Jorum Preservation Watch Report

Synopsis

This is the second Preservation Watch report produced as part of the Jorum Service-in-Development work (the first report was produced under Jorum R&D August 2004 - July 2005). It builds upon the first report, providing a summary of some of the key initiatives that are currently taking place within digital preservation and focussing on a number of key issues that need to be addressed prior to a decision being taken about the implementation of a preservation process for learning objects within Jorum. It sets out in detail the arguments for preservation and discusses some of the issues surrounding costs and the measurement of the value of learning objects. It argues that roles must be clearly defined and this requires clarification of responsibilities for learning objects and for the preservation of learning objects. It sets out the criteria that may be used in order to appraise and select objects for retention, acknowledging the vital importance of user expectations and requirements as well as the value of uniqueness, the importance of the concept of trustworthiness and the difficulties of dealing with a wide range of formats and dealing with complex aggregate learning objects. Metadata is seen as key to the preservation process, and preservation metadata elements need to be carefully considered.

The report provides recommendations for further work, breaking this down into distinct sections relating to functional and informational models for learning objects, archival responsibility, appraisal, metadata and formats. The recommendations also include the creation of case studies to illustrate methods and approaches and the continuation of a preservation watch to report on significant developments within the digital preservation community.

Author
Jane Stevenson

Date
24 July 2006

Version
Final

Status
Final

Authorised
Signed off by JISC July 2006

Contents

1. Executive Summary

There are currently a whole host of initiatives underway within digital preservation and this Preservation Watch report outlines some of those that may be of relevance for the preservation of learning objects. The Research Libraries Group online newsletter, DigiNews, has compiled a list of ten promising initiatives[1] which are a reflection of the range of activities that is happening in the digital world, many of which have broad potential. JISC are funding a substantial amount of work in the area of digital preservation within institutions[2], but several of these projects were not complete at the time of writing this report, so it will be useful to continue a preservation watch to consider their final outcomes and conclusions.

The arguments for the preservation of learning objects centre on the concept of objects as important knowledge assets. It may be that providing long-term access to objects is one reason for creators to deposit them in a repository, and there does seem to be some evidence to suggest that creators are looking to ensure that their materials are housed in a safe environment. A survey recently conducted by the JISC, Rights and Rewards in Blended Institutional Repositories, demonstrates that preservation is one of the main reasons why participants contribute teaching materials to an institutional repository. When asked which reasons would encourage participants to contribute to a repository, seventy percent said that if repositories helped them to 'manage and preserve resources' they would be 'much more likely' or 'likely' to deposit.[3] There is an investment of time and resources in making a quality, widely used learning object, and if it is only accessible for a short time it may not be considered as worthwhile or indeed as prestigious to the creator as one that lasts, and is used, for a number of years. The requirements of the user community are key to the decisions about what to preserve and how long to preserve it for. As e-learning becomes embedded into the curriculum, there needs to be more research into user requirements over the short and longer term and whether users will benefit from continued access to materials over time.

It is important to secure the maximum benefits from the investment of time, resources and skills made in the creation of materials. The potential for reuse of learning objects provides a means to maximise the return for the original investment in creation, and preservation will support the principle of materials being repurposed for different contexts over time. Reuse may be encouraged by the creation of smaller more flexible chunks of content along with comprehensive metadata that facilitates discovery and provides comprehensive information about the materials.

As intangible assets, it can be difficult to measure the value of learning objects and establish the costs of maintaining them. The decision to fund the long term preservation of learning objects is effectively an investment decision based on the probability of

[1] RLG DigiNews Staff (2005). Watch This Space: Ten Promising Digital Preservation Initiatives. RLG DigiNews 9(4), 15 August 2005. http://www.rlg.org/en/page.php?Page_ID=20744#article1 [May 2006]
[2] http://www.jisc.ac.uk/index.cfm?name=programme_404 [May 2006]
[3] http://rightsandrewards.lboro.ac.uk/files/resourcesmodule/@random43cbae8b0d0ad/1137423150_SurveyReport.pdf [June 2006]

producing benefits over time. There are many factors within preservation that are difficult to predict, which makes costing a complex issue, but if we want to preserve in order to continue to enable access to materials for teaching and learning then we need to find ways of costing and of measuring the value of learning objects. The level of automation that can be achieved is a key factor in the calculation of cost and it is likely that the costs of producing learning objects will fall as the tools to create them are developed and improved and as methods of preservation progress.

The JISC-funded Espida project[4] is looking at how we can measure the value of digital materials over time, and is creating a sustainable business-focused model for digital preservation. ArXiv.org,[5] which provides open access to eprints, has worked out costs at a highly detailed level, which helps to clarify the exact costing implications. The National Archives[6] have similarly carried out detailed analysis of costs and have found that the highest costs are associated with ingest and preservation planning. Efficiencies can be achieved with a collaborative research and development approach on preservation planning. Automation is widely seen as the key to sustaining preservation activities. The LOCKSS Program[7] (Lots of Copies Keep Stuff Safe) has been very successful and provides a strong case for the pursuit of interoperability and for preservation to be carried out as a collaborative, community effort.

There is a need to clarify the roles of information providers, owners, users, funders and repository service providers. In terms of responsibility for preservation, the situation within the Jorum repository is complicated because the ownership of objects is not straightforward and there are issues surrounding the rights of individuals and institutions as well as the existence of both stored objects in Jorum and virtual objects referenced by Jorum. Identification of the responsible institution for a virtual object might be possible via automated metadata generation. However, this would problematic for aggregate objects that include virtual objects.

The SHERPA project is using the OAIS reference model (the Open Archival Information System) to develop a preservation environment,[8] assigning rights and responsibilities, establishing protocols and workflow processes. Whilst there is a strong case for a national repository taking on a national role to provide an archival function, there is also a strong case for a collaborative effort that involves contributing institutions as well as other bodies with expertise and experience in this area.

An analysis of the nature of learning objects may help to clarify responsibilities as well as the preservation approaches that can be taken. The concept of an information object as the basic unit for preservation is worth further exploration. This refers to the data object and information required to render the data object and make it meaningful. It excludes the

[4] http://www.gla.ac.uk/espida/ [May 2006]
[5] http://www.arxiv.org/ [May 2006]
[6] http://www.nationalarchives.gov.uk/preservation/digital.htm [May 2006]
[7] http://www.lockss.org/lockss/Home [May 2006]
[8] Knight, G. (2005) SHERPA-DP OAIS Report: An OAIS Compliant Model for Disaggregated Services. http://ahds.ac.uk/about/projects/sherpa-dp/sherpa-dp-oais-report.pdf [May 2006]

additional metadata that is provided and the structure that is provided for an aggregate learning object, so it effectively strips the learning object down to the basics, which may be a useful approach when considering preservation. It would be worth exploring this further, looking at the relationship between the learning object and the information object and the implications for preservation.

A rigorous process of appraisal should result in the selection of high quality, pedagogically sound learning objects, which should help to ensure that the investment in infrastructure pays off in the educational experience of the users. The Digital Preservation Coalition (DPC) have created a Decision Tree[9] that may help in the creation of an appraisal process. Factors to consider when creating an appraisal policy include the quality of objects, the level of use, the metadata that is provided, ease of migration and whether there are dependencies, which may be technical, external or internal (where an object forms part of a set of objects). The Digital Asset Assessment Tool[10] that has been created under a JISC-funded project may help prioritise what to preserve, as it is intended to be used to assess assets for preservation risk and produce assessment criteria for storage media and formats. The process of appraisal is made more complex by the existence of aggregate learning objects as well as complex information objects. Smaller chunks of content will be easier to appraise and easier to preserve and indeed arguably more likely to be reused, but aggregate objects will be difficult to handle. Historical importance is clearly a strong argument for some level of preservation, and even if objects are not preserved in order to maintain them as usable teaching materials, a sample of learning objects should be kept for historical purposes. In addition, the value of most assets will increase if they are seen as unique, and this is a concept that requires further consideration as far as learning objects are concerned.

There has been a great deal written about the concept of a trusted digital repository but it may be that the concept of trust for a learning object repository is not the same as the trust required for repositories of research materials. Repositories may be expected to be trustworthy, regardless of the type of material that they hold, in terms of financial sustainability and human aspects of competency, but the significant properties of learning objects are distinct from other types of materials. The concept of accuracy of a learning object is problematic because this is partly subjective. What may be more important is that the metadata provides an accurate description of the object. In terms of the long-term perspective, digital preservation as a process should be seen to be trustworthy and this will be facilitated by a clear preservation policy.

Metadata is often considered to be the key to successful preservation and a Data Dictionary for Preservation Metadata[11] has been created by the PREMIS working group (Preservation Metadata: Implementation Strategies) which identifies core digital preservation metadata elements and includes practical examples. It would be useful to look at these core elements alongside the current Jorum metadata, as outlined in the

[9] http://www.dpconline.org/graphics/handbook/dec-tree-select.html [May 2006]
[10] http://www.ulcc.ac.uk/daat [May 2006]
[11] http://www.oclc.org/research/projects/pmwg/premis-final.pdf [May 2006]

application profile, in order to make some conclusions about the preservation metadata required for learning objects and the resources that this might require.

It may be that a practical approach to preservation will necessitate selecting a limited number of the most common formats to preserve and carrying out a periodic review of formats. The costs and risks associated with digital preservation tend to grow when a digital collection includes a large number of diverse file formats, as was found in a study commissioned by JISC.[12] Criteria can be developed for the treatment of each format, in terms of a migration strategy. It may be useful to identify file formats at particular risk in terms of access and integrate this into the appraisal process.

It is important to continue to monitor developments within digital preservation and gather more evidence of user requirements, patterns of use of learning objects and how the community perceptions of Jorum develop. In the end, whilst the arguments for preservation may be persuasive, the choices that are made now may not be proved to be valuable or otherwise for some time. But if the choices are made on the basis of empirical evidence, best practice and responding to user requirements, then they can be justified. Sound policies should set out roles and responsibilities and ensure that appraisal decisions and preservation processes are shown to be trustworthy and transparent.

[12] James, H, Ruusalepp, R, Anderson, S, Pinfield, S (2003), Feasibility and Requirements Study on Preservation of E-Print. http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf [July 2006]

2. Digital Preservation Recommendations

The recommendations set out here have been identified as key research areas to be progressed.

2.1 Preservation Watch

2.1.1 It is important to continue to monitor developments within digital preservation, as it clearly makes sense to benefit from the approaches taken by others and seek to adopt tried and tested solutions that enable preservation and long term access to be efficient and effective.

There are currently a number of projects running that may have relevance for a digital preservation solution for Jorum, in particular the espida and LIFE projects. Espida (http://www.gla.ac.uk/espida) is developing a sustainable business-focused model for digital preservation. The LIFE project (http://www.ucl.ac.uk/ls/lifeproject/index.shtml) aims to look at the life-cycle of the collection and preservation of digital materials and deliver information about the management of material throughout the life-cycle. It is about investigating methods and developing models to attribute costs of digital preservation to identified stages within the lifecycle of digital collections.
There are several recent developments and more established initiatives of interest. Some of these are continuing to evolve and develop and any evaluation work or test-bed projects carried out that refer to such initiatives should be monitored. Of particular interest are the DPC Decision Tree, the Digital Archive Assessment Tool, the DCC Digital Curation Manual and Pronom, the TNA's file format registry.

It will be useful to continue to monitor use and evaluation of the OAIS Information Model as part of the watch, and monitor the development of the DCC's Representation Information registry.

The preservation watch should continue to monitor developments in preservation processes in order to work towards recommendations for processes that should be used for learning objects. It is likely that a number of options will become feasible for institutional repositories, as more work is carried out on preservation solutions.

2.2 OAIS Reference Model

The OAIS Reference Model essentially provides a functional model and an information model. It addresses the full range of preservation functions but it is not implementation specific.

A functional model for learning objects could set out the full range of preservation functions that would be involved in maintaining an archival repository for learning objects, and also how this might relate to the functions within the current Jorum

repository. This will help to clarify the architecture and practical operation of an archive and also the information flows between the functional entities. The OAIS functional model is a suitable starting point, but it may be that it requires modification in order to be suitable for learning objects. The SHERPA-DP project found that modifications needed to be made to the model in line with the practical operation of eprint services.[13]

A functional model would help in clarifying exactly what is involved in setting up an archive, facilitating further discussions on responsibilities, organisational issues and costs.

The OAIS information model provides a means to describe the types of information that are exchanged and managed within an OAIS. It would be worth exploring the OAIS concept of an Information Object and how this might relate to the concept of a learning object and the content packaging specifications. The conceptual separation of information objects from learning objects may provide a useful way of thinking about the preservation process. This approach essentially means separating the actual data object and the information required to render it and make it meaningful from the metadata which forms part of the content package and relates to the pedagogy and organisation of the resource. The preservation of the metadata is a relatively straightforward task, and a common solution for this can be implemented, but the preservation of the actual information objects is a far more difficult task, with a wide range of formats and different levels of complexity and interaction.

2.2.1 Produce a functional model for learning objects

2.2.2 Consider the OAIS Information Model and how this relates to learning objects, in particular the separation of the component parts of a learning object.

2.3 Responsibilities for Learning Objects

In terms of clarifying the responsibility for individual objects, this is not straightforward because the objects are owned by contributing institutions and also because Jorum contains deposited objects and also provides access to virtual objects. In addition to this, learning objects can be aggregates of information objects and learning objects use content packaging to wrap up the information objects with the necessary metadata that provides pedagogical information and information about the organisation of resources. The particular nature of packaged learning objects therefore means that identifying responsibilities for preservation and long term access may be difficult.

Archival responsibility needs to be identified for each object. Jorum cannot take on responsibity for materials that are not deposited in the Jorum repository (i.e. they are described within Jorum but held elsewhere) and it would not make sense to preserve objects that are deposited in Jorum but already preserved elsewhere. A metadata field that provides this information would be a very useful addition to Jorum learning objects,

[13] http://ahds.ac.uk/about/projects/sherpa-dp/sherpa-dp-oais-report.pdf [May 2006]

possibly as an extension to the Jorum application profile. It would be useful to automatically identify virtual objects that are referenced by Jorum but are not actually stored in the Jorum repository as the responsibility of the institution that holds the actual object. At minimum it is important to establish that the preservation of virtual objects is not the responsibility of Jorum.

2.3.1 Set out the archival responsibility for learning objects and for the preservation processes, encompassing virtual objects stored elsewhere and objects deposited in JORUM

2.3.2 Consider the addition of an 'archival responsibility' metadata field and whether this could be automatically populated for virtual objects

2.4 Appraisal

Unless all learning objects are to be preserved, a method of appraisal is necessary in order to make decisions about which resources are going to be kept long term. A small number of objects may be kept for historic purposes and a larger number may be kept to provide continuing access to them for learning and teaching.

Part of the appraisal process should take into account depositors' and users' expectations about the resources that are available to them over time, and this requires further investigation into user requirements.

2.4.1 Monitor the use of Jorum over time in order to learn more about depositor and user expectations and patterns of use.

2.4.2 Investigate the 'Decision Tree' (http://www.dpconline.org/graphics/handbook/dec-tree-select.html)created by the Digital Preservation Coalition as a means to help inform an appraisal process for learning objects, possibly making modifications to it that would be appropriate for an appraisal process for learning objects.

2.4.3 Investigate the Digital Asset Assessment Tool (http://ahds.ac.uk/about/projects/daat/) to assess its relevance for learning objects.

2.4.4 Analyse the feedback users provide on the quality of materials via the comments section within the system in terms of how useful it might be for future appraisal.

2.4.5 Undertake a survey of a small number of users to find out more about their expectations and attitudes towards long-term access and the sorts of objects that they might continue to find valuable over time. It would be preferable if such a survey was carried out after people had been using Jorum for some time, so they were familiar with the service.

2.5 Metadata

Metadata is generally seen as the key to the successful management of digital records. A number of metadata elements required for preservation will already be included in the administrative, descriptive and technical metadata provided for Jorum learning objects, but additional metadata fields will inevitably be necessary. A decision on the metadata to use will need to consider how necessary or useful it is, whether 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, which should be clarified in the Jorum preservation strategy.

Whilst there are a large number of preservation metadata sets available, the PREMIS Data Dictionary (http://www.oclc.org/research/projects/pmwg/premis-final.pdf) is arguably the most definitive publication, produced by a team of 30 experts from five countries and likely to become an international open-source standard. This therefore provides a useful source for looking at preservation metadata. It does have limitations, for example, it does not focus on descriptive metadata, it does not directly consider rights and permissions that are not directly associated with preservation actions and it does not deal with technical metadata for all types of digital file. But it does deal with metadata that is required for maintaining the viability, renderability, understandability and authenticity in a preservation context.

2.5.1 Investigate additional metadata requirements for the preservation of learning objects. Use the PREMIS data dictionary as a basis for comparison with the preservation metadata that currently exists in Jorum and report on the recommended PREMIS metadata elements and their relevance for learning objects as well as issues relating to resource requirements for the creation of the metadata and a risk analysis on the omission of additional metadata fields.

2.5.2. Explore the issues involved in metadata attached to aggregate objects, which may not include metadata at the individual information object level. Consider whether information objects within aggregate objects require the same metadata, whether certain aggregate level metadata can be applied to individual objects, and whether this can be automated.

2.5.3 Investigate the automation of preservation metadata. Look at findings of the Metadata Generation Research project being carried out by The Metadata Research Center (http://ils.unc.edu/mrc/mgr.html)

2.5.4 Jorum already contains metadata that will be useful for preservation, but it would be useful to carry out an analysis of the consistency and completeness of the metadata as this will impact upon preservation requirements.

2.6 Formats

For preservation purposes, it is essential to have detailed information about the format of a resource. Whilst the Jorum metadata does usually include format information, it is often limited to identification of the name of the access software, and it may not list all of the formats that exist within an object. It may not include the software version, and more fundamentally, the format may simply be described as a zip file, as this is the format for a content object. (NB: Jorum requirements now include list of formats within zip file).

2.6.1 Preserv, the preservation Eprints service, will link Eprints to PRONOM software (http://www.nationalarchives.gov.uk/pronom/), which is maintained by The National Archives, for identification and verification of file formats. The outcomes of the Preserv project should be monitored to ascertain whether this would be potentially useful for learning objects.

2.6.2 As preservation processes are expensive, an archive may decide to limit the number of file formats that it accepts. It would be worthwhile looking at how various archives have approached this and identifying a minimum set of preservation formats that are necessary to manage the range of data types within learning objects. The UK Data Archive has a preservation policy document for converting to stable formats which would be worth a closer look.

2.7 Case Studies

2.7.1 Take a sample of learning objects from within Jorum, selecting objects having different levels of organisational complexity and containing a range of file formats. The exploration of each of the recommendations given above should then include reference to these specific learning objects within the case study. For example, setting out the metadata that should be added to each learning object to enable them to be preserved; recording the formats that are given for each learning object and whether this is adequate for preservation; looking at how the appraisal process might be used for each of the example objects.

3. Introduction

The first report on digital preservation[14] was intended to set out the current digital preservation landscape and raise many of the issues that will need to be addressed if the decision is made to implement a preservation process for learning objects in Jorum. It was a broad report, encompassing work being undertaken in many different domains.

[14] Stevenson, J. (2005) Preservation Watch Report.
http://www.jorum.ac.uk/docs/pdf/Digital_Preservation_Report.pdf [May 2006]

This second report (part of the Jorum Service-in-Development Phase 1 Research and Development) builds upon the initial report. It covers some of the initiatives and project developments that have been happening in digital preservation since the publication of the first report and refers to some of the key issues that need to be addressed for the preservation of learning objects. It is not intended to be a concluding report, as more work remains to be done, but it does contain recommendations for further more focussed work.

The preservation process itself still raises challenging questions and issues, although great strides have been made in seeking and implementing solutions. At the Workshop on Research Challenges in Digital Archiving and Long-Term Preservation organised by a number of U.S. organisations in April 2002, it was acknowledged that

'No acceptable methods exit today to preserve complex digital objects that contain combinations of text, data, images, audio and video and that require specific software applications for reuse.'[15]

The report went on to say that 'effective and affordable digital preservation methods for multi-media do not exist today'. However, developments over the last few years have addressed this challenge. A survey of organisations carried out by Cornell in 2005[16], looked at their digital preservation efforts in terms of scope, priorities, resources, and overall readiness. The survey indicated that many institutions have digital preservation policies, there is a growing awareness among senior managers, more resources are being committed and the technical infrastructure is present as well as direct practical experience with some preservation activities. Over the last few years there does seem to have been a shift in focus from technology concerns to organisational ones, reflecting the fact that significant technical developments have been made, but organisational challenges remain.

The e-learning community is having much discussion and debate about the use and reuse of learning objects, about cultural and social barriers to reuse, and about collaboration between institutions, but long-term access to learning materials has not been a priority. The challenge of archiving complex digital resources, packaged up into learning objects and with a pedagogical context, is clearly a substantial one. The preservation of learning materials is likely to raise particular issues, but where there is a commonality of interests with the digital preservation community in general it is important to take advantage of this. It is clear that Jorum will need to collaborate with others and benefit from solutions being worked out and implemented at a national and international level. The preservation of digital materials and associated metadata is common to many digital preservation initiatives, but there are particular issues surrounding learning materials, such as the granular nature of learning objects, the pedagogical context and the differences

[15] It's About Time: Research Challenges in Digital Archiving and Long-term Preservation. August 2003. Final report of the Workshop on Research Challenges in Digital Archiving and Long-Term Preservation, April 2002. http://www.si.umich.edu/digarch/NSF%200915031.pdf [March 2006]

[16] Kenney, A.R. and Buckley, E. (2005) Developing Digital Preservation Programs: the Cornell Survey of Institutional Readiness, 2003-2005 RLG DigiNews August 15 2005. http://www.rlg.org/en/page.php?Page_ID=20744#article0 [January 2006]

surrounding the preservation of teaching materials as opposed to research materials, which may influence approaches to intellectual property rights, authenticity and appraisal.

The Digital Preservation Coalition report on preservation needs in the UK[17] identifies 18 core needs and gives recommendations. Among the key needs are that the awareness of digital preservation issues needs to be more commonplace, particularly amongst data creators, and projects need to be funded from the outset with the long-term value of the information produced and the cost of retention taken into account.

Repositories generally have a high profile within the learning and educational communities at the present time. Campbell et al, in their report on Repository Management and Implementation state that 'there is an implicit assumption that all these assets require management and preservation.' They go on to say that more consideration needs to be given to 'the degree of management and preservation that different types and manifestations of resources require'.[18]

4. Current Digital Preservation Initiatives and latest developments

This section is intended to summarise some of the key initiatives that are currently underway within digital preservation and which may be of relevance for the preservation of learning objects.

4.1 RLG Feature Ten Promising Digital Preservation Initiatives

RLG (Research Libraries Group) DigiNews have reviewed digital preservation and digitisation research and development projects around the world and compiled a list of 10 initiatives that they believe to be particularly significant.[19] This report cannot attempt to cover all of the outcomes of these projects, but it is worth providing a brief summary of each of them, as they represent some of the latest research developments within digital preservation, indicating how far the digital preservation community has come in terms of research and solutions.

4.1.1 The ECHO DEPository Project, 2004-2007

The National Digital Information Infrastructure and Preservation Program (NDIIPP) of the Library of Congress has developed eight cooperative digital preservation partnerships.

[17] Digital Preservation Coalition (2006). Mind the gap: assessing digital preservation needs in the UK. http://www.dpconline.org/docs/reports/uknamindthegap.pdf [Feb 2006]
[18] Campbell L.M. et al, (2004) Repository Management and Implementation, a white paper for alt-i-lab 2004. http://www.jisc.ac.uk/uploaded_documents/Altilab04-repositories.pdf [May 2006]
[19] RLG DigiNews Staff (2005). Watch This Space: Ten Promising Digital Preservation Initiatives. RLG DigiNews 9(4), 15 August 2005. http://www.rlg.org/en/page.php?Page_ID=20744#article1 [January 2006]

These include The ECHO DEPository Project, 2004-2007 [20] which is working on Web archiving methodology and practice. This may not be directly relevant to Jorum, but many of the learning objects in Jorum link to web pages, and therefore there are issues surrounding the persistence of these web pages. Other useful web archiving work is going on within the Internet Archive and UK Web Archiving Consortium as well as the the International Internet Preservation Consortium (IIPC), which is an important international collaboration in preserving internet content. It comprises the National libraries of France, Australia, Canada, Denmark, Finland, Iceland, Italy, Norway, Sweden, United Kingdom, The Library of Congress (USA) and the Internet Archive.[21]

4.1.2 Incentives for Data Producers to Create Archive-Ready Data Sets, 2005-2006

A project funded by the Digital Archiving and Long-Term Preservation (DIGARCH) program, sponsored by the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP) and National Science Foundation (NSF). This project[22] is investigating ways to increase cooperation between producers and archives. It will be looking at incentive mechanisms for producers to deposit archive-ready data sets.

In terms of learning objects, it may be that incentives to deposit in Jorum and incentives to deposit in a learning object archive need to be further considered in terms of the relationship between the two. If learning objects are considered for preservation once they are deposited in Jorum, then the incentives for deposit would be both for current access and for future access and preservation. If an archive were to operate as a separate entity, then incentives to deposit in an archive would be distinct.

4.1.3 Metadata Generation Research project

The Metadata Research Center, sponsored by the University of North Carolina Chapel Hill, is working on a Metadata Generation Research project,[23] developing a model to facilitate the most efficient and effective means of metadata production by integrating human and automatic processes. The possibility of reducing metadata generation costs is attractive for a service such as Jorum, where metadata generation is time consuming and costly. It will be useful to monitor progress in the automatic generation of metadata, and whether any of the solutions could be applied to Jorum.

4.1.4 Digital Academic Repositories (DARE), 2003-2006

SURF, the higher education and research partnership organisation for network services and information and communications technology (ICT) in the Netherlands, have funded the DARE[24] initiative, which will develop a shared infrastructure to manage academic information and to make research results more accessible.

[20] http://www.ndiipp.uiuc.edu/ [May 2006]
[21] http://netpreserve.org [July 2006]
[22] www.si.umich.edu/research/project.htm?ResearchID=73 [April 2006]
[23] http://ils.unc.edu/mrc/mgr.html [May 2006]
[24] http://www.surf.nl/en/themas/index2.php?oid=7 [April 2006]

4.1.5 Australian Partnership for Sustainable Repositories (APSR), 2004-2006

The APSR project[25] is a wide-ranging project covering: the integration of preservation metadata into open source repositories; a GIS metadata capture tool; guidelines, such as for trusted digital repositories in higher education; sustainable models for common file formats; risk analysis including a risk notification tool implementation; 'Fedora in a Box'; and documentation focusing on core issues and standards.

The initial discussion paper[26] covers many of the issues that the project is addressing. It recognises that repositories are products of the technologies of their time and are at risk themselves, it is therefore essential to consider the issue of sustainability for a repository, as well as for the data objects. It considers the preservation of the byte-stream, which may not be as straightforward as is often proposed, as well as the sustainability of access to meaning, and also makes the point that 'preservation metadata is the key to the survival of the content'. The importance of identification of the file formats is considered, as technical decisions for long term preservation are generally made on a format-specific basis. Also, the value of the data over time is a particular aspect of the economic value, and is integral to sustainable planning.

4.1.6 Digital Curation Centre

The Digital Curation Centre in the UK is playing a key role in the development of tools and research deliverables, including the creation of a digital curation manual. Currently (April 2006) the two sections that have been completed are on open source software and metadata. The metadata section[27] does not deal in detail with preservation metadata, as it is intended to provide an overview and introduction to metadata in a general sense. It refers to the importance of preservation metadata to allow the 're-creation and interpretation of the structure and content of digital data over time' which means that it needs to support a wide range of different functions including descriptive, structural and administrative functions. It provides a summary description of preservation metadata initiatives and an overview of the OAIS information model and the archival information package, which is effectively at the heart of the OAIS conceptual approach to preservation.[28]

4.1.7 Auditing and Certification of Digital Archives, 2005-2006

The Andrew W Mellon Foundation has awarded a grant to the Center for Research Libraries for digital archives certification for a project to develop processes and activities for certification and audit of digital archives.[29]

---

[25] www.apsr.edu.au/ [March 2006]
[26] Bradley, K (2005) APSR Sustainability Issues Discussion Paper. http://www.apsr.edu.au/documents/APSR_Sustainability_Issues_Paper.pdf [May 2006]
[27] Day, M (Nov 2005). DCC Digital Curation Manual: Installment on Metadata. http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata/metadata.pdf [April 2006]
[28] *See* CCSDS Reference Model for an Open Archival Information System (2002). http://public.ccsds.org/publications/archive/650x0b1.pdf
[29] www.crl.edu/content.asp?l1=13&l2=58&l3=142 [April 2006]

4.1.8 The Digital Preservation Cluster (2004-2006)

DELOS (Network of Excellence on Digital Libraries) Digital Preservation Cluster (WP6) has a work plan for digital preservation. The Digital Preservation Cluster[30] is looking at developing a methodological framework and theory for digital preservation. Deliverables for the first phase include a study of metrics for testing and validating digital preservation strategies and frameworks, a survey of digital repository systems and storage models, a report on file format registries and their relationship to preservation strategies, and a repository functionality analysis.

4.1.9 The Virtual Archives Laboratory

The National Archives and Records Administration (NARA) has an active research programme. The Virtual Archives Laboratory[31] is a digital preservation research test bed. It focuses on projects pertaining to scalability, persistence, and authenticity. This research will feed into the development of the Electronic Records Archive that NARA is building.

4.1.10 European Commission Digital Preservation projects

The European Commission has funded Digital preservation projects under the 5th call of the Sixth Framework Programme. Projects led by UK partners Include Permanent Long-term Access through Networked Services (PLANETS) Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR) and CCLRC Digital Preservation Europe (DPE).[32]

4.2 Supporting Digital Preservation and Asset Management in Institutions

JISC are funding a substantial amount of work in the area of digital preservation within institutions, and some of the projects that are running under this title may be of particular relevance.

4.2.1 Digital Asset Assessment Tool (DAAT)

The Digital Asset Assessment Tool (DAAT)[33] is a project to develop a tool for determining the preservation needs of digital holdings. It will cover areas such as physical risks, format risks and organisational risks. It may help in addressing the prioritisation of the allocation of resources for preserving learning materials. The effectiveness of the tool is due to be tested in several trial institutions.

4.2.2 The Assessment of UK Data Archive and The National Archives compliance with OAIS/METS

[30] www.dpc.delos.info/cluster/index.php [April 2006]
[31] www.archives.gov/era/research/virtual-archives-lab.html [April 2006]
[32] http://ec.europa.eu/research/fp6/index_en.cfm?p=0 [July 2006]
[33] http://www.ulcc.ac.uk/daat [April 2006]

This was a JISC-funded project to assess how the operational structures of The National Archives (TNA) and UK Data Archive (UKDA) could be informed by OAIS and exploring the potential for interaction between existing metadata standards utilised within the two institutions and METS (Metadata Encoding and Transmission Standard).[34]

This project looked at the advantages of conforming to the OAIS, and how relevant it is for organisations such as the UKDA and TNA, as well as how structures can be informed by OAIS. The UKDA and TNA have service functions as well as preservation functions, in the same way that Jorum would if the decision was made for Jorum to act as a service and archive.

The METS format aims to provide a standard to hold metadata associated with a digital object in a form to facilitate sharing, cross searching, exchanging and displaying. It is essentially a 'container for all metadata necessary to describe, navigate and maintain a digital object (descriptive, administrative and structural metadata).'[35] As the OAIS is a conceptual framework, compliance is somewhat open to interpretation. However, the OCLC/RLG work on trusted digital repositories placed some emphasis on compliance with the OAIS as being a major attribute of a trusted repository.

The findings of this project were that the UKDA and TNA do comply with the mandatory responsibilities set out in the OAIS, and in fact it was felt that most properly set up archives would be likely to do so. However, the issue of authentication did require further thought (the dissemination of authenticated copies is one of the mandatory responsibilities of an OAIS).

Many of the OAIS responsibilities are already adhered to by Jorum. For example, there is a legally valid transfer agreement that enables Jorum to carry out preservation processes, possibly migrating learning objects to a variety of formats and distributing them in different formats, as well as giving Jorum the right to change the metadata as appropriate. Also, Jorum successfully makes the information available (something that may be set out in a dissemination policy that states the conditions of access and any conditions of use, such as acknowledgement of the original data creators). Availability is related to the effectiveness of the metadata and searching capabilities and the ease of access, and this is something that Jorum will monitor over time.

The JISC report concludes that the UKDA and TNA show a high compliance with the OAIS responsibilities. It makes an interesting observation about the relationship between the producer and the archive:

'The biggest discrepancy with the OAIS's concept of an archive's responsibilities was revealed to be the understanding of the Producer and the Designated Consumer Community. In an ideal world the archive would exercise strong control over the producers of material that is later handed over to the archive – digital preservation is said

[34] Beedham, H., Missen, J., Palmer, M., Ruusalepp, R. (2005), Assessment of UKDA and TNA Compliance with OAIS and METS standards. http://www.data-archive.ac.uk/news/publications/oaismets.pdf [May 2006]
[35] Ibid

to begin with the creation of the object that is to be preserved and archival requirements should be considered at the time of creation of a digital object. In the case of TNA, the archive has in reality very little control over the format in which material is created and later received from their donors. This situation has its roots in the legal framework, as well as in the way the material is created and its creation funded.'

The Designated Community is a key part of the OAIS model. Ensuring that the information is 'independently understandable' to the Designated Community should ensure that the usability is maintained. The community should be able to understand the information without needing the assistance of experts who produced the information.

'A general conclusion by both partners to this project is that the OAIS has an inbuilt limitation in that it overly assumes both an identifiable and relatively homogeneous consumer (user) community.'

The Jorum repository is specifically targeted at tutors and researchers, and not directly at students, and in terms of understanding the learning objects, the tutors and support staff form the Designated Community as they are the ones deciding which objects to use and how to use them for their students, and they will modify them as they see fit. The Designated Community can be made more specific by grouping tutors working with different levels of students and in different subject areas. If a tutor designs a learning object, they will make it understandable to their own students and not necessarily consider a wider community of users. But it may be reasonable to suggest that a learning object created for an undergraduate physics degree is effectively aimed at this particular level of expertise and this particular subject and it should be understandable to another physics teacher working at this level, though they may see the need to modify it for it to be suitable for their students. Whereas individual tutors may design objects specifically for their own courses, learning objects developed by projects (such as the JISC-funded X4L projects) may think more carefully about the community of users that they are designing objects for and ensure that the metadata accompaying the objects is as informative as possible, specifying such things as an educational level and category of user.

The Designated Community should understand the digital object, and the key to this is the Representation Information, which is all of the information required to make resources meaningful. Essentially it is everything that is required other than the actual data object, to enable users to utilise and understand the resource. This means everything from being able to interpret the bits to understanding what the figures in a table actually represent, so that information is understandable in terms of interpreting the data and in terms of preserving the meaning over time. The report on the UKDA and TNA does not address this issue in detail, but it should be noted that this issue is probably more relevant for scientific data, where Representation Information is key to understanding the raw data. The report does state that the UKDA seeks to ensure that 'data are accompanied by adequate documentation to enable their use for secondary analysis', but it concentrates more on the digital preservation strategies of UKDA and TNA, and summary information on these strategies is provided at the end of this section.

For an archive of learning objects, the producer and user community are going to be the same as the community that currently use the Jorum service, although the user community may expand to include historians and educational researchers. However, the Designated Community is still very wide ranging if it is taken as a whole, as all tutors and researchers will all have differing needs and requirements.

The report on the UKDA and TNA compliance with the OAIS also concludes that although the OAIS offers useful guidance concerning responsibilities, it is also important to consider other responsibilities and standards, which may take priority. One example of regulation that is important, for example, is access to disabled users and compliance with SENDA.

UK Data Archive Preservation Strategy
The UKDA has identified the minimum number of preservation formats that are necessary to manage the range of data types in the collections and carefully chose the migration paths for these formats. There is a proposal to develop a fully comprehensive and open XML standard that will provide a single definitive preservation format.

The UKDA strives to ensure that the:

- materials it acquires and accessions are suitable for scholarly use
- data are accompanied by adequate documentation to enable their use for secondary analysis
- data are checked and validated according to strict data processing procedures
- data are professionally catalogued and indexed according to appropriate standards
- data are effectively preserved for future use by converting them to several standardised formats and retaining multiple copies on different storage media
- format of materials is changed as necessary to preserve access to their intellectual content, reducing the risk of losing access to them over time
- materials are kept in conditions suitable for long-term archival storage

The UKDA has a preservation policy document.[36] This refers to the conversion of information to stable formats that are as software and hardware independent as possible.

The National Archives Preservation Strategy
The National Archives has adopted a strategy based on migration:

'Original bit streams transferred to the archives are held in perpetuity, along with all previous preservation manifestations and current presentation surrogates. Full metadata histories are maintained, even for obsolete manifestations and surrogates in support of the presumption of authenticity.'[37]

[36] http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf [January 2006]
[37] Beedham, H., Missen, J., Palmer, M., Ruusalepp, R. (2005), Assessment of UKDA and TNA Compliance with OAIS and METS standards

TNA's preservation policy outlines its strategy of migration.[38] The storage management system provides for bit-level integrity checking and multiple copies, and additionally archive systems are provided for backups.

4.2.3 SHERPA

SHERPA[39] is developing open-access institutional eprint repositories in a number of universities. A report on preservation standards[40] looked at the OAIS model for preservation, the rights that are necessary for preservation and the current standards of metadata for this purpose. SHERPA DP[41] is about creating a persistent digital preservation environment for institutional repositories.

The SHERPA-DP OAIS Report[42] acknowledges that eprint archives do not currently have the capabilities to ensure long-term preservation. The report puts forward a model whereby essential preservation services are outsourced to a third-party. The preservation services would be responsible for the construction and maintenance of the archival version of the eprint and the creation of additional metadata.

Whilst the report is essentially looking at eprint archives, it provides a useful perspective for using the OAIS reference model for any disaggregated preservation service. This type of service would be one option for the preservation of learning materials. The JISC Continuing Access and Digital Preservation Strategy (2002-2005) states that preservation processes could be carried out in-house, via collaborative approaches or via third-party agencies. The Arts and Humanities Data Services (AHDS) works on the latter approach, preserving eprints for a number of institutional repositories.

The SHERPA-DP report sets out a particular approach to a digital preservation information model. The OAIS model assumes that information comes into the OAIS and is archived before it is disseminated to consumers. The SHERPA-DP report argues that this does not reflect eprint repositories, and the reality is that the information is made available to the consumer following deposit, and only subsequently transferred to the archive store. This is likely to be the model that Jorum would adopt.

The SHERPA-DP project is only accepting deposits in a limited number of file formats. Whilst this clearly has advantages for management and preservation, it is not currently considered practical to limit the file formats that Jorum contributors can use, though the reality may in fact be that the vast majority of materials are in a limited number of formats. It may be decided to limit the file formats that will be accepted into the archive. Learning objects may then be taken into Jorum where there is no strategy for preservation

[38] http://www.nationalarchives.gov.uk/about/pdf/preservation_policy.pdf [January 2006]
[39] http://www.sherpa.ac.uk/ [January 2006]
[40] http://www.sherpa.ac.uk/documents/D4-5_Report_on_Preservation_Standards.pdf [January 2006]
[41] http://ahds.ac.uk/about/projects/sherpa-dp/ [April 2006]
[42] Knight, G. (2005) SHERPA-DP OAIS Report: An OAIS Compliant Model for Disaggregated Services. http://ahds.ac.uk/about/projects/sherpa-dp/sherpa-dp-oais-report.pdf [May 2006]

of the particular format, but they could be transferred to archival formats. With limited resources it seems almost unavoidable that an archive will have to limit the file formats. This issue is generally simpler for eprints, where the vast majority of materials will be in formats such as Microsoft Word, plain text, HTML and rich text format, with some images such as tiff, png, gif and jpg. The text-based formats can often be migrated to XML. But for interactive, multi-media content created using applications such as Macromedia Flash, the challenges of archiving are far greater.

4.2.4 Paradigm: Personal Archives Accessible in Digital Media

This is a collaboration between the libraries of the Universities of Oxford and Manchester to explore the issues involved in preserving digital private papers of contemporary politicians 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.[43] This project runs to December 2006 (extended to February 2007), and has been working on the creation of a workbook providing advice and guidance on the preservation of private papers. The parts of this workbook that would be of most relevance to the preservation of learning materials have not yet been completed (April 2006), for example, appraisal, preservation metadata, preserving digital objects and the use of the OAIS model. The practical testing of repository software and tools may also be of interest.

4.2.5 ESpida: An Effective Strategic model for the Preservation and Disposal of Institutional Assets

A project based at the University of Glasgow that is developing a sustainable business-focused model for digital preservation at FE/HE institutions.[44] *See* later section of this report on Costs and Value.

4.3 JISC Capital Programme

Digital preservation will be addressed as an integral part of digital repositories under the new JISC Capital Programme, which runs from April 2006 to March 2009.[45]. The Digital Repository and Preservation Strand will develop tools and automation, and models to clarify roles and responsibilities for digital preservation. A data curation project under the e-Infrastructure Strand will investigate disciplinary differences and impact on curation.. There will also be calls for collaborative projects, domain model development and technical model development within e-learning.

[43] http://www.paradigm.ac.uk [April 2006]
[44] http://www.gla.ac.uk/espida/ [April 2006]
[45] http://www.jisc.ac.uk/capital_roadmap.html [April 2006]

4.4 Preserv: Preservation Eprint Services

Preserv[46] is a JISC project investigating and developing infrastructural digital preservation services for institutional repositories. The project is looking to adapt the Eprints software to allow the collection and dissemination of preservation-oriented metadata. It is implementing an ingest service based on the OAIS and using PRONOM to identify file formats.

The project should provide very useful and relevant information to inform the future decisions made about the preservation of learning materials. It is not due for completion until February 2007 and as yet there are no substantial outputs.

4.5 The National Archives

The National Archives (TNA) continues to develop PRONOM, a file format registry.[47] The latest version includes a software tool to identify file formats automatically.

Version 4 of the PRONOM technical registry was released in October 2005, together with the DROID automatic file format identification tool. In January the PRONOM Persistent Unique Identifier (PUID) Scheme was published[48]. This is an extensible scheme for providing persistent, unique and unambiguous identifiers for file formats held in the PRONOM technical registry. These are fundamental to the exchange and management of digital objects, by allowing human or automated user agents to unambiguously identify, and share that identification of, the Representation Information required to support access to an object. At present the scheme is confined to one particular class of Representation Information: the format in which a digital object is encoded. PUIDs have been assigned to over 130 of the most common formats, and more being added on a regular basis.

The e-Government Metadata Standard[49] has now adopted this scheme as the recommended encoding scheme for describing file formats.

4.6 US National Archives and Records Administration (NARA): Electronic Records Archive

The Electronic Records Archive (ERA)[50] provides a useful example of a very large scale attempt to create an archive of digital records.

---

[46] http://preserv.eprints.org/ [April 2006]
[47] http://www.nationalarchives.gov.uk/aboutapps/pronom/ [January 2006]
[48] http://www.nationalarchives.gov.uk/aboutapps/pronom/puid.htm [January 2006
[49] Office of the E-Envoy, E-Government Metadata Standard (version 3, April 2004) http://www.govtalk.gov.uk/documents/eGovMetadataStandard%2020040429.pdf [January 2006]
[50] http://www.archives.gov/era/ [January 2006]

The US National Archives and Records Administration (NARA) has a huge challenge on its hands – saving the records of the federal government. It is confronting thousands of formats (maybe in the region of 16,000) and unprecedented volumes of data. The scope of the problem is unlimited and open-ended because formats keep developing and changing. NARA has hired Lockheed Martin Corporation (September 2005) to build a system for a permanent Electronic Records Archive. However, this is a long term project and the outcomes will not be known for some time.

Work currently being undertaken includes learning how to extract data from old formats and make the data available in useful, modern formats: 'For example, San Diego researchers took a collection of data on airdrops during the Vietnam War…and reformatted it so it could be displayed using non-proprietary versions of digital mapping programs known as geographic information systems, or GIS.'[51]

4.7 A Technology Analysis of Repositories and Services

A group at Johns Hopkins University in the US is conducting a technology-based analysis of repositories and services with funding from the Mellon Foundation.[52] They are including DSpace, Fedora, ePrints, Digital Commons (ProQuest), and applications include Sakai, and various e-publishing systems such as Open Journal Systems (OJS), and DiVA. They have a strong emphasis on digital preservation capabilities for the repositories. The main purpose of their analysis is to examine each of these systems with a transparent, clearly defined methodology, intending to define functional requirements for repositories to support various types of content and uses. They are using a wide variety of content with multiple file formats. They will be creating use cases to compare with the functionality of different repositories. They have collected a number of scenarios from various authors and are currently creating use cases from these before developing functional requirements and then mapping the capabilities of repositories to the functional requirements. Whilst there was a willingness to include systems such as WebCT and Blackboard in the analysis, the proprietary nature of these systems is problematic. The outputs will include a set of best practices and recommendations. The initial findings suggest that even individual institutions may require multiple repositories and applications, especially when considering service needs such as digital preservation.

4.8 LOCKSS

The LOCKSS system was designed to archive electronic journals. A LOCKSS box collects and stores everything that a reader can see on a publisher's website and stores it bit-for-bit as it is received. The system collects content by crawling the web and preserves it in the format supplied by the publisher. The LOCKSS system preserves the look and feel, or the 'performance' of a work. Each LOCKSS box collects its content

[51] Talbot, D. (2005) The Fading Memory of the State. http://www.technologyreview.com/InfoTech/wtr_14583,258,p3.html [January 2006]
[52] https://wiki.library.jhu.edu/display/RepoAnalysis/ProjectRepository [January 2006]

independently, and then compares the content with others that have collected the same content, in effect carrying out mutual audit. Each library collects and preserves content for its own readers alone, and the system does not support file sharing, so there can be many copies of any individual work.

An interview with the Director of the LOCKSS Programme in a recent issue of RLG DigiNews[53] provided a useful insight into the LOCKSS system. Whilst it has proved to be successful, with over 80 libraries worldwide, as well as many publishers, using the software, that does not mean that there are not failures at ingest, during storage and in the hardware used. It is true to say that all components of an information system are unreliable in the timescales needed for digital preservation and therefore it must be 'a fundamental design principle of any digital preservation system that it be capable of tolerating failures of any of its components.'

The article included a useful list of recommendations to include in policies and procedures for digital preservation:

- human intervention is expensive and prone to error: automate
- aim for full transparency and audit-ability in everything
- legal frameworks change and paper trails are not reliable so preserve legal rights and restrictions with the content to the greatest possible extent
- guard against technical arrogance by using open source software and nurturing a critical and contributing technical community
- collection development is a local activity

LOCKSS carries out on-access format migration which enables content collected in one Web format to be transparently presented to readers in another Web format.

The CLOCKSS (Controlled Lots of Copies Keep Stuff Safe) Initiative[54] has been set up to test the feasibility of a large, community-managed dark archive. The member librarians and publishers are working together to develop frameworks for publisher disaster failover systems and processes for providing public access to materials. There will be full public disclosure of CLOCKSS operations, governance, and technology. During the initial two-year project (2006-2008), CLOCKSS members will be working towards implementing a production system.

[53] Editors' interview with Victoria Reich, Director, LOCKSS Program. RLG DigiNews Vol.10 No.1 http://www.rlg.org/en/page.php?Page_ID=20894#article0 [February 2006]
[54] http://www.lockss.org/clockss/Home [April 2006]

5. Initiatives in Repositories for Learning Materials

5.1 TrustDR

The TrustDR project[55], running from June 2005 to May 2007, is 'a project to examine the practical issues in setting up digital rights management systems (DRM) in repositories of learning objects'. It is concerned with exploring the legal, organisational, cultural and technical aspects of operating an institutional digital repository of learning objects. The legal dimensions of e-learning, particularly those affecting the sharing and reuse of learning materials in the form of learning objects, are felt to present serious obstacles to future development, so this project is very timely. Preservation presents a further dimension to the issue of digital rights management, and this project may provide some conclusions on this issue.

The issue of trust is seen as centrally important and a means to reduce running costs:

'So, if trust reduces transaction costs in an economy how can we build and maintain it in the context of digital repositories? Some of the main barriers to the success of such repositories are not technical but legal and cultural.'

The project will be considering the cultural issues that need to be addressed in developing Digital Rights Management systems and aiming at an agreed legal expression of rights in the form of licences and user agreements from various groups of stakeholders. The project will also be looking at how these expressions of rights can be included in rights metadata using a Digital Rights Expression Language (DREL). It proposes that there are benefits from adopting a simple licensing regime and incorporating it into policy and DRM systems. TrustDR have produced a simple framework built on previous work by project Romeo and the DRM report produced for JISC by Intrallect.

5.2 Edinburgh Learning Object Repository (LORE)

LORE is a learning object repository for the University of Edinburgh.[56] It is currently limited to a specific number of e-learning projects, but there are plans for a university-wide repository. The website includes advice on the creation of reusable learning objects, emphasising such things as careful selection of format, the advantages of self-contained resources, the significance of granularity and the importance of standard metadata. It is conducting a pilot study to investigate if teaching staff are willing to share learning and teaching resources with colleagues and it is also conducting a survey to find out how staff manage their digital teaching materials.

[55] http://www.uhi.ac.uk/lis/projects/trustdr/ [April 2006]
[56] http://www.lore.ed.ac.uk/ [April 2006]

5.3 Community Dimensions of Learning Object Repositories

This project 'aims at identifying and analysing the factors that influence practical uptake and implementation of learning object (LO) repositories'.[57] It will look at drivers, barriers and enablers, which may then inform the recommendations on wider policy issues. The project runs until May 2007. The project has begun the process of reviewing institutional policy and strategy, which will feed into a set of recommendations on institutional strategy and policy, and a set of structured guidelines for those setting up repositories.

The first deliverable of the project was the 'Report on Learning Communities and Repositories'.[58] This report looks at socio-cultural, pedagogic, organisational and technological barriers and possible solutions. It recognises that preservation may be one of the drivers for setting up a repository, but it does not address preservation specifically. However, long-term access to learning objects is not essentially about preservation, it is really about user demand and user expectations over time. It will therefore be useful to find out whether the findings of the CD-LOR project provide any evidence for long term user demand.

6. The Case For and Against Preservation

The question of why educational practitioners should use repositories in the first place is still an important one to address:

'The "'Why" question is perhaps one of the most significant problems relating to the use of repositories that the learning technology community has failed to adequately address.'[59]

The 'why' question is outside the scope of this report, and the assumption is made that it is worthwhile to create and use learning object repositories. The question is whether it is worthwhile maintaining access to learning objects over time, and what that time span might be.

It is important to remember that commitment to preservation does not guarantee long term access, whilst leaving preservation effectively to chance does not mean that the data will automatically be lost. In the end, the efficacy of preservation can only be assessed after the fact. We could decide not to actively preserve learning objects in Jorum, and simply to leave this to chance, a position that could in itself be seen as an active strategy. Users will deposit objects, others will download them and may use newer software to access them. In this way they may continue to be accessed for some time, and it is likely

[57] http://www.ic-learning.dundee.ac.uk/projects/CD-LOR/ [March 2006]
[58] Margaryan, A. et al (2006) CD-LOR Deliverable 1: Report on Learning Communities and Repositories http://www.ic-learning.dundee.ac.uk/projects/CD-LOR/CDLORdeliverable1_learningcommunitiesreport.doc [April 2006]
[59] Campbell, L., Blinco, K., Mason, J. et al (2004) Repository Management and Implementation. White paper for alt-i-lab 2004. http://www.jisc.ac.uk/uploaded_documents/Altilab04-repositories.pdf [May 2006]

that the more popular and widely used objects will survive for the longest time, as well as those created in more stable and common formats, with any number of people accessing them over time with different software versions.

This section sets out the case for actively carrying out preservation. Not all materials need to be retained for long periods, and a later section will discuss appraisal methods. Preservation will always be about prioritising, in just the same way as archivists have always made decisions about what to retain within archive repositories. Unless everything is kept, which is likely to be a practical impossibility, there will always be some form of selection.

6.1 Learning Objects as Intellectual Capital

'We are shifting from an industrial economy to a knowledge economy, where repositories of digital information and the tools to mine, analyse and repurpose them represent a society's intellectual capital.'[60]

Learning objects are one of the knowledge assets of the e-learning economy. By creating and using them we are benefiting from the expertise of others and sharing knowledge. If learning objects are valued in this way, then the case for preservation is made stronger.

There seems to be some feeling that tutors within further education value teaching materials more highly than those in higher education, having come from a different tradition of teaching. FE tutors have very little time to develop their own teaching materials, and are generally more willing to share. The Espida project at the University of Glasgow[61] found that academics view teaching materials as fairly ephemeral, but it may be that a shift in thinking will gradually take place, as there is more realisation of the potential of digital materials. It is often the case that individuals and organisations do not appreciate that their personal and business records might have long term value to an archive because they are only thinking from the perspective of current use and they have no need to think about possible future use and other potential uses of the materials.

Inevitably, the intellectual value depends partly upon the person doing the valuing. One person may see learning materials as ephemeral, another may see a great deal of potential. So, with a mixture of perceptions, it may depend largely upon who is consulted and which questions are asked. It could be argued that if the community sees learning objects as worth sharing, then they are worth preserving. The proliferation of institutional repositories seems to validate the principle of sharing many types of digital materials, and it is hoped that this will be echoed within repositories that deal more specifically with learning materials.

[60] Digital Preservation Coalition (2006). Mind the gap: assessing digital preservation needs in the UK. http://www.dpconline.org/docs/reports/uknamindthegap.pdf [Feb 2006]
[61] http://www.gla.ac.uk/espida/index.shtml [April 2006]

6.2 Course Requirements

It is clearly highly desirable within a local repository that there is a responsibility to keep objects whilst the course that they are being used for is active, and there may also be arguments for keeping supporting material. In a paper about maintaining learning objects for the future,[62] Harvey refers to the requirement to keep early versions of learning objects, such as working notes and drafts, and the need to keep the final materials for the life of a course, which may be up to 10 years. This responsibility would have to rest with the institution providing the course, and therefore it would be the role of institutional repositories. A national repository such as Jorum cannot practically adopt course requirements as one of the criteria for preservation.

6.3 Depositors' Expectations

It is important to consider depositors' expectations in terms of access to materials they deposit.

The expectations of Jorum depositors will need to be assessed over time. It may be that there are no particular expectations about long term access to the materials that they deposit, but this needs to be ascertained, possibly by surveying depositors or providing other methods of feedback. Depositors may not have considered the long term perspective, although they are advised about selection of file formats and encouraged to create self-contained learning objects, both of which may facilitate long term preservation.

The motivation for producing learning objects may be increased if creating them improves the profile of a teacher within their subject area, and maybe beyond. This is certainly taken as read for research outputs, and is one of the drivers for deposit into eprints archives. Many repositories sell themselves on the basis of secure storage and similarly, within an e-learning repository, long-term access may help with motivation to deposit. If the materials are only available for a few years, they are less likely to become widely used. Making a quality, widely used learning object that only lasts for a few years may not be considered as prestigious as one that lasts, and is used, for 10 years.

One project being funded under the JISC Digital Repositories Programme that may be of particular interest is Rights and Rewards in Blended Institutional Repositories[63] which runs for two years from July 2005. Amongst other things, this project is looking at the rights academics would like to retain in their teaching materials and research outputs and the rewards that may act as drivers for deposit and use of repositories. This project should be very relevant for Jorum, providing 'transferable models of expressing and protecting rights for teaching objects, and of motivating and rewarding academics appropriately for deposit.' It will be interesting to see whether it does conclude that preservation is a factor

[62] Harvey, Professor R. (2002). Now You See It, Now You Don't: Maintaining Digital Learning Objects for the Future. http://www.usq.edu.au/electpub/e-jist/docs/Vol5_No2/Harvey%20-%20Final.pdf [Jan 2006]
[63] http://rightsandrewards.lboro.ac.uk/index.php?section=1 [March 2006]

in encouraging deposits. A national survey has already been conducted on whether academics feel they should be paid for making their teaching materials available to others. This is not directly relevant to preservation, but it may provide useful information on the value academics place on their teaching outputs and the way that repositories need to promote themselves:

'emerging repositories need to make it clear to contributors and users what they are trying to achieve in a language that can be understood. The advantages must be clearly outlined so that people can see why there is a need for repositories and sharing of teaching material amongst the HE and FE sectors. It is important to remember that people will not change because others want them to; they will only change if they see a reason for doing so, such as seeing a personal benefit or to help others with their learning or teaching.'[64]

A particularly interesting observation was made in the discussion section relating to contributions to repositories in section 5.2, where people put forward the reasons for depositing their materials. Over one-third of participants who had contributed to a repository before said that one of the reasons was to preserve their materials. This shows that people are keen to ensure that their materials are housed in a safe environment. This can also be seen as a benefit for institutions, as their employees are preserving institutional assets. In answer to the question: 'If you were to contribute to a repository in the future which of these reasons would make you more or less likely to do so?', about 30% said that they would be much more likely to deposit if the repository helped them to manage and preserve resources.

The lack of a long-term business model may deter depositors, as they may not feel confident about the financial viability of a repository. Funding for repositories may not explicitly be for the long term, and even if the current funding is not for a specific period of time, there is no guarantee that it will continue indefinitely. Depositors are probably not concerned about learning materials to the same degree that they are concerned about research output. But they might be attracted to invest in a repository that they believe to be sustainable. Current debates surrounding the deposit of research output in institutional repositories refers to the reluctance of academics to change behaviour. This may also be an issue for learning materials. The Preserv Project Manager points out that preservation will follow commitment to implementing an institutional repository. As institutional repositories develop and their needs are clarified, the preservation community is likely to respond:

'I'm confident that there will be preservation services that support the needs of IRs at various levels of cost when those institutions decide what they want from such services, and that has to emerge from coherent policy and consequent projections about content development.'[65]

[64] Bates, M et al (2006) Rights and Rewards Project Academic Survey: Final Report http://rightsandrewards.lboro.ac.uk/files/resourcesmodule/@random43cbae8b0d0ad/1137423150_SurveyReport.pdf [March 2006]
[65] Hitchcock, S (Preserv Project Manager), Jisc Repositories Jiscmail list, 31 March 2006.

This proposition strengthens the case for concentrating on developing the current Jorum repository and maintaining clearly defined policies for deposit, management, access, etc. Arguments for preservation are going to be immeasurably strengthened if it can be shown that Jorum is widely used, that materials are frequently downloaded over time, and that the teaching community start to refer specifically to materials held in Jorum, in a similar way that articles in journals are cited.

6.4 Users' Expectations

The requirements of the user community are key to the decisions about what to preserve and how long to preserve it for. As e-learning becomes embedded into the curriculum, there needs to be more research into user requirements over the short and longer term and into whether the user community would actually change over time. This is important in order to build evidence of the requirements and expectations of tutors. Jorum users are part of the same community as the depositors, namely teaching and support staff in higher and further education, although depositors also include those engaged in JISC funded projects to create learning materials. The benefits of preservation of a whole range of learning materials will be capable of appropriation by the whole teaching community. However, it is useful to think in terms of depositors and users as separate concepts for the purposes of preservation.

The need to focus on the user community is seen as key within the OAIS model (which refers to the specified users of an OAIS as the Designated Community), as this will affect what is kept and how it is kept. It might be useful to think in terms of different user communities, maybe categorised in terms of educational level or subject specialisms. The user communities are not necessarily static, and over the long-term learning objects may become useful for different communities. It is not easy to predict future expectations as well as take into account changing user communities. Whilst one community may cease to use an object, it may have potential to be repurposed, though this may depend upon appropriate changes to the metadata designed to facilitate its discovery within different contexts. It is interesting to note that the earliest papers in arXiv.org, from 1992, are still being downloaded from the UK arXiv mirror site.[66]

User feedback should be seen as essential in making decisions about preservation, as the key issue is whether users will benefit from continued access to materials over time. As far as Jorum is concerned analysis of users and user feedback will be taking place:

(1) Jorum will gather statistics on download, and though this will not necessarily reflect actual use of materials, frequent downloads of an object do infer that it is being widely used. This could help to inform decisions about preservation.

(2) Users will be able to add comments about objects, and these can be collated and used to help ascertain the value of individual objects to users.

[66] http://uk.arXiv.org [March 2006]

(3) Jorum incorporates a star rating system, enabling users to identify resources that they find particularly useful.

6.5 The Original Learning Objective Remains Valid

It may be that the resource continues to effectively meet the learning objective that it was originally designed to fulfil. This may be true in some subject areas where changes over time are less frequent and less dramatic. A learning object that remains valid in this way may not require alterations to the pedagogical metadata. For example, a learning object that seeks to analyse a passage of Shakespeare or one that explains the structure of an oxygen molecule is likely to continue to meet its objective and be valid within the context that it was created in the long-term. If a self-contained learning object seeks to explain a simple immutable concept, then there may be a stronger case for preservation over the long term than for a learning object that is rooted more strongly in current thinking or current events. It is, however, quite possible that the materials would come to be seen as old-fashioned in terms of functionality and design, because they were created using older versions of software.

6.6 Reuse

The term Reusable Learning Object (RLO) is now widely used, and this is seen as a key factor in the potential take-up of learning objects in teaching. The potential for reuse provides a means to maximise the return for the original investment in creation. Reuse of research materials is common, with researchers adapting material and adding their own findings to create new output. Reuse of teaching materials has tended to be more limited, although sections from academic text books may be utilised or lecturer's notes and lesson plans may be used in modelling new courses.

In terms of long term preservation, reuse is a key concept, as preservation enables use over time, and this implies repurposing materials for different contexts.

'If constructed appropriately, warehoused wisely, and catalogued accurately, a learning object might find usage beyond its original audience'.[67]

Reuse is about developing a culture for using learning objects produced elsewhere, for sharing knowledge and adapting it to specific circumstances. Collaboration should help to encourage reuse and encourage the idea of raising the profile of tutors and institutions as a result of sharing resources. It may be that in the future there is academic peer review of learning objects in the same way as we now have with academic research papers.

[67] Richards, G., McGreal, R, Hatala, M., Friesen, N., The Evolution of Learning Object Repository Technologies: Portals for On-line Objects for Learning. Journal of Distance Education, 2002, vol.17, no.3. http://cade.icaap.org/vol17.3/richards.pdf [April 2006]

Materials that are most suitable for reuse are generally considered to be those that are stand-alone and self-contained:

'In the language of digital education resources, an image of a molecule is usable in far more instructional contexts than an entire lesson on molecular bonding.'[68]

A learning object that is essentially an image of a molecule, with metadata attached, could be used in many different contexts. It could be used as part of a course on molecular bonding, but if you take the course as a whole, it is usable in only a limited number of contexts. Conversely, it could be argued that context is very important in the learning experience. Wiley describes a recognised reusability paradox, which is that whilst learning theorists may emphasise the pre-eminence of context in learning, the principle of reuse would seem to encourage context independence. It is yet to be seen whether in practice decontextualisation is a hindrance or an aid to the continued use of learning objects. It could certainly be argued that a high level of decontextualisation makes indexing for human discovery more difficult. Wiley concludes:

'…while the most decontextualised learning objects are reusable in the greatest number of learning contexts, they are also the most expensive and difficult for instructional designers to reuse'.[69]

Jorum encourages depositors to create smaller chunks of learning, as these are generally considered to be more flexible and easier to incorporate into some kind of aggregate that a teacher might create for a course. In terms of aggregate objects, depositors are being encouraged to consider depositing both the object as a whole and the dis-aggregated individual assets. These individual assets would then link to the larger, aggregate object. At present, metadata held at individual asset level, where assets are within aggregate objects, cannot actually be viewed in Jorum. However, as part of the ongoing development activity this situation is set to change so that the metadata will become visible, which will facilitate the principle of dis-aggregation. There may be an option for a template to be produced so that depositors can effectively copy the metadata at the aggregate level across to individual items, and then they may modify it as appropriate. Whilst this initiative would greatly facilitate preservation, it is not a mandatory obligation, and it remains to be seen whether depositors take up this option.

There is no mandate about the type of objects that should be deposited in Jorum, and therefore, from the point of view of preservation, it will be necessary to have strategies for coping with both contextualised and decontextualised objects. The complexity of the objects will vary widely, there will be a wide range of formats used and the quality of metadata will also vary.

Whilst reuse of learning objects is a widely recognised concept, Tompsett argues that in reality the idea of creating new courses from existing reusable learning objects cannot be

[68] Wiley, D. (2005) Learning Objects: Difficulties and Opportunities. http://wiley.ed.usu.edu/docs/lo_do.pdf (Jan 2006)
[69] Ibid

seen as 'a simple and unbounded benefit of repositories of RLOs.'[70] He examines the problems of reconfiguring a large number of RLOs to integrate them into another learning object. He states that there is little evidence for building a new course in this way:

'If repositories of RLOs are to support reconfigurability, then it must be possible to discover a set of 'consistent' RLOs that will integrate together to form a substantial part of a course. Further, if a sufficient number of RLOs are available, it should be possible to create an appropriate sequence to use for the materials from metadata alone. Both problems are considered as fundamental to reconfigurability but do not occur in simple reuse (at any level).'

Tompsett examines these problems from a mathematical perspective and concludes that the creation of repositories alone is insufficient to support the creation of new courses and far more research is needed into the accuracy and sufficiency of metadata and the potential for RLOs to be reconfigured.

A paper by Verbert, et al,[71] presents an ontology based framework for repurposing learning object components. The framework 'disaggregates learning objects and provides direct access to their components, enabling their reuse in dynamic compositions of new learning objects.' It supports slide presentations, which are a common learning object type, enabling the presentation to be disaggregated into clear segments which are categorised into more meaningful content objects such as definitions, references and examples, and they are then annotated using the Automatic Metadata Generation framework, which automatically describes each component. The idea is to combine metadata from the object with metadata about the context in which it is used. The aggregation process searches for components and adds them to a slide presentation. This work allows on-the-fly repurposing of learning object components for slide presentations, but the efficiency and effectiveness of this approach is yet to be tested.

The reusability paradox shows that there are competing demands to consider. The priority for tutors will be to create good quality learning objects that meet the needs of their students, which may mean creating learning objects that are complex and pedagogically specific. There are strong arguments for placing heavy emphasis on the principles of reuse, but standards can only be mandatory for JISC-funded projects that are creating learning objects, not for individual tutors.

The potential for reuse is not easy to calculate. The decision can be made that a successful learning object may satisfy the objectives for which it was created, i.e. for a particular course within a particular institution, and then it may be deleted because it does not have much potential for reuse. It will be difficult to establish criteria for judging what is likely to be reused. The JISC report on Long-Term Retention and Reuse of Learning

[70] Tompsett, C. (2005) Reconfigurability: creating new courses from existing learning objects will always be difficult! Journal of Computer Assisted Learning Vol.21 No.6.
[71] Verbert, K. et al (2005) Ontology based Learning Content Repurposing: the ALOCoM Framework http://www.cs.kuleuven.ac.be/~hmdb/ProlearnIClass/papers/Verbert.pdf [May 2006]

Objects and Materials[72] recommended that guidelines should be created for best practice in creating reusable learning objects for UK higher and further education. However, despite best attempts to predict future use it is inevitable that tutors will combine and use learning objects in ways unimagined by the original creators.

6.7 Teaching Methods and Learning Design

Learning design is concerned with the principles of teaching strategies and educational goals. It is concerned with the 'who does what when and using which materials and services in order to achieve particular learning objectives.'[73] One teacher may look to fulfil the same educational goals as another, but have a very different strategy for doing so. Conversely, two tutors may use the same strategy for different teaching purposes. It is clearly useful to be able to analyse and judge the efficacy of different methods.

The preservation of the learning design is partly about being able to reconstruct and interpret the history of teaching methods and teaching content. Learning objects may be part of a teaching strategy and preservation could seek to capture this, providing evidence of the ways that tutors have used learning objects over time. It may be argued that an archive of learning design should preserve not just the learning object and metadata, but all of the documentation surrounding the learning objects. For example, it may be important to know the syllabus of a course in order to understand the context of the learning object.

In addition to their historical importance, preserved complex learning objects may also have a role to play in helping tutors to create their teaching strategy, effectively basing it upon past pedagogical approaches.

Whilst learning design is something to think about, at present any preservation strategy could only address the metadata that is attached to the objects within Jorum, and learning design is not in reality a consideration for the current repository.

6.8 Historical Importance

The arguments for keeping learning objects for their historical importance are the same arguments that might be used for any historical documents. It is important to value our heritage, be it cultural, social or intellectual. Just as we find great value in the records of teaching in Victorian classrooms, so future researchers will want to research and understand teaching in the early 21st century.

---

[72] Barker, E., et al. (Nov 2004), Long Term Retention and Reuse of E-Learning Objects and Materials (JISC commissioned report) http://www.jisc.ac.uk/index.cfm?name=project_elo [January 2006]

[73] IMS Learning Design Frequently Asked Questions (2003) http://learningnetworks.org/downloads/IMSLD/IMS%20Learning%20Design%20FAQ%201.0.pdf [March 2006]

Historical importance is clearly a strong argument for some level of preservation, and even if objects are not preserved in order to maintain them as usable teaching materials, a sample of learning objects should be kept for historical purposes. The issue then becomes one of methods of selection. *See* under Appraisal.

6.9 Uniqueness

If a learning object that is deposited into Jorum is known to be unique, there may be a stronger argument for preservation. This would seem self-evident in some ways: it is an essential characteristic of traditional archives, and it is what separates them from printed works. However, this is a problematic issue when dealing not only with electronic records, but also with learning objects that are deposited by numerous individuals into a national repository. How can we know which objects are in fact unique? At present, there is no means to identify unique learning objects when they are deposited into Jorum. So, a learning object may be stored in Jorum and may also be stored elsewhere, with no means to identify which is the 'master' copy, or whether the copy held elsewhere is going to be preserved.

Once materials are downloaded from Jorum, new copies are effectively created, and these may then be modified, effectively creating similar learning objects that may be uploaded into different VLEs at different institutions.

The value of most assets will increase if they are seen as unique, but this is a concept that requires further consideration as far as learning objects are concerned, and it is by no means straightforward. Work in this area is ongoing as part of the Jorum Service-in-development activities.

6.10 Costs of Creation and of Re-creation

Learning materials are an investment of time, resources and skills on the part of individuals and institutions:

'Learners and instructors find the Web a convenient medium for educational and administrative transactions and have spurred an unprecedented investment in time and resources in the creation of materials and network infrastructure for distance and augmented learning.'[74]

It is important to secure the maximum benefits from the investment made in the creation of materials. It makes no sense to recreate materials that have already been created; it makes much more sense to modify existing materials for new purposes.

74 Richards, G., McGreal, R, Hatala, M., Friesen, N. (2002) The Evolution of Learning Object Repository Technologies: Portals for On-line Objects for Learning. Journal of Distance Education, vol.17, no.3. http://cade.icaap.org/vol17.3/richards.pdf [May 2006]

If arguments for preservation are to include arguments about costs, then much more work needs to be done in this area. *See* section on Costs and Value.

6.11 Trusted Digital Repository

It could be argued that Jorum should be seen as a 'trusted' digital repository:

'The concept of a trusted digital repository, providing reliable long term access, may be important in developing a good reputation that in turn encourages deposit.'[75]

If Jorum is to be a trusted repository, then this must surely be a strong argument for taking on the responsibility for preservation, and fulfilling the other attributes of a trusted repository.

6.12 Rights Issues

Preservation will have implications in terms of rights issues. Where material is preserved, it requires preservation copies to be made, and modifications may be needed (restricted under the Copyright and Digital Preservation Act 1988: such actions must be agreed to by the copyright owner).

The TrustDR project[76] is dealing with legal issues in learning object repositories (see earlier references to this). One of the project outcomes is 'taking a more long term and strategic attitude to the management of digital teaching and learning materials'.

6.13 Technological Advances

It is inevitable that new and more efficient means to preserve materials will be developed over the next few years, and it may be that materials kept now will be easier to preserve in the future. It would be feasible to simply keep the original format and the bit stream for the present, on the basis that research work being undertaken now will find better ways to preserve and better mechanisms for using the data in the future. This is a controversial approach, because it effectively passes responsibility on to future custodians. However, if preservation becomes easier in, say five years time, then the cost/benefit analysis may move in its favour. It is important to continue to monitor developments within preservation, in order to be able to take advantage of common solutions.

[75] Stevenson, J. (2005) Jorum Preservation Watch Report. http://www.jorum.ac.uk/docs/pdf/Digital_Preservation_Report.pdf [May 2006]
[76] http://www.uhi.ac.uk/lis/projects/trustdr/ [March 2006]

6.14 Reasons not to Preserve

There are a number of arguments against preservation:

- It may simply be cheaper to create new learning objects than invest in the necessary infrastructure to manage a preservation process
- It may be easier and more practical to create new learning objects, especially simpler objects that are suitable for reuse
- Newer learning objects will benefit from using newer software, allowing greater functionality, and users may find them easier to access
- Newer learning objects may be more suited to current pedagogical approaches
- Avoiding long term preservation may mean avoiding complexities with rights issues over time

By selectively preserving learning objects, some of these arguments may be addressed, as the selection can take into account issues such as format and access software and also the costs involved in creating particular learning objects.

Apart from any other arguments that may be put forward against preservation, it may be that reuse over time does not happen in reality, but as we are still in the early stages of an online learning culture, we do not yet have the evidence for this.

6.15 Conclusion

Learning objects provide new ways of visualising, thinking about, interacting with and understanding complex topics. It seems likely that teaching techniques will continue to integrate learning objects, which can provide students with a more flexible, personalised and satisfying way of learning, more in tune with the ways that they now embrace new technologies. New tools will facilitate creation and tutors are likely to develop skills in the creation of learning objects as they become more integrated with mainstream teaching. Ideally learning objects should provide a stream of benefits over time.

It is necessary to gather more evidence of costs, user requirements, expectations, the reality of patterns of use of learning materials and how the community perceptions of Jorum develop. In the end, whilst the arguments for preservation may be persuasive, the choices that are made now may not be proved to be valuable or otherwise for some time. But if the choices are made on the basis of empirical evidence, best practice and responding to user requirements, then they can be justified.

7. Costs and Value

'The single greatest threat to materials being preserved over the long term is money. Societies will have good times and bad. Keeping content safe must be a marginal expense

in order to decrease the threats during bad times as well as to maximize available funds for new acquisitions during good times.'[77]

The costs of preservation cover a range of areas, many of which also relate to the implementation of digital repository functions in general. These include management, appraisal, selection, acquisition, ingest, metadata creation, preservation processes, access documentation, technical co-ordination and maintenance of hardware and software.

Learning objects are a form of capital that have no physical existence. As intangible assets, it can be difficult to measure their value and the costs of maintaining them. A learning object may be seen as an intangible asset that has a reasonable probability of producing benefits at some future time. The decision to fund the long term preservation of learning objects is effectively an investment decision. It is assumed that tutors and students in the future will benefit from being able to access and use materials created now.

In order to consider the value of preserving learning materials over the long term, it is necessary to think in terms of a reasonable probability of producing benefits and look at present costs in expectation of future returns. Benefits are not just monetary and could include such intangibles as the creation of goodwill and the furtherance of academic research. Prospects are always affected by risk and uncertainty. One of the difficulties is that the value of learning objects may not be apparent immediately, and it is difficult to predict what the level of use will be over time. Economic viability would need to be calculated over a reasonable time period.

We can appreciate that an archive (such as The National Archives or a local record office) will invest time and money to preserve resources even though their value and usefulness may not become apparent until well into the future. It is hard to estimate the value because the value of an archive, which is historic value, is not the same as the value of a current resource: they will be used by different user groups for different purposes. However, learning materials may continue to have value for teaching and learning, as well as taking on an historic value, so that they continue to have value for the original purpose that they were created, and may be used by the same community of users.

We do not have experience of preserving online learning materials, and therefore it is difficult to predict whether the value of the materials as teaching aids will change over time. If we decided to preserve a selection of learning objects purely for historical purposes, the task would be easier. But if we want to preserve in order to continue to enable access to materials for teaching and learning then we need to find ways of measuring that value. Whist it seems likely that the value decreases as the materials become older, conversely it could be that the level of use increases, maybe if the metadata is enhanced, or if the subject area that the learning object is relevant for becomes more popular within the teaching curriculum, or if the learning object is recommended as being of particularly good quality.

[77] RLG DigiNews Vol.10 No.1 http://www.rlg.org/en/page.php?Page_ID=20894#article0 [March 2006]

The highest cost of the creation of learning objects is staff time:

'To make high-quality learning objects in collaboration with faculty requires staff with a wide-range of skill-sets, including graphic design, programming, 3D modelling, instructional design, metadata production, and (obviously!) subject knowledge expertise.'[78]

For objects that are being deposited in Jorum, the costs would include the additional metadata that needs to be added to most objects. There are also resources involved in the general management and maintenance of the Jorum repository, such as training, negotiation and promotional activities, and these should be factored in when considering the total cost of each individual learning object. These costs should be taken into account when calculating the value that is placed on the actual content. For example, if you have 100 learning objects that have been deposited, then the cost of each of those objects should include the total costs for running the repository divided by 100 for each object. If a learning object is deleted then its total cost is the cost of its creation and management, right up to the time taken to check and select it for deletion. If an object is in the Jorum repository for five years and has received minimal use, then its costs per use are high, especially if it is quite complex and took some time to create in the first place. It could be argued that objects that are rarely used should be deleted, but if the costs of creation were high, this may be an argument in favour of preservation. In essence we need to look at the cost to create and maintain the learning object up to the point of either preservation or deletion, and the value of the learning object to the teaching community, which would be based upon the amount of use. We then need to look at the cost of preservation of different types of objects, taking into account the level of complexity of the object and the formats that have been used. A learning object archive would exist in order to provide benefits to the whole learning and teaching community. However, the benefits must be seen to be real and substantial.

One of the uncertainties that we have to deal with is the risk of the demise of current hardware and software. This is notoriously difficult to predict. A learning object may not require any preservation processes for some time, because it remains easy to access. So, if an object is created on a stable platform that remains accessible for say ten years, then the cost of the object is probably going to be low, because very little maintenance is needed and the object continues to be available for use. However, it may be the case that the performance of the object is altered over time when new versions of software are used to access it.

The level of automation that can be achieved is a key factor in the calculation of cost. It is likely that the costs of producing learning objects will fall as the tools to create them become easier to use and more widespread, and it may be that the addition of metadata becomes automated to a greater degree than at present. This is an important issue, as the time required for the addition of metadata is significant. Metadata may be added by the producer and by the Jorum cataloguers when the object is deposited. There will also need

[78] Roy, M. (2004) Learning Objects (Educause Evolving Technologies Committee). http://www.educause.edu/ir/library/pdf/DEC0402.pdf [Feb 2006]

to be changes to the metadata over time, in order to ensure that the learning object remains relevant and in order to ensure that any changes made are documented.

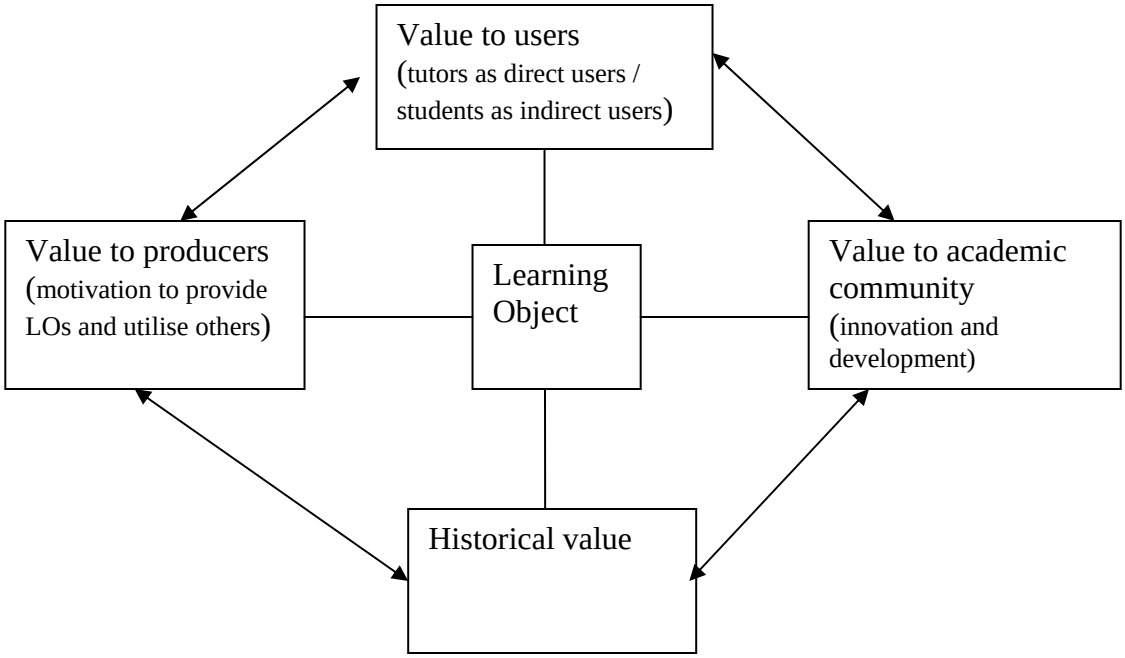
Having information on the rate of use would help us to calculate the costs per use. Statistics for the number of downloads of learning object provide some idea, but download is not the same as use. Furthermore, even with statistics on use, actual value to students can only be inferred. It is essential to monitor use and encourage feedback from users. If there is high confidence in Jorum as a national repository, and satisfaction from use, then the archival function is more likely to be effective. Whilst further research on the creation of a preservation function is being undertaken, the use of the current Jorum repository can be monitored and its value as a national repository established.

Balanced Scorecard

A Balanced Scorecard provides a method for measuring intangibles. Typically this divides into the financial perspective; the internal business perspective; the innovation and learning perspective; the customer perspective. For each of these there will be goals and measures.

It is important to be clear about the strategic aims in order to create a scorecard. The most critical factor is put at the top of the scorecard, e.g. user satisfaction, and then other factors would be added, such as operational efficiency and finance, which are both linked to learning and innovation.

A Balanced Scorecard for learning materials would put the information assets at the centre. It would need to look at the value of the assets to the learning community.



Balanced scorecard for learning objects

The above scorecard only provides a starting point, and would need more detailed consideration. In general the financial value of the asset is part of the scorecard, but currently the learning objects in Jorum are free at the point of use, and JISC are not looking for a financial return on use but an academic benefit.

7. 1 Espida – an Effective Strategic Model for the Preservation and Disposal of Institutional Assets

Espida[79] is a JISC-funded project that is seeking to develop a sustainable business-focused model for digital preservation, based on the University of Glasgow's diverse electronic holdings. Whilst this project is looking at the assets of an institution, its findings will certainly be relevant to the issues surrounding costs, risks and benefits for the storage of learning materials, as it is looking most crucially at how we can measure the value of digital materials over time. It should also help to inform a strong argument in favour of sustainability to put forward to funders. Whilst the costs of sustainable digital preservation are often seen as the main barrier to be confronted, it is just as important to answer the questions of why digital materials should be maintained at all. It is important to emphasise that preservation is about knowledge, and preservation of the object is simply a means to preserve information as knowledge.

[79] http://www.gla.ac.uk/espida/index.shtml [April 2006]

The espida work identifies the key players as the asset creators and the senior management. The situation for Jorum is slightly different, with asset creators and a national repository that is funded by JISC on behalf of the whole community, but the principle of selling the case for preservation is the same. The key players need to understand the value of the assets, the technological fragility of the assets and the need for sustained support for preservation:

'It is important to state that we are not selling "digital preservation" to our senior management and asset creators, we are selling them the motivation to manage, reuse and preserve assets that are of value to them and the University.'[80]

The main outcome of espida will be an overarching model as the culmination of a number of models that will be created during the project. The final model will be completed by September 2006. There are currently no interim models available from the espida website (April 2006).

7. 2 LIFE: Life Cycle Information for E-Literature

This project is a collaboration between University College London (UCL) Library Services and the British Library. LIFE[81] is looking to develop generic life cycle model applicable to all digital collections. Once individual stages have been established, these will then be costed. There are many life cycle models available for all sorts of information, and LIFE is looking at an amalgamation of traditional library cost model and the digital preservation model. The idea is also to come up with decision points on the model, i.e. what, when, where, how.

A report has been produced on the first stage, which was a literature review to help form the basis of the life cycle model. The main areas covered were life cycles in libraries, life cycle costings in libraries and life cycles in the context of digital preservation. The report summarises many life cycle models, including the stages in the management outlined in the Digital Preservation Handbook (Jones, Beagrie) and the life cycle approach recommended by the OAIS reference model, both of which have been cited in the Jorum digital preservation watch. It also looks at literature referring to roles and responsibilities.

'LIFE aims to provide life cycle costings … to include all stages of the life of an item, including preservation. Very few of the costings outlined in this review provide metrics for digital preservation, and the ones that do, approach the problem from a variety of different angles…. LIFE will synthesise the information… into a new model, generic

[80] Currell, J, McKinney, P (2006) Investing in Value: A Perspective on Digital Preservation. D-Lib Magazine April 2006, vol.12 no.4. http://www.dlib.org/dlib/april06/mckinney/04mckinney.html
[81] http://www.ucl.ac.uk/ls/lifeproject/ [April 2006]

enough to provide the flexibility to provide costings for different sorts of electronic collections.'[82]

7.3 arXiv.org

arXiv.org[83] based at Cornell University Library, provides open access to scientific eprints, acting as an archive for research papers. It had 45,000 submissions in 2004/05 and has had over 350,000 in total.

The cost for ingest into arXiv.org is borne by the producers, and it uses a quasi automated vetting process in that potential contributors have to be championed by someone who is already contributing.

The costs of running the service have been worked out in great detail. Providing costings at a highly detailed level helps to clarify the exact costing implications. They have looked at capital equipment, in terms of annual expenditure on the purchase of hardware and software, the amortization rate (i.e. writing off intangible assets over the projected life of the assets, generally a 3-5 year range), annual maintenance, licences and development fees. Direct operating costs are based on the OAIS functions of ingest, data management, storage and preservation planning. Server costs have been worked out in terms of equipment cost $10,600, an amortization rate over 5 years, annual fees of $2,400 and annual equipment costs of $4,500.

The total can be divided by current database numbers to get a cost per submission. A greater number of similar objects leads to decreased costs, as well as greater automation. In 2004/05 arXiv.org received 45,000 submissions, and the cost worked out at $0.10 for each submission.

Personnel costs were worked out at $113 per hour staff time when all components were factored in (a weighted hourly rate, looking at the number of productive hours worked in a day). There are also service costs involving fees paid for outsourcing and related services as well as contingency and overheads.

The annual costs for maintenance/preservation worked out at $305,333 per year, which is $6.79 per submission in one year, and $0.94 per submission worked out over the lifetime of arXiv.org.

Staff costs are the greatest expense by far, and so the total costs will only decrease with more automation of human effort. In general arXiv.org is highly automated and is therefore relatively cheap to maintain.

[82] Watson, J. (Nov 2005) The LIFE project research review Mapping the landscape, riding a life cycle (final draft). http://www.ucl.ac.uk/ls/lifeproject/documentation/review.pdf [April 2006]
[83] http://www.arXiv.org [April 2006]

7.4 The National Archives

The National Archives (TNA) preservation service[84] has a great deal of experience in the management of digital materials and deals with a huge amount of data. The Archive deals with a large number of formats and range of complexity and quality, as it does not have any control over the creation of materials (though it may offer advice to government departments).

Ingest involves appraisal, transfer, pre-accession processing, cataloguing and loading. TNA have attempted to cost ingest, and it has been calculated at £18.76 per file, factoring in operational costs, external costs, and annual number of files ingested. Costs are linked to the complexity of transfer, not the volume of material. Manual processing increases costs. The creation of metadata is a high cost element, so automated metadata extraction will always prove very beneficial.

In terms of data and storage management, the current cost is £3.34 per megabyte, taking into account capital costs, annual operational costs and volume of records. Costs relate primarily to storage capacity rather than use, but it is difficult to predict future transfer volumes. Current costs for online delivery are 13p per item.

Future costs will need to take into account the preservation strategy, research and development costs and future operational functions. Preservation planning includes risk assessments, identifying and testing migration pathways and looking to automate more in the future.

As far as migration is concerned, the major cost is the development of a migration pathway. Subsequent automation means that implementation costs are relatively low. For emulation the major cost is the development of an emulator. The cost differential is not clear at the present time.

The highest costs are associated with ingest and preservation planning. Efficiencies can be achieved with a collaborative R&D approach on characterisation and preservation planning. Automation is seen as the key to sustaining preservation activities.

8. Roles and Responsibilities

Responsibilities within digital preservation are currently not well defined. There is a need to clarify the roles that surround learning object repositories, which include information providers and information creators, owners, users, funders and repository service providers. In terms of responsibility for preservation, the situation within the Jorum repository is complicated because the ownership of objects is not straightforward and there are issues surrounding the rights of individuals and institutions as well as the existence of both stored objects in Jorum and virtual objects referenced by Jorum.

[84] http://www.nationalarchives.gov.uk/preservation/ [April 2006]

Looking at roles relating to a learning object repository and a preservation function in more detail, there are a number of areas to consider:

- financing
- creating objects
- creating metadata
- acknowledging copyright
- taking in objects
- obtaining and using objects
- providing access
- building a collection (institutional, national, specialist, subject based)
- storing a collection
- administration
- preservation planning
- carrying out preservation
- managing preserved objects
- advocacy
- awareness raising
- provision of advice and support
- preservation watch

An archive of learning materials would represent a significant store of intellectual capital for learning and teaching, preserving useful, informative ways of representing knowledge in order to teach new generations. This is a responsibility that all parties should have an interest in. In seeking to identify and clarify responsibilities, the OAIS model may prove useful. The SHERPA project, for example, is using OAIS to develop a preservation environment, assigning rights and responsibilities, establishing protocols and workflow processes.

On a broader scale, it may be that a roadmap is needed to define the roles and responsibilities of national organisations, funding bodies and institutions. It would be beneficial to investigate relationships between institutional repositories and external preservation service providers and to look at models for cooperation that enable the sharing of knowledge, expertise and infrastructure across traditional domain boundaries.

Responsibilities can be hard to define within repository architectures, which can be anything from highly distributed models to self-contained repositories providing the whole range of services. Distributed models require levels of interoperability and the easy exchange of content and services. According to the Director of the LOCKSS Program, digital preservation is best done as a collaborative, community effort.[85] This allows

[85] Editors' Interview with Victoria Reich, Director, LOCKSS Program, RLG DigiNews 15 Feb 2006, Vol.10 No.1. http://www.rlg.org/en/page.php?Page_ID=20894#article0 [May 2006]

community control over the direction of the technology and applications. This approach has worked very well for LOCKSS, which is an alliance of libraries and publishers. The members direct development and build local collections, preserving a wide variety of formats and genres. However, further investigations would need to be made to ascertain whether this approach would be suitable for learning materials.

Virtual Objects

Whilst Jorum does act as a physical repository for learning objects, it also includes the metadata for virtual objects, whose resources are stored elsewhere. It would be problematic for Jorum to have responsibility for maintaining access to materials that are not physically within the repository, though it would be reasonable to take responsibility for ensuring that the metadata that is necessary to describe the learning objects is preserved, and provide information about where the responsibility lies for the learning object. Identification of the responsible institution for a virtual object might be possible via automated metadata generation. However, this would problematic for aggregate objects that include virtual objects. The RLG report on trusted digital repositories refers to accepting responsibility for long term maintenance of digital resources on behalf of depositors. This would effectively exclude resources that are not actually deposited.

Materials stored in Jorum

How much responsibility should Jorum be expected to take on for materials that are stored within the repository? Does Jorum need to be seen as a trusted digital repository, and is this an essential part of that role? Does it need to maintain access to objects in order for users to see it as trustworthy? If the learning objects continue to be owned by contributing institutions, should they have the responsibility for preservation?

There are many difficult questions surrounding preservation, and these will need to be explored further. Other than for specific JISC-funded projects, there is no compulsion to deposit materials in Jorum, and depositors may deposit the same materials into their own institutional repositories, so it may be argued that they should they take on the responsibility for preservation if they provide access to materials. After all, they have intellectual ownership of the materials.

Whilst there is a strong case for a national repository taking on a national role to provide an archival function, there is also a strong case for making it a collaborative effort that involves contributing institutions as well as other bodies with expertise and experience in this area.

Costs and Ownership

The cost of creating the learning objects is currently borne by the institutions that create them. If Jorum were to take on preservation, where would the costs of preservation lie? The ownership remains with the individual institution, so what implications does this have? Would it be preferable, or indeed feasible, to agree that after 5 years, or once

preservation is carried out, the rights pass to Jorum, so that the learning objects are in effect preserved for the benefit of the whole academic community?

Analysis of Learning Objects

A learning object may be broken down in order to think more clearly about the constituent parts from a preservation perspective.

Learning objects may be single information objects with metadata, or they may be more complex, with a number of information objects that are structured in a certain way to create an aggregate object with metadata to describe the overall object and possibly with metadata at the individual information object level.

An information object may be defined as the actual data object itself along with the Representation Information necessary to make it meaningful. The data object and its associated Representation Information need to be clearly identified and understood, as it is the Representation Information that is needed to interpret the data, which is essentially bits, and make it into meaningful information. The information object is the basic unit of access, containing all of the relevant information required to reproduce the document, which will include metadata that is necessary in order to understand the object. A learning object is an information object with additional metadata to provide structural and descriptive information. However, the distinction between metadata that is part of the information object and metadata that is part of the learning object is not necessarily obvious, and therefore it is not clear how useful this distinction might be. It would be worth exploring this further, looking at the relationship between the information object and the accompanying metadata.

An analysis of the nature of learning objects may help to clarify the preservation approaches that can be taken and who has responsibility for the parts of a learning object. The party responsible for the information object may not necessarily carry out preservation, but it is their decision whether to do so or not. It may be that the responsibility is jointly shared, and then the decision for preservation would need to be jointly made.

For example, if the metadata for a virtual object is deposited, the depositing institution may be identified as the responsible party for preservation; if an object is actually deposited in Jorum, it may be that Jorum takes responsibility for preservation of the information object and accompanying metadata or it may be that the responsibility is identified as being held elsewhere by a third party. It is vital that this is clearly documented, preferably within the metadata that is attached to the object.

9. Version Issues

The reality for learning materials is that they will be reused in a number of environments, creating different versions. Where an institution has its own learning object repository,

materials may be deposited into this and into Jorum, and the likelihood of confusion over versions increases. A range of individuals and agencies will use and possibly alter the materials, and this adds complexity to roles and responsibilities. A central repository acting as an archive would alleviate the problem of version control.

If an object is migrated, or if an object is improved by utilising new technologies, making it more sophisticated and more interactive, should the older object also be kept? The newer one might be tied more closely with current technologies and it might be that in a few years time it is not very accessible, but the older one is either more accessible or easier to migrate to newer platforms. If different versions are kept, there needs to be a system for version control.

10. The Open Archive Information System (OAIS)

The OAIS was described in some detail in the previous digital preservation watch report. It is a high-level framework for understanding the structural organisation of a repository and for applying common concepts and terminology. It explains in great detail all of the entities required in the process of ingest, management and access.

An OAIS-compliant repository exits to preserve information, defined as data objects that can be interpreted by the user community. The SHERPA-DP project has produced a report looking at an OAIS compliant model for disaggregated services.[86] In terms of learning objects, it would be possible to break down preservation in a similar way to that proposed in the report, with suitable modifications, for example:

- preservation of the bit-stream that represents the information within the learning object
- preservation of the information content, embodied in a file or format
- preservation of the experience of interacting with the information content
- preservation of the structure of the learning object
- preservation of the pedagogical context

Jorum could seek to incorporate the OAIS functional entities of ingest, archival storage, administration, data management and access within its existing structure. Alternatively, if an archive is seen as a separate concept to the Jorum repository, an OAIS-compliant structure could be created as a separate entity to Jorum and domain specific terminology could then be applied.

Within the OAIS, the essential unit of data object and Representation Information is known as an Information Package, comprising the data object itself, the Representation Information required to interpret and understand it and the preservation description information. If the OAIS model was to be followed, an Archival Information Package

[86] Knight, Gareth (2005), SHERPA-DP OAIS Report. http://ahds.ac.uk/about/projects/sherpa-dp/sherpa-dp-oais-report.pdf [April 2006]

(AIP) would need to be created before dissemination via a Dissemination Information Package (DIP). The SHERPA-DP report proposes a workflow whereby the dissemination copy of the eprint and associated metadata is derived from the original Submission Information Package (SIP), and the Archival Information Package is constructed later in the workflow, after the information package has been transferred to the preservation service. This model would seem to be suitable for Jorum, where the learning objects are deposited and made available for download, and the creation of archival copies would come later in the workflow, when the information package is transferred to the preservation service.

The JISC Digital Repositories Programme Support team are currently preparing an evaluation report on the Reference Model for an Open Archival Information System (OAIS) as a basis for a generic reference model for repositories, examining the arguments for and against using this model across different types of repositories.

10.1 Representation Information

Representation Information typically maps the bits of a data object into commonly recognised data types, such as character and integer, and into groups of these data types. So, it provides the data structure concepts that are applied to the bit sequence. This is the Structure Information of the Representation Information. There is a further type of Representation Information, known as Semantic Information, which gives further meaning, such as specifying the language used and special meanings associated with the elements of the Structural Information.

Once you have the Representation Information, you want to display it in understandable forms, and Representation Rendering Software enables you to do this. For example, it may be possible to use the Acrobat application software to make the Representation Information within a PDF humanly readable. Whilst access software presents the information contents of an Information Object in human readable forms, and therefore incorporates some understanding of the Representation Information, it is not considered advisable to see Access Software as a replacement for the Representation Information. There is a danger that the Access Software will cease to function over time, and then it becomes necessary to have a full description of the Representation Information.

The DCC Development Team has put up a prototype registry/repository for Representation Information.[87] Within this there are examples of Representation Information such as the FITS standard dictionary and EAST standard. This registry could provide part of the infrastructure surrounding the preservation of learning materials.

Representation Information and the Designated Community

[87] http://dev.dcc.rl.ac.uk/dccrrt/ [July 2006]

According to the OAIS, a digital object should include the Representation Information necessary to render it and for the Designated Community to understand it. But Jorum is not responsible for quality control of the objects, and cannot make judgements as to whether all deposited objects will be understandable to the Designated Community. For example, a contributor may deposit a learning object including tables of data, and it is up to the contributor to ensure that the tables are labelled meaningfully. If the axes of a table are labelled with abbreviations or units that are not universally understood then the efficacy of the learning object is reduced and it is less likely to be used by others. This may come to light via user feedback on the resource and it may be taken into account when considering whether to preserve the object, during an appraisal process.

11. Appraisal

A trusted digital repository should implement sound appraisal processes. Unless there is a decision to preserve everything, there must be an appraisal process to decide which learning objects (or information objects) are going to be preserved. There will always be a level of guesswork when deciding what to keep, even when principles and expertise are applied, but a rigorous process should result in the selection of high quality, pedagogically sound learning objects, which should make the investment in infrastructure pay off in the educational experience of the users.

The relationship between responsibility for preservation and appraisal needs to be clear. It may be that the responsibility for preservation is identified as belonging to a depositing institution or third party, in which case they may also undertake the appraisal process. However, it will be useful for Jorum to consider appraisal methods and make recommendations about the approach to and criteria for appraisal.

Recommendations may be made about who should be responsible for carrying out appraisal, which might include tutors, learning technologists, subject experts, librarians or archivists. By setting out the criteria for appraisal, a systematic approach can be used that helps to justify the decisions taken.

The Digital Preservation Coalition (DPC) have created a Decision Tree[88] for the long term preservation of digital materials:

"Clearly defined selection policies will enable cost savings in terms of time taken to establish whether or not to select and also potential costs further down the track of needing to re-assess digital resources which are either in danger of becoming or are no longer accessible."[89]

[88] Interactive Assessment: Selection of Digital Materials for Long-term Retention http://www.dpconline.org/graphics/handbook/dec-tree-select.html [March 2006]
[89] http://www.dpconline.org/graphics/handbook/dec-tree.html [March 2006]

This will be a useful tool to help construct a selection policy for learning objects. The principle is that the flow of the questions represents a logical order of evaluation. If the response to early questions is not favourable there is little point in accepting preservation responsibility for the resource or continuing its evaluation. There may be some irrelevant or inappropriate questions within the Decision Tree as far as Jorum is concerned, but it could form a useful basis to construct an appropriate tree for learning objects.

11.1 User Demand for Long-term Access to Learning Objects

One of the most important questions to ask when considering the whole issue of preservation is whether there will continue to be users of the material over the longer term. When considering this issue there are two main categories of users:

(a) those continuing to use learning objects for teaching purposes
(b) those using learning objects for historical purposes

As far as category (a) is concerned, more work needs to be done to find out about patterns of use of learning objects. This can only happen over time, once Jorum has become firmly established and widely used. It will be important to understand what users require from a learning object repository and whether they are likely to continue to expect access to materials over the course of several years. For those in category (b) who might use learning objects for historical purposes, the responsibility is to keep a selection of objects that would provide evidence for teaching methods and pedagogies to future generations.

Cheap digital storage and sophisticated retrieval tools is shifting the balance of costs such that keeping everything is not such a daunting prospect. It may be considered cheaper to keep everything than to select, and cheaper to search than to organise, and there is always an argument that every resource will prove to be useful to at least one potential user. However, the sheer quantity of resources over time would make this a problematic option, not least from the point of view of maintaining effective access, and the fact that all resources would eventually require preservation treatments would mean that costs would start to escalate.

Preservation could be treated as a passive activity; something that is effectively user-led. Users download the objects from Jorum and integrate them into their own environments, which may effectively mean migrating resources from an older version of the software to a newer version. For example, they may take something created in Macromedia Flash 3 and use it with Flash 5. However, there are clear problems with this approach. It means there would be no consistency, and that preservation would be left to chance. The user may access the object using newer software, but the old version is still in the Jorum repository. Also, migration to a newer version of the software may involve changes in the performance of the object that are not necessarily obvious. If users could redeposit materials back into Jorum, then it would be possible for them to effectively deposit an updated version of an object, but this is not possible at the moment, and it also creates major issues around version control and authenticity. It would seem preferable to ensure

that users are a central part of the appraisal process but have a systematic and controlled process where everything is clearly documented.

11.2 Appraisal Criteria

11.2.1 Quality

Objects may be preserved on the basis that they are considered to be of high quality. If objects that are of high quality are not preserved, then the risk is run of deleting good quality objects and replacing them with recently deposited lower quality objects, although measurements of quality themselves will always be open to argument. At present, there are no general quality assurance procedures for learning objects. These would be useful as they would help to encourage a higher level of quality in creating of materials. Whilst the Jorum workflow does provide guidance for depositors, there is currently no formal quality control at the point of deposit into Jorum, although this can be built into the workflow on request from the institution.

There is an argument that quality control should not be carried out by a repository: the role of a repository is to provide access to materials that people wish to deposit and others wish to use. However, if it is accepted that an archive cannot preserve all materials, then a selection process must be introduced, and quality may be one of the criteria. It may be considered appropriate to appoint a range of appraisers, who have experience and subject expertise, to take as balanced a view as possible.

Within Jorum there is a comments and star rating system, enabling users to determine the quality of resources. The feedback provided in this way by users could form part of the appraisal process. This would be a means to introduce peer review, which has always played a key role in assessing quality in the bibliographic world. It may be that users would 'only use materials that are quality controlled and peer reviewed'.[90]

MERLOT[91] is a resource for learning and teaching that holds descriptions of materials for cross-searching. It has an Awards Program for Exemplary Online Learning Resources which 'recognizes and promotes outstanding online resources designed to enhance teaching and learning and honors the authors and developers of these resources for their contributions to the academic community'.[92] A MERLOT editorial board (these are discipline based) selects an outstanding, peer-reviewed resource to receive its annual MERLOT Classics award according to program criteria. This is one way of encouraging users to think of learning materials as high quality resources. The peer-review process itself is carried out by at least two higher education faculty members who compose a composite review that is posted to the MERLOT website.

[90] Campbell L.M. et al, (2004) Repository Management and Implementation, a white paper for alt-i-lab 2004. http://www.jisc.ac.uk/uploaded_documents/Altilab04-repositories.pdf [May 2006]
[91] http://www.merlot.org [Feb 2006]
[92] http://taste.merlot.org/catalog/awards/ [February 2006]

11.2.2 Level of Use

Level of use is central to any appraisal process and it is important to consider this alongside quality. It may be that objects that are perceived to be high quality are not being used. This may be because they are too complex, too inflexible for many teaching situations or because the metadata is not adequate for users to identify them. For this reason it will be important to examine patterns of use and find out why some objects are more widely downloaded than others.

11.2.3 Metadata

The quality and completeness of the metadata will impact upon appraisal and preservation, and also upon use, which affects its value and quality. For example, if all of the formats within an object are not specified in the metadata this will cause problems for the evaluation process. Accurate file format metadata is considered to be essential when carrying out preservation processes

11.2.4 Technical Dependency

Appraisal may take technical dependency into account, as realistically those materials that are highly technically dependent on other resources will be far more problematic in terms of reuse.

11.2.5 External Dependency

Objects may reference external information and their efficacy may be reduced if the external links cease to function. It is a higher risk to keep objects that reference external information.

11.2.6 Set of Objects Dependency

Learning objects may effectively form a set, and it may be considered necessary to keep the whole set in order to give the proper context and allow the user to continue to take advantage of the intended use of the objects in a sequence. It may be that one object effectively depends upon the continued existence of a whole sequence in order to remain effective.

11.2.7 Ease of Migration

Some learning objects will be easier to preserve than others. This may be due to the format used or because of the structure of the object. It is not ideal to base appraisal on this, because content, quality and level of use are more worthwhile factors to use, but as preservation is a resource intensive activity, it could be argued that ease of preservation should be one criteria for appraisal. It would need to be considered as just one factor, alongside the other criteria.

11.3 Appraisal Schedule

An appraisal process would normally involve creating an appraisal schedule. The schedule provides an important means of documenting all decisions. It is vital to know what has been kept and what has been deleted and to appreciate the how and why of the decisions made. Learning objects may undergo a periodic review, to decide whether they should be kept for a further defined period, maybe every five years.

11.4 Digital Asset Assessment Tool

The Digital Asset Assessment Tool (DAAT)[93] is being developed by the University of London Computing Centre (ULCC) and the Arts and Humanities Data Service (AHDS) to help prioritise what to preserve, to assess assets for preservation risk and produce assessment criteria for storage media and formats. It would be useful to consider using this tool to help inform an appraisal process for learning objects. Its effectiveness is due to be tested in a number of trial institutions.

11.5 Appraisal and Granularity

Whilst there have been debates about the merits of smaller chunks of content versus more complex learning objects in terms of the ease of reuse, the Jorum repository is designed to service both. There is an additional complication, in that learning objects may be aggregates of a number of information objects or they may just include single information objects.

In reality there is a gradual scale of complexity, from a small chunk of non-aggregate content, such as a single image, to complex aggregate objects with very specific learning objectives. The aggregation levels outlined in the IEEE LOM are similar:

1. The smallest level of aggregation, e.g. raw media data or fragments.
2. A collection of level 1 learning objects, e.g. a lesson.
3. A collection of level 2 learning objects, e.g. a course.
4. The largest level of granularity, e.g. a set of courses that lead to a certificate.

The process of appraisal is made more complex by the existence of aggregate learning objects as well as complex information objects. Smaller chunks of content will be easier to appraise and easier to preserve and indeed arguably more likely to be reused. But aggregate objects will be difficult to handle, whether they contain smaller, simpler information objects or more complex information objects. It will be necessary to investigate more thoroughly the implications of carrying out the appraisal process on aggregate objects. Aggregate objects may have taken a significant amount of time and effort to create, and so looking at appraisal from the point of view of cost and value, a

[93] http://www.ulcc.ac.uk/daat.html [May 2006]

complex object is high cost and therefore there is a motivation to increase its value by maintaining it over time so that more tutors can benefit from using it.

12. Trustworthiness

The RLG report on Trusted Digital Repositories was referred to at length in the first Preservation Watch Report. It has become an important reference document for considering the issues in involved in the concept of trust and trustworthiness within digital repositories.

In essence, a trusted digital repository: 'accepts responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users' and meets 'expectations of trustworthiness'.[94]

The RLG report sparked a great deal of debate about the concept of a trusted digital repository and also the concept of trustworthiness. A distinction exists between the idea of a repository that is seen as trusted, in the sense that users believe that the repository will look after the resources that it has in a reliable manner, and the idea of trusted resources, which refers essentially to the resources being what they purport to be.

The RLG have now produced an Audit checklist for the certification of Trusted Digital Repositories.[95] It would be useful to measure an archival repository for learning objects against the criteria laid out here. For example, the checklist includes the following statements:

- A repository must demonstrate an explicit, tangible, and long-term commitment to compliance with prevailing standards, policies, and practices.
- A repository must have designated staff with requisite skills and training and must provide ongoing development.
- The repository must demonstrate reliability in all its operations and support to its range of users. Reliability and sustainability are essential to establishing trust in the repository.
- Long-term preservation is a shared and complex responsibility. A trusted digital repository contributes to and benefits from the breadth and depth of community-based standards and practice. Regular review is a requisite for ongoing and healthy development of the repository.

Use of the checklist would demonstrate the areas where a repository is fulfilling its responsibilities and the areas that might be addressed, or that are maybe not relevant to learning objects. It may raise issues or suggest areas of responsibility that have not yet been fully considered within Jorum.

[94] RLG/OCLC (2002) Trusted Digital Repositories: Attributes and Responsibilities. http://www.rlg.org/legacy/longterm/repositories.pdf [May 2006]
[95] http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf [March 2006]

Concepts of trust are important for a learning object repository and there needs to be more investigation of this. It may be that the trust people place in a learning object repository and in the learning objects themselves is not the same as the trust required for resources such as research papers and archives of people and of organisations. It would seem to be important that a potential user can trust that the learning object is accurate, but this can be a problematic issue. Accuracy is partly subjective. To take a simple example, a learning object make seek to explain how to build a wall, but it may only explain one method where there may be other equally valid methods, and it may leave out some steps that others would consider to be important. Use of the instructions within the learning object may not result in a good solid wall, but this does not necessarily mean the object is at fault. A learning object will tend to reflect the author's opinions and experiences. Maybe what is more important is that the purpose and content of the learning object is accurately reflected in the metadata, and therefore the metadata reliably describes the information object.

If a learning object is popular it may indicate that it is seen as trustworthy, successfully fulfilling its stated function. But this cannot be taken as read without further investigation. Users may utilise the object any way that they wish, so even the original learning intention is not vital to the success of the object.

In terms of the long-term perspective, digital preservation itself should be seen to be a trustworthy process. This should be facilitated by a clear preservation policy, which should set out the technical infrastructure.

There has been a great deal written about the importance of authenticity and the importance of the user having trust in the digital object. However, this needs to be considered specifically from the point of view of learning objects: whether future users will be concerned about the authenticity of a resource and what authenticity actually means in this context. Authenticity is about something being what it purports to be - for a learning object this may mean that it should fulfill the description given in the metadata so that a user can trust that it is relevant to them.

13. Persistence

Long term persistence of links within resources is clearly a central issue. Link failure is a common problem. For example, a study of the persistence of URLs cited in articles in D-Lib magazine between 1995 and 2004 found that approximately 30% failed to resolve.[96]

The National Library of Australia refers to resources that are linked to by other resources as requiring persistence (this is specifically for web materials). Jorum includes virtual

[96] McCown, F, et al (2005) The Availability and Persistence of Web References in D-Lib Magazine http://www.iwaw.net/05/papers/iwaw05-mccown1.pdf [May 2006]

objects that are referenced by a URL. It is vital that these links continue to identify and provide access to the resources in order for the learning materials to remain useful:

'The persistence of links between resources and of links in resource discovery services is essential to ensure long-term public access to web-based materials. It is, therefore, an important aspect of the archiving and preservation strategies adopted for these resources.'[97]

A resource may be removed by the institution providing it, probably because it no longer considers the resource to be useful or current. It may have been updated and provided with a new URL. A broken link within a Jorum learning object will reduce the credibility of the object so it is vital that institutions alert Jorum if they wish to remove or change a URL. Alternatively (or maybe in addition to this) Jorum can check links on a periodic basis.

The situation is made more complex where there are virtual resources that form part of a larger learning object that also includes non-virtual resources, but in terms of the persistence of the location, the same principles apply.

MERLOT checks all links to the materials described its catalog on a monthly basis. The result is a list of materials found to have invalid URLs.[98] Each material on the list is tested by following the URL provided on its Detail View page. Each verified invalid URL is then investigated. The investigation process is thorough and would take some time to complete. If the material cannot be located and the submitter of the material does not respond to an email within 4 weeks, the material is deleted. If a new URL is located, MERLOT is updated accordingly.

14. Metadata

The PREMIS Working Group was awarded the prestigious Digital Preservation Award for 2005[99], showing the importance that is placed upon preservation metadata. PREMIS have created a data dictionary, identifying core digital preservation metadata elements and including practical examples as well as a software protocol. PREMIS should encourage consensus building and collaboration, which is crucial in so many digital preservation issues. The core metadata elements identified by PREMIS are increasingly seen as an international standard, and it would be useful to look at the core elements alongside the current Jorum metadata, as outlined in the application profile, in order to make some conclusions about the preservation metadata required for learning objects and the resources that this might require.

[97] http://www.nla.gov.au/guidelines/persistence.html [May 2006]
[98] http://taste.merlot.org/policies/removing.htm [February 2006]
[99] http://www.dpconline.org/graphics/advocacy/press/award2005.html [January 2006]

A key question is whether creating and maintaining preservation metadata on the scale needed for the preservation of learning objects is realistic or sustainable. The PREMIS Data Dictionary refers to metadata for the objects needing preservation, metadata for the actions carried out and metadata for the people or software controlling the actions. Even after the questions of resources to create and maintain metadata have been considered, there is the question of the quality of the metadata in terms of consistency and completeness. The metadata itself will need to be preserved over time, although this should be relatively straightforward as it can be created in a way that avoids application dependence.

There continue to be many developments within this area that will be worth monitoring. The National Library of New Zealand have been developing a software tool to automatically extract preservation metadata and output it as XML[100]. These sorts of tools may prove to be relevant for Jorum, where metadata creation is a particularly resource intensive activity.

15. Preservation Process

The preservation process refers here to the actual processes used to maintain access to the data. The process could be defined in a broader sense as being the whole lifecycle:[101]

- appraisal – the determination of which materials are to be preserved
- accession – a controlled process of import into the archive and evaluated for completeness
- description – provenance, integrity, structural and behavioural characteristics of the material
- arrangement – possibly aggregating entities into containers for storage management, or arguably this is addressed by the content package.
- preservation – migration or emulation processes
- access – supporting discovery, retrieval and display

LOCKSS (Lots Of Copies Keep Stuff Safe) have a substantial amount of experience in carrying out preservation, and whilst this is only one possible approach, via use of the LOCKSS boxes, it can provide some useful conclusions. Some of the key recommendations made by LOCKSS (outlined by Victoria Reich, the Director of the LOCKSS programme) are: minimise any processing as much as possible; migrate on demand; preserve presentation and look and feel; use open source software; guard against dependence on limited, centralised technical expertise; allow no single points of failure; strive for diversity in administration, funding, and technology.[102]

[100] www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction [March 2006]
[101] A Rajasekar et al, (2005) Digital Preservation Lifecycle Management for Multi-Media Collections. San Diego Supercomputer Center, University of California. Lecture notes in computer science 3815
[102] RLG DigiNews Vol.10 No.1. http://www.rlg.org/en/page.php?Page_ID=20894#article0 [March 2006]

Migration is a common strategy for preservation, and it may happen on deposit, on demand or according to a pre-defined strategy. Migration on deposit can have advantages, as materials can immediately be migrated to a limited number of formats, and subsequent preservation may not be needed for some time. Even with such migration, the formats chosen will eventually become obsolete. In reality, the content is usually preserved in its original form as well as a migrated form.. Where a migrated copy is used, there may be some loss in performance, but access to the original is an exception rather than the norm. Batch migration is a useful method to save time and resources, converting a whole batch of materials to a new format in one operation.

On-demand migration requires the ability to convert dynamically from the obsolete to the current format, a method used by LOCKSS:

' When a format is thought likely to become obsolete, the digital preservation system is enhanced with the ability to present the reader, upon request, with the requested content in a current format. In effect, the migration tool is integrated into the dissemination pipeline of the preservation system rather than being applied to the preserved content.'[103]

This should have the advantage that the latest technology can be employed and the actual preserved content is only migrated when absolutely necessary. Format converters that are developed can themselves be preserved to document the original format. However, migration on demand may slow down user access to materials.

In the D-Lib Magazine article on formation migration[104], a proof-of-concept implementation of migration on access is described using a GIF to PNG conversion. Where a client request was made for a GIF image, a GIF to PNG conversion was made and the converted content was delivered. The migration was not perceptible to the user and there was no noticeable delay in access. LOCKSS are currently working on a full implementation of the API (application programming interface) for plug-in format converters, and a broader set of converters than the proof-of-concept converter.

The retention of the byte-stream in its proper and logical order is widely regarded as an essential part of digital preservation and is sustainability as the base level. It can then be delivered identically to how it was first deposited.[105] Cedars recommends against repeated format conversions, and instead argues for keeping the preserved byte-stream, while tracking evolving technology by maintaining the technical metadata. When access to the material is required, the preserved byte-stream can be converted to a file system and software to render access to the material can be identified via the technical

[103] Rosenthal, D et al (2005), Transparent Format Migration of Preserved Web Content. D-Lib Magazine, January 2005 Vol.11 No.1. http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html [May 2006]
104 Ibid
[105] A byte can be defined as a contiguous sequence of a fixed number of bits, generally accepted as 8 bits, and a bit is just one binary digit, which is the basic unit of information in computing

metadata.[106] Preservation of the byte-stream is regarded as a relatively routine IT operation (though this assertion is open to question).

It is advisable to treat the storage of bits (or bytes) as a separate activity:

'High level models for persistent repositories indicate that digital archiving and long-term preservation is best handled by separating archival storage of bits (storage management) from data management, logical representations, and higher level services that can be built on top of a persistent storage architecture.'[107]

The OAIS distinguishes the byte stream from the object itself, enabling the different aspects of the issue of preservation to be considered separately. As well as maintenance of the raw data, the meaning must be rendered in future technical environments.

Physical preservation of digital information can be managed by copying a byte stream from one medium to another. The more challenging problem is logical preservation, in order to ensure that the information remains meaningful so that computers and humans can interpret and use it.

The Interactive Telecommunications Program at New York University carried out a project called The Archive Ingest and Handling Test (AIHT)[108]. A moderately complex digital archive was given to a number of participants with the idea of testing the feasibility of transferring archives from one institution to another to test the stresses involved in transfer, ingestion, management and export of a digital archive. It was found that even the initial transfer of the archive (from George Mason University to the Library of Congress), which was just the start-up phase, was fraught with problems. There were problems with different sorts of identifiers, such as URLs, file names, file extensions, query strings and digests, and URLs were not easily convertible to file names. Extensions were variable, such as .jpg or .JPEG and in some cases labels were incorrect. One of the conclusions was that identifiers must be carefully checked and verified and files labelled with extensions must similarly be verified. Tools meant to validate formats may themselves have variable implementations, which makes it almost impossible to have an absolute truth. After the ingest exercise, it was found that the desirability of a digital object is more closely related to its quality than the quality of its metadata, and so many of the metadata fields were desirable rather than requisite, which may be significant where donors are creating the metadata:

'if donors had the energy and ability to canonicalize the metadata around the…objects, they would also be equipped to preserve it themselves.'

---

[106] Cedars Guide to: Digital Preservation Strategies (2002) http://www.leeds.ac.uk/cedars/guideto/dpstrategies/dpstrategies.html [March 2006]

[107] It's About Time: Research Challenges in Digital Archiving and Long-term Preservation. August 2003. Final report of the Workshop on Research Challenges in Digital Archiving and Long-Term Preservation. http://www.si.umich.edu/digarch/NSF%200915031.pdf [Feb 2006]

[108] Shirky, C. (2005) AIHT: Conceptual Issues from Practical Tests. D-Lib Magazine December 2005. http://www.dlib.org/dlib/december05/shirky/12shirky.html [Feb 2006]

So, the conclusion was that a 'high and rigid bar for metadata production' was likely to be counterproductive. As far as errors on ingest were concerned, the project found that even a small percentage of exceptions in operations on a large archive, such as the incorrect reading of MIME types, can create a large problem. It may be necessary to 'apply human effort to those exceptions in which fixing one error saves either a large number of files or files of great significance; and be prepared to declare some number of files beyond the current economic threshold of preservation.'

One option would be to store these files without actively preserving them, which will essentially delay the problem of preservation, though new technologies may be created in the future that allow more effective preservation methods.

Whilst Cedars recommends byte-stream preservation, the Archive Ingest and Handling Test report states that bit storage is not a solution. There are many examples of perfectly stored but difficult to read bits. The AIHT report classifies this as 'playback drift', described as 'the tendency of a fixed set of binary data to stop functioning or being interpreted in the expected or hoped for manner, because of the complex ecosystem of applications, operating systems, and hardware changes, even though the data may be stored perfectly over decades.'

The overall conclusions of the report include the observation that employing a whole host of strategies for preservation is the best defence against systematic failure. It is better to think in terms of data-centric strategies for shared effort rather than tool-centric or environment-centric strategies. So, standardisation efforts can revolve around the data rather than the tools and processes.

Professor Harvey[109] cites a number of reasons for access problems:

- degradation of the media
- loss of functionality of access devices
- loss of manipulation capabilities
- loss of presentation capabilities
- weak links in the documentation chain
- loss of contextual information (the need for Representation Information)

Research is still being carried out on the benefits of migration versus emulation, and in terms of preserving the functionality as much as possible, it may be concluded that emulation is the best option, preserving the software that was originally used to create and access the resource.

109 Harvey, Professor R., Now You See It, Now You Don't: Maintaining Digital Learning Objects for the Future. http://athene.riv.csu.edu.au/~dharvey/ [January 2006]

15.1 Preserving Complex Objects

Static self-contained objects present fewer issues for preservation, but with complex objects such as simulations and visualisation tools, there are a number of challenges in terms of what to preserve and how to preserve the functionality, look and feel and utility.

Learning objects may be static, dynamic (dependent upon data that might have variable instantiations) and multimedia (providing audio and visual experiences). Whilst persistence is less of a problem for static objects, it is more problematic for dynamic and multimedia objects.

An article by Harvey[110] provides a useful example of genuine learning material used for audiovisual archiving. The following table summarises the information, software used and software required to access the information:

Type of information	**Software used to create or edit**	**Software required to access**
Text	Word 2000	Word (recent version)
Videoclips	Video editing software	Quicktime
Soundclips	Microsoft Sound Recorder	WAV file player or MP3 player
Images	Picture editing software	Web browser (e.g. Explorer 5)
Web sites	HTML editor	Web browser (e.g. Explorer 5)
Web forum	Various, e.g. database software, text editors, etc.	Web browser (e.g. Explorer 5)

Table taken from Professor Harvey's article

This is a relatively uncomplex example, but it still requires the application of some complex strategies and practices. Preservation has to be approached by categorising information and then having strategies for preserving each category.

16. Formats

It may be that a practical approach will necessitate selecting a limited number of the most common formats to preserve. Mandating the use of a limited number of file formats for the actual creation of digital materials would facilitate management and preservation. The US state of Massachusetts, for example, has made the decision that all the documents

[110] Harvey, Professor R. (2002). Now You See It, Now You Don't: Maintaining Digital Learning Objects for the Future. E-Journal of Instructional Science and Technology, vol.9 no.2. http://www.usq.edu.au/electpub/e-jist/docs/Vol5_No2/Harvey%20-%20Final.pdf [January 2006]

created by its employees must be in a data format called OpenDocument, which is an open format.[111] However, learning materials will continue to be deposited in Jorum in a whole range of formats, but a restricted set of formats could be used within the archive. This may mean that certain objects are under greater risk of being lost because they are not in a preferred format on deposit into Jorum, and the migration process will be too resource intensive, but with limited resources it simply may not be possible to prevent this.

'There is an extraordinarily large range of potential file formats and versions which may be submitted for sustainable storage in a repository. Pragmatically, it is probably unavoidable for any repository...to make a decision about supporting a finite range of those formats'.[112]

A preservation policy for learning objects could incorporate a periodic review of formats. Criteria can be developed for the treatment of each format, in terms of a migration strategy. The archive should consider the implications of providing long-term access to each format and ensuring that the intellectual content remains meaningful:

'Potential criteria for assessment include the status of the file format specification, the extent to which the internal structure of the format is known, the availability of third party conversion tools to migrate the format, and existence of free viewers on different operating systems.'[113]

The conclusion is that open, portable formats such as XML are to be preferred to proprietary formats.

It may be useful to identify file formats at particular risk in terms of access, and combine this with an appraisal process that encompasses other criteria, in terms of the actual content of the learning materials, quality of the metadata, level of granularity, perceived potential for reuse or historic value.

Harvard University Library has been awarded a Mellon grant to develop the Global Digital Format Registry[114] to provide sustainable distributed services to store, discover, and deliver Representation Information about digital formats. This will run for two years from February 2006. In terms of the Open Archival Information System (OAIS) Reference Model, the format typing of a digital object is Representation Information about that object; that is, it provides 'information that maps the Data Object into more meaningful concepts.' The format of a digital object must be known in order to interpret the information content of that object properly. Without knowledge of its format, a digital

[111] Times Online, 18 October 2005 http://business.timesonline.co.uk/article/0,,9075-1831039,00.html [January 2006]
[112] Bradley, K (2005) APSR Sustainability Issues Discussion Paper. http://www.apsr.edu.au/documents/APSR_Sustainability_Issues_Paper.pdf [April 2006]
[113] G. Knight, AHDS, Report on Preservation Standards, for SHERPA project. http://www.sherpa.ac.uk/documents/D4-5_Report_on_Preservation_Standards.pdf [January 2006]
[114] http://hul.harvard.edu/gdfr/about.html [January 2006]

object is merely a collection of undifferentiated bits. Thus, format typing is fundamental to the effective use, interchange, and preservation of all digitally-encoded content.

DROID (Digital Record Object Identification) is a software tool developed by The National Archives to perform automated batch identification of file formats.[115] It has a public API for ease of integration with other systems. Jorum does already automatically detect the technical format of any uploaded objects, and this is added to the metadata, but the problem is that objects are often deposited as zip files, and therefore the formats within the zip file will not be identified. The requirement is to be able to identify all formats within a content package. Jorum may benefit from using JHOVE[116], the object validation environment, as the validation of formats is equally important as identification. The format may purportedly be Microsoft Word (doc), Macromedia Flash (swf) or TIFF, but does it actually comply with the specification for the format? Furthermore, JHOVE performs characterization, which determines format-specific significant properties of objects. These properties are known as the object's Representation Information:

' The standard Representation Information reported by JHOVE includes: file pathname or URI, last modification date, byte size, format, format version, MIME type, format profiles, and optionally, CRC32, MD5, and SHA-1 checksums'.[117]

JHOVE could be integrated into the workflow as part of the Submission Information Package creation and ingest into an archive.

17. Key Definitions

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.”

Digital preservation involves the application of technologies and methods to ensure that information remains accessible and meaningful in the long term.

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

[115] http://www.nationalarchives.gov.uk/aboutapps/pronom/droid.htm [May 2006]
[116] http://hul.harvard.edu/jhove/ [May 2006]
[117] http://hul.harvard.edu/jhove/ [February 2006]

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."[118]

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.

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".[119]

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.

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.

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.

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.

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