resolution, compression,etc)
6	**Known system requirements** (OS, plug-ins, Java, etc)
7	**Installation Requirements**
8	Storage Information (capacity and format, eg 1.3MB on CD)
9	**Access Inhibitors** (encryption, watermark)
10	**Finding and Searching Aids, and Access Facilitators** (system or method to enhance access)
11	Preservation Action Permission (have permission to copy)
12	Validation (validation mechanism to verify authenticity eg. checksum)
13	**Relationships** (between this manifestation and other objects – metadata, finding aids, etc)
14	**Quirks** (loss in functionality or change in look and feel)
15	**Archiving Decision: work** (whether work should be archived and date)
16	**Decision Reason: work** (why decision to archive or not made)
17	**Institution Responsible for Archiving Decision: work**
18	**Archiving Decision: manifestation**
19	**Decision Reason: manifestation**
20	**Institution Responsible for Archiving Decision: manifestation**
21	**Intention Type** (intended use – preservation master/access copy)
22	**Institution with preservation responsibility** (agency responsible, date of start responsibility)
23	**Process** (process applied to produce current manifestation) • name of process • name of agency responsible • critical hardware used • critical software used • how process carried out • guidelines specified to implement process • date and time • result • process rationale • changes • other
24	Record Creator (have contributed data to this record)
25	Other (cover anything else)

Appendix (5) METS

METS is a metadata interoperability standard, a format for encoding metadata necessary for management and exchange of digital objects. It was developed by the Digital Library Federation and is maintained by the Library of Congress.[116] It is a wrapper for descriptive, administrative and structural metadata for textual and image based resources. It can be used to keep bit stream and metadata together.

A METS document can be used as a Submission Information Package, Archival Information Package or Dissemination Information Package within the Open Archival Information System.

The IMS Content Packaging specification is similar to METS and therefore it may be felt that there is no need to consider METS, but there may be some benefit in making a comparison between the principles of both standards, looking at the advantages of each. Within the IMS Content Packaging Specification learning objects are represented as the element 'resources'. This implies that the resource and the metadata are closely coupled. However, the coupling is only enforced in transport and when the content and metadata exist in repositories they can be managed independently. METS may be useful when considering the question of what information should be captured in the resource metadata and what information should reside in the content package. A preliminary crosswalk from METS to IMS Content Packaging has been proposed by Yee and Beaubien (2003). This sort of work may help to foster links between the e-learning community and the library and archiving community.

A METS document consists of seven section:
Header: metadata describing the METS object itself
Descriptive metadata: may consist of internal metadata and/or may point to the descriptive metadata provided by the JORUM Application Profile or other element sets such as Dublin Core
Administrative metadata: may consist of internal metadata and/or may point to the administrative metadata provided by the JORUM AP
File Section: lists all of the files that form part of the resource, which may be grouped by type
Structural Map: outlines the hierarchical structure of the object and links the elements of that structure to content files and metadata
Behaviour: used to associate executable behaviours with content in the METS object

METS could be deployed to wrap the metadata provided by the JORUM Application Profile, and additional metadata could be provided within the METS object. Related files within an object could be grouped together, so that if a learning object included thumbnails and PDF images, these could be put into separate groups.

METS also includes a Structural Links section, which is useful for recording the existence of hyperlinks between items.

The primary use of METS is to exchange metadata and digital resources. It is ideal for resources that have multiple parts and layers, and this will only apply to some of the JORUM learning objects. It is also useful as a container to encapsulate all metadata and data for an object for placement in a digital archive

[116] http://www.loc.gov/standards/mets/

Appendix (6) Key Organisations

This is a list of the key organisations relevant to this report. It is not intended as a comprehensive list of organisations working in the area of digital preservation.

UK organisations/bodies:

- JISC www.jisc.ac.uk
- Digital Preservation Coalition (DPC) http://www.dpconline.org
- Digital Curation Centre http://www.dcc.ac.uk/index.html
- British Library http://www.bl.uk/about/collectioncare/digpresintro.html
- Arts and Humanities Data Service (AHDS) http://ahds.ac.uk/
- DigiCult http://www.digicult.info/
- UKOLN http://www.ukoln.ac.uk/
- CETIS http://www.cetis.ac.uk/
- National Learning Network (NLN) http://www.nln.ac.uk/index.asp
- The National Archives (TNA) http://www.nationalarchives.gov.uk/preservation/
- e-Learning Research Centre – http://www.elrc.ac.uk/

International bodies:

- National Archives of Australia http://www.naa.gov.au/
- National Library of Australia – PADI (Preserving Access to Digital Information) http://www.nla.gov.au/padi/
- Erpanet http://www.erpanet.org/
- Research Libraries Group (RLG) http://www.rlg.org/en/page.php?Page_ID=552
- OCLC (Online Computer Library Centre) http://www.oclc.org/
- RLG: Research Libraries Group http://www.rlg.org/
- NEDLIB – Networked European Deposit Library http://www.kb.nl/coop/nedlib/
- NARA – US National Archives and Records Administration http://www.archives.gov/
- International Association of Sound and Audiovisual Archives (IASA) http://www.iasa-web.org/
- International Council on Archives http://www.ica.org/
- IEEE Learning Technology Standards Committee http://ieeeltsc.org/

19. Bibliography

The bibliography is grouped into broad subject areas.

All web sites were accessed between February and May 2005 and all URLs were checked in July 2005.

Overviews/General:

- Digital Preservation Coalition. (Sept 2004) 'Directory of Digital Preservation Repositories and Services in the UK'. Available at: http://www.dpconline.org/docs/guides/directory.pdf
- Jones M., Beagrie, N. (2003) 'Preservation Management of Digital Materials: A Handbook'. Available at: http://www.dpconline.org/graphics/handbook/
- Jones, M. Aug (2003) 'Digital Preservation Activities in the UK – Building the Infrastructure', *World Library and Information Congress: IFLA General Conference and Council* Available at: http://www.ifla.org/IV/ifla69/papers/129e-Jones.pdf
- Digital Preservation Coalition/PADI. 'What's New in Digital Preservation'. Available at: http://www.dpconline.org/graphics/whatsnew/
- Thomas, D. (2004) 'Digital Preservation at The National Archives'. Available at: http://www.nationalarchives.gov.uk/preservation/digitalarchive/pdf/dpattna.pdf
- Giaretta, D., Lyon, L., Robinson, B. [2004]. 'Curating for the Future – The work of the Digital Curation Centre'. Available at: http://www.allhands.org.uk/2004/proceedings/papers/167.pdf
- Lavoie, B.F. (OCLC). (April 2003) 'The Incentives to Preserve Digital Materials: Roles, Scenarios and Economic Decision-Making' Available at: http://www.oclc.org/research/projects/digipres/incentives-dp.pdf
- US National Science Board. (2005) 'Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century' National Science Board draft report. Available at: http://www.nsf.gov/nsb/meetings/2005/LLDDC_draftreport.pdf
- International Council on Archives. (April 2005), 'Electronic Records: A Workbook for Archivists' Available at: http://www.ica.org/biblio/Study16ENG_4.pdf
- Carpenter, L. (April 2005), 'Supporting Digital Preservation and Asset Management in Institutions' *Ariadne No.43.* Available at: http://www.ariadne.ac.uk/issue43/carpenter/
- Coutts, M. et al (Cedars). (April 2001), 'Cedars Summative Evaluation: final report'. Available from Available at: http://www.nla.gov.au/padi/metafiles/resources/3047.html
- Cedars. (March 2002), 'Cedars Guide to the Distributed Digital Archiving Prototype' Available at: http://www.leeds.ac.uk/Cedars/guideto/cdap/guidetocdap.pdf
- Russell, K., Sergeant, D. (June 1999), 'The Cedars Project: Implementing a Model for Distributed Digital Archives'. *RLG DigiNews*, June 15 1999, vol.3, no.3. Available at: http://www.rlg.org/legacy/preserv/diginews/diginews3-3.html
- Bellinger, M. (OCLC). (Oct 2002), 'Cost and Business Models for Digital Preservation'. PowerPoint presentation Available at: http://www.dpconline.org/text/events/elearning.html
- Royan, B. (Oct 2002), 'eLearning and the Business Case for Digital Libraries' Available at: http://www.dpconline.org/graphics/events/presentations/pdf/elearning1a.pdf
- MANDATE - Managing Digital Assets in Tertiary Education Available at: http://www.jwheatley.ac.uk/mandate/
- The Digital Archiving Consultancy/Digital Data Curation Task Force. (November 2002), 'Report of the Task Force Strategy Discussion Day' Available at: http://www.jisc.ac.uk/uploaded_documents/CurationTaskForceFinal1.pdf
- Beagrie, N. (April 2005), 'Digital Preservation: Best Practice and its Dissemination'. *Ariadne*, issue 43. Available at: http://www.ariadne.ac.uk/issue43/beagrie/
- International Council on Archives. (April 2005), 'Electronic Records: A Workbook for Archivists', Available at: http://www.ica.org/biblio.php?pdocid=285

Institutional Repositories:

- British Libraries/CURL DPC Forum. (Oct 2004), 'Digital Preservation in Institutional Repositories' Available at: http://www.dpconline.org/graphics/events/041019forum.html
- RLG/OCLC. (May 2002), 'Trusted Digital Repositories: Attributes and Responsibilities' Available at: http://www.rlg.org/longterm/repositories.pdf

- PREMIS Working Group. (Sept 2004) 'Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community' Available at: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf
- MacColl, J., University of Edinburgh. (Oct 2004), 'DSpace Institutional Repositories and Digital Preservation' Available at: http://www.dpconline.org/docs/events/041019maccoll.pdf
- Wheatley, P., University of Leeds. (March 2004), 'Institutional Repositories in the context of Digital Preservation' A Technology Watch Report for the Digital Preservation Coalition. Available at: http://www.dpconline.org/docs/DPCTWf4word.pdf
- Florida County Digital Archive. (July 2004), 'DAITSS Overview' Available at: http://www.fcla.edu/digitalArchive/pdfs/DAITSS.pdf

Policies, strategies, costs:

- Beagrie, N. (2002), 'A Continuing Access and Digital Preservation Strategy for the JISC 2002-2005' Available at: http://www.jisc.ac.uk/index.cfm?name=pres_continuing
- Barker, E., et al. (Nov 2004), 'Long Term Retention and Reuse of E-Learning Objects and Materials' (JISC commissioned report) Available at: http://www.jisc.ac.uk/index.cfm?name=project_elo
- Erpanet. (January 2003), 'Policies for Digital Preservation: Seminar Report' *ERPANET Briefing Paper, Training Seminar, Paris.* Available at: http://www.erpanet.org/events/2003/paris/ERPAtraining-Paris_BriefingPaper-e.pdf
- AHDS. (2001), 'A Strategic Policy Framework for Creating and Preserving Digital Collections' (v5.0, July 2001) Available at: http://ahds.ac.uk/strategic.pdf
- The National Library of Australia. (Feb 2002), 'Digital Preservation Policy' Available at: http://www.nla.gov.au/policy/digpres.html
- Oltmans, E, Kol, N. (April 2005), 'A Comparison between Migration and Emulation in Terms of Costs', *RLG DigiNews* 9(2) Available at: http://www.rlg.org/en/page.php?Page_ID=20571#article0
- Cedars. (April 2002), 'Cedars Guide to: Digital Preservation Strategies' Available at: http://www.leeds.ac.uk/Cedars/guideto/dpstrategies/dpstrategies.html
- PADI. (November 2003), 'Digital Archives: Who keeps them and who pays?: a response to the study commissioned by the UK's National Preservation Office in August 1997'. Available at: http://www.nla.gov.au/padi/response.html

Learning Objects and E-Learning projects:

- Valderrama, R.P. et al. (2005), 'Development of intelligent reusable learning objects for web-based education systems' *Expert Systems with Applications Vol.28 No.2.* Available at: http://www.informatik.uni-trier.de/~ley/db/journals/eswa/eswa28.html
- Currier, S., Campbell, L. (July 2004), 'Evaluating Learning Resources for Reusability: The "DNER & Learning Objects" Study'. Available at: http://www.unitec.ac.nz/ascilite/proceedings/papers/059.pdf
- JORUM. (2004), 'Scoping and Technical Appraisal Study'. Available at: http://www.jorum.ac.uk/research/archive/research/publications.html
- JORUM. (2005), 'Deposit Licence Agreement'. Available at: http://www.jorum.ac.uk/contributors/cregister/index.html
- Downes, S. (Nov 2002), 'Design Principles for a Distributed Learning Object Repository Network' http://education.qld.gov.au/learningplace/onlinelearning/courses/sdownesnov.html
- Beagrie, N. (April 2003), 'National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity' (commissioned for and sponsored by the National Digital Information Infrastructure and Preservation Program, Library of Congress). Available at: http://www.clir.org/pubs/reports/pub116/contents.html
- Koppi, T., Bogle, L. and Bogle, M. (Feb 2005), 'Learning objects, repositories, sharing and reusability' *Open Learning, vol 20, no.1.*
- Long, P. (2004), 'Learning Object Repositories, Digital Repositories, and the Reusable Life of Course Content' *Campus Technology Magazine.* Available at: http://www.campus-technology.com/article.asp?id=9258
- Barker E. et al. (June 2004), 'Long term Retention and Reuse of E-Learning Objects and Materials', (JISC commissioned report). Available at: http://www.jisc.ac.uk/index.cfm?name=project_elo

- IMS. (June 2004), 'IMS Question and Test Interoperability Framework' Available at: http://www.imsglobal.org/question/index.cfm
- Dolphin, I., Miller, P. (2002), 'Learning objects and the information environment', *Ariadne issue 32*. Available at: http://www.ariadne.ac.uk/issue32/iconex

Persistent Identifiers:

- Powell, A. (UKOLN), (April 2003), 'Identifiers for learning objects - a discussion paper'. Available at: http://www.ukoln.ac.uk/distributed-systems/lo-identifiers/
- Powell, A., Johnston, P., Campbell L., Barker, P. (April 2005), 'Dublin Core Metadata Initiative: Guidelines for using resource identifiers in Dublin Core metadata and IEEE LOM'. Available at: http://www.ukoln.ac.uk/metadata/dcmi-ieee/identifiers/
- The Stationary Office (TSO) and JISC. (May 2004), 'Digital Object Identifiers for Publishing and the e-Leaning Community'. Available at: http://dx.doi.org/10.1786/430153514841

Preservation Metadata:

- Day, M. (UKOLN). (Aug 2001), 'Metadata in a Nutshell'. Available at: http://www.ukoln.ac.uk/metadata/publications/nutshell/
- PREMIS Working Group. (Sept 2004), 'Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community'. Available at: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf
- PREMIS Working Group. (May 2005), 'Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group'. Available at: http://www.oclc.org/research/projects/pmwg/
- Cedars. (March 2002), 'Cedars Guide to: Preservation Metadata' Available at: http://www.leeds.ac.uk/Cedars/guideto/metadata
- RLG. (1993), 'Final Report of the Working Group on Preservation Issues of Metadata' Available at: http://www.rlg.org/preserv/presmeta.html
- Erpanet. (Sept 2003), 'Metadata in Digital Preservation' training seminar report. Available at: http://www.erpanet.org/events/2003/marburg/finalMarburg%20report.pdf
- Caplan, P. (Oct 2004), 'PREMIS – Preservation Metadata Implementation Strategies Update 1. Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community' *RLG DigiNews Vol.8 No.5.* Available at: http://www.rlg.org/en/page.php?Page_ID=20462#article2
- Guenther, R. (Dec 2004), 'PREMIS — Preservation Metadata Implementation Strategies Update 2: Core Elements for Metadata to Support Digital Preservation' *RLG DigiNews Vol.8 No.6*
- OCLC/RLG Working Group on Preservation Metadata. (Jan 2001), 'Preservation Metadata for Digital Objects: a Review of the State of the Art'. Available at: http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf
- OCLC/RLG Working Group on Preservation Metadata. (June 2002), 'Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects'. Available at: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf
- U.S. Government Printing Office. (June 2004), 'Report on the Meeting of Experts on Digital Preservation: Metadata Specifications'. Available at: http://www.gpoaccess.gov/about/reports/metadata.html
- National Library of Australia. (Oct 1999), 'Preservation Metadata for Digital Collections'. Available at: www.nla.gov.au/preserve/pmeta.html
- Manjula Patel (UKOLN). (July 2004), 'Scoping Study for a DCC Metadata Standards Registry' (draft). Available at: www.dcc.ac.uk/docs/Scoping_Study_for_a_DCC_Metadata_Standards_Registry.pdf
- ERPANET. (May 2003), 'Metadata in Digital Preservation' An Erpanet Briefing Document. Available at: http://www.erpanet.org/www/products/marburg/documents/erpaTraining-Marburg_BriefingPaper.pdf
- PREMIS Working Group. (May 2005), 'Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group'. Available at: http://www.oclc.org/research/projects/pmwg/premis-final.pdf
- Day, M (UKOLN). (Sept 2001), 'Metadata for digital preservation: a review of recent developments'. Available at: http://www.ukoln.ac.uk/metadata/presentations/ecdl2001-day/paper.html

- Day, M. (UKOLN). (Nov 2001), 'Metadata in support of subject gateway services and digital preservation'. Available at: http://www.ukoln.ac.uk/metadata/presentations/rome-2001/paper.html
- Heery, R., Patel, M. (2000), 'Application profiles: mixing and matching metadata schemas', *Ariadne issue 2*5. Available at: http://www.ariadne.ac.uk/issue25/app-profiles/

Metadata for E-Learning:

- IEEE Standard for Learning Object Metadata. Available at: http://www.ieee.org/
- 'UK Common Metadata Framework: Draft X4L Application Profile' (draft 0.2.2). Available at: http://www.cetis.ac.uk/profiles/uklomcore
- 'UK LOM Core v0.2' (May 2004). Available at: http://www.cetis.ac.uk/profiles/uklomcore
- Stevenson, A. (April 2005), 'JORUM Application Profile' (draft). Available at: http://www.jorum.ac.uk/docs/pdf/japv1p0.pdf

Formats:

- Stanescu, A. (OCLC). (Nov 2004), 'Assessing the Durability of Formats in a Digital Preservation Environment, The INFORM Methodology' *D-Lib magazine Nov 2004 vol.10 no.11.*
- The Representation and Rendering Project, University of Leeds. (2003), 'Survey and Assessment of Sources of Information on File Formats and Software Documentation'. Available at: http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf
- Adobe (White Paper). (2002), 'PDF as a Standard for Archiving'. Available at: http://www.adobe.com/products/acrobat/pdfs/pdfarchiving.pdf
- LeFurgy, W.G. (Dec 2003), 'PDF/A: Developing a File Format for Long term Preservation' *RLG DigiNews Vol.7 No.6*. Available at: http://www.rlg.org/legacy/preserv/diginews/diginews7-6.html#feature1
- PRONOM, The File Format Registry of The National Archives. Available at: http://www.nationalarchives.gov.uk/pronom/default.htm
- Brown, A. (The National Archives). (April 2005), 'Automating Preservation: new developments in the PRONOM service', *RLG DigiNews 15 April 2005*. Available at: http://www.rlg.org/en/page.php?Page_ID=20571#article1
- The Global Digital Format Registry. Available at: http://hul.harvard.edu/gdfr/
- Brown, A. (The National Archives). (June 2003), 'Digital Preservation Guidance Note: Selecting File Formats for Long term Preservation'. Available at: http://www.nationalarchives.gov.uk/preservation/advice/pdf/selecting_file_formats.pdf
- Brown, A. (The National Archives). (June 2003), 'Digital Preservation Guidance Note: Selecting Storage Media for Long term Preservation'. Available at: http://www.nationalarchives.gov.uk/preservation/advice/pdf/selecting_storage_media.pdf
- Abrams, S.L. (Harvard University), Seaman, D. (Digital Library Federation). (August 2003), 'Towards a Global Digital Format Registry'. Paper given at the World Library and INformation Congress, Berlin, August 2003. Available at: http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf
- Erpanet. (May 2004) 'File Formats for Preservation:final report'. Available at: http://www.erpanet.org/events/2004/vienna/Vienna_Report.pdf
- Erpanet. (October 2002), Urbino Workshop report 'XML for Digital Preservation: Final Report'. Available at: http://www.erpanet.org/events/2002/urbino/Urbino%20Workshop%20Report.pdf
- Erpanet. (April 2003), 'Long-term Preservation of Databases' workshop report. Available at: http://www.erpanet.org/events/2003/bern/Bern_Report_final.pdf
- Abrams, S.L. (Harvard University). (2004), 'The role of format in digital preservation', *Vine, vol.34, no.2.* [available via Emerald Fulltext]
- Gladney, H.M. (2005), 'Preserving Dynamic Digital Records', *Digital Document Quarterly vol.4, no.2*. Available at: http://home.pacbell.net/hgladney/dynamic.pdf

Storage

- Linden, J. et al (British Library). (February 2005), 'Technology Watch Report: the large-scale archival storage of digital objects' (DPC Technology Watch Series report). Available at: http://www.dpconline.org/docs/dpctw04-03.pdf

Audio-visual archives:

- International Association of Sound and Audiovisual Archives. (Aug 2004), 'Guidelines on the Production and Preservation of Digital Audio Objects'. Available at: http://www.iasa-web.org/tc04/index.htm
- Edmondson, R. (April 2004), 'Audiovisual Archiving: Philosophy and Principles' (UNESCO publication). Available at: http://unesdoc.unesco.org/images/0013/001364/136477e.pdf
- Library of Congress Digital Audio-Visual Preservation Prototyping Project. Available at: http://www.loc.gov/rr/mopic/avprot/avprhome.html

Frameworks and the use of frameworks:

- CCSDS, (Jan 2002) 'Reference Model for an Open Archival Information System. Available at: http://www.ccsds.org/documents/650x0b1.pdf
- Lavoie, B.F. (OCLC). (Jan 2004), 'The Open Archival Information System Reference Model: Introductory Guide' (A Digital Preservation Coalition Technology Watch Report). Available at: http://www.dpconline.org/docs/lavoie_OAIS.pdf
- Holdsworth, D., Sergeant, D.M. (University of Leeds). (1999), 'A blueprint for Representation Information in the OAIS model'.Available at: http://www.personal.leeds.ac.uk/~ecldh/Cedars/ieee00.html
- Metadata Encoding and Transmission website, 'METS: An overview and tutorial'. Available at: http://www.loc.gov/standards/mets/METSOverview.v2.html
- Yee, R., Beaubien,R. (Nov 2003), 'A preliminary crosswalk from METS to IMS content packaging. Available at: http://raymondyee.net/files/2003/09/Library.Hi.Tech.latest.pdf
- METS: An Overview and Tutorial. Available at: http://www.loc.gov/standards/mets/METSOverview.v2.html

Authenticity

- Digicult. (August 2002), 'Integrity and Authenticity of Digital Cultural Heritage Objects'. *Thematic issue 1*. Available at: http://www.digicult.info/downloads/html/7/7.html

International:

- National Archives of Australia, (May 1999), 'Recordkeeping Metadata Standard for Commonwealth Agencies'. Available at: http://www.naa.gov.au/recordkeeping/control/rkms/rkms_pt1_2.pdf
- Heslop, H. et al (National Archives of Australia). (2002), 'An Approach to the Preservation of Digital Records' (Green Paper). Available at: http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf
- National Archives of Australia. (May 2004), 'Digital Recordkeeping: Guidelines for Creating, Managing and Preserving Digital Records' (exposure draft) . Available at: http://www.naa.gov.au/recordkeeping/er/guidelines/DigitalRecordkeeping.pdf
- Alemneh, D.G. et al. (Oct 2004), 'A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience, *First Monday Journal*. Available at: http://www.firstmonday.org/issues/issue7_8/alemneh
- Victorian Electronic Records Strategy Standards (Australia) . Available at: http://www.prov.vic.gov.au/vers/standard/
- US National Archives and Records Administration, Electronic Records Archive. (Dec 2003), 'Introduction to Preservation and Access Levels Concepts'. Available at: http://www.archives.gov/era/pdf/preservation-and-access-levels.pdf

Related areas:

- James. H. et al. (Oct 2003), 'Feasibility and Requirements Study on Preservation of E-Prints' (JISC commissioned report). Available at: http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf
- Day, M. (Feb 2003), 'Collecting and Preserving the World Wide Web' (UKOLN feasibility study for JISC and the Wellcome Trust). Available at: http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf
- Fitzgibbon, A., Reiter, E. (May 2003), 'Memories for Life: Managing Information over a Human Lifetime'. Available at: http://www.nesc.ac.uk/esi/events/Grand_Challenges/proposals/Memories.pdf