The Cedars Project Report

April 1998 – March 2001

Prepared By
The Cedars Project Team

June 2001

Executive Summary

The past two decades have seen a massive increase in the range and volume of digital information resources, and their acquisition by libraries and research repositories. As yet in the UK there is no formal mechanism for the long-term preservation and continued accessibility of this material. This is a growing concern for the research community because the future of scholarship will rely on our ability not only to acquire and provide access to a wide variety of digital content but also to preserve it into the future through changing technological regimes. The Cedars project was established by the Consortium of University Research Libraries and was funded under phase three of the JISC Electronic Libraries Programme (eLib). Over the past three years from April 1998 – March 2001 Cedars has conducted a focused investigation into the issues surrounding digital preservation and the responsibilities that research libraries would have to assume to ensure continued accessibility to digital materials. The Project work has included a practical component to establish a prototype or "demonstrator" digital archive distributed across the Cedars partner sites – the universities of Oxford, Cambridge and Leeds. Ultimately the purpose of this technical work was to inform the development of both practical and strategic guidance for the HE community on how best to preserve digital resources.

The Cedars Project demonstrator archive was developed based on the Open Archival Information Systems (OAIS) reference model and many of the recommendations found in this report are informed by work with this emerging ISO standard. The OAIS model presents a useful approach for the establishment of digital archives – particularly in a distributed environment – and also describes a standard vocabulary. This report provides an overview of the basic functions necessary for a digital archiving service and this is largely based on the OAIS reference model and the Cedars Project experience with implementing it. (Chapter 2.)

An overview of the Cedars demonstrator archive forms a main part of this document (Chapter 3). The report explains how it was envisaged and implemented – but importantly also includes discussion of the many organisational and strategic issues that have been raised by attempting to establish a distributed digital archive. The Cedars demonstrator archive includes a wide variety of digital materials and the specific technical approaches to digital preservation employed by Cedars have been tested on both current and older digital materials already under threat. Case studies of these materials are included within the report. Based on the practical work of the demonstrator archive and other testing, the Cedars project has been able to provide some basic recommendations on the use of migration and emulation as strategies for providing continuing access to digital materials over time. (Chapter 5.)

As is clear from work done in the Cedars Project, the challenges presented for digital preservation are not limited to technical concerns and problems. Even the technical issues cannot be fully understood without some reference to collection management policies, skills requirements, intellectual property rights or costs. Although the technical issues remain key ones, many of the organisational and management issues are as important and often equally (if not more) complex. The preservation of digital materials will impact on many areas of an organisation, including finances, staffing, selection of material, and intellectual property rights. The Cedars project has explored many of these issues and produced guidance and recommendations. (Chapter 4 and Chapter 6.)

The Cedars Project has provided the academic community with a critical opportunity to explore the long-term preservation of digital materials within the context of research libraries, who will be key stakeholders in taking this important work forward. The Project has completed its first three years and has moved into a fourth and final year. This report documents the first three years from April 1998 to March 2001. The final year of the project will focus on taking the learning and outcomes described in this report and disseminating them to the wider HE community and beyond. The final year will be complete at the end March 2002.

Contents

Chapter One : Introduction and Project Overview

1.1 Introduction

1.1.1 Purpose of this Document

The purpose of this document is to provide a general framework into which the Cedars project findings and results can be incorporated. To this end the report will provide:

- A broad structure for presenting the project findings and results in an organised and contextualised manner;
- A general framework including vocabulary and concepts for understanding digital preservation and its implications for information access into the future;
- An exemplar of a digital archive based on the Open Archival Information System (OAIS) reference model developed by the International Standards Organisation (ISO).

It is expected that the framework of this report will allow the results of the project to be disseminated both as a full report as well as using a series of web pages with links to the relevant deliverables. Most project deliverables and outcomes exist as entities in their own right – such as guidelines and recommendations, reviews and working papers. However, as a whole, the results will sit within this framework for delivery to JISC (through this and the final report) and to the wider community (through this report and via the Web).

1.1.2 Scope and Applicability

The Cedars Project has focused its work on the long-term preservation of digital materials. Throughout this report, the focus will remain on this long-term function. In this context long-term means, broadly speaking, long enough to be concerned with changing technologies. Given the pace of changing technology in some cases this may not be very long at all but the Cedars Project has chosen to focus on a timetable which accommodates the short-term as well as decades and even centuries. It is important that readers understand the distinction between long-term preservation and "continuing access", and hence avoid blurring these two different functions.

Although continuing access is a key part of a digital archive, mechanisms for providing this for archived materials are likely to change through time. As a safety net digital preservation *always* involves the long-term preservation of the original byte-stream as it was deposited into the archive. Disseminating a copy of the archived digital object and some mechanism(s) for providing access to its content are a critical component of any digital archive but do not, in themselves, ensure effective long-term preservation of the object.

The guidance and recommendations included in this report may be applicable to any information organisation concerned with long-term preservation and continuing access to digital materials. They are specifically designed for academic or research organisations, in particular libraries and archives. The report does not assume a certain level of technical knowledge but does assume a general understanding of the application of technology to information services and some general awareness of digital preservation issues.

In addition the Cedars outcomes may be of interest to publishers and information producers/suppliers who create information that may need long-term preservation. This is particularly relevant in the light of ongoing work in the development of legislation for the deposit of non-print materials in national libraries and other designated depositories.

1.2 The Cedars Project Overview

1.2.1 Project Background

In early 1998, the Cedars Project began as part of phase 3 of the JISC[1] Electronic Libraries Programme (eLib). As part of eLib Phase 3 Cedars sat alongside other digital library projects all of which focused on hybrid library developments or on the implementation of the Z39.50 standard – Cedars was the only project funded under the "Digital Preservation" strand of eLib Phase 3.

Cedars is led by the Consortium of University Research Libraries (CURL) and is managed as a partnership between three CURL institutions – the universities of Cambridge, Leeds and Oxford. Its main broad objective has been to explore issues relating to the long-term preservation and continued accessibility of digital materials for research libraries. Cedars focused primarily on digital resources that are typically held by research libraries and used sample digital library content as part of a demonstrator digital archive distributed across the three partner sites. The sample content was intended to represent the range of digital materials that might be found in this type of academic library. The production of a demonstrator archive on a very practical level was intended to inform the development of both practical and strategic guidance of value to CURL and its membership, the wider HE community and beyond. The Project has now completed its first 3 years of JISC funding and its outcomes and recommendations will be described and discussed in this report. Cedars has now begun a fourth and final year of JISC funding which will focus primarily on thorough dissemination of the project's results.

1.2.2 The Consortium of University Research Libraries and Digital Preservation

Digital preservation is a key issue for CURL and all its members, who between them hold a significant proportion of the nation's research collections. CURL libraries, whose members include the national and other legal deposit libraries, have long assumed curatorial responsibility for the long-term preservation of research resources in traditional paper formats, to assure continued access to their intellectual content for the benefit of the wider academic community. They have a similar commitment to the preservation of digital resources. Once a framework for legal deposit of electronic publications has been established, it will be libraries such as those in CURL who will have an obligation to deal with their long-term preservation. CURL is committed to reaching understanding and building expertise in the issues involved in long-term preservation and continued accessibility of digital content. Such work will be of benefit to the whole academic community and remains critical to the future of scholarship.

1.2.3 Project Methodology

The Cedars Project was established to explore strategic issues and although it included an important practical component, it was primarily a research or exploratory project. The basic project methodology was as follows:

- rapid prototyping of a 'proof of concept' or demonstrator distributed digital archiving system;
- testing of that system across a range of different organisations to assess usability and scaleability issues;
- and finally, synthesis of the findings from this work into guidance and recommendations for the HE community about how best to preserve digital scholarly materials.

Although the original project proposal suggested that work at each partner site focus on a specific type or class of digital materials, it was agreed early in the planning stages that a programme of work should not emphasise the separate partner

[1] Joint Information Systems Committee of the Higher Education Funding Councils for England, Scotland and Wales and the Department for Education in Northern Ireland.

sites but attempt to focus on the project as a whole – the planning team felt that as originally envisaged, the project could evolve into three separate projects rather than one cohesive programme of work. The Cedars demonstrator archive was therefore organised as a single coherent demonstrator archive system distributed across the partner sites. The project partners agreed that this would provide a useful “proof of concept” for a distributed archive or a federation of archives. This notion was motivated by a shared vision across the partners that a strategy for digital preservation in the UK on a national level would need to be based on a distributed model and would necessarily involve a number of different key stakeholders including any number of research libraries. Over the course of the project’s first three years this assumption has been confirmed by work done in the area not only by Cedars but by other organisations as well.

Based on the above methodology, the organisation of the project[2] involved the initial establishment of three working groups – each consisting of members from the partner sites and focused on a particular issue which was to be addressed at both strategic and practical levels. These groups were established for a fixed term and ran for the first year of the project while planning was completed and staff appointed. The groups were as follows:

Content Issues
The Content Issues Working Group focused on the practicalities of obtaining rights clearance for materials to use in the demonstrator archive. This involved overseeing the negotiations between the copyright holders and the project. Cedars status as a short term pilot project and the fact that testing the Cedars demonstrator archive did not involve public access to materials meant that most rights holders were happy to provide content to the project. The project received a wide range of digital materials including online databases, digitised materials electronic journals and other multimedia materials. In addition to rights clearance, the CIWG also focused on strategic issues associated with how materials could be preserved without violating intellectual property rights and included a number of representatives from the publishing industry.

Access Issues
The Access Issues Working Group was established to focus on a specific technical issue surrounding digital preservation – metadata. In conjunction with UKOLN, this working group focused on a practical level on what metadata would be necessary for materials within the demonstrator and what formalisms might be appropriate for including the metadata in the demonstrator archive. This led to the development of an outline metadata specification for digital preservation which had important strategic ramifications on an international level. It was one of the first (and most comprehensive) documents of its time.

Preservation Strategies
A third Cedars working group was established and its main focus was on the technical issues related to developing a distributed archive. The Preservation Strategies Working Group focused on both the design of a distributed architecture for the demonstrator as well as on its implementation.

After the first year of the project, most of the initial planning was complete and *the project team* was in place led by the Project Manager. The project team was then responsible for taking the main project work forward. For the final two years of the project, the project team then took the recommendations and work done by the working groups and focused on meeting the project’s main objectives and deliverables.

1.2.4 Project Objectives

In its first three years, the Cedars project aimed to investigate issues relating to the long-term preservation and continued accessibility of digital information resources typically included in library collections.

It order to achieve this aim the project had three main objectives:

[2]For a full description of the Project’s management structure please see the Cedars Project Plan (July 1999). Available at http://www.leeds.ac.uk/cedars

- promote awareness about the importance of digital preservation, both among university libraries and their users, and among the data creating and data supplying communities upon which they depend;
- identify, document and disseminate strategic frameworks within which individual libraries can develop collection management policies which are appropriate to their needs and which can guide the necessary decision-making to safeguard the long-term viability of any digital resources which are included in their collections;
- investigate, document and promote methods appropriate to the long-term preservation of different classes of digital resources typically included in library collections, and to develop costed and scaleable models.

There is an enormous range of digital resources (e.g. text, sound, pictures, moving images). In focusing on the following categories, the project aimed to identify issues and make recommendations which could be generalised and extended to the full range of digital materials:

- ▷ digitised primary resources
- ▷ electronic journals
- ▷ large online databases
- ▷ electronic ephemera
- ▷ digital resources in which the intellectual content is bound to structure, form and behaviour (e.g. multi-media CD ROMS)

1.2.5 Project Deliverables

To meet the above broad objectives, the Project specified a number of key deliverables including:

- a demonstrator project to test and promote the technical and organisational feasibility of a chosen strategy for digital preservation (based on the Open Archival Information Systems reference model);
- guidelines for developing collection management policies;
- guidelines developed by the demonstrator project on how to preserve different types of digital resource;
- focused investigation of the metadata necessary for long-term preservation[3];
- clearly articulated preferences about data formats etc. which are most readily and cost-effectively preserved;
- publications of benefit to the whole higher education community, available on the WWW.

1.2.6 Project Outcomes

Cedars has met its original objectives through three key areas of work. The results from these main strands encompass the deliverables listed above as well as a number of other results that were not formally anticipated when the project was funded.

The key outcomes for the project over the past three years are:

[3] Initial work on preservation metadata was led by the UK Office for Library and Information Networking (UKOLN) and produced a survey of relevant work. Building on this work the project then worked closely with UKOLN on the development of an outline specification for digital preservation metadata

1.2.6.1 The Cedars Demonstrator Archive

The project has focused a great deal of effort on technical development and on the establishment of a working demonstrator archive. It was this practical work that was to be the building block for the project's guidance and recommendations. The demonstrator is based on an implementation of the Open Archival Information Systems (OAIS) reference model and is distributed across all three partner sites. Although the original project proposal did not refer to the OAIS model specifically, soon after the project began it became clear that work on OAIS was relevant to the project's plans for a demonstrator. Uptake of OAIS across the library and archives communities has been significant over the past three years and it continues to be a key standard in this area. The Cedars Project was one of the first to attempt to adopt OAIS and members of the Cedars team have used this experience to feed directly back into OAIS reference model as it has developed. OAIS is now undergoing the ISO process and its acceptance as a standard is anticipated later in 2001.

1.2.6.2 Metadata for Digital Preservation: the Cedars Project Outline Specification

As proposed in the original project bid UKOLN[4] has played a key role in developing the "Preservation Metadata: The Cedars Project Outline Specification"[5] which has proved to be one of the key deliverables of the Cedars project. This work has had both a practical focus (through the creation of a formalism using XML for the demonstrator project) as well as a high profile international status. This was one of the first documents of its kind and has been widely recognised as an important first step for the development of a standard for preservation metadata. As a result of Cedars work in this area, the project has been invited to participate in an international working group on Preservation Metadata (co-ordinated by RLG and OCLC).

1.2.6.3 Guidelines for Collection Managers

Dialogue (mainly through focus groups) with library and archive representatives proved early on that guidance was desperately needed if institutions were to implement a digital preservation strategy. The Cedars project has now produced a preliminary set of basic guidelines – another of the project's deliverables. This work also includes information on cost elements as well as guidance on intellectual property rights issues.

1.2.7 Project Resources

1.2.7.1 Staffing

The Cedars Project was overseen by a Project Director but managed by a Project Manager who was full time for the first 1.5 years and then .80 FTE for the remaining 1.5 years. The Project Manager was supported by three FTE Project Officers – one based at each of the lead sites – and these four posts comprised the complete project team. The Project Manager was based at Leeds.

The Cedars Project Team, at its full complement, consisted of the following people:

Project Director (April 1998 – present), Clare Jenkins, now Director of Information Services, Imperial College
Project Manager (April 1998–March 2001), Kelly Russell, University of Leeds
Cedars Project Officer ((May 1999 – present), Ellis Weinberger, University of Cambridge
Cedars Project Officer (April 1999 – present), Derek Sergeant, University of Leeds
Cedars Project Officer (November 1998 – November 2000), Andy Stone, University of Oxford

[4] The UK Office for Library and Information Networking, based at the University of Bath. See http://www.ukoln.ac.uk
[5] The Cedars Project Team. June 2000. See appendix

Overall direction of the Project was determined by a broad-based Advisory Board that met bi-annually and directed on a more detailed level by a project-wide Management Group that met quarterly. Membership and terms of reference for both the Cedars Advisory Board and the Cedars Management Group can be found in the appendices.

1.2.7.2 Funding

The Cedars Project was funded under the eLib Phase 3 programme which placed a strong emphasis on matching JISC funding with some institutional funding from project partners. Over 3 years Cedars was granted 377K of JISC funding. This included funding for staff, equipment, travel, evaluation and dissemination. As is the case with all JISC funding, this does not cover office costs or "overheads" which are an expected contribution from the project partners. The project budget for JISC funds was divided as follows:

Total staff costs (over 3 years):		
Project Manager	£	100,000
Other project staff (5 person/years fte at AR1/2)	£	143,000
External specialist consultancy, as required	£	15,000
Technical requirements		
Hardware		
3 well-configured networked PCs for Project staff	£	6,000
Other miscellaneous equipment	£	5,000
Software		
Miscellaneous standard PC software	£	1,000
Special-purpose software for testing	£	5,000
Other costs		
Dissemination/conference costs	£	25,000
Contingency (toward conference)	£	15,000
Travel and subsistence	£	22,000
Evaluation	£	40,000
Total JISC funding	£	**377,000**

The Cedars partner sites all committed a significant proportion of institutional support for the project. This was to be in the form of office costs as well as technical expertise from local staff and storage space on the local archiving systems. Institutional contribution to the project was divided across all three partners as follows:

2 person years (technical expertise)	£	56,000
Other staff contribution (e.g. focus groups, advisory board, CURL contribution, etc.)	£	15,000
Access to large file storage on each site	£	60,000
Office costs	£	90,000
Total institutional contribution	£	**221,000**
Total Project Resources	£	**598,000**

1.2.7.3 Institutional Contributions

Although many funded projects promise institutional contributions, it is well known that this does not always translate into real resource contributions over the course of the project. During the course of the three project years, the Cedars Project kept an estimate of the time and resources each partner contributed to the project in the form of staff time and office costs (it was more difficult to estimate value for money in terms of access to the large scale file storage but it was certainly provided). At the end of the first year, the three project partners and CURL had contributed approximately 30K of local staff time and, although local staff contributions were reduced somewhat as the project team was established in the second and third years, the partners had still exceeded their promised contribution of local expertise. By the end of the project's first three years the partners had exceeded their proposed contribution by an estimated 15%. The office costs or overheads for the salaried project staff (2 at Leeds, 1 at each Cambridge and Oxford) were also slightly more than the 90K estimated in the original project proposal.

1.3 Structure of the Report

Each chapter of this report focuses on a different aspect of the Cedars Project work. The report will be best understood when read as a single document because each chapter builds on the previous chapters.

Chapter One is this introduction.

Chapter Two provides an overview of a generic digital archive. This has been based on the Open Archival Information Systems reference model. This chapter will outline the key functions of an archive as defined by the OAIS model (making clear where the Cedars project has amended or enhanced the model for the purposes of the project) and provide some discussion of the key issues Cedars has addressed in the design of a demonstrator archive.

Chapter Three will then focus in detail on the Cedars Project implementation of OAIS and describe the Cedars Project Demonstrator Distributed Archive. As one of the Project's key deliverables, this "proof of concept" system has been developed and tested over the course of the past 3 years.

Chapter Four provides some analysis of the relevant organisational and management issues as they have been considered as part of the Cedars Project. This chapter will explore and provide recommendations on issues to do with collections management, staffing, and copyright.

Chapter Five is an overview of the technical specifications developed as part of Cedars and will focus primarily on the Cedars Project Outline Specification for Preservation Metadata. Although this was not originally envisaged a key deliverable of the project, it has become the cornerstone of the Project's demonstrator archive and continues to influence the development of metadata standards in this area. Chapter 5 will also cover work Cedars (and the CAMiLEON Project[6]) has done on specific technical approaches such as migration and emulation for providing continuing access to archived materials.

Chapter Six is devoted to a detailed discussion of costs and cost issues for digital preservation.

Each chapter will include a summary and references to further reading, as well as specific recommendations.

[6] The CAMiLEON Project is a "sister project" to Cedars. It is funded under the JISC/NSF International Digital Libraries Programme and involves a partnership of the University of Leeds and the University of Michigan. CAMiLEON focuses on the investigation of using emulation as a technique for providing long-term access to digital materials. CAMiLEON stands for Creative Archiving at Michigan and Leeds: Emulating the Old on the New, and is funded for three years from November 1999.

Chapter Two : The Functions of a Digital Archive

2.1 Introduction

As suggested in the previous chapter, this chapter provides an overview of the generic functions that constitute a digital archive. A more detailed practical discussion of these functions will be given in Chapter three which describes the Cedars Project Demonstrator Distributed Archive.

The functions of a digital archive share many similarities with traditional archives. These functions, broadly speaking, describe how materials are submitted to the archive, how they are organised and managed once within the depository, and finally how continuing access to the archived materials is provided. The Cedars Project has adopted a reference model developed by the Consultative Committee for Space Data Systems called the Open Archival Information Systems model[7] (hereafter called OAIS). The OAIS model has become widely accepted by the library and archives communities both within the space data community and outside of it. The OAIS model can be applied both to traditional or digital archives. Cedars only applies it in a digital context. The OAIS model describes an archive as "consisting of an organisation of people and systems, that has accepted the responsibility to preserve information and make it available..."[8]

Specifically the OAIS model has provided the Cedars project with a welcome set of well articulated concepts and a comprehensive vocabulary. These have allowed the project to communicate across the partner sites (and across disparate technical backgrounds) and discuss the implementation of a demonstrator archive.

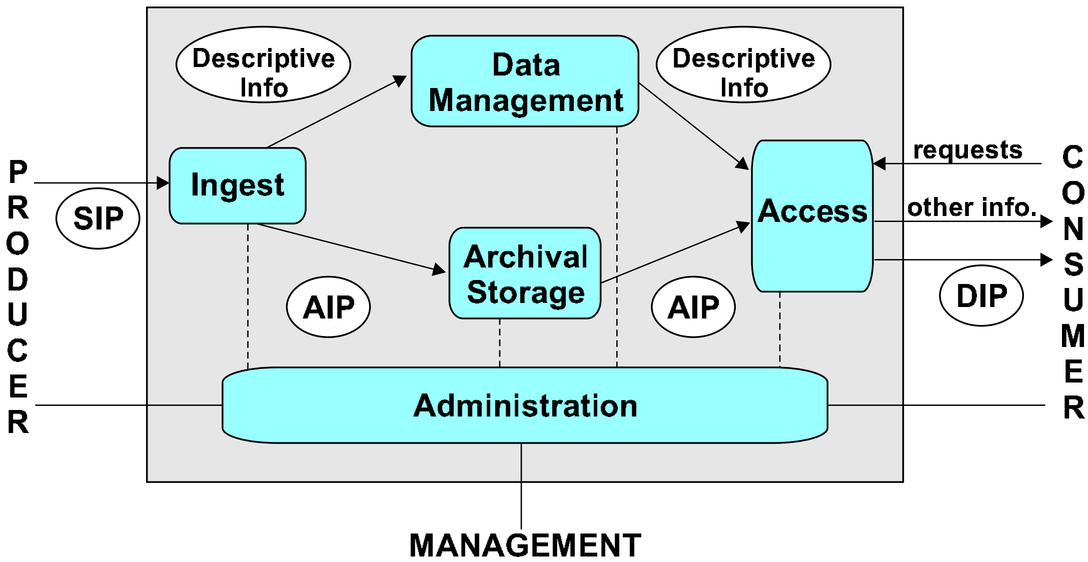


Figure 2.1: Overview of functional entities in the OAIS reference model

Figure 2.1[9] represents the functions described in the OAIS reference model. The main functional entities within the OAIS archive are **Ingest**, **Archival Storage**, **Data Management**, **Archive Administration**, and **Access**. Figure 2.2 expands on figure 2.1, based on the Cedars view of its demonstrator archive implementation.

The **Information Package** forms the basic unit of currency of an OAIS digital archive. This includes both a digital

[7]Consultative Committee on Space Data Systems. *Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK*. May 1999

[8]Ibid.

[9]Ibid. OAIS figure 4-1, OAIS Functional Entities

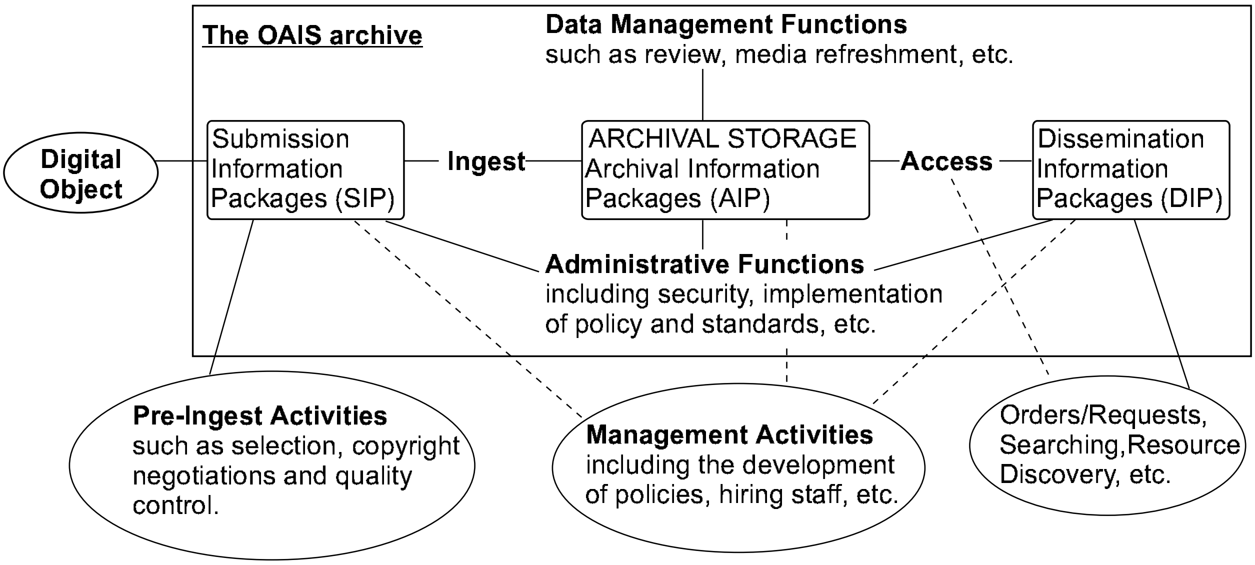


Figure 2.2: Overview of functional entities in the OAIS reference model

object and the necessary associated metadata. There are different types of Information Packages: as objects are submitted to the archive they arrive as **Submission Information Packages** (SIP), containing the primary digital object and any other information (if available) supplied by the content provider; the **Ingest** function encapsulates this information and other metadata as an **Archival Information Package** (AIP); when a copy of the object is provided by the **Access** function, a **Dissemination Information Package** (DIP) is created – this DIP may contain support systems to use the object as well as the digital object. **It is the AIP which is the cornerstone of the digital archive, as this is the storage unit for long-term preservation**. These functions are reviewed below, and discussed in further detail in chapter three.

2.2 Submission and Pre-Ingest

Before an archive can accept responsibility as a trusted archive service, there are a number of management tools that must be in place. The archive should be covered by an agreed and well documented collections policy document. Chapter four covers policy issues in some detail. Digital materials, in the sense of collections and collections development, are not always different than their non-digital counterparts – many of the same criteria will apply to both types of materials. However, for digital materials the following elements will be the most critical:

- Collections development strategies (including evaluation criteria for assessing potential submissions)
- Technical strategies for providing continuing access
- Collections development procedures (including review procedures pertaining to retention and de-accessioning of materials)

Where appropriate the archive will need to ensure availability of copyright and intellectual property rights information including licences, schedules for deposit (where regular updates will be forthcoming), appropriate documentation, and may even include details of specific preferred formats and media. (For further information on copyright and intellectual property rights see chapter four.)

As part of the **Pre-ingest** activities the archive will also need to:

- Check against any existing deposit schedules to ensure everything expected has been received
- Assign unique identifier(s)[10] to the digital object[11]
- Virus check and validate the integrity of the digital object
- Assess in detail the **Significant Properties** of the digital object that the archive (and/or content provider) would like to preserve
- Validate, improve, and create the documentation
- Where appropriate, reformat the digital object according to existing archive policies
- Ensure all necessary metadata for preservation and continuing access accompanies the object (or has been identified)

Recommendation 2.1: An organisation assuming preservation responsibility for digital materials should have a well documented policy covering all Pre-ingest and Ingest criteria.

2.2.1 Assessing an Object's Significant Properties

The archive will need to make decisions about what level of preservation is appropriate for each digital object (or each class of objects). This involves assessing which properties of a particular digital object are regarded as significant. These decisions influence the levels and methods of access that will be possible for the object, and the level of preservation metadata required for long-term retention. For traditional materials access and preservation were mainly the same thing and so were generally handled by the same organisation. If an archive was custodian to a set of papers they would continue to be readable and therefore continue to be useable. However for digital materials simply maintaining a bytestream does not ensure the digital material will be preserved in a manner that continues to be useable. For digital materials *Access* in the technical sense can be at a variety of levels, the level at which it is maintained will depend on value judgements made by the archivist and/or collection manager. In Cedars, the level of content and functionality retained are referred to as the digital object's **Significant Properties**. Determining the Significant Properties will also dictate the amount of **Metadata** (including **Representation Information**) that must be stored alongside the **Primary Digital Object**, as these ensure that the object is accessible to the agreed level. The Significant Properties therefore determine the underlying technical components that need to be documented and supported to ensure preservation of those Significant Properties. The Cedars project refers to this as the **Underlying Abstract Form**[12] (UAF) of the object.

A digital object's Significant Properties are not empirical; archives will make judgements at levels appropriate to fulfil their preservation responsibilities and meet the needs of the archive's user communities. Decisions about an object's Significant Properties should be reflected in the archive's collection management policies. For example, materials deemed to be part of the collection's core might retain all of the original functionality where other materials deemed as peripheral will not include the full complement of these as Significant Properties. For Cedars, the creation and maintenance of the detailed metadata associated with the object's Significant Properties is the backbone of an archive's preservation function.[13]

Significant Properties: A simple example An archive takes deposit of a PDF electronic journal. The judgement is made that the text within this journal is the only Significant Property. There is no need to store information about

[10] The assignment of unique identifiers is part of both **Pre-ingest** and formal **Ingest**. Cedars assigns a unique identifier as soon as negotiations for submission begin, so all objects are given identifiers at Pre-ingest. At Ingest the object is assigned a unique identifer as part of the AIP, which may or may not be the same as the previous identifier

[11] Where physical objects are received, including documentation and digital media, these must be labelled and archived until their retention is proven to be unnecessary

[12] See Chapter three for further information on UAFs

[13] The Cedar's metadata framework, *Metadata for Digital Preservation: The Cedars Outline Specification* (Appendix 5) is receiving international recognition, and will contribute significantly to the development of standards in this area

the PDF environment, but only to include information sufficient to retrieve (or render) the ASCII text from the journal. (This results in a loss of layout and formatting, but these are not deemed “significant”.)

Significant Properties: A more complex example An electronic journal which is published via the web as HTML. The Significant Properties are deemed to include the hypertext links (maintaining the web journals internal structure) as well as the multimedia functions (such as support of sound and video clips). Preservation must capture this (full) level of functionality. Although end-users (or ‘journal subscribers’) currently read the journal in HTML, these pages are actually created on the fly from SGML. The archive chooses to preserve the SGML files, and so the Representation Information of the AIP includes robust technical descriptions of the digital object including information about software to run the video as well as how to retrieve the marked up text and images.

The detailed descriptions and the technical systems necessary for rendering the Primary Digital Object ensures that long-term preservation supports continuing access. As stressed earlier, providing continuing access should be kept conceptually separate from the preservation function. Users of the archive’s contents will not have to be aware of either the workings of the preservation function or of the **Representation Network** that supports Access to the Primary Digital Object’s Significant Properties. Representation Networks are discussed in detail in Chapter three.

Within an archive, decisions taken about an object’s Significant Properties, and hence the Underlying Abstract Form, underpin the long-term preservation function. The creation of Representation Information to support the UAF is a critical component of the archive. However, the skills required to do this for each object during the Ingest function are not likely to be covered in the human resources of most research libraries and traditional archives. By automating (to some degree) the assessment of Significant Properties and the creation of Representation Information, the efficiency of Ingest and the costs associated with it will improve.[14]

Recommendation 2.2: Further work should be undertaken in the assessment of Significant Properties and the development of Representation Networks.

2.3 Digital Preservation Metadata

In the OAIS model, the AIP is the cornerstone of the archive. Within the AIP the original Primary Digital Object is stored (as a bytestream) along with metadata necessary for making that bytestream into a meaningful (and useable) digital resource. In the future what is known about preserved digital materials will come from the information stored with them within the AIP. The Primary Digital Object is meaningless without some description of what it is and how it works. In OAIS there are two main types of metadata **Representation Information** and **Preservation Description Information** (PDI). The Representation Information supports the transformation of the Primary Digital Object into a usable (rendered) digital object. The PDI is all of the other supporting metadata that is deemed necessary for purposes of long-term preservation. Chapter five provides more detail on the Cedars project work with preservation metadata.

2.4 Ingest into the Archive

Ingest is described by the OAIS model as the functional entity that contains the services and functions to: accept **Submission Information Packages** (SIPs) from content providers; prepare **Archival Information Packages** for storage; and ensure the AIP is effectively established within the archive. As Ingest is the point of acquisition it is at this stage that effective long-term preservation becomes manifest.

[14] During testing, the Cedars demonstrator archive hand-constructed a Representation Network for the objects received (see Chapter three). More work is needed on automated analysis of digital objects in order to move towards a service environment

2.4.1 Ingest Concepts

Cedars has developed the concepts of the Ingest function, and then tested these practically. Chapter three describes Cedar's work with **Test Sites** that examined the technical processes involved in Ingest. These are the steps which were taken at Ingest to both store a digital object and to enable the use of its intellectual content in the future:

1. The unique identifier assigned at Pre-ingest is validated, and then assigned to the Ingest object or, if necessary, a new identifier is assigned. The identifier for the object must be entirely unique.
2. Based on the Significant Properties of the object, a suitable Underlying Abstract Form is chosen.
3. The object is converted into the bytestream (the Primary Digital Object).
4. Use the Significant Properties and UAF to determine which Representation Information (and which Representation Networks) are needed in order to enable the transformation of the Primary Digital Object back into an object a scholar can view.

> For example, the object is a PDF file with layout and embedded multimedia gizmos as Significant Properties and so it is preserved as a PDF file. The UAF of the object should be a PDF file.
>
> The Representation Network in this case will need a **Render-Analyse-Convert Engine** which can turn the stored bytestream back into a PDF file, and a Render-Analyse-Convert Engine which can enable the PDF file to be viewed. The Representation Network will also need to know which platforms the Render-Analyse-Convert Engines need.

5. PDI metadata needs to be created to provide basic contextualisation, provenance, and catalogue description of the Primary Digital Object.

The activities above are not strictly ordered, and in a functional service archive more administration activities would need to be added.

2.5 Archival Storage and Maintenance

OAIS describes this as "the services and functions for the storage, maintenance and retrieval of Archival Information Packages."[15] Such functions include adding AIPs from Ingest into permanent storage, managing the storage hierarchy, refreshing the storage media, and allowing AIPs to be disseminated from the archive. After completing the Ingest function, the AIP is built out of the Primary Digital Object, the Representation Information, and the Preservation Description Information.

2.5.1 Storage

To successfully place an AIP in permanent storage there are several factors to consider. These should be laid down in a policy document, which is placed under regular review. The major consideration is whether to have an in-house archival store, or whether to use a third party archival store. If the latter is chosen a detailed service level agreement needs to be

[15] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK

drawn up, and subsequently the vendor needs to be monitored to ensure that the service level is, indeed, being delivered. The criterion which the **Archival Store** has to meet is: it will store a named bytestream in such a way that at any given point in the future it can return an identical bytestream. These are two different models from the OAIS for an Archival Store:

- A single, central repository
- A federation of repositories

Where a federation is adopted, each member can either use their own Archival Store or bring in a third party vendor.

Once a bytestream has been handed to the Archival Store it should be verified and validated. Mechanisms also need to be in place to judge the authenticity of the original digital object. Once in the Archival Store, the AIP's integrity should be regularly checked (see 2.5.3 below).

Recommendation 2.3: An archive's policy for Archival Storage should provide for integrity safeguards.

2.5.2 Archival Storage and Preservation Metadata

A decision should be taken about how much metadata needs to be stored in the AIP. When the metadata is needed for as long as the Primary Digital Object it should be placed into the Archival Store. As such, some Representation Information always needs to be stored in the AIP. Some PDI also needs to be retained for the lifetime of the AIP, for without this the AIP becomes anonymous – and so users would no longer be able to discover whether the AIP was of use to them or not.

One question arises relating to storage of the AIP. When, and how tightly do you bind metadata to the Primary Digital Object? Phrased another way, should all components of the AIP be stored in one logical place? This is a policy decision, and Cedars chooses to store all of the AIP packed as a single bytestream (so all components are in the same place). Storing the components together has two definite advantages: management of the AIP only has to deal with one storage object; and if all of the archive's infrastructure is lost, the storage object can still be understood (to some degree) as 'standalone'.

To aid the 'standalone' understandability Cedars also chooses to use human readable formats for most parts of the PDI. It is also prudent to record the AIP's unique identifier as part of the PDI.

Recommendation 2.4: An AIP can operate as 'standalone' if human readable formats are used for some metadata fields.

Of course, there may be metadata elements related to an AIP (but not necessarily part of the PDI or RI) where it makes sense to manage and store these separately. An example of one such metadata element is 'access statistics'. These could be managed in a separate database, and never be stored as part of the AIP. Three factors which may also influence which components are stored in the Archive are:

- costs, where you may only want to store a minimal textual description in the Representation Network rather than storing complete software packages;
- intellectual property rights, you may not be able to store third-party software or existing catalogue entries;
- value – a judgement as to what is useful to the future user, and what is not.

One strategy to managing "dynamic" metadata sets is to store a snapshot of this with the AIP in the long-term **Archival Store** and keep a database that maintains the current set. Periodically a new AIP is placed into the Archival Store which snapshots the current set, and in this new AIP a record of the primary AIP's unique identifier is kept.

2.5.3 Refresh and Disaster Recovery

Part of the Archival Storage policy should include the provision for maintenance of the Archival Store, and for recovery from major disaster. Whichever medium the archive uses to hold the bytestream, it will have a predicted data retention longevity. Well within this boundary a rolling process of media refresh should take place, and regular data integrity checks should be used to confirm that the bytestreams are intact. To ensure the capability of disaster recovery all bytestreams need to be backed up on media held at a geographically separate location. In order to protect holdings from corruption, it is best to hold at least three copies of every resource. It is not essential to use the same medium for all the copies of a bytestream.

The Archival Store must also perform replication migration from ageing media[16] to modern media. This will be informed by the level of Data throughput and the Quality Control requirements of the archive.

Recommendation 2.5: The Archival Storage policy should include the provision for maintenance of the Archival Store and for the recovery from major disaster.

2.6 Dissemination and Access to Archived Materials

In the OAIS model, access to archived materials is done through a **Dissemination Information Package** (DIP) – that is, a copy of the Primary Digital Object is disseminated along with the necessary metadata (perhaps including software if necessary) to 'render' it usable. The DIP is different from the AIP because firstly it may only contain part of the Primary Digital Object (or a partially processed/analysed derivative); and secondly, only contains metadata necessary for providing the appropriate level of access. Once again, this illustrates the clear separation between the Preservation function and providing continuing access to the digital materials.

One of the least constructive mantras used in relation to the preservation of digital material has been "you cannot have preservation without access." While it is true that there is no preservation of digital material without the possibility of meaningful retrieval, it is misleading to imply that Access to digital material is intrinsically linked to its preservation. There are important reasons (both legal and technical) why Access and Preservation should be considered separately. Publishers and other rights holders are often cautious about the preservation of their materials if they perceive that this is linked to providing unlimited public access. In the interests of actually preserving our scholarly heritage for future generations it is helpful to understand and discuss preservation and continuing access in the appropriate context. The distinction between continuing access and long-term digital preservation needs to be better understood to ensure interests in the former do not jeopardise the latter.

After a recent meeting in the UK the authors of the OAIS Reference Model are now planning to include a separate logical component, called the "Preservation Planning" function. Cedars welcomes this development, as it should emphasise planning for the allowance of dissemination as separate from the actual provision of dissemination.

2.6.1 Dissemination Information Packages - Examples

An Image File of the Surface of Mars For this still image, the AIP will contain a reference to detailed technical information about the file format and the systems necessary to render the picture in particular environments (e.g. the format definition of TIFF, and specifications about how to render this on a PC and on a Postscript printer). The AIP also contains much sensitive metadata describing the space programme which acquired the digital image. The user receives a DIP which simply contains a copy of the TIFF file, a public domain viewer, and PDI indicating the year the picture was taken.

A Database File A more complex example might involve a database where the Representation Information provides an

[16] Ageing incorporates a sense of technology ageing, as obsolete technology presents problems for reading its media

emulator of the original database environment. The end user needs a copy of the Primary Digital Object, a copy of the emulator, and instructions on how to use these. These would therefore form the DIP.

Dissemination of objects from the archive is covered in more depth in Chapter three.

2.7 Archive Administration

This function lies within the archive, and administers everything which is at one remove from the direct path of an object through the archive from Ingest through Archival Storage (plus the related data management) to Dissemination (Access). The OAIS reference model defines the tasks of "Administration" as follows.[17] Administration will:

- Control physical access to the archive's plant
- Manage system configuration (including auditing system operations, performance, and usage)
- Develop standards and policies
- Negotiate submission agreements
- Quality audit of Archive Information Packages once ingest is complete (the OAIS model implies this will often be contracted out)
- Interact with "Management"

To this list the Cedars Project has agreed to add the following activities:

- Review and maintain Representation Information (and Representation Networks)
- Review data types (formats) stored in the archive to ensure continuing access if needed
- Negotiate Access agreements with Service providers or others
- Liaise with other archives

2.8 Data Management

Data Management covers all aspects of an OAIS archive, and is essential for long-term preservation and day to day administration. It represents good record keeping of every stage described in this chapter. Data Management also includes the management of the overall holdings of the archive. Some of the components that the OAIS model[18] suggests including in Data Management are:

- Pricing information and availability constraints
- Tracking of user requests
- Security information
- Statistical information to improve archive operation

[17] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK
[18] Ibid.

- Customer profiles
- Accounting information

Cedars would augment the list to include:

- Records from Pre-ingest negotiations
- Monitoring the allocation of Unique Identifiers (and designing allocation policies)
- Maintaining records of holdings for use with finding aids

2.9 Recommendations

1. An organisation assuming preservation responsibility for digital materials should have a well documented policy covering all Pre-ingest and Ingest criteria.
2. Further work should be undertaken in the assessment of Significant Properties and the development of Representation Networks.
3. An archive's policy for Archival Storage should provide for integrity safeguards.
4. An AIP can operate as 'standalone' if human readable formats are used for some metadata fields.
5. The Archival Storage policy should include the provision for maintenance of the Archival Store and for the recovery from major disaster.

2.10 Further Reading

1. A blueprint for Representation Information in the OAIS model, David Holdsworth, Derek M. Sergeant, http://esdis-it.gsfc.nasa.gov/MSST/conf2000/PAPERS/D02PA.PDF
2. CCSDS 650.0-R-1: Reference Model for an Open Archival Information System (OAIS). Red Book. Issue 1. May 1999, http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html
3. Ingest Standards (and others) in the OAIS model, David Holdsworth, http://gps0.leeds.ac.uk/ ecldh/cedars/AWIICS.html
4. The Archive Ingest Process, Mike Martin, http://ssdoo.gsfc.nasa.gov/nost/isoas/awiics/ingestmethodology.html

Chapter Three : The Cedars Demonstrator Archive and Related Work

3.1 Introduction

While much can be learned from theories and discussions about long-term digital preservation, there is a deficit of practical implementations to work through and test these. Without testing theories in a practical way, their weaknesses and short-comings are never revealed. In this field of digital preservation there is always the sobering reminder that, if valuable digital objects are not preserved soon, or even now, it may not be possible to preserve them at all. Fortunately, even when a resource is preserved in a bad manner, it can be detected and improved steps can be applied to ensure the longevity of the resource (this is not possible where no preservation steps have been taken). However preservation of digital materials is an urgent issue because of the speed with which new technologies become obsolete. The first lesson learned from practical work is this: even when a resource is preserved, this is not fail safe, so the resource needs monitoring in order to ensure the preservation is effective for the long-term. Cedars has built some computer systems, mainly using web technology (Java, HTML, cgi), to test several of the OAIS stages of an archive. These are combined in a distributed architecture, and have been tested by a group of six test-site institutions. Part of the Cedars project's investigation has used material from an already obsolete environment in order to understand how the proposed solutions can reach forward into the future.

3.2 Cedars Digital Architecture

Cedars has designed an architectural model for a distributed archive. This architectural model is categorised as a "federated archive" by the OAIS model. A federated archive "includes both Local and Global communities, and has both Local and Global access."[19] As the practical work of Cedars was based at three geographically separated sites, the distributed architecture consists of three archives. Figure 3.1 shows the local systems in place for one of these sites.

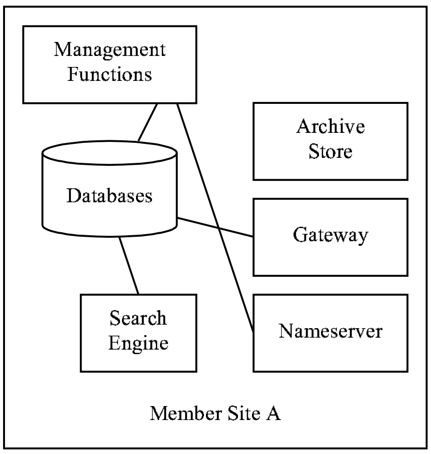


Figure 3.1: The Architecture of a Single Site

The criteria for this architecture is to

- Ensure that a preserved digital object can be found

[19] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK

- Ensure that the found object can be understood

For these criteria to be met, it is vital that each preserved digital object has a unique identifier.

3.2.1 Unique Identifiers and the Nameserver

All of the digital objects archived by Cedars are allocated a **CRID** (Cedars Reference ID). Early in the project, the decision was taken to adopt the (then emerging) Unique Resource Name (URN) convention for all of the CRIDs.[20] The URN convention imposes a colon separated structure for domains, sub-domains, and identifiers. Responsibility for assigning these CRIDs was distributed between the local sites, so each archive site is allowed to allocate any name within their domain although they must adhere to local policy. Software checks ensure that the same identifier cannot be allocated twice.

Here are some valid CRIDs that have been allocated by Cedars:

- `CEDARS:LDS:t2:Ctsmpr111706`
- `CEDARS:OX:cup:hopwood`
- `CEDARS:LDS:rao:MimePPT`
- `CEDARS:CAM:user18Q3:sci:HldSer12x02x2001`
- `CEDARS:CAM:user82Wt:photo:AmCam10x09x2000`

At each site a nameserver program is used to translate CRIDs into the location of the resource. Each nameserver has its own database listing the correlation between CRID and resource location. The nameserver is bound onto a specific socket port, and listens for a CRID request from a web-browser. When such a request is received, the nameserver looks up the CRID in its database, and returns the location of the resource.

In this way existing AIPs can be relocated, either for efficiency of Access, or as the result of phasing out a particular data store. Existing AIPs can also be replaced with functionally equivalent objects. Using a **Nameserver** in this way allows metadata to be associated unambiguously with AIPs, and so search engines will utilise CRIDs.

3.2.2 Management Functions

Management functions are the tools used by the local site manager to manipulate the archive site's holdings. These are employed for a wide variety of tasks, although in the current demonstrator some of these tasks are not yet fully automated. Here is an example of the use of some management functions as they are applied during the lifecycle of a resource:

> A ViewStore file (a BBC microcomputer database circa 1983) is received for ingest.
>
> - A tool scans the formats currently understood by the archive
>
> Since ViewStore files do not appear in the current Representation Network work starts in this area, and a general investigation is made to find out as much as possible about this file format.
>
> - Tools are used to add a section into the Representation Net to handle this format

[20] Although CRIDs conform to the URN syntax, there is no reliance on any URN technology

- A CRID is allocated to the Primary Digital Object (PDO) and the Nameserver database is updated
- The PDO (the ViewStore file) and its metadata are packaged into an AIP
- The AIP is transferred into the local data store
- A tool is used to specify an access contract
- Some PDI metadata is placed in a database and made accessible to finding aids and search engines

Later, during a routine Representation Network update, a better emulator for the BBC microcomputer is added. A user downloads the resource and chooses to download the emulator to render it.

- An invoice is sent to the user, to cover the license for the emulator.

3.2.3 Archive Store

Two of the three project sites were able to utilise local archives at the local institution. These meet the following requirements of a Cedars archive.

- A named bytestream can be placed in the store
- Replication is used to prevent losing the stored bytestream
- The named bytestream can be retrieved from the store

A common interface was built between the Store and the Management Tools and Gateway software by using a Java Abstract Class. This passes requests for data retrieval to the data store, and provides general facilities as needed by the various management tools (including the ability to place a bytestream in the store). The two institutional archives provide the necessary storage using quite different internal methods.

3.2.3.1 LEEDS File Archive (at the University of Leeds, ISS)

The LEEDS (Low cost everlasting data store) archive was produced as an in-house system which came into operation in 1991, and aimed to meet a strict set of requirements.[21] In 2000 the system was updated from an Exabyte-120 robot tape library (116 8mm helical scan tapes, giving a capacity of 1 TByte) to an Exabyte-X200 robot tape library (up to 200 Mammoth 8mm SCSI tapes, and a capacity of 8 TBytes) with the possibility of adding additional tape libraries (hence further increase to storage capacity). Data is stored and retrieved from the LEEDS store by using a custom software tool, which hides the internal mechanisms from the user. As well as storing user data, the system is resilient to hardware malfunctions and protects against:

- loss of its index partition
- loss of a data cache
- corruption of the master table of tape locations

[21] See http://www.leeds.ac.uk/ucs/systems/archive/req.html

- tape failures

The primary ethos for the LEEDS system is adequately presented by its lead architect in “The Medium is NOT the Message.”[22]

The Java class `LeedsAIPStore` uses knowledge of the internal mechanisms of the LEEDS store and generates requests directly, cutting out the user interface software.

3.2.3.2 HFS archive (at Oxford University, OUCS)

In 1995 Oxford University chose an IBM system to meet both its backup and archival requirements.[23] This system went into productive use later that year. The first priority of this system was to provide some form of “unlimited” disk-space. It employs an integrated suite of software (Adstar Data Storage Management, or ADSM) and a large automated tape silo (2 RS/6000-R40 systems running AIX 4.1.5/4.2.1 and a 3494 Automated Tape Library, with a capacity of 1300 cartridges and 8 tape drives, giving over 20 TBytes of storage).

After installation, they discovered some difficulties that had not been anticipated. When ADSM was installed, some essential software was not yet available. Such schemes would have also incurred a number of difficult policy decisions. Archiving (as distinct from migration) also needed careful consideration. The backup facilities of ADSM were good and fully integrated into the system, but users have been running the system to its limits. This required constant attention and administrative effort in order to keep pace with the demand. Despite having expected a turn-key solution, a surprisingly high level of system errors has been encountered.

As well as an archiving, migration, and backup service, the ADSM offers an ftp service. This is the service that has provided the fewest snags. Cedars adopted this interface with the ADSM, and manage a small file area in the Hierarchical File Store (HFS) through the ftp interface.

The Java class `OxfordAIPStore` uses a restricted functionality FTP client to pass requests on to the HFS.

3.2.4 Gateway

The Gateway is a web interface that provides a customised view of locally held AIPs. Using this interface a Dissemination Information Package can be requested. It also allows the user to traverse through the nodes in the representation network. At the moment the view is selected based on the AIP, but this could be changed so that different users are given views appropriate to their user type.

Currently the gateway automatically traverses the appropriate sections of the Representation Network for the AIP and consolidates this into a list of available methods for rendering the AIP. Where possible, the Representation Network also contains an indication of the Mime-Type for each data format. Current platforms such as web-browsers and email readers readily understand these Mime-Types so the **Consumer** is able to access the intellectual content of an AIP without needing to understand the Representation Information metadata. An advantage with Mime-Types is that, even where a current platform does not know how to render a particular Mime-Type, several well established repositories exist on the Internet for finding applications to handle these.

[22] Holdsworth, David. The Medium is NOT the Message. Available at http://esdis-it.gsfc.nasa.gov/MSST/conf1996/A6_07Holdsworth.html

[23] See http://users.ox.ac.uk/ alex/hfs-AXIS-paper.html

3.2.5 Search Engine

The Search Engine is a simplistic Perl program that uses a hand crafted database which lists the important resources in Cedars holdings. Given a search string, the program search engine compares this to the author and title fields in the database. A web-form is then used to present the findings and a summary of the hits on these fields. The results from the search engine is a list of CRIDs that the web-form uses to call the Nameserver which then directs the user to the correct gateway page for the AIP chosen.

3.2.6 Distributed Architecture

Distribution of AIPs between the different archive stores in Cedars is achieved through the CRIDs and Nameserver databases. The Management Tool for Ingest allows an Ingest Manager to select which site is used for the storage of the AIP. At this stage the AIP is assigned its CRID, with the first section of this indicating the site of ingest (which is also the site allocating the Unique Name). Each Nameserver database also contains the location of every other Nameserver program, so any Cedars Nameserver can be used to locate any of the CRIDs within the `CEDARS:` domain. Figure 3.2 illustrates this.

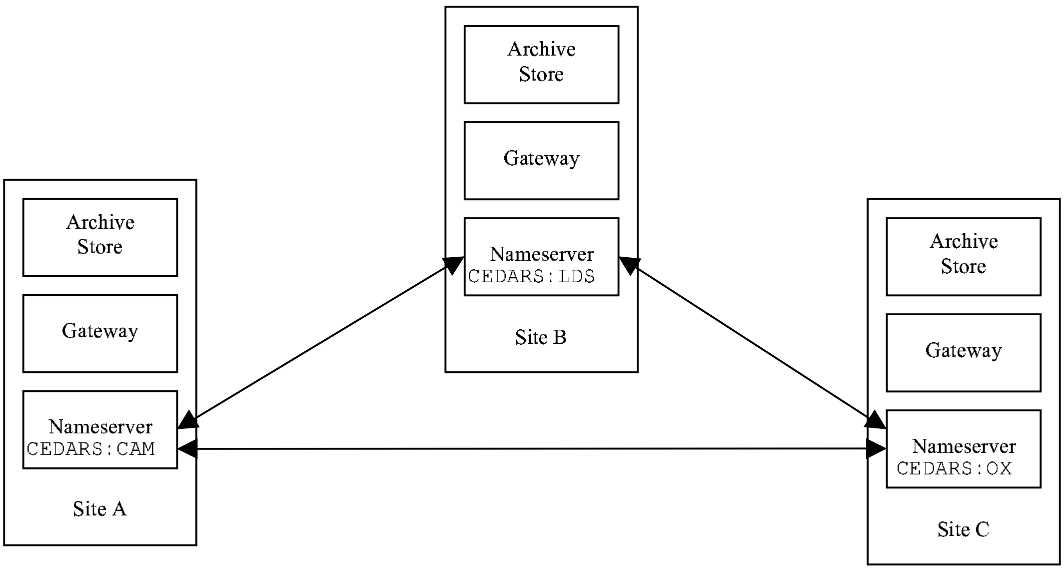


Figure 3.2: Nameserver interactions across the Distributed Archive

A request is made to the name server of Site C for the resource with name:

- `CEDARS:CAM:user18Q3:sci:HldSer12x02x2001`

Site C looks this up in its Nameserver database and passes the request on to the nameserver at Site A (based on the CAM domain). When the Nameserver at Site A interrogates its database it finds that the resource currently resides in the Archive Store at Site C and returns the resource location so that a user can use the Gateway at Site C to access the resource. Even though the resource has a CRID that is in the `CAM` domain, a policy change at some point has meant that Site A no longer keeps resources of this type so it has been transferred to Site C. Since the CRID is permanent, the resource keeps the same identifier, and only the Nameserver databases are updated. In reality, all of the Nameserver interactions are transparent and so a web-browser call to a Nameserver with a valid CRID will automatically deliver the gateway page for the correct AIP (irresepective of how many Nameservers are involved in the actual resolution of the CRID).

3.3 Handling Archival Information Packages

Fundamental to the OAIS model is the AIP. This is the smallest unit of storage and choosing its composition will affect the success of preservation. Figure 3.3 shows the hierarchy involved in the composition of the AIP in the OAIS model and figure 3.4 shows these components from another perspective.

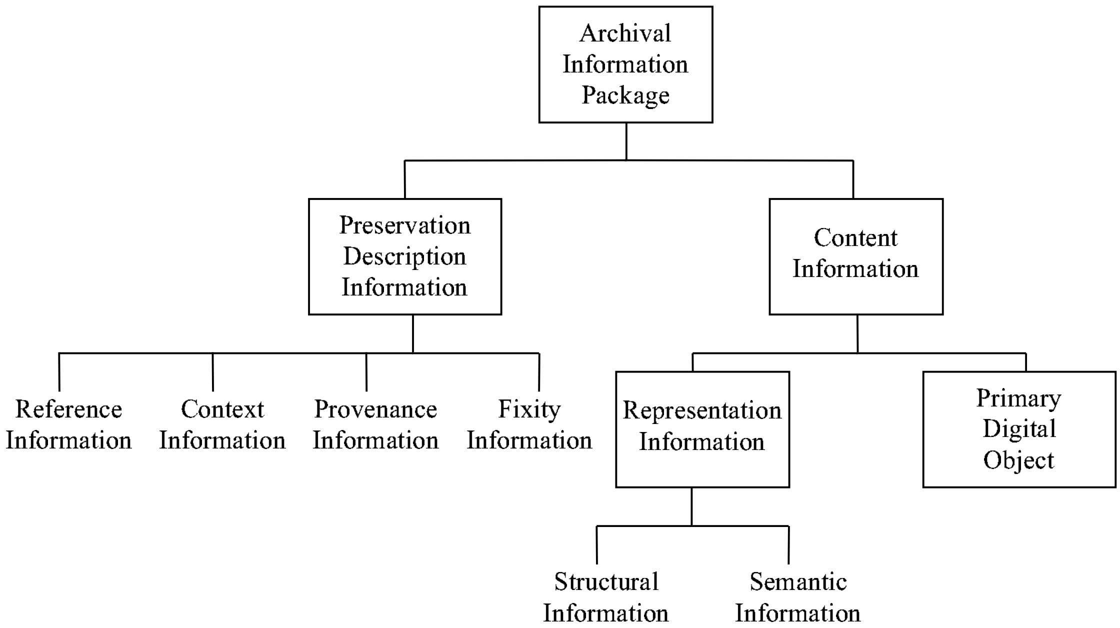


Figure 3.3: The OAIS hierarchy of an AIP

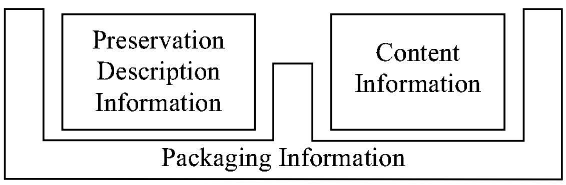


Figure 3.4: Combining the components of an AIP

3.3.1 Metadata: Specification and Internal Structure

In its demonstrator, the primary concern of Cedars was to assess whether the information captured under the elements of the Cedars Project metadata specification (see Appendix 5 for the complete specification document) would be sufficient to allow for *full* preservation for the long-term requirements of future users. For the purposes of the demonstrator, it was assumed that preservation would address complete retention of an object's Significant Properties. To do so, a suitable method had to be chosen to represent the metadata electronically as part of the AIP. The constraints for choosing this method were:

1. The nature of the metadata
2. The current (and future) rapport of the method
3. The simplicity of the representation
4. The practicalities of inter-operability within the distributed archive
5. Understandable by humans

Due to the PDI component of metadata being highly textual XML was chosen. This successfully matches each of the criteria. However, as will be discussed later, RI consists mainly of indirections to nodes in the Representation Network (which are specialised AIPs). XML does not match criterion 1, 3, or 4 for RI. A simpler approach of using property files lent itself to RI metadata, and also matches the approach used for Gateway files and cached AIPs by the Cedars archive.

3.3.1.1 Preservation Description Information

Here is a simplified section of the Cedars XML DTD for PDI:

```
<!ELEMENT informationPackage
(preservationDescriptionInformation?,contentInformation?)>

<!ELEMENT preservationDescriptionInformation
(referenceInformation?,contextInformation?,provenanceInformation?,
fixityInformation?)>

<!ELEMENT referenceInformation
(resourceDescription?,existingMetadata?)>

<!ELEMENT resourceDescription
(DCtitle | DCcreator | DCsubject | ...)*>

<!ELEMENT DCtitle  (#PCDATA | reference | list | p)*>
<!ELEMENT DCcreator (#PCDATA | reference | list | p)*>

<!ELEMENT existingMetadata (existingRecord+)>

<!ELEMENT existingRecord (#PCDATA)>
     <!ATTLIST existingRecord scheme CDATA #IMPLIED>

<!ELEMENT contextInformation (relatedInformationObject+)>

<!ELEMENT relatedInformationObject (relationship,reference)>
<!ELEMENT relationship (#PCDATA)>

<!ELEMENT provenanceInformation
(historyOfOrigin?,managementHistory?,rightsManagement?)>

<!ELEMENT historyOfOrigin
(reasonForCreation*,custodyHistory*,changeHistoryBeforeArchiving*,
originalTechnicalEnvironments*,reasonForPreservation*)>
```

```
<!ELEMENT reasonForCreation (#PCDATA | reference | list | p)*>
<!ELEMENT custodyHistory (#PCDATA | reference | list | p)*>
<!ELEMENT changeHistoryBeforeArchiving (#PCDATA | reference | list | p)*>
<!ELEMENT originalTechnicalEnvironments
(prerequisites*,procedures*,documentation*)>
<!ELEMENT reasonForPreservation (#PCDATA | reference | list | p)*>

<!ELEMENT managementHistory (ingestProcessHistory?,administrationHistory?)>

<!ELEMENT rightsManagement (negotiationHistory?,rightsInformation?) >

<!ELEMENT negotiationHistory (#PCDATA | reference | list | p)* >

<!ELEMENT rightsInformation (copyrightStatement*,actors?,actions?) >
```

All of the elements can be populated with free text. Naturally, in a service environment, individual elements would have further sub-divisions and some format constraints (it is certainly desirable to standardise the formats of common fields). Indeed, this was identified by the Cedars Project test sites. Precise specification of formats is secondary to the process of testing whether the set of elements in the specification is ample.

One technical issue did arise with using XML for the PDI. Where one or more catalogue records already exist for a resource, how do we keep copies of these in the AIP? Other than discarding an existing catalogue record completely, it could be mapped onto the relevant elements in the PDI – however four ideas were expounded about how to retain the actual record:

- **(A)** Copy the whole record into the XML as free text data
- **(B)** Mark the record up as XML and store it in another namespace
- **(C)** Define the record as a separate (non-XML) entity
- **(D)** Store the record as an AIP in its own right

Both **(A)** and **(B)** would only apply to a limited set of possible existing records, and XML is not suitable for several existing records. Option **(D)** would imply that the record either needed to be shared by several AIPs or that it was an interesting resource independently of the AIP it catalogues. Only **(C)** is generally applicable to all possible existing records.

The complete DTD corresponding to the Cedars Metadata Specification, together with some populated XML records (from the testing phase) can be found in Appendix 6.

3.3.1.2 Representation Information

Representation Information is separated into two types, **Structural Information** and **Semantic Information**. Figure 3.5[24] shows the division of Representation Information into the two components. To ensure that this technical metadata is chosen well and will therefore provide for future access to the intellectual content of the Primary Digital Object, some key concepts must be addressed.

The first of these is Significant Properties – a concept introduced and discussed in Chapter two. What constitutes the intellectual content of the Primary Digital Object? For each resource entering the Cedars archive, its Significant

[24] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK. OAIS figure 4.11

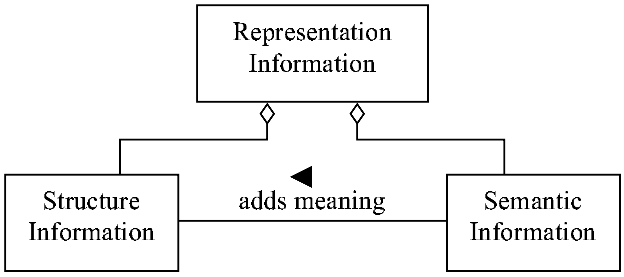


Figure 3.5: Components of Representation Information

Properties must be identified. These can range from a decision that every aspect of "look and feel" of operation with the resource must be maintained, to the decision that only the text is significant. For the whole "look and feel" the archive might have to preserve a description of the sounds made by the printer and keyboard, as well as all of the software tools that were used to create and edit the resource. Whereas, for the "text only" example, embedded pictures, fonts, sounds, and movies can all be thrown away as insignificant. As part of the submission agreement, these Significant Properties need to be identified, in light of the technical feasibility and the costs (including the cost of obtaining the same intellectual content if the Significant Properties are chosen wrongly). Such decisions are traditionally taken by the collection managers (in conjunction with systems managers as necessary).

The second concept is that of the Underlying Abstract Form (UAF) which is based on the agreed Significant Properties. This term is used to encapsulate the recognition that the data (Primary Digital Object) has an existence and an intellectual content separate from the medium upon which it is written. The UAF captures all the Significant Properties of the Primary Digital Object, and is independent of the medium upon which the data is written. Any given digital object is likely to have a number of possible UAFs. Choice of the UAF for preservation is part of the Ingest process.

Some examples of Underlying Abstract Form:

- Many CDs actually contain a file system, and successful operation (and access to intellectual content) only relies on that file system. Copying such a file system onto a partition on a hard disk delivers an equivalent working representation. Physical file placement is unimportant. For this, 'file system' is a viable UAF.
- In some cases it is only important to have a file tree, and the CD contents can be copied into a directory within any existing file system.
- Data held in a relational database can equally well reside in a variety of database engines, and still deliver its original content. Exporting a set of comma-separated files holding the content can be used as a system-independent representation of that content. So, '(system-independent) relational database' is a viable UAF.
- A plain text document consisting of lines of characters drawn from the ASCII character set is meaningful in a variety of environments. In this case, 'ASCII text document' can be a UAF.

The Underlying Abstract Form informs the choice of Structure Information and provides the formalism for the Semantic Information to "add meaning" to the Structure Information.

- Structure Information provides the information necessary for generating the UAF from the preserved bytestream

(that is, the Primary Digital Object part of the AIP)

- Semantic Information is the means to obtain the intellectual content from the UAF. There may be multiple entries in the Semantic Information, corresponding to different platforms and different ways of rendering the intellectual content.

Above, some examples were given of possible UAFs. For many digital objects there are multiple levels of abstraction, ranging from the UAF to higher level abstractions. Providing Semantic Information at higher levels of abstraction than the UAF makes it possible to share tools common between resource types, even though these are not applicable to all objects of the same UAF. Choosing the truly *underlying* UAF is not always trivial. Cedars currently adopts a policy of choosing the highest level abstraction that discards no significant information.

See the following illustration:

- A single PDF file held on a diskette can be considered as a file system, a file tree, a single file, or a PDF file. Cedars contends that the PDF file is the most useful level of abstraction for its Representation Network.
- A set of PDF files might be treated as a set of objects, each of which has the above level of abstraction, or perhaps a set of PDF files is a valid abstract form.
- A set of PDF files with a plain ASCII text `READ.ME` file, and a copy of the Acrobat reader on a CD. This raises more questions. The Abstract Form (AF) chosen is a file tree, but should the Representation Information record that a set of PDF files is incorporated within that file tree?

The archive needs to keep track of all of the different UAFs that need to be understood to use the AIPs in its holdings. When a resource is obtained by the archive, and the UAF for this is identified, the UAF must be mapped into a single bytestream. As well as keeping track of the UAF, the archive must keep track of the routines to map the bytestream back into the UAF.

The first example above is mapped into a bytestream by replicating it. The other two examples can be mapped into a bytestream using an open source algorithm such as `tar` or `zip`. For more complicated UAFs, such as a relational database driven web-site, a new algorithm for the mapping may need to be developed, or an existing technique extended. It is vital that the algorithm provides a fully reversible mapping process, as any information lost while mapping into the bytestream can never be recovered in the future.

Where possible, open source tools should be used for Semantic Information as well as Structure Information. Sometimes the use of a proprietary tool to view the intellectual content of a resource cannot be avoided. Where this is the case, when the platform that supports this tool becomes obsolete the tool also becomes obsolete. In such a situation, when a request for a resource with no "current" tool for Semantic Information is made, there are several options:

1. Deny the request, the information is no longer accessible
2. Provide the Primary Digital Object, and leave the recipient to invent a way to obtain the information from the UAF
3. Develop an emulator that resurrects the proprietary tool
4. Develop a new tool that replaces the proprietary tool
5. Develop a tool to migrate from the proprietary format to a "current" format (which should maintain the Significant Properties).

In the Semantic Information part of the technical metadata, there may be reference to software tools that are specific to only this resource. There may also be other software tools which are relevant for all resources of the same format. When other tools become available to work with one of these shared formats common to several AIPs, these are added to the appropriate node in the Representation Network and hence become available to all the relevant AIPs. Representation Networks are used to keep track of the tools, algorithms, and technical descriptions for all the AIPs in the Cedars archive. (A Representation Network is "the set of representation information which fully describes the meaning of a data object,"[25] these are discussed in detail below in section 3.5.)

There are four kinds of tools for semantic information, and a single AIP can refer to any number of any type of tool.

Rendering Tool a piece of software that runs on a computational facility and displays (in some manner) the intellectual content of the UAF.

Emulator a piece of software that mimics another (often obsolete) computational facility. For some resources a **Rendering Tool** may need to be combined with the **Emulator** in order to obtain the intellectual content of the UAF.

Migration On Request Tool a piece of software that understands the UAF and produces the same intellectual content in a different format. This new format can be understood using an appropriate **Rendering Tool** or transformed into yet another format by another **Migration On Request Tool**.[26]

Format Description a rigorous description of the format of the components of the UAF, to allow a human being to interpret the Primary Digital Object. Such a **Format Description** also allows new tools to be developed that understand this type of primary digital object. Good open source Rendering Tools can be thought of as a **Format Description**.

Migration does not appear in this list, because any decision to migrate into a standard (or preferred) format happens either at Pre-Ingest or Ingest. Such format migrations are recorded in the PDI as part of the `ingestProcessHistory` (or in `changeHistoryBeforeArchiving`).

The Representation Information is stored as a list of properties. For example:

```
#RI
StI = @OX:sti:singleFile
SeI.1 = @LDS:rao:powerpoint
SeI.2 = @LDS:rao:MimePPT
```

or

```
#RI section of metadata
StI = @LDS:stiTarFS
SeI.1 = @LDS:raoWin100/.!ContentRAE=WOB:\\install.exe
SeI.2 = @LDS:raoMac100
SeI.3 = @LDS:raoDyna100
```

The StI is the Structural Information, and the SeI is the Semantic Information. All of the metadata are references to nodes in the representation network, and these are assigned CRIDs to uniquely identify them. For the first example, the resource does not specify a specific hardware platform but can be understood on any computational platform that can handle `powerpoint`. The second example has Semantic Information for two hardware platforms and one general computational platform. The `Win` platform requires a parameter specific to this resource in order for it to use the in-built **Rendering Tool**.

[25] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK
[26] This technique is described further in Wheatley, P. R. "Migration – a CAMiLEON discussion paper". www.leeds.ac.uk/CAMiLEON/index.html

3.3.2 The Storage Format of the Primary Digital Object

This is directly related to the Structure Information part of the Representation Information. As this references the algorithm to convert the bytestream into the UAF (in a completely reversible way), the storage format of the Primary Digital Object is simply a single bytestream.

All of the AIPs in the Cedars archive store this bytestream directly. This means that the packaging information for the Primary Digital Object part of the AIP is always the same.

3.3.3 Packaging the AIP

The format of the internal components of the AIP in the Cedars demonstrator archive are as follows:

Preservation Description Information an XML record

Representation Information a list of properties (in Java notation)

Primary Digital Object an unmodified bytestream

Existing Catalogue Records several named bytestreams

Each component of an AIP is packaged together using an ASN.1 (Abstract Syntax Notation) wrapper. Only the ASN.1 Basic Encoding Rules are used. ASN.1 was chosen because it is well established, the Basic Encoding Rules are simple, and it can handle very large binary data files.[27]

Figure 3.6 illustrates the ASN.1 structure for packaging an AIP. The packaging information is a version number and a CRID. The object referenced by the CRID contains details of the internal structure of the AIP and documents the Semantic Information to obtain the components from the ASN.1 packed AIP.

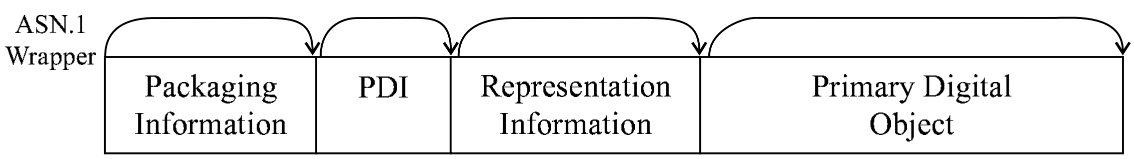


Figure 3.6: The Package Structure of an AIP

The first three components of the AIP are human readable, which provides a backup in case of archival system failure. The following is an example of a (highly abridged) ASN.1 packed AIP from the Cedars archive.

```
0^F=h^B^A^C^S^^N^e<?xml version="1.0"?>
<!DOCTYPE informationPackage SYSTEM "http://cedars/cedars.dtd"><!-- 4/12/2000:12:40:41 --
>
<informationPackage>
<preservationDescriptionInformation>
```

[27]The ASN.1 BER are an ISO standard, the specification can be found at http://www.sdct.itl.nist.gov/ ftp/l8/other/Standards/iso8825/X680.pdf

```
<referenceInformation>
<resourceDescription>
<DCtitle>George3 Source - tape 1</DCtitle>
<DCsubject>Source code of George3</DCsubject>
<DCdescription>Source code of the George3 operating system for ICL 1900 computers
and for ICL 2900 machines running DME</DCdescription>
<DCcontributor>D.Holdsworth</DCcontributor>
</resourceDescription>
</referenceInformation>
<contextInformation>
<relatedInformationObject>
<relationship>Same Subject</relationship>
<reference>DH:g3mt:123451</reference>
</relatedInformationObject>
</contextInformation>
<provenanceInformation>
<historyOfOrigin>
<reasonForCreation>This is the source released to installations running the
system.</reasonForCreation>
<custodyHistory>EarlierCustodian:Dave Higgins--CAP/Gemini</custodyHistory>
<originalTechnicalEnvironments>
<prerequisites>ICL1900 computer</prerequisites>
<documentation>available from ICL</documentation>
</originalTechnicalEnvironments>
<reasonForPreservation>Historic system prevalent in the UK in 1970s and early
80s</reasonForPreservation>
</historyOfOrigin>
<managementHistory>
<ingestProcessHistory>PDI Populated
when=4/12/2000:11:08:19 actor=D.Holdsworth@leeds.ac.uk action=requestIngest
when=4/12/2000:11:09:08 actor=CedarsIngestManager action=AuthoriseIngest
when=4/12/2000:11:09:08 actor=new_ingest.perl action=autoPDIscript
when=4/12/2000:12:35:35 actor=testsite11 action=EditXMLRecord
</ingestProcessHistory>
</managementHistory>
<rightsManagement>
<negotiationHistory>Official permission obtained from technical director of
ICL.</negotiationHistory>
<rightsInformation>
<copyrightStatement>
<nameOfPublisher>ICL</nameOfPublisher>
<dateOfPublication>1984?</dateOfPublication>
<placeOfPublication>London - England</placeOfPublication>
<rightsWarning>This is only made available as a historic object.  It is not to be
used for any other purpose.</rightsWarning>
</copyrightStatement>
</rightsInformation>
</rightsManagement>
</provenanceInformation>
</preservationDescriptionInformation></informationPackage>
^SN#RI
StI = @LDS:stiByteStream
SeI.1 = @DH:raoTapeImage
SeI.2 = @DH:g3:tapeprint^D<binary data>
```


The example above does not contain any existing catalogue records. These are incorporated using ASN.1 as part of the PDI component. This is shown in figure 3.7.

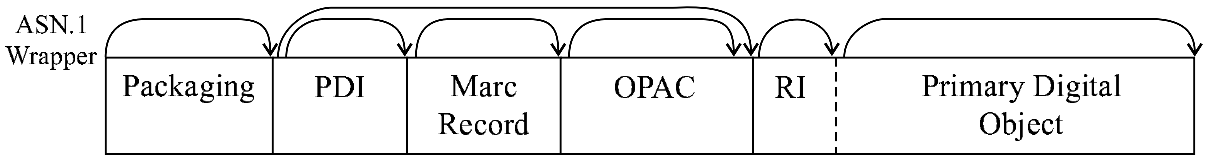


Figure 3.7: Existing Catalogue Records in the ASN.1 AIP

3.4 Internal Archival States of the AIP

During the life time of an AIP, it becomes available in many states. These range from a stored object in the archive being available for users served from several caches to an unavailable object which is yet to be committed to long term storage.[28]

3.4.1 Maximum Availability

A fully ingested AIP of modest size which is in high demand would be in this state.

- The AIP is on-line ready to be delivered on demand
- The AIP exists in the long-term store and is given an internal Local Long-Term ID (LLTID)
- The gateway data is held in a local file for immediate delivery by the gateway
- The nameserver translates the CRID into the correct call to the gateway program

3.4.2 Medium Availability

A fully ingested large object which is in high demand would be in this state.

- The AIP exists in the long-term store and can be delivered in due course
- The gateway data is held in a local file for immediate delivery by the gateway. It contains the LLTID
- The nameserver translates the CRID the correct call to the gateway program

3.4.3 Minimum Availability

A fully ingested AIP which has not been requested for some time would be in this state.

- The AIP exists in the long-term store and can be delivered in due course
- The nameserver translates the CRID into a special call containing the LLTID, which can be used by the gateway to then recreate the local gateway file, by extracting the correct metadata from the AIP

[28] More information about these states can be found in Holdsworth, D. "States of Being" and Holdsworth, D. "Implementing Changes of State". www.leeds.ac.uk/cedars/demonstrator

3.4.4 Interim Availability

A recently ingested object of modest size would be in this state, awaiting further processing into a long-term state (even so, in this state it is available for delivery by the archive).

- The AIP is on-line ready to be delivered on demand
- The gateway data is held in a local file for immediate delivery by the gateway. It does not contain the LLTID (which has yet to be allocated)
- The nameserver translates the CRID into a call for the gateway file

3.4.5 Initial Availability

A newly ingested object of modest size would be in this state, awaiting further processing into a long-term state (in this state it is available for delivery by the archive).

- The bytestream version of the original digital object is on-line ready to be delivered on demand
- The gateway data is in the form of a file for immediate delivery by the gateway. It does not contain the LLTID (which has yet to be allocated).
- The nameserver translates the interim CRID into the call to the gateway
- The AIP is available to be generated on demand

3.4.6 Stages Between the States

3.4.6.1 Ingest

Stage One

The first stage of Ingest is in the production of the bytestream (of the Primary Digital Object), and a gateway data file. This includes allocation of a CRID for the object, as it is part of the gateway data. At this stage the object cannot be seen by the outside world, but a Nameserver entry can now be created to give access to the object at this stage, even though it is not yet in long-term storage. This places the resource into the Initial availability state.

Stage Two

The AIP is generated from the bytestream and the gateway data. This is an ASN.1 package. A new gateway file is generated, referencing the AIP by its local name in the interim staging filespace. At this stage the Stage One bytestream is destroyed. This is the Interim availability state.

Stage Three

The AIP is copied into the long-term storage, and its LLTID is allocated. A new gateway file is generated, now containing the LLTID. This is the state described as Maximum availability above.

3.4.6.2 Maturity

Stage Four

The AIP is deleted from on-line storage. It can be recovered from the gateway when needed, or the end-user can be supplied with the facility for direct recovery from the long-term store. This is the state described as Medium availability.

Stage Five

The gateway file is deleted. The nameserver entry is updated to deliver the correct special call to the gateway. This contains the LLTID, from which the gateway can recover the information necessary to create the gateway file. This is the Minimum availability state.

Stage Six

Copies of the AIP are held by other members of the archive federation. Local Nameservers get entries to address their copies.

3.4.6.3 Expiry

Once the AIP has been committed to the long-term store it is preserved permanently. The nameserver responsible for the CRID keeps continual record of the LLTID call to the gateway. In short: Expiry? Never!

3.5 Representation Networks

As mentioned earlier, Cedars stores Representation Information as a property file listing indirections that point to nodes in the Representation Network. Our Representation Nets involve three main types of node:

Data Format Definitions (DFD) which define a data format. These are sometimes actual bytes and sometimes a more abstract entity such as an API. There are also DFDs for describing a real magnetic tape, or a human being interacting with a desktop WIMP interface.

Render/Analyse/Convert Engines (RACE) take in data in one format, and deliver it in another format. This may be a new format from a Migrate On Request type RACE, or be a document viewer from a Rendering Tool type RACE.

Platforms are typically computer systems, usually with storage, and are necessary for the execution of RACEs. They must contain or have access to storage suitable for storing the data that they are processing.

A Data Format Definition includes 2 lists of Render/Analyse/Convert Engines. One list enumerates those RACEs capable of accepting the defined format as input, and of delivering another format as output. The other list enumerates those RACEs capable of delivering the defined format as output. An Underlying Abstract Form (UAF) is a specialisation of a DFD node that includes RACEs that can accept the raw bytestream of the Primary Digital Object as input and deliver the defined format as output. For example:

- If a filesystem has been preserved as a tar file, it could be described as having a UAF of a filetree. The UAF node referenced by the Structure Information metadata could include a link to a RACE describing UNIX `tar` for generation of the UAF on a UNIX system, and also an RAE describing `WinZIP` for generation of the UAF on a PC system.

For any AIP the RI has two components at the top level, Structure Information and Semantic Information. Structure Information is concerned with describing (and especially regenerating) the UAF and consists of:

- The CRID of the UAF object
- Any parameters needed by the UAF object to regenerate the specific UAF

Semantic Information is concerned with interpreting the UAF, and consists of a list of CRIDs of RACEs. With each CRID is held the parameter values needed by the particular RACE. Each RACE contains information (often in the form of software) for some particular processing of this AIP. In particular, the RACE includes a reference to the platform upon which it is to operate.

In some cases the Semantic Information may be empty, and rely entirely on rendering facilities accessed via the UAF description. This depends on the extent to which the UAF is defined as a high level abstraction. If we have a multimedia object containing `HTML` and `JPEG` files, the UAF object will indicate how to unpack the file-tree. The Semantic Information will reference a tool that understands the association of the filename extension with the different data formats. These associations are also references to the Representation Network.

There is a value in attaching rendering capability to the data format level, rather than to the resource, because new facilities can be recorded at the data format level. These immediately become relevant to many preserved objects. This is particularly important in regard to following technological evolution.

3.5.1 Gödel Ends

As Gödel's theorem tells us, any logical system has to be incomplete. There must be truths which the system cannot itself deduce.

The Representation Networks must have end nodes corresponding to formats that are understood without recourse to information in the archive, for example plain text using the ASCII character set. All references to such a format must be via the same CRID, and the management of the archive must have an inventory of such objects. As a format becomes obsolete, the object referenced by the CRID can then be updated to permit understanding of the obsolete format in terms of current practice.

In the Cedars scheme, the platforms upon which the render/analyse/convert engines (RACEs) run are the things that become obsolete as a result of the march of technology. A data format becomes less and less accessible as the platforms of the relevant RACEs become obsolete. As a fallback there is the documentation of the format, which we might consider as a special RACE whose platform is human (e.g. a programmer). In the case of proprietary formats we may not have this information.

Thus the platforms (e.g. Win32) are the things that are outside the archive, and as such are the true Gödel ends of the system. One suggestion[29] is that emulation of the original computational environment gives the very best hope for recreation of the experience of a preserved digital object. This can be a laborious process, and technology shifts may render a true recreation impossible. It is quite possible that the technology of the time actually limited the access to the intellectual content, and a far better access to archived material can be achieved by implementing viewers that operate on the obsolete format directly.

Therefore the archive administration keeps an inventory of the platforms which occur in the archive's Representation Nets, and the Gödel end platforms each contain the CRIDs of the data formats that rely on that platform. The administration then needs to keep this inventory under review, and add new RACEs to data format objects in order to maintain accessibility.

Some platforms may not be true ends, as they may have been realised by emulation. An emulator is a type of RACE, and as such depends on a platform.[30]

3.5.2 Proprietary Formats

Any archive has an understandable reticence about keeping data which is held in undocumented formats. However, reticence must not be seen as a synonym for rejection. A current proprietary format may not have any publicly available documentation, but may have readily available rendering facilities on current platforms.

As obsolescence threatens, the commercial value of the documentation is minimal, and there is real prospect of being able to add this into the Representation Net later.

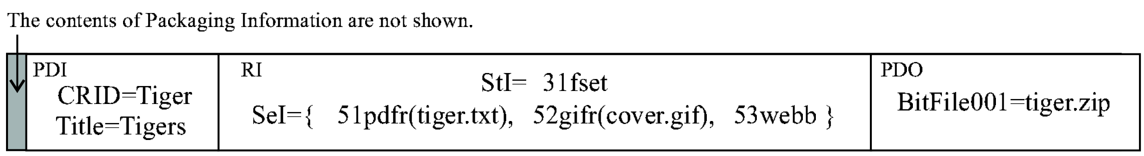
Indeed, where an open source rendering tool exists for a proprietary format, its source code acts as a fairly rigourous documentation of the format. Keeping documentation in the form of source code reduces the cost of implementing new rendering tools in the future.

3.5.3 An example of a Representation Network

Here is an example of one of the AIPs in the Cedars archive. The Primary Digital Object consists of a PDF book and a colour GIF image of the original front cover. The CRID for this AIP is:

```
CEDARS:LDS:pubarea:Tiger
```

The PDI component of the AIP is simplified for this example, and so is only a title and the object's CRID. The CRIDs are the arcs of the Representation Net. Only the tail of the CRIDs are shown.



The structure information (StI) in the RI is referenced indirectly and is an AIP which describes the data format (in this case a set of files) and facilities for interpreting it. The semantic information (SeI) in the RI is a list of CRIDs for RACEs that can give access to the intellectual content.

[29] See Rothenberg, J. "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation" www.clir.org/pubs/reports/rothenberg/contents.html

[30] The CAMiLEON project is investigating emulation, and provides critiques of other work in this field. See http://www.si.umich.edu/CAMILEON/

The StI has the following specialised fields in its Primary Digital Object: UAF, a description of the underlying abstract form; TOI, a *t*ransformer object instance (a special RACE which can generate the UAF from the preserved byte-stream); RACE, a list of render/analyse/convert engines that can generate other representations (or provide APIs).

PDI CRID=31fset Title=FileSetUAF	RI StI= 32uafp SeI={ 50gate, 54asciir }	PDO UAF= 71fset RAE= 80flist TOI= { 81pczip, 82maczip, 83unixzip }

The Representation Net node (↑31fset) conveying information on how to deal with the UAF of the "Tigers" AIP also carries RI to interpret its own Digital Object component. The object referenced by ↑32uafp conveys this information, and only deals with the UAF specialisation of a DFD. Although two choices are provided to render our fileset AIP in the SeI, while the Cedars gateway (↑50gate) can trace through the internal indirections of the Representation Net and present an overall decision framework, the ASCII renderer (↑54asciir) can only display the property file in the PDO verbatim.

PDI CRID=81pczip Title=PCunzip	RI StI= 34raep SeI={ 50gate, 54asciir }	PDO Platform= 91pc engine=winzip.exe Params=extract OutForm= 72ftree

Examining the contents of the fileset AIP (↑31fset) the UAF field references an AIP containing an ASCII description of a fileset (↑71fset), whose RI shows how to render the ASCII file. The RACE field provides access to a tool (↑80flist) that displays the structure of the UAF, in this case by listing the filenames (tiger.pdf and cover.gif) which make up the set. The TOI list field provides methods for transforming the content file of the "Tigers" AIP into these two files on different computational platforms. Only the RACE for the PC platform has been shown (the others are similar). This RACE (↑81pczip) has a dedicated StI (similar to ↑32uafp). ↑91pc references an AIP containing an ASCII description of the PC platform, and ↑72ftree is the DFD for a filetree.

PDI CRID=32uafp Title=uafUAF	RI StI= 32uafp SeI={ 50gate, 54asciir }	PDO UAF= 73uafpf RAE= 80flist TOI= { 84asciicp }

PDI CRID=51pdfr Title=PDFreader	RI StI= 35rael SeI={ 50gate, 54asciir }	PDO InForm= 74pdf RAE={ 87pcpdfV, 88macpdfV }

PDI CRID=87pcpdfV Title=pcPDFviewer	RI StI= 34raep SeI={ 50gate, 54asciir }	PDO Platform= 91pc engine=acroread.exe Params=none OutForm= 75gui

These last three nodes illustrate some important features of our Representation Net. ↑32uafp is a Gödel end for UAF nodes, ↑87pcpdfV is a Gödel end for the Rendering Platform. We show PDF as a platform (↑51pdfr), which is one of a number of choices. As the Gödel end platforms become obsolete this formalism offers the choice of adding PDF rendering capability on new platforms or emulation of the obsolete platform upon which the existing rendering software will run. The parameter, shown in brackets in ↑01trav, is passed through to the software engine of ↑87pcpdfV via the ↑51pdfr. If desired, (tiger.pdf) could be rendered via ↑88macpdfV instead. All new rendering software for pdf is added to the list of RACEs in ↑51pdfr.

3.6 Access and Access Contracts

The dissemination of resources and metadata from the archive to its end users is termed Access by the OAIS model. As mentioned a number of times already, this must be kept very separate and distinct from preservation and preservation activities. The unit of delivery for a resource at Access is a **Dissemination Information Package** (DIP).

The Cedars gateway provides an interested party with a sample of the metadata to describe the resource (served from the gateway file) and with the means to check the technical requirements that are needed to obtain the intellectual content from the Primary Digital Object. From this starting point, the Representation Network can be followed to find out extra technical information that may be useful in making the decision about whether to request a DIP for this resource. The gateway also allows the interested party to request a DIP.

At the moment two types of DIP are served from the Cedars demonstrator archive, the first is a replica of the AIP. This contains the Primary Digital Object together with all of the metadata. The second kind of DIP is a replica of the Primary Digital Object, this contains no metadata whatsoever. These two types of DIP are shown in figures 3.8 and 3.9.

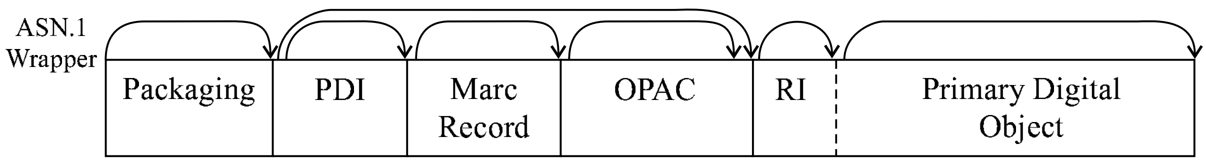


Figure 3.8: A DIP which replicates the whole AIP

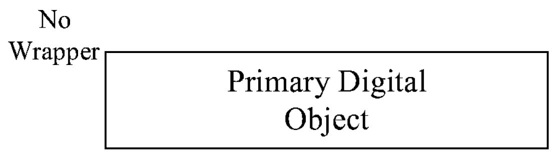


Figure 3.9: A DIP containing only the Primary Digital Object

To develop the Cedars demonstrator into a service archive, new types of DIPs would be needed. For example, some of the metadata would not be relevant to a general user of the resource and would not be placed into the DIP.

Representation Information would probably be translated into a standardised set of instructions in the DIP providing an easy way to obtain the intellectual content from the DIP. In general there is no need for an end-user to have access to all the complex metadata stored within a Representation Network. In some instances a DIP could be built from several AIPs, for example if a whole series of PDF articles were ordered, in which case the DIP would consist of a subset of the PDI metadata describing each article together with a common RI description about PDF and a copy of each Primary Digital Object (each article).

Part of the agreement between the test content providers and the Cedars project was that the material would not be made generally available, but should be carefully controlled. The Cedars metadata specification includes an extensive section on Rights Metadata, which records both the rights holders (and original copyright statements and warnings) and the allowed actions (and the actors permitted to perform these). This section of PDI metadata is designed to provide a qualified human being with enough information to decide whether or not a user should be given access to a resource.

To enforce this decision, the Cedars demonstrator allocates a digital certificate to each user. Where an AIP is placed in restricted access, a reference to an access contract is placed in its gateway file. When a user makes a request to download one of the DIPs for this resource, this is done using a secure web call (via `https://`). If the user has a Cedars digital certificate that is listed in the access contract then the request is granted, otherwise the download is prevented. Whenever a new user is given permission to access resources, their digital certificate number is added to the relevant access contracts. Because a CRID is used to reference an access contract object, several resources can share the same access contract.

3.7 Testing

As part of the Cedars Project Demonstrator Archive a number of “test sites” were invited to test different aspects of digital archiving.[31] Six test-sites used a custom built web-interface to the demonstrator in order to ingest and access various types of digital material (image files, text files, electronic journals, large online databases, and other complex multi-media materials). These six test-sites were chosen from organisations with a breadth of expertise and interests. For each test-site, Cedars conducted an in-depth interview to learn more about their current situation with regard to digital preservation. The six test-sites were:

- The University of Birmingham
- The University of Exeter
- University College London
- The British Library
- MIMAS (Manchester Computing)
- Birmingham Central Library (a public library)

The custom built web-area allowed each test-site to conduct a series of tests with the Cedars Demonstrator Archive. At the end of each test feedback was elicited using a constrained set of questions. These looked at the facilities offered by the demonstrator and the knowledge requirement to use these. The bullet points in this section summarise the feedback comments obtained from the testing.

3.7.1 Initial Negotiations for Ingest

This task is to notify the archive that a content provider intends to submit a piece of digital content to be preserved in its holdings.

- A larger set of fields for metadata would allow a content provider to adequately describe their resource
- Knowledge of how the information is treated once it reaches the archive would help the provider know which information is truly useful
- Guidance and training would be valuable

3.7.2 Providing PDI

This task allows the test-site to populate the descriptive metadata.

- The metadata set was good
- Sometimes it was not obvious how to find values to populate the fields
- A smarter interface with “what’s this?” popups would make it easier to use
- Constraining choices and formats for the fields is necessary
- Import and export facilities to quickly populate fields would improve the edit capabilities

[31] See http://www.leeds.ac.uk/cedars/testsites.htm for further details of the testing

3.7.3 Linking to the Representation Network

At this stage the RI is updated by choosing from a suitable description of structure and underlying format.

- The good interface made this daunting task quite straight forward
- A mechanism for automatically extending the choices is needed
- Worked examples would help

3.7.4 Uploading the Primary Digital Object

The content provider is given instructions enabling the use of FTP to upload their bytestream.

- Upload was very successful
- Some institutions do not allow FTP, alternative methods need to be found.

3.7.5 Retrieving a resource

This task allows any of the test resources to be downloaded.

- PDI was easy to view (although it could have been presented better)
- Downloading the resource was straight forward
- Once web-browser filenaming foibles were addressed, the resources could be viewed without much difficulty

3.8 Using the Past to Inform the Future

As the techniques developed for the Cedars demonstrator are intended to enable preservation of a digital object forever, it is difficult to test in practice whether they are successful or not. To enable assessment of these techniques, say ten years into the future, we actually look ten years into the past. By simulating the application of our techniques to a resource from 1990 (under the initial constraints of hardware and software available in 1990) and then evolving the Representation Network to reflect shifts in the technical environment we can test whether the techniques stand the test of time. If this simulation shows the resource is successfully preserved and could be retrieved today, then we can assume that the preservation of a resource today using the same techniques would ensure it was available for at least ten years. In this way these simulations use the past to inform the future of digital preservation. Obviously, the assumption is also made that new inroads developed in digital preservation would be adopted and this would again add to the longevity of the archival holdings. As well as content from the “current” technology the Cedars project has permission to use some materials from earlier, now obsolete, systems. These include software from the BBC microcomputer and from the George3 mainframe system. Cedars is also working closely with the CAMiLEON project[32] and its study of using emulation as a strategy for digital preservation.

While it may be tautological, it is worth mentioning that the past also informs us of techniques that are a bad idea to adopt. That is why Significant Properties need to be identified before any non-reversible approaches are applied.

[32] Creative Archiving at Michigan and Leeds, Emulating the Old on the New (www.si.umich.edu/CAMILEON/)

3.8.1 BBC Microelectronics Education Programme (MEP)

Cedars has two resources that were used as part of the BBC Microelectronics education programme. Both of these were from 1986, and developed in the West and North Yorkshire region of MEP. The overall programme ran from 1983 through the mid 1980s and used the BBC microcomputer, which was then being established in most schools in the UK. The MEP covered a wide range of subjects ranging throughout the primary and senior schools.

The BBC microcomputer is an 8-bit computer using a 6502 central processing unit. This computer had between 32Kbytes and 64Kbytes. While it is not a completely obsolete platform, the hardware has not been manufactured for a decade and working examples are becoming difficult to find.

The first resource, “Kitchen Planner in 3D”, presented immediate difficulties for long-term preservation. While the handbook can be read, PDI metadata assigned, and some RI metadata, the 5¼ inch disk (80 track, double-sided) cannot be read. Unless a means to read the disk can be found, or another copy of the resource found, the preservation effort started too late and the resource is lost. The metadata without the resource is of limited value.

The second resource, “Human Digestion Simplified”, consisted of two 5¼ inch disks (both 40 track, single-sided). Fortunately, these disks were read successfully. This was done using a BBC microcomputer with a twin disk drive. Once these disks were read, the Underlying Abstract Form (UAF) of the data on these needed to be mapped into a bytestream. A Serial ↔ RS-423 cable was built to allow data to be spooled from a BBC disk to a PC. This was used to detach the Significant Properties of the UAF (a BBC disk image in this case) from the 5¼ inch media and map them into a bytestream on the local network. The process is reversible, so the data can be sent back to the BBC and written to a 5¼ inch disk if necessary. At the moment we suspect that one of the 5¼ inch disks is now unreadable, so this resource was preserved only just in its available lifetime.

The “Human Digestion Simplified” resource is a multi-media workbook intended to guide a pupil through the biology of the human digestive system and, at the end of each learning section, test the knowledge of the pupil. Due to the complexity of this resource, the best way (currently) to adequately access the intellectual content is through an emulator. The CAMiLEON project has used this resource to test several emulators, and also written a discussion paper that explores the implications of migration strategies for this resource (compared to other classes of resource).[33]

3.8.2 George 3

We have a number of objects that are images of data taken from a system in wide use in the 1970s. The data for each of these was the contents of a magnetic tape from the George3 system. This system was in widespread use on multi-user computers (ICL 1900) in the 1970s and early 80s. Although the machines and their architecture do not exist any more, we have an emulation that allows access to the intellectual experience of 1970s computation.

Three particular tapes have been fully archived and extensive Representation Information built that enables them to be rendered on a current platform. The contents of these tapes is plain text data, but represented in ICL 6-bit characters, and packaged in an ICL proprietary data format.

Each of these has been preserved as an AIP containing a tape image, which can be extracted via the CEDARS gateway. Two of the tapes hold the source code of the system. This arrangement preserves the form in which the original source text was issued to installations running George3. The data has been preserved as a byte-for-byte images of the two tapes. It is possible to recreate a copy of the original that would work with a real ICL 1900, if we still had any of them (but we do have an emulation).

For today’s audience, we also have tools in the Representation Network that can render the source text from these AIPs into ASCII (handled easily by modern computational platforms). The tool is stored in the Representation Network as a

[33] Wheatley, P. “Migration – a CAMiLEON discussion paper” www.leeds.ac.uk/CAMiLEON

Render/Analyse/Convert Engine, and is available as a source program written in C. The Representation Information for this source program describes how to generate an executable program on a current platform. Both tapes can be visualised in ASCII by using this executable program.

A third tape from the ICL George3 system is a tape holding a system dump from the system, originally obtained in the system's conventional manner (by writing a file onto magnetic tape in an ICL specific format). This can be accessed and printed in the same manner as the other two tapes. This is indicated by the resource referencing the same nodes in the Representation Network.

This method of preservation preserves the original way that a user would access digital information, even though that information has been held in a form that is no longer current, by appropriate rendering tools. Because the tools are simple and written in widely available languages, and are still available in source code form, these tools need little (or no) modification to make them available on current platforms.

As well as the three tapes described above, Cedars also has several other tapes from the George3 system. These are of various kinds. These did not contain a simple data structure and plain text, however we have been able to preserve the data from them. To read them we need to use the George3 emulation system.

3.8.3 CAMiLEON and Domesday

The CAMiLEON project is focusing on investigating emulation as a strategy for digital preservation. The project is looking at two computer systems from the 1980s, the BBC microcomputer from the UK, and the Apple II from the USA.

As part of their investigations they are looking at the Cedars resources from the MEP, and are also looking at how appropriate emulation is for different categories of digital resources. The first phase of user testing in CAMiLEON looked at the preservation of a classic 2D platform game. For the long-term preservation of their resources, CAMiLEON is using the techniques developed for the Cedars demonstrator and is storing its AIPs containing resources at the Leeds site.

One resource of huge interest in the UK is the BBC Domesday videodisk. CAMiLEON is working in consultation with Cedars to design a method to preserve this material. This resource used a BBC microcomputer to control a videodisk player, with several multi-media applications across 4 sides of 2 videodisks.[34]

In keeping with Cedars strategy of using open source tools to render AIPs, CAMiLEON aims to develop an emulator written well in a new open source language called C–.[35] Where possible findings are shared between CAMiLEON and Cedars to improve the preservation techniques of both projects.

3.9 Recommendations

While looking at all of the aspects of implementing an archive for in-perpetuity preservation of digital material, it is easy to become too concerned with trivia and minutiae and miss a single step that makes the whole process simpler. For that reason, it is important to understand the concepts underpinning this chapter rather than just using the recommendations.

3.9.1 General Recommendations

1. Build the archive so that it can be easily incorporated and interoperable within a larger distributed federation of archives

[34] Finney, A. "Domesday" www.atsf.co.uk/dottext/domesday.html
[35] Holdsworth, D. "Emulation: C-ing Ahead" www.leeds.ac.uk/CAMiLEON/

2. Adopt a standard for unique names. Assignment of names and maintenance of them is of paramount important to preservation
3. Always make a decision before discarding anything (including metadata) – never discard by default
4. Metadata specifications should be extensive and allow for free text commentary
5. Identify which elements of metadata may change, and which will remain static
6. Choose the right structural formalism for each component of the AIP. Generally simplest is best (e.g. XML for the PDI). Also choose a good formalism to pack the AIP.
7. Test the archive and its interface with likely users of the system

3.9.2 Ingest Recommendations

8. Significant Properties need to be identified and clearly understood prior to ingest
9. Choose the highest level of abstraction that discards no significant information
10. Use freeware and open source tools where possible, especially to map between the bytestream and the UAF.
11. Standard formats are preferable, although well documented proprietary formats are better than non-documented formats

3.9.3 Storage Recommendations

12. Long-term preservation storage should be kept distinct from access
13. Make sure that storage providers meet minimum service requirement guarantees
14. Use a formalism that allows AIP to be self-standing. Keep metadata human readable where possible

3.9.4 Access Recommendations

15. Preserve the tools and monitor the Gödel ends of the Representation Network
16. Where open source programs are used for rendering tools and other semantic tools, this also serves as a good document for the format description
17. Monitor available technical options for providing access

Chapter Four : Preservation Issues for Managers of Digital Collections

4.1 Introduction

As is clear from work done in The Cedars Project, the challenges presented are not limited to technical concerns and challenges. Even the technical issues cannot be fully understood without some reference to collection management policies, skills requirements, intellectual property rights or costs. Although the technical issues remain key ones, many of the organisational and management issues are as important and often equally (if not more) complex.

Preservation of digital materials cannot be considered as an isolated activity conducted by a single isolated department within an organisation. Digital preservation needs to be an integral part of digital collection management and must, therefore, be incorporated into the overall management of an organisation's digital assets from acquisition through to preservation. Two critical reasons for this are:

1. Because digital materials rely on technology that rapidly becomes obsolete, their preservation must be considered at the time of creation (ideally) or upon acquisition to the collection. Preservation decisions for digital items cannot wait until their continued utilisation has proved their value. Delays in taking preservation decisions can (and most often will) result in making these decisions harder to implement.
2. Because preservation and management of digital materials will involve expenditure (especially in the short term), integration of preservation into the everyday management and organisation of the library or archive will ensure the necessary skills and knowledge are embedded within the organisation and achieve more effective economies of scale.

The costs involved in digital preservation are not easy to establish, and this is discussed in detail in Chapter six. Preservation of digital materials will impact on many areas of an organisation, including finances, staffing, selection of material, and intellectual property rights.

4.2 Staffing For the Preservation of Digital Objects

For many libraries and archives assuming the preservation responsibility for digital materials means re-examining existing resource allocation - both financial and human resources. Work on the Cedars Project has shown that as organisations accept and understand preservation management, it is often the additional skills and expertise requirements that are the most daunting. It is often unclear to what extent this expertise will come from existing staff within the organisation (either re-deployed or re-trained) and to what extent the organisation may need to hire or outsource.

Institutions will need to have access to various skills:

- to assist in the selection of the most appropriate format, when creating or preserving an object;
- to create and edit metadata, in a form which is compatible with existing resource discovery systems, and which supports the functions of the stored digital object;
- to assist in the selection of the Significant Properties of an object (see chapter three);
- to manage a technical preservation programme which ensures continuous access to the stored objects as a managed collection;

- to assist in the selection of physical storage options;
- and to regulate scholarly access to the archive, governed by current Intellectual Property Right legislation and legal precedent.

In addition, institutions may need to employ the skills to negotiate the right to preserve digital objects, and to draft and exchange agreements with the various holders of intellectual property rights; to validate the completeness of the object to be preserved; and many other aspects such as planning the installation of successive generations of Preservation systems.

4.3 The importance of Collection Management guidance

Good vegetable ink, on good acid-free paper, stored in a cool, dry, dark room, will last a thousand years.[36] Vellum will last longer. Ink on paper is very stable. If, however, we take a digital object, stored on any kind of medium, and put it in a dark room, walk away, and come back in perhaps five years, we will probably not be able to use the digital object. Either the medium will have deteriorated, or the hardware to read the medium will have been superceded or broken, or the software to interpret the information on the medium will have become unavailable. To ensure preservation, we will have to actively employ an ongoing rolling programme to transfer the digital object to new platforms and media. There is a difference between benign neglect and serendipity.

Collection management policies for digital materials will be akin to those used for any other type of material. Current library staff can modify existing policies to ensure the preservation of digital objects. The same skills needed to develop effective policies for subject areas can be used to modify current collection management policies to ensure preservation of digital objects. As has been said of digital preservation:

> *the skills and judgement developed in preservation professionals – the ability to discover the original form of an object and the intent of its creator, and to prolong the life of the object or return the objects as nearly as possible to its state at the time of its creation – are precisely the same skill sets that are needed for the future, albeit practised in a radically different context.*[37]

The preservation responsibilities of an institution are based on its position in a local structure, or on responsibilities due to membership of regional or national bodies. Each institution needs to decide which subjects it specialises in, and which subjects it is responsible for. This includes both current subject holdings and subject holdings the institution wishes to develop. This will depend both on local needs and on regional or national agreements.

- Each institution will need to determine what responsibilities it is seen as having by internal and external bodies. They will also need to discover where the digital objects which need preserving are, and use a typology (see Appendix 3) to classify them.
- Once responsibility has been accepted, it has to be decided whether the organisation will carry out the preservation itself or use an outside body to preserve the object.[38]
- An institution will have to take responsibility for the preservation of material produced by the institution, and may need to take responsibility for the preservation of commercially published material of importance to the institution which falls outside legislation guidelines.
- An electronic records management policy, which flags significant documents early in the lifecycle of the document, will ease the selection of digital objects for preservation.

---

[36] Where the hypothetical room survives for a thousand years

[37] Smith, Abbey, "Preservation in the Future Tense" in CLIR Issues Number 3, May/June 1998 Washington D.C. Council on Library and Information Resources

[38] If legislation for the deposit of published digital material is introduced in the UK, legal deposit libraries will preserve some commercially published material. However, some commercially published material may fall outside legislation guidelines.

- Where an institution cannot or will not preserve the digital object for which it is seen as having responsibility, it needs to ensure that a reliable structure for preservation of the object is available.
- The preservation of published material may be undertaken by an individual institution, or by a group of institutions which feel that preserving such an object is important.
- Larger groupings may be appropriate for preservation responsibility. It seems likely, that in such an event, CURL libraries will be key players.

4.4 Intellectual Property Rights

This section aims to assist librarians and archivists when they need to address the intellectual property rights issues which arise during digital preservation activities. Digital preservation activities may include actions which infringe intellectual property rights. Cedars offer suggestions and advice which arise from the Cedars experience. These will look at general rights issues in the United Kingdom, with an emphasis on copyright, rights negotiation, and preservation licence issues.

4.4.1 General Rights Issues

4.4.1.1 What is Copyright?

Copyright can be thought of as a bundle of economic rights and moral rights. The basic framework of these rights is statutory, although the explanatory case law is of great importance.

Copyright can cover many types of creative effort, which may include plays, paintings, sound recordings, and the typographical format of certain published editions. Several copyrights may subsist simultaneously in a single item, and care should be taken to establish the exact nature and ownership of each.

For example:

1. The words of a song will be protected as a literary wok, and its music as a musical work: these copyrights might be held by different people, and might have different commencement and expiry dates.
2. In a book of essays contributed by different authors, each essay would be a separate literary work, the compilation itself would also be protected, as would the typographical format.
3. In a multi-media digital object, each of the content elements may be subject to separate copyrights, and the rights might be owned by different people.

The usual term of protection for copyright is 70 years from the end of the year of the death of the author.

4.4.2 Other Rights Issues

4.4.2.1 Rights in Databases

Certain databases, which by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation, are protected by copyright as a literary work.

A new right, called database right, to prevent unauthorised extraction, has been introduced for all databases, whether they enjoy copyright or not. Database right applies where there has been a substantial investment in obtaining, verifying or presenting the contents of the database. The term of protection in this case is only 15 years, but may be renewed if there is a substantial change to the database. Therefore, in principle, the database right can last indefinitely.

4.4.2.2 Legal Deposit

Legislation requiring legal deposit of published digital objects is under discussion in the United Kingdom. A voluntary deposit scheme for published digital objects is being used to explore the issues involved.

Legislation in the United Kingdom for the legal deposit of published digital objects will not solve all digital preservation rights problems. Many libraries will still need to negotiate rights for the preservation of commercially published material which is of importance to their institution, since it is certain that some digital objects will fall outside the legislative boundaries for compulsory legal deposit.

4.4.2.3 European Union Copyright Directive

At the time of writing (March 2001) a draft European Union Copyright Directive is currently being negotiated. This directive is likely to change what libraries are allowed to do in order to preserve items under copyright.

The proposed Directive seeks to offer protection to copyright works where these are distributed electronically. The aim is to facilitate cross border trade in copyright works. Member States will be permitted to apply various exemptions from the basic restrictions on copyright, although they may also have to provide fair compensation for copyright owners.

The draft Directive permits member states to limit the copyright owner's exclusive right of reproduction "in respect of specific acts of reproduction made for archiving or conservation purposes by establishments which are not for direct or indirect economic or commercial advantage, such as, in particular, libraries and archives and other teaching, educational or cultural establishments."[39]

It is not yet clear whether the United Kingdom intends to take advantage of this exemption, or, if it does, what form such an exemption would take. The wording of any exemption will be a contentious matter, and considerable lobbying from all affected parties is to be expected.

Until these issues are resolved, Cedars can do no more than flag the approaching debate.

4.4.3 Who Owns the Copyright for This Work?

The basic principle is that the author of a work is the first owner of any copyright in it. The main exception is where a work is made by an employee in the course of employment, where the copyright is first held by the employer. This may

[39] Article 5(2)(c).

not necessarily be the case for work produced by scholars employed in institutions of higher education. Copyright may be transferred to others in subsequent transactions.

4.4.4 Has Copyright Been Infringed?

The copyright owner has the exclusive right to do certain acts, including copying the work, and renting or lending the work to the public. Infringement occurs when a person does, or authorises another to do, one of the restricted acts, to all, or a substantial part of the copyright work, without the licence of the copyright owner.

4.4.4.1 Copying

Copying involves reproducing the work in any material form. For films and TV programmes copying includes a photograph of any image. This would include, for example, posters and postcards of a single frame.

4.4.4.2 Lending

Lending can be thought of as making a copy of the work available, through an establishment which is accessible to the public, for use by a person, without charging the person who is borrowing the work. Certain libraries have special privileges regarding lending, as will be explained.

4.4.4.3 Renting

Rental can be defined as making a copy of the work available for use by a person, and charging the person who is renting the work.

4.4.5 Is There an Excuse for the Copyright Infringement?

Copyright law has to balance the interests of the copyright owner against the needs of the public for access to copyright material. Some acts – which would otherwise constitute breach of copyright – are allowed. Some are expressly permitted by statute. These are known as permitted acts. Other permitted acts have developed at common law.

4.4.5.1 General – fair dealing

Fair dealing with certain types of works for the purposes of research or private study does not infringe copyright in the work. Fair dealing with databases is not permitted for commercial research. There are similar provisions to allow fair dealing for the purpose of criticism or review, and to allow the reporting of current events.

The rules are founded in statute law, but the case law is of considerable significance in determining the boundaries in each case.

4.4.5.2 Education and libraries

The provisions dealing with educational establishments and libraries exemplify the Act's attempt to balance the interests of the copyright owner against the legitimate needs of the public. We will discuss certain provisions of particular relevance.

4.4.5.2.1 Libraries

Only prescribed libraries benefit from these provisions. Libraries are prescribed by statutory instrument. For example, all university libraries will be prescribed libraries.

Copying Under certain prescribed conditions, prescribed libraries may make copies of articles in periodicals, and of parts of published works. These conditions include the librarian's being satisfied that the copies are only supplied to persons requiring them for the purposes of research and private study.

Lending Copyright in a work is not infringed by the lending of copies of the work by a prescribed library or archive which is not conducted for profit.

There is no similar exception for rental, so it is important that libraries should not inadvertently or deliberately charge sums which convert lending into rental.

4.4.6 Rights Negotiation for Submission

Preparing for Negotiation

1. Discover who can authorise the preservation of the digital object in question. Obtain accurate contact details for this person, taking into consideration the possibility that there may be several rights owners.
2. Discover and document actions one needs to be able to take in order to preserve access to the intellectual content of the digital object.
3. Find out why the person who could authorise a request to preserve an object might refuse to do so. Prepare reasoned arguments, and explanations of technical preservation methods, to persuade them that authorising preservation will not cause them, or their institution, commercial harm.
4. Request permission to preserve access to the intellectual content of the digital object.
5. Keep detailed records of all steps in the negotiation process.

> **Sample letter**
> Sending this letter may be sufficient in order to obtain rights to preserve an object. If the rights owners agree to the terms in this letter, the institution will be able to carry out any actions necessary in order to preserve the digital object.
>
> ---
>
> Dear (name of negotiator for owner of rights)
>
> In order to ensure continued access to the (name of the digital object), (name of the archiving institution) requests permission to preserve the (name of the digital object) in a digital preservation archive. The sole reason for preservation of the (name of the digital object) is to ensure continued access to the digital object in the long-term. Access to the object would continue to be granted only in accordance with the terms of applicable licences (if any), and in accordance with prevailing law.
>
> I would be happy to answer any further questions you may have, and may be contacted directly at the address below. I look forward to hearing from you.
>
> Yours sincerely,
>
> (Name and contact details of negotiator for archiving institution)

We may be saved all this trouble in some cases, because preservation licences may be negotiated with rights owners by national higher education bodies.

If the rights owners are unwilling to agree to a licence on the terms indicated in this letter, then an individually drafted licence will have to be negotiated before preservation can take place. This will allow us to take account of the particular concerns of the rights owners. This brings us to the licensing agreement.

4.4.7 The Licensing Agreement

4.4.7.1 The Archiving Institution

The archiving institution should seek a licence allowing it to take all necessary steps to preserve access to the intellectual content of the digital object. The nature of these steps should be clearly specified. They will include the technical and other actions defined in pre-negotiation preparation. It may be helpful to reiterate that the archiving institution does not seek any alteration to the existing conditions under which an authorised user is given access to the digital object, and that these will continue to be respected.

4.4.7.2 Points for Consideration

The following general points should be considered when a licence is being drafted. The detailed drafting of the licence should be carried out with the assistance of a legal advisor.

1. If the digital object is protected by intellectual property rights, the archiving institution must seek a licence before any attempt to preserve the digital object.

2. During the preservation process the archiving institution may need, for example, to store, translate, copy, or re-arrange the electronic form of the digital object to ensure preservation. Care should be taken to ensure that all actions necessary for the preservation process are permitted under the terms of the licence.
3. If necessary, it is possible to draft a very limited licence which allows access for the purposes of preservation only, and for no other purposes.
4. It may be appropriate to display a notice to authorised users accessing the preserved object. The notice could warn them of the rights status of the object, or permitted actions, or both.
5. Right owners will be concerned to ensure the integrity of the digital object. The archiving institution needs to consider the technical and administrative means by which this can be guaranteed. An institution which wishes to preserve digital objects will have to respond in a flexible manner to the needs of the rights owners.

4.4.8 Summary of IPR Issues

An institution should ensure that purchase of a digital object, or the licence to use it, includes the right to preserve access to the intellectual content of the digital object. It may need to negotiate the right to preserve a digital object, and it must maintain current information, relating to the owners of intellectual property rights in the preserved digital object, in the metadata concerning the preserved digital object.

In order to ensure the preservation of digital objects, librarians and archivists must consider the relevant aspects of current intellectual property rights. They will have to assure the rights owners that the commercial value, and the integrity, of the preserved digital object will not be lost. By attending to rights issues, they will reduce risk for their institutions and encourage the preservation of digital objects.

4.5 Selection of Material for Digital Preservation

It is difficult to discuss preservation management for digital materials without reference to how materials will be chosen for long-term preservation. Although mass storage technology is advancing, allowing more data to be stored for lower costs, the sheer volume of digital material created means decisions need to be taken about what is preserved. Currently, selection decisions should be pragmatic. The decisions have to be based upon the estimated value of the material, the cost of storage and support mechanisms, and the production of metadata to support the material.

4.5.1 Based on What Criteria?

Selection of material has always been a fundamental element in collections management. For traditional materials for instance, different types of material are selected for different reasons – e.g. paperback copies might be selected for access purposes (where their eventual destruction through use is a given) while a leather bound hard copy of the same work would be chosen for the purposes of long-term retention. Libraries and archives are both familiar with these decisions.

Although libraries will not choose to preserve everything, they are still in the process of identifying what to preserve and what to ignore. We know that there is some digital material that is

- currently “high use”
- the type of material (typically commercially published scholarly works) that we would expect to preserve if it were published in traditional printed format
- tied to the long-term or cultural interests of the organisation.

Explanation and caveats need to be made.

4.5.1.1 High Use

As a starting point preservation of material needs to include material which is of high value now. Material which is currently low use and may have a long period of low/no use after preserving may represent an investment in preservation that an organisation is not willing to take. Libraries should be concerned with identifying High Use materials and taking safeguards to ensure they are preserved. It has to be recognised that low use materials may be valued in the future. Hard decisions may need to be taken about what is systematically kept and what is not.

At present, there is material which is purchased by organisations with a view to a "short lifecycle" – material for which demand is expected to be (possibly very) high but for only a relatively short period of time. There are also of course, materials bought with a view to retention and use over the long-term.

4.5.1.2 Cultural Interests of the Organisation

Selection for preservation will need to be closely tied to the long-term research or cultural interests of the organisation. Where possible, selection of materials for long-term preservation should reflect the broader institutional mission if policy makers and paymasters are expected to include digital preservation on the strategic agenda.

Although libraries currently apply selection criteria for acquisitions and (increasingly) for digitisation we must avoid the pitfalls and limitations of policies which work for printed materials but may not be appropriate for new digital materials. Despite this fact, it is undoubtedly useful for institutions to learn from existing policies.[40] Digital materials often extend far beyond the scope of traditional materials and, as our society becomes increasingly "informationalised", digital data will continue to grow exponentially.

> Consider the data that is collected each time you visit a supermarket: in future, this information could be of real value to researchers studying our society as consumers.
>
> Who will store this potentially valuable research data?
>
> If selection criteria are based too much on current research interests and needs we could stand to lose a great deal of valuable data.
>
> What material will scholars will be using in 20 years time?
>
> It could be interesting to ask collection mangers to look at a particular research collection within their purview and ask them to imagine what resources will be included in that collection in 20 years time.

[40] Such as the selection criteria used by the National Library of Australia for their PANDORA project and UK based work ongoing at Oxford on selection of material for digitisation. Policies also exist for many of the research councils (e.g. NERC) and for the Arts and Humanities Data Service. Librarians can also learn a great deal from the policies in place in other archives or repositories (e.g. the Public Record Office).

4.5.2 Why is Selection of Resources Difficult?

Because the digital future is so potentially volatile and unpredictable, it is very difficult to make value judgements about digital materials. However, if discussions about digital preservation begin by trying to consider everything it can be overwhelming and may inhibit any action at all. A balance needs to be struck between work that is achievable and work that takes into account the broader picture and is, therefore, applicable and scaleable to different situations/materials.

There is a pressing need for some action now. The reactive approach that many libraries take to the acquisition of digital material may be dangerously short sighted. In a print environment, a decision to acquire can be implicitly a decision to preserve. However, for non-digital materials generally libraries have the luxury of making a separate preservation decision later. Some key points with respect to digital materials are:

- Acquisition and preservation need to be closely linked because if a preservation decision is not made at the time of acquisition, it may be too late to consider preserving it later on.
- There are risks associated with making acquisitions decisions based solely on the desire for immediate access to the material rather than also considering the long term view and long term access.
- Decisions about format for deposit and arrangements for licensing material will need to consider preservation issues at the time the material is acquired. For example, if a library agrees to a licence which is based on access only, it may find that when the licence expires, there is little, if any, scope for preserving the material to provide long-term access for scholars of the future.
- Ownership of digital material is sometimes difficult to understand. Responsibility for preserving material has traditionally been inextricably linked to ownership. If we aren't clear on who actually owns a resource, it stands to reason that responsibility for its preservation is uncertain.

4.5.3 How Can We Select?

Discussions about how material should be selected invariably relate to the role of the organisation and the level of responsibility institutions take regarding particular digital materials. Understanding the role and therefore the responsibility each organisation takes in relation to resources within its collection is a key factor in moving forward with digital preservation. One approach to the selection of resources suggests that the decisions should be driven by the type of digital asset to be considered. An audit of an institution's resources, for example, might include detailed consideration of what material there is based on the following initial assessment:

- What material is owned by the organisation through purchase agreements?
- What material is created by the organisation?
- Assets related to the management of the organisation (e.g. electronic records)

This initial assessment would then be followed with consideration in all of the above categories of what will be around in future. The following criteria could be applied to this inquiry:

- Financial value
- Legal value
- Value to the organisation's business/mission

- Value to others

All organisations are likely to include resources in the first three broad categories (owned, created etc.) and will be well placed to analyse their future value based on the criteria list provided (financial, legal and so forth). Following on from this examination, organisations would need to engage in a detailed cost/benefit analysis which should focus on the inherent risks if material is not preserved.

Digital preservation is an issue for more than just libraries and this must be reflected in any digital archiving model. Although traditionally the curatorial professions have kept separate agendas (for reasons often associated with funding) it is time to bring archives, libraries, museums, galleries etc together to develop a shared agenda for long-term preservation of digital materials. In any model for the preservation of resources, there are economies of scale associated with a consortial approach. Existing co-operative organisations such as CURL and RLG are potentially in a very powerful lobbying position to mobilise the research community as well as the policy makers – action must focus on both levels.

4.6 Recommendations

1. Further work needs to be done on criteria for selection of material. However this cannot be done in isolation, This will need to be part of a programme of digital preservation policy development – selection criteria should reflect the overall aims of collection development.
2. If there is material which is currently of high use which is at risk of being lost, libraries should be concerned with identifying that material and taking safeguards to ensure it is preserved.
3. Effort invested now will make it possible to save our waning digital heritage.

Chapter Five : Technical Standards and Related Guidance

5.1 Introduction

Digital preservation, as a discipline, is in desperate need of standards and guidelines. This will allow for consistent preservation activities and confidence that previous mistakes made by others will be avoided. Most digital preservation solutions to-date have been built in an ad hoc manner to meet specific institutional needs. Cedars has attempted to uncover which strategies are available and, where possible, to indicate the shortcomings or applicability of these to a good long-term preservation archive.

The foremost source of ideas and terminology used by Cedars is the CCSDS Reference Model for an OAIS. This OAIS model is well conceived and flexible enough to allow the most suitable solution for a designated community.

At this stage Cedars, very deliberately, avoids providing absolute recommendations or a step by step solution for digital preservation. The Cedars project maintains that such a solution can only be achieved once a good understanding of the issues involved and the pitfalls has been gained. As such, it is best that institutions still wrestle with this complex field so that the eventual solution will ensure that an AIP is kept in a way that makes it useful forever.

In this chapter the Cedars metadata for digital preservation is outlined, and then various technical methods are given for ensuring that a preserved digital resource will survive the march of both technology and time. By providing the correct metadata and the correct technical solutions (and the ongoing archival maintenance) an AIP can truly remain accessible.

5.2 Preservation Metadata

A major component of the work encompassed by the Cedars project was the development of a metadata framework to enable the long-term preservation of digital materials. This metadata is required to support every aspect of preservation as well as meaningful access to the archived digital content. This metadata includes descriptive, administrative, technical, and legal information. A copy of the complete metadata specification document can be found in Appendix 5.

5.2.1 Background

In 1998 the Cedars project produced a preliminary review of metadata initiatives that were identified as being relevant to digital preservation[41]. This review noted that metadata is needed for digital preservation, regardless of the particular preservation strategy chosen. Clifford Lynch has described the function of some of this metadata:

> *Within an archive, metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information. This metadata will also be maintained and will be migrated from format to format and standard to standard, independently of the base object it describes.*[42]

The OAIS model has identified and distinguished various types of metadata needed to support a digitally preserved resource. Each resource is packaged together with its metadata, as an **Information Package**. An Information Package

[41] Day, M. "Metadata for Preservation" 1998 http://www.ukoln.ac.uk/metadata/cedars/AIW01.html

[42] Lynch, C. "Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information" 1999 http://www.dlib.org/dlib/september99/09lynch.html

combines two things: **Content Information** and **Preservation Description Information** (PDI). The Content Information groups the **Primary Digital Object** (the preserved digital resource) with **Representation Information** (RI): the RI is the metadata needed to retain meaningful access to the preserved data object. The PDI is sub-divided into groups of different kinds of descriptive metadata, so that what the Primary Digital Object actually is can still be understood indefinitely.

5.2.2 The OAIS Structure for Digital Preservation Metadata

The generic structure of preservation metadata in the OAIS model is illustrated below in figure 5.1 (This is essentially a copy of figure 3.3).

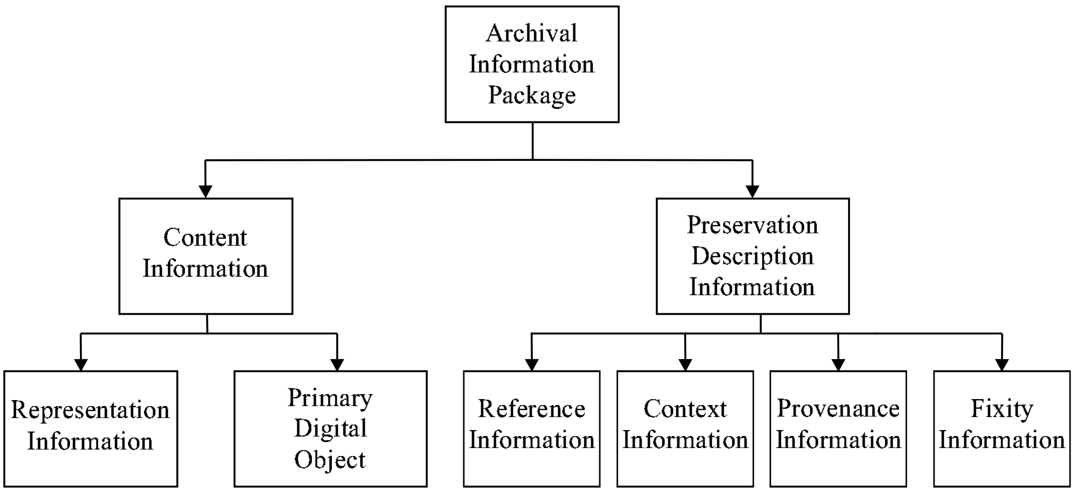


Figure 5.1: Hierarchy structure of metadata in the OAIS model

5.2.2.1 Content Information

The information that is the primary object of preservation. Containing a Primary Digital Object and any Representation Information (technical metadata) needed to transform this object into meaningful information.

5.2.2.2 Primary Digital Object

This is the digital resource held in a bytestream for storage purposes. This contains all of the Significant Properties of the original digital object.

5.2.2.3 Representation Information

Representation Information contains all the metadata needed to obtain and render the intellectual content of the stored digital object.

5.2.2.4 Preservation Description Information

Any other information (metadata) which will allow the understanding of the Content Information over an indefinite period of time.

5.2.2.5 Reference Information

Reference metadata contains all identifiers (e.g. ISBNs, SICIs, UPCs) for the resource and additional information describing the resource. This also includes existing catalogue metadata records (e.g. MARC).

5.2.2.6 Context Information

Information that documents the relationships of the Content Information to its environment. This includes how the Content Information relates to other Content Information objects existing elsewhere, as well as how it relates to non-archived objects (e.g. documentation or finding aids or links to other objects within the collection).

5.2.2.7 Provenance Information

Provenance Information is metadata that documents the history of the Content Information. This information tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. For Cedars this element also contains information about the reason a resource was created and why it was preserved.

5.2.2.8 Fixity Information

This element will be used to prove the authenticity of an Archived Information Package, for example, by use of a checksum or a digital signature.

5.2.3 Granularity Issues

It has been widely recognised that the metadata required for long-term digital preservation is complicated by the levels of "granularity" that can occur within a single digital object or collection of objects. Metadata may be assigned, for example, at the level of a complete digital collection, a single digital object or even (in the case of complex digital material) at the individual file level. In part the granularity of the metadata will be determined by the digital object itself and the level of description necessary to ensure preservation, but it will also be influenced by collection management policies in place at the archive. In addition the granularity of the metadata may be influenced by concerns about rights management of some more complex digital objects (e.g. where different parties own different components of the content and/or systems[43]). How an archive chooses to assign metadata, and at what level of granularity, are not decisions imposed by a metadata specification. The Cedars preservation metadata specification allows for description at any level but ultimately the decision resides with the archive. While the Cedars Outline Specification makes no assumptions about the level at which metadata will be assigned it does assume that archives will do so at levels appropriate to fulfil their preservation responsibilities and meet the needs of the archive's user communities.

[43] For example, a digital sound recording with associated copyright and performance rights for individual tracks

5.2.4 Obligation

In early iterations of the outline specification the Project attempted to define all metadata elements as either mandatory or optional. However, due to the rigid and over-prescriptive appearance of these designations, some material might have been excluded from preservation because they lack specific metadata elements. To prevent this the specification provides guidance on the “significance” of each element determined by the extent to which the element may be usefully applied across a wide range of digital materials. In addition, specific types of digital material should consider some elements mandatory while these are not applicable to other types of digital materials. The terms used are as follows:

- **Very Significant** an element deemed to be very useful for preservation across a wide range of digital objects and therefore necessary if available
- **Significant** an element deemed to be of use to most digital objects
- **Less Significant** an element deemed to be of some use but not strictly necessary.

Structured “walk-through” meetings to assign metadata to different types of real digital materials did conclude that there should be a minimal set of metadata that was deemed mandatory. This was suggested to be at least a date of some kind, either the official timestamp of ingest or the date of creation. Such a metadata element also enables a future user to establish the environment (both technical and political) of the original object.

5.2.5 The Cedars Project Stance on Metadata

The OAIS model provides an important separation between metadata elements based on the purpose that the metadata elements serves. The Cedars project sees this distinction as very helpful. While the Cedars outline metadata specification is intended to aid in the preservation of digital resources, it is hoped that organisations will add elements that address aspects of PDI or RI needed by their resources. One idea that was useful in the Cedars demonstrator archive was that of incorporating a free text running commentary into the metadata. This proved useful for both PDI metadata and RI metadata. The emphasis on metadata is to adequately describe the resource for any potential uses in the future, so any metadata requirements should not preclude (or even discourage) the preservation of the resource itself.

5.3 Technical Strategies for Continuing Access

Digital preservation refers to the series of managed activities necessary to ensure that an AIP remains safely in storage, and where necessary can be accessed using current technology. No preservation strategy is complete without a technical strategy to ensure continuing access (which by necessity includes ensuring that the bytestream is successfully stored). Broadly speaking, there are three main types of technical strategy currently employed for providing continued access to archived materials. These are technology preservation, emulation, and data migration from one format into another. Often it is asked, which media is best to ensure a successful long-term storage, and should the storage bytestream be compressed or not. Each of these technical strategies and these questions are addressed in turn below.

5.3.1 Technology Preservation

The strategy of technology preservation means that to provide continued access to a primary digital object, a system that originally understood the primary digital object is also preserved. This means that specific hardware and software needs to be conserved. A method is also needed to transfer the bytestream (in long-term storage) from current (and unknown future) technologies to the conserved system. (Unless the original media is also conserved, see the later section on media longevity.) For example, this might require a working 32-bit Intel architecture configured to run the Windows 95 operating system. This is like maintaining an old record player in order to continue listening to an album collection. All the mechanical parts and power supplies need to be kept in a working condition (which means ongoing maintenance, testing springs and cleaning rust), and the vinyl disks need preserving. At best, this is a short-term solution providing access to a valued resource. If all of the hardware, software, and configurations are conserved then this does ensure that the material is accessible.

Technology preservation can never be seen as a true long-term preservation strategy. Some of the issues that make this strategy problematic are:

- Storage space requirements for the hardware (never decreases)
- Maintenance expertise needed for hardware (and software configurations)
- Very few disaster recover strategies (hope to make or find a working system)
- Prohibitive costs for replicating a hardware system
- The use of the resource is restricted to physical access to the hardware system

To consider a bleak (and yet real) example:

> The BBC Domesday videodisk system requires
>
> - **Hardware**
>   - A British Broadcasting Corporation Master Series Microcomputer
>   - A BBC specific colour monitor (possibly an RGB monitor at the correct sync rate)
>   - A BBC LV ROM player (a modified Phillips LV videodisk player)
>   - A 5¼″ disk drive
>   - An Acorn tracker ball
>   - Connecting cables (several rare cables, including fragile ribbon connectors)
> - **Software (mostly in Rom or cartridge forms, so is this actually hardware?)**
>   - The Video File System (VFS)
>   - A BCPL interpreter
>   - The Domesday software
>   - Data and software on 4 sides of 2 video disks

There are too many critical components that can fail. Even if a moving part like, for example, the platter for holding the video disk malfunctions (so the disk cannot be turned over or swapped), the entire system is inoperable. To provide a recovery solution for when one of these components fails, the hardware design (every device) must be adequately documented. Unfortunately the inadequacies of such documentation are only discovered when a major recovery relies on these.

Although Cedars does not consider this strategy as part of long-term preservation, there may be circumstances where it provides the best short-term stop-gap until long-term preservation can be applied. If the only environment in which a valued resource can be understood is on a single hardware system (configured in a particular way) then to preserve the resource the hardware system must also be preserved. If the resource is truly valuable, effort must be invested to move from this short-term solution to a longer term preservation strategy.

5.3.2 Technology Emulation

The role of technology emulation as a long-term digital preservation strategy remains uncertain, as little practical work has yet been done to test this strategy. The gist of this strategy is that, to preserve access to a digital resource, an emulation of its native platform is developed for current platforms. Figure 5.2 (borrowed from "Emulation, Preservation and Abstraction"[44]) shows how an emulator mimics the original platform while running on a different host platform. First the original digital object is preserved in an AIP. The primary digital object component of the AIP can then be run under emulation to give adequate access to the Significant Properties of the original. Once Emulator 1.00 has been developed, access to the primary digital object will be ensured even as the original platform becomes obsolete. The CAMiLEON project recommends that emulator 1.00 is implemented in such a way, that the transition from emulator 1.00 to emulator 1.01 (as host platform 1 also becomes obsolete) can be implemented economically. Note that the equality sign indicates that the access to the Significant Properties given by emulator 1.01 is equivalent to the access given by emulator 1.00.

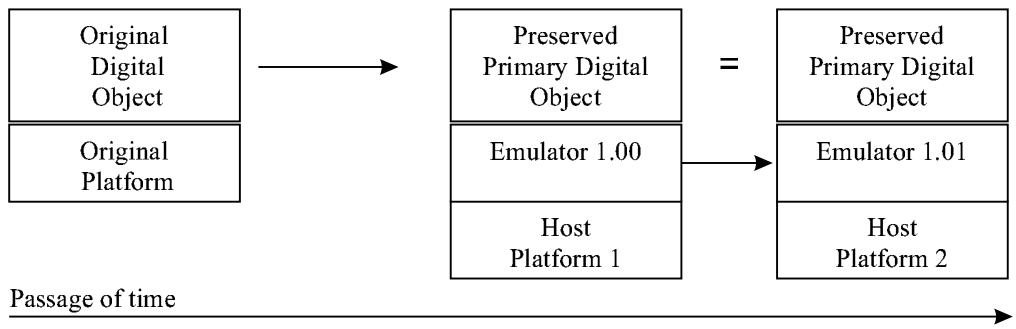


Figure 5.2: Using an emulator to mimic an obsolete platform

Holdsworth and Wheatley[45] summarise the debate on emulation so far.

> *"Bearman has cited emulation as a dangerous strategy that fails to actually preserve digital objects and is not a realistic approach due to the enormous costs of emulator development."*

> *"Rothenberg has proposed that the work involved in the production of an emulator can be considerable, so it can be postponed until resources are available by instead producing an emulator specification at the time of platform obsolescence."*

However there is no practical evidence for this lack of optimism. Reliance on an adequate specification of a platform is a very risky approach to preservation, for the same reasons as it is risky for the technology preservation strategy. The appendix of Holdsworth and Wheatley[46] gives evidence of such risks for the George3 platform.

Much work has already been done by hobbyists and enthusiasts on emulation, and this seems contrary to the astronomical costs implied. Indeed, there are some commercial products that emulate Wintel type platforms on non-Intel platforms.

The development of emulators is a skilled task, as is their maintenance and support. However, the requirements are less arduous than technology preservation. Certainly, in technical terms, emulation truly is a realistic and practical approach to the preservation of some digital materials. As with any tools referenced by the Semantic Information component of the AIPs Representation Information (see chapter three) there is still the need to monitor these tools to protect from the threat of creeping obsolescence. This is all part of managing the Gödel ends of the representation networks, which is a standard archive administration activity.

[44] Holdsworth, D. and Wheatley, P. "Emulation, Preservation and Abstraction" www.leeds.ac.uk/CAMiLEON
[45] Ibid.
[46] Ibid.

Unlike Rothenberg, both CAMiLEON and Cedars see the importance of taking early account of the prospects of an emulation strategy, even when the original platform is still available. This may involve implementing an emulator to allow operational comparisons between a resource on the emulator and on the original platform.

The CAMiLEON project asserts that choosing an appropriate abstraction is vital to successful emulation. The extent to which the emulation mimics the original technical environment or emulates only those components necessary to access the data also needs to be decided. CAMiLEON identifies a number of factors that should influence the choice of the abstract emulation interface.[47]

- Complexity (or lack of it), in the hope that the emulator implementation is straight forward
- Availability of documentation of the chosen interface
- Mapping of peripherals into easily specified abstractions
- Retention of Significant Properties

Further recommendations are:

- Emulator code should be produced using standard Software Engineering techniques. These include the use of a good code structure, informative and plentiful commenting and good documentation.
- Emulators should be written in open source, preferably in a subset of C chosen to be compatible with future programming language semantics.[48]
- Where non-standard code (for example to render the raster display) is incorporated in an emulator, this should be modularised and well documented.

Cedars suggests that an emulation strategy is very suitable for some classes of digital resource. All of our current BBC content relies on emulation to provide access to the intellectual content. The example of the BBC Domesday video disk given above lends itself to emulation.

For some classes of resource emulation is not necessary, or indeed appropriate. Examples of classes of resource where emulation is not necessary are ASCII text documents, TIF images, and PDF documents. All of these have rendering tools on current platforms, and also have open source implementations of these rendering tools. Where an open source rendering tool exists for a format this can be transferred onto new platforms as old platforms become obsolete.

5.3.3 Migration

Unlike the two previous strategies, which provide an environment in which the digital resource can be understood, migration moves the digital resource into a format that is easier to process on current technology (and indeed future platforms). There is some confusion over the definition of migration for digital preservation. A report commissioned by the Research Libraries Group and the Commission for Preservation and Access in the US, helpfully distinguishes between migration and what has been termed refreshing. The report suggests that

> *"…Migration is a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation."*[49]

[47] Ibid.
[48] Holdsworth, D. "Emulation: C-ing Ahead" www.leeds.ac.uk/CAMiLEON/
[49] The Research Libraries Group (RLG) Final report of Working Group on Preservation Issues of Metadata. 1998

Here, we are concerned with transformation migration, defined in the OAIS as:

> *"A Digital Migration where there is some change in the Content Information . . . while attempting to preserve the full information content."*[50]

This is further broken down into two categories, reversible migration and non-reversible migration. While this is a crucial distinction, it is also necessary to judge whether a migration can actually be reversed once the original migration software and platform are unavailable.

Many libraries and archives are already involved in migration strategies to ensure continuing access to digital resources, and believe that this is the most practical approach – at least for the short to medium term. Rather than focussing on the technology, this strategy tends to focus on intellectual content and making it accessible using current technology.

The Cedars project takes a more cautious (but we believe sensible) approach to migration. This aims to preserve both the original digital object and a tool for understanding or interpreting the digital object.

For access, rather than preservation, it is possible to migrate a resource into a preferred format and to store this new version of the resource. However, this does not imply that the original format should be discarded – indeed it should be preserved in case a new migration or emulation tool becomes available that provides a new way of interpreting the intellectual content.

The most critical assessment that needs to be made in light of a migration is whether the Significant Properties of the digital resource are maintained. Paul Wheatley[51] looks in more detail at the different strategies that are available for migration, and the applicability of these to different classes of digital resource. In this paper Wheatley also introduces a new concept for digital preservation which is related to migration, that of resource recreation. This involves recoding and perhaps retyping and typographical entry to manually create an equivalent resource on a current platform.

Migration provides the potential benefit of reducing the number of formats that an archive needs to maintain access to. Where the chosen target formats are very widely used, and have well documented definitions or open source implementations of rendering tools, long-term preservation and accessibility becomes much easier. Widespread formats are likely to motivate the development of rendering tools or migration tools (into current formats) when modern platforms no longer understand these formats.

Cedars also suggests the use of **Migration On Request Tools** (see Chapter three) to preserve the original digital object in its original format, and then to provide for access in whichever format is desirable.

5.3.4 Standard File Formats

Migration is closely linked to the notion of standard file formats. For certain classes of digital resource there are standards that capture all of the Significant Properties. An archive could request that all resources that can be captured in a standard format must be submitted to the archive in this format, for example:

- Text documents with layout, including emails (RTF or PDF)
- Database tables (CSV)
- Colour images (TIF)
- Sound (AIFF)

[50] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK
[51] Wheatley, P. "Migration – a CAMiLEON discussion paper" www.leeds.ac.uk/CAMiLEON/

- Family history (GEDCOM)

However, these formats will not capture all of the Significant Properties of all of the digital resources in these categories. There are also problems with versioning, where with PDF the format has changed across different versions. Another reason why an archive should not be so prescriptive is that resources not fitting into any of the standard file formats will not be preserved (the list above does not mention multimedia web pages or interactive educational software).

At the moment there is a risk involved with not knowing the 'futuricity' of any standard file format. As akin to migration, where there are open source tools or a widespread userbase for any file format the longevity of that format is favourable, but this will not necessarily be continued.

As always, there needs to be the infrastructure to monitor the Representation Network to ensure that these formats can still be understood.

Only once the impact of standard file formats on preservation of Significant Properties is assessed, and the longevity of tools to work with these is established, should this approach be taken.

5.3.5 Media Longevity

The Cedars position on media for data storage is that longevity is not the issue. The timescale over which the Cedars project is looking far exceeds the lifetime of any technology. Even where the medium itself has a life of decades, the devices to read a particular type of storage volume tend to become unavailable in much less than a decade. There are exceptions, which come about when a technology acquires an importance for data transfer, as has happened with 3½″ disks and 9-track ½″ magnetic tape. Even where this longevity happens, the attractiveness of the medium for data storage evaporates quite quickly as higher density devices become available.

The Cedars stance is that a bytestream can be stored forever. This storage will involve copying as part of the management of a data archive. Systems with this property are already around.

The essence of the storage aspect of Cedars is to concentrate on the representation information, so that the bytestreams of yesteryear can still deliver meaningful information long after their native formats are in common use.

5.3.6 Compression Techniques and Long-Term Archival Storage

While it seems a tempting cost model to ensure that the bytestreams placed in storage are as small as possible, this is deceptive.

Before a bytestream is compressed the impact of this action must be analysed. When the compression process is reversed (by uncompressing the bytestream), are the Significant Properties preserved? There are two types of compression algorithm[52]: lossy, for example JPEG, MPEG, and Indeo; and loss-less, for example RLE and Huffman coding (both based on LZ compression). Only loss-less compression is suitable for preservation storage.

As a caveat, lossy compression is no good for preserving critical data (such as an AIP) however where a resource contains components which are already in a lossy compression format (such as a HTML file tree containing JPEG images) these can still be preserved. File formats (and lossy compression) are decisions at pre-ingest, ingest, or decisions about format migration – these are separate from bytestream storage.

[52] Examples of compression algorithms are taken from Ladino, J. N. "Data Compression Algorithms" http://www.ccs.neu.edu/groups/honors-program/freshsem/19951996/jnl22/jeff.html

Loss-less compression is the only viable option for compression in the long-term data store. Loss-less compression may actually provide more advantage when applied to data transfer, either at ingest, when the resource is uploaded to an archival store, or at dissemination, when a resource is transferred to a consumer's computer.

If a policy is made to compress a bytestream in storage, then the loss-less algorithm should be entered into the Representation Network. Choosing the algorithm needs to be done on the basis of its potential longevity:

- Non-proprietary
- Open-source
- Well established (proven to be reliable)

However, the Cedars project found that the best policy was to not apply compression. Since both storage facilities that Cedars used (see Chapter three) use multiple tape libraries, these have loss-less compression built into the firmware of the tape decks. As the stored data is automatically (and invisibly) compressed and uncompressed any compression applied to the bytestream (AIP) prior to it being written to its storage medium actually makes this firmware compression less effective.

5.4 Recommendations

1. Follow the overall structure for metadata provided by the OAIS model
2. Allow free text commentaries in the metadata to allow anything that is not covered to be captured
3. Technology preservation is only a short-term strategy as a stop gap for true preservation
4. Emulator code should be well produced and well documented
5. Where migration is used, use reversible migration
6. Migration on Request tools should be developed
7. Standard File Formats need to be chosen well
8. Do not get side tracked by media longevity or compression

5.5 Further Reading

1. The Research Libraries Group (RLG) Working Group on Preservation Issues of Metadata – final report (RLG 1998)
2. The University of Pittsburgh Recordkeeping Functional Requirements Project, funded by the US National Historic Publications and Records Commission (Bearman and Sochats 1996)
3. The Recordkeeping Metadata Standard for Commonwealth Agencies developed by the National Archives of Australia (1999).
4. The logical data model (based on entity-relationship modelling) developed by the National Library of Australia (NLA) to help identify the particular entities (and their associated metadata) that needed to be supported within its PANDORA proof of concept archive (Preserving and Accessing Networked Documentary Resources of Australia, Cameron and Pearce 1998).

Chapter Six : The Costs of Digital Preservation

6.1 Introduction

Not a great deal is known about the costs of preserving complex digital objects over time. However, there is a perceived wisdom within the library community that it will be more expensive and more intensive than preservation of traditional library materials. This may not prove to be the case, as the costs involved in traditional libraries are also fairly unknown quantities.

Meaningful comparisons of the costs of digital vs traditional preservation may not be possible, or even helpful. It is certain that the costs of preservation of digital materials will be different than preserving other materials. The ongoing costs of digital preservation are also likely to span a different timeframe than traditional preservation and different approaches to digital preservation will prescribe quite different costs and timeframes.

This chapter will identify some of the main costs elements that libraries can expect to encounter when considering digital preservation as part of their ongoing collection management function. It is divided into two parts: the first provides an introduction and overview of some of the general issues associated with the costs of digital preservation, whilst the second breaks down the cost elements of digital preservation.

6.1.1 Long Term Preservation and Continuing Access

Again, it is reiterated, that long-term preservation and continuing access are different. The Cedars Project has focused its work on long-term preservation. This chapter looks more at the costs involved in long-term preservation than it does at the costs of providing continuing access.

6.1.2 A Timeframe for Digital Preservation

The costs of preservation always represent an ongoing commitment – whether for digital or traditional materials. However there is growing realisation that the time between an object's "creation" and its preservation is shrinking rapidly for digital materials. Preservation needs to be addressed increasingly at the time of acquisition, or even creation, of the digital resource. Although costs for digital preservation remain obscure, we can be sure that if the preservation of materials is considered earlier these costs will be reduced.

For digital materials, commitment must be made over the long-term. The meaning of this commitment will depend, in part, on the archiving model in which the preservation occurs. Regardless of variations in the archiving models, digital preservation will require ongoing resources. It is important to recognise that different approaches for long-term preservation will have different cost timeframes. For example, if an archive adopts a migration strategy which will move the digital object into current software, action (and therefore resources) will be necessary each time a software upgrade occurs. If this migration happens when the object is acquired by the archive, this is an up-front cost. However if migration is chosen as a method for providing continuing access, and the object is continually migrated to new software then this is likely to be an ongoing cost. By comparison if another type of migration strategy is adopted where materials on **Ingest** into the archive are migrated into standard formats then action (and therefore resources) to migrate that object will be required less frequently.[53]

Recommendation 6.1: Preservation of materials must be considered as early in the lifecycle of a resource as

[53] Similar comparisons can be done for emulation strategies. For example, where emulators are built and maintained within the archive at the time of Ingest; or where metadata describing systems requirements is stored at the time of Ingest and the emulator is engineered on demand.

possible. The point of creation or acquisition is ideal. Early preservation management will reduce costs.

6.2 The lifecycle of a digital resource

The lifecycle of traditional library materials looks something like this: a book is published; put on a shelf to be accessed; and preservation occurs only when the object begins to deteriorate. Digital materials have a different lifecycle. Ongoing activity is needed to ensure continuing access. The way a digital object is created influences how (or indeed whether) it can be preserved. Likewise, decisions taken at the start-point of preservation can impact on future access.

The preservation lifecycle of a digital object starts at the time its creation. For libraries involved in digitisation projects this means preservation of the digital files must be considered when the project begins – and must be budgeted for! Where no preservation actions are performed before acquisition, the length of time of inactivity increases the complexity of preservation activity needed (and hence the costs associated with such activity).

6.3 Cost/Benefit

The costs of preserving digital materials need to be considered in light of the relative benefits. Digital preservation costs inevitably involve trade-offs. Decisions to reduce costs could compromise the completeness of the preservation. However enormous outlay to preserve a complex digital object which is never accessed is also undesirable.

This suggests an approach to preservation which is appropriate to the perceived (estimated) value of the digital object. The other observation is that digital objects of less value can only be preserved where the initial outlay is cheap enough. The long-term value of digital materials is difficult to determine, as the decisions are required before the value becomes self-evident. Analysis of the benefits of preservation is also linked with collection management and selection policies.

One trade-off that can be made is in the quantity and quality of metadata gathered for a digital object. The better the metadata, the more ‘valuable’ and ‘useful’ the preserved object, but this will be more costly to produce and assess (quality assurance). The costs associated with metadata production can be shifted and reduced, either by taking metadata from the **Content Provider** at Ingest, or by developing automated techniques for metadata production.

The proof argument to cost/benefit analysis is: demanding the maximum, absolute ultimate in preservation will be self-defeating – when the process is made too expensive it will not be done.

6.4 Technical Considerations

The level at which a digital object is archived and maintained will impact on the costs associated with its preservation. The **Significant Properties** of a digital object as they have been agreed (see Chapter three) may alter the costs associated with preserving those specific properties. A digital object’s Significant Properties are not assumed to be empirical. The preservation of an object’s full functionality may prove more costly than just a bare-bones preservation of basic intellectual content. The question that has to be asked is whether the object’s long-term value is worth the long-term expense of preserving the ‘bells and whistles’.

A major factor in the cost implications of technical considerations is the economies of scale. Preserving many similar digital objects (whether complex or not) will be less costly (pro rata) than preserving several digital objects of unique types.

6.5 Breaking Down Costs of Digital Preservation

The cost of preserving an object will depend on many factors. There will be a multitude of considerations that impact on how these costs become manifest. Once the elements that influence cost are identified, these can be used to generate a decision-matrix.

6.5.1 A Conceptual Taxonomy of Archives

Using the OAIS taxonomy of digital archives,[54] this section considers which aspects of the archives activities will have significant bearing on the costs.

The main aspects in which archives vary is in the quantity, nature, and the relationship to the data it deals with. There are three main aspects to this:

Content, data types and formats Clearly archives may be devoted to a particular subject area. From a preservation perspective it is less important whether there is diversity in this respect than whether there is diversity in the types (and formats) in which the data is held. Where a greater diversity of data types must be supported, the greater the preservation challenge (and likely costs).

Access The quality and variety of Dissemination methods to an archive's holdings has implications on the costs of providing these. Where a particular community demands a particular interface and access mechanism these are likely to be obligatory costs.

Authority An archive can also be classified by its authority with respect to its digital objects. This concerns the archive's relationship with the Content Provider rather than the user of the Access function. There are two aspects to this: rights, whether the archive owns the rights to the Primary Digital Object or not; and control, can the archive dictate the types of digital objects which arrive for Ingest or must it accept whatever types or formats the Content Providers can produce.

This taxonomy shifts the costs goal-posts, for example with Authority, where an archive does not own the rights to the data it may incur higher costs of negotiation (and periodic re-negotiation). Clearly expertise and experience in particular domains reduces costs associated with familiarising staff with the target Primary Digital Objects. With regard to Access, certain Dissemination methods can also be the means of receiving revenue – in this case the outlay is recouped. Of course, these three aspects are not independent. Figure 6.1 shows a simplified table of these aspects.

	Lower costs →	Higher costs
Data types & formats	Limited number	Many and various
Rights	Ownership	Non-ownership
Control	High degree	Low degree

Figure 6.1: Simplified costs associated with Data taxonomy

[54] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-R-1 RED BOOK. OAIS figure 4.11

6.5.1.1 Relationships between archives

The OAIS model discusses the issue of interoperability between archives. It distinguishes between four models for this. The choice of interoperability model will have associated costs, but also allows resources and costs to be shared between archives. These are the four models:

Independent Archives This serves a single designated community and is designed solely to meet the needs of that community. It may or may not be distributed.

Co-operating Archives Archives can exchange information, but do not necessarily require common Ingest or Dissemination methods. Higher degrees of co-operation are obtained where such common methods are adopted.

Federated Archives These allow search sessions across the whole federation of archives. The common catalogue (and manager) binds the federation together. Other types of federations (with different binding functions) are also possible.

Archives with shared Functional Resources This is the highest degree of interoperability and permits use of shared resources.

The degree to which co-operating increases or decreases overall costs varies with both the model of co-operation and the scales involved.

6.6 Cost Elements of Digital Preservation

The elements listed below are ordered in approximately the order they tend to occur within a collection manager's workflow. These elements are directly associated with preservation and ensuring long-term accessibility. It should be recognised that the costs of preservation and continuing access are inter-related. Having acknowledged that, it is not always possible to distinguish between preservation costs and costs for providing access, this list attempts to focus specifically on preservation.

1. **Selection for preservation** Collection Managers (e.g. archivists, subject specialists) and Systems Managers need to act in joint consultation to decide on policies, including relative values of digital objects and costs of specific technical issues.
2. **Negotiation for the right to preserve** This includes the staffing time for the negotiator and for the drafting and exchange of agreements. It also includes detailed consideration of the Primary Digital Object to ascertain the relevant rights holders (including those of software and underlying technology components).
3. **Negotiation for the right to provide access** Some time will also be required for negotiating Access arrangements, especially where short-term Access is envisaged. This may not apply to all archives (or all digital material).
4. **Applying the appropriate strategy for preservation** This includes time taken to ensure the Primary Digital Object is adequately prepared for archiving. As part of this, the Significant Properties need to be determined.
5. **Quality control and validation of AIPs** The completeness of any digital object needs to be checked. Time may also be required to check the documentation, and its correlation to the Primary Digital Object.
6. **Metadata production** All aspects of the metadata need to be populated, although not necessarily at the same time. This will range from the study of documentation and the PDO itself to automated analysis and harvesting of metadata. Metadata costs will also need to accommodate the gathering of rights management information.

7. **Archival storage** This includes maintenance and purchase of hardware and software as well as the periodic inspection of stored AIPs. All aspects of backup and movement between generations of media are also included in this element.

8. **Administering the archive** As well as general administration of the Ingest, Archive Storage, and Dissemination functions, this element includes the costs involved in keeping abreast of developments. Developments in technology and law will change the preservation and access requirements for some digital objects.

9. **General Costs** These include staffing (salaries, overheads, training, etc.), insurance, building overheads, meeting certification and compliance standards, etc.

6.7 Staffing Implications

Clearly one of the major costs of a digital archive will be the cost of the staff required to run and maintain it. Equally clearly this will be need to be a team which will posses several different kinds of skills. Based on Cedar's experience, there is a need for (at least) four major kinds of skills:

- **Curatorial** involved for subject specific decisions or decisions which involve collection management policies.
- **Technical** involved in implementing and maintaining the technical aspects of the archive and also involved in monitoring technological change
- **Legal** primarily involved when dealing with rights negotiation, although can also influence Access provision
- **Managerial** overall management of the collection and the preservation process

Some of these roles will involve senior staff whilst other roles may be suitable for more junior staff.

6.8 Recommendations

1. Preservation of materials must be considered as early in the lifecycle of a resource as possible. The point of creation or acquisition is ideal. Early preservation management will reduce costs.
2. Costs can be reduced by co-operating with other archives.

6.9 Further Reading

- "Scoping the Future of Oxford's Digital Collections," a study funded by the Mellon Foundation which includes criteria for the selection of materials for digitisation. http://www.bodley.ox.ac.uk/scoping/
- "Guidelines for the Selection of Online Australian Publications Intended for Preservation," The National Library of Australia http://pandora.nla.gov.au/scoap/guidelines.html
- "The Berkeley Digital Library Sunsite," http://sunsite.berkeley.edu/Admin/collection.html
- "Digital Electronic Archiving: The State of the Art and the State of the Practice," April 1999. http://www.icsti.org/icsti/whats_new.html
- "The Case for New Economic Models to Support Standardization Efforts," Clifford A. Lynch, http://www.niso.org

Definitions of Terms

Action: any activity associated with resources requiring preservation

Resources: Funding commitment either in the form of direct payment or human time and expertise

Collection Manager: used broadly to mean librarian, archivist, subject specialist, etc.

Systems Manager: used broadly to mean technical specialist

Preservation Strategy: the set of policies and guidelines that provide a regime suitable for long-term preservation of digital materials. This covers all areas of management and practical/technical implementation, and incorporates a strategy for **Continuing Access**.

Continuing Access: ensuring that technical approaches to obtain the intellectual content of a Primary Digital Object are available for current technologies.

Technical approaches to Continued Access: Currently there are three broad technical approaches to preserving digital materials: technology preservation, technology emulation, and data migration. Technology preservation and emulation focus on the technology itself. Data migration strategies focus on the need to maintain the digital files in a format which is accessible using "current technology". The **significant properties** indicate how appropriate each approach is for a digital object.

Significant Properties: Those characteristics (both technical, intellectual, and aesthetic) agreed by the archive or by the collection manager to be the most important features to preserve over time. A digital objects significant properties are not assumed to be empirical. For Cedars, the creation and maintenance of the detailed **metadata** associated with the objects significant properties are the backbone of an archives preservation function.

Metadata for Digital Preservation: A robust system of resource description – for the purposes of resource discovery, managing access, and ensuring preservation of the digital objects. Specifically preservation metadata will take two forms (from the **OAIS reference model**):

- **preservation description information** which includes general resource description, rights management information, provenance and context descriptions, and histories of actions taken for the purpose of preservation.
- **representation information** which maps the stored data into meaningful concepts – i.e. systems information and references to formats and software tools.

The Open Archival Information System (OAIS) Reference Model: This Reference Model has been developed by the Consultative Committee for Space Data Systems to provide a conceptual framework and reference tool for defining a digital archive. It describes a specific functional model of requirements (both people and systems) for implementing a digital archive. The OAIS is undergoing the ISO process and its publication as a standard is expected later in 2001. The Cedars project has provided a demonstrator project based on it. The importance of OAIS to the archiving community is undeniable but its usefulness to research libraries and archives is still largely unexplored.

Glossary of Selected Terms from the Open Archival Information Systems (OAIS) Reference Model

Archival Information Package (AIP): Content Information and the associated Preservation Description Information that is needed to preserve the Content Information over the Long Term. It has associated Packaging Information.

Archival Storage: The OAIS entity that contains the services and functions used for the storage and retrieval of Archival Information Packages.

Archive: An organization that intends to preserve information for access and use by a Designated Community.

Content Information: The set of information that is the primary target for preservation. It is composed of a Primary Digital Object and its Representation Information.

Dissemination Information Package (DIP): An Information Package, derived from one or more AIPs, that is distributed to the Consumer in response to a request.

Ingest: The OAIS entity that contains the services and functions that accept Submission Information Packages from Producers, prepare Archival Information Packages for storage, and ensure that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.

Long Term: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future.

Long Term Preservation: The act of maintaining information, in a correct and independently usable form, over the Long Term.

Preservation Description Information (PDI): Information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.

Primary Digital Object (PDO): The digital object that is the primary target for preservation.

Representation Information (RI): The information that maps a Primary Digital Object into more meaningful concepts, and describes how to obtain the intellectual content of these.

Submission Agreement: An agreement reached between an OAIS and the Producer that specifies a data model for the Data Submission Session.

Submission Information Package (SIP): The Information Package identified by the Producer in the Submission Agreement with the OAIS.

Appendix One : Membership of Management Group and Advisory Board

1.1 Cedars Advisory Board

Membership:

Richard Blake, Public Record Office

Lynne Brindley, British Library[55]

Reg Carr, University of Oxford

Lorcan Dempsey, UKOLN[56]

John Dolan, Birmingham Central Library

John Duke, University of Leeds

Nancy Elkington, Research Libraries Group

Mirjam Foot, British Library[57]

Peter Fox, University of Cambridge (Chair)

Peter Graham, Syracuse University

Dan Greenstein, Arts and Humanities Data Service[58]

Margaret Hedstrom, University of Michigan

Clare Jenkins, CURL Executive Secretary, then Imperial College London (Project Director)

Michael Lesk, National Science Foundation[59]

Vanessa Marshall, National Preservation Office

Charles Oppenheim, JISC/CEI nominee[60]

Alasdair Paterson, University of Exeter

Emma Robinson, University of London

Alex Reid, University of Oxford[61]

Chris Rusbridge, University of Glasgow

Kelly Russell, Cedars Project Manager

Mike Sayers, University of Cambridge

Helen Shenton, British Library

Jan Wilkinson, University of Leeds

[55] From April 1998 – December 2000
[56] From April 1998 – June 2000
[57] From April 1998 – July 1999
[58] From April 1998 – November 1999
[59] From April 1998 – April 1999
[60] From April 1998 – December 2000
[61] From April 1998 – October 2000

Terms of Reference

The Advisory Board is established with the following terms of reference:

- To oversee the project on behalf of the HEFCE and its territorial partners, and the Joint Information Systems Committee(JISC), providing reports back to JISC as necessary.
- To represent the best interests of the wider HE community in the UK in advising the Cedars project how to best develop systems and resources commensurate with the needs of research and scholarship, information services and archives in UK universities.
- To receive regular reports from the Cedars project on its progress, future project plans and associated milestones and deliverables; to comment on such plans in light of available resources and the need for economy and efficiency; and to ensure the project has in place an appropriate management structure and project management methodology to secure the realisation of project plans.
- To support the Cedars project and to act as advocates for the project and its staff in furthering the project's aims, with particular reference to the need to maintain for the project a very high level of visibility in the UK HE community and beyond.
- To commission such additional activities and advice as may be necessary to advance the project, evaluate its outcomes and provide the requisite level of feedback to the funding agencies.

The Cedars Advisory Board has been established to support the Cedars project and will continue initially for the period of project funding which is three years from March 1998, It is expected that the group will meet twice yearly.

1.2 Cedars Management Group

Membership:

Paul Ayris, University College London

Chris Bailey, CURL Executive Secretary

Lou Burnard, University of Oxford

Nancy Elkington, Research Libraries Group

Peter Fox, University of Cambridge

Dave Holdsworth, University of Leeds

Clare Jenkins, Project Director (Chair)

Patricia Killiard, University of Cambridge

David Price, University of Oxford

Kelly Russell, Cedars Project Manager

Terms of Reference

The Cedars Management Group has been established to assume responsibility for the Cedars project work on behalf of CURL and the Cedars Advisory Board and the Cedars Project Partners.

It has the following terms of reference:

The Cedars Management Group will:

- receive regular reports from the CEDARS project on its progress, future project plans and associated milestones and deliverables
- comment on such plans in the light of available resources and the need for economy and efficiency;
- ensure the project has in place an appropriate management structure and project management methodology to secure the realisation of project plans;
- approve of any proposed changes to the original proposal and subsequent work plan;
- agree staffing arrangements at each partner site for the project as proposed by the Project Manager;
- arbitrate on any conflicts within the project or negotiate a solution to such conflicts.

It is expected that the Management Group will meet quarterly.

Appendix Two : A Digital Preservation Bibliography

[1] Mark Fresko. *Digital Preservation Study Page USING TELEMATICS TO SUPPORT PERSONAL*. Applerace Ltd, http://easyweb.easynet.co.uk/˜marc/, 1998. This page is intended for the Digital Preservation Study being conducted by Applerace Limited for the European Commission. The information on this page is presented without any guarantee of its correctness. It shows Guidelines and potential Guidelines which have been suggested by survey respondents. Entries on the list have not been validated in any way.

[2] *Digital Library Net*. Libronix Corporation, http://www.digitallibrary.net/, 1999. This site hosts links relating to digital libraries and related technologies like information retrieval and metadata.

[3] Hartmut Weber, Marianne Dörr, and Andrew Medlicott. *Digitization as a Means of Preservation?* CLIR, http://www.clir.org/pubs/reports/digpres/digpres.html, 1997. Final report of a working group of the Deutsche Forschungsgemeinschaft (German Research Association).

[4] Doron Swade. *The Problems of Software Conservation*. CCS, http://www.cs.man.ac.uk/CCS/simulate/sim_home.htm, 1992. The purpose of this article is to explore museological aspects of software. The issues raised are concerned less with computing as an historiographic tool than with computing as an object of historical study.

[5] Einat Amitay. *IR and IE on the web*. Einat Amitay, http://www.mri.mq.edu.au/˜einat/web_ir/mailinglists.html, 2000. This site is a collection of online resources for research in the field of information retrieval and information extraction from the web.

[6] Neil Beagrie and Daniel Greenstein. *Creating A Viable Scholarly Data Resource*. AHDS, http://ahds.ac.uk/deposit/viable.html, 1998. This Information Leaflet is intended to inform arts and humanities researchers and teachers about steps they can take to ensure that data resources they create today are accessible to them,and possibly to others in future despite numerous and unpredictable changes in computer technologies.

[7] Anne Nieutin, Delphine Weber, and Lionel Micault. *Preserving collections*. BNF, http://www.bnf.fr/institution/anglais/consergb.htm, 1997. Conservation past and present.

[8] Stuart Lee. *Scoping the Future of Oxford's Digital Collections*. Oxford University, http://www.bodley.ox.ac.uk/scoping/, 1999. This is the home page of the Mellon-funded 'Scoping the Future of Oxford's Digital Collections' Report which ran from November 1998 to July 1999.

[9] NHPRC. *The Ten Questions*. UMICH, http://www.si.umich.edu/e-recs/Research/NHPRC.html, 1991. In 1991, the NHPRC (National Historical Publications and Records Commission) funded a Working Meeting on Research Issues in Electronic Records, held in Washington, D.C. One of the main results of this meeting was a set of ten questions that future research in electronic records should try to reach a consensus on.

[10] Peter Graham. *Intellectual Preservation: Electronic Preservation of the Third Kind*. CPA, http://www.clir.org/cpa/reports/graham/intpres.html, 1994.

[11] Karla Youngs and Oliver Moss Jane Williams. *TASI: Digital Preservation*. ILRT, University of Bristol, http://www.tasi.ac.uk/building/dig_pres1.html, 1998.

[12] Peter Scott. *Electronic Journals Resource Directory*. University of Saskatchewan Libraries, http://library.usask.ca/˜scottp/links/, 1999.

[13] Thom Shepard. *UPF VIEWS*. WGBH, http://info.wgbh.org/upf/UPF_VIEWS.HTML, 1998.

[14] *PRESERVING AND ARCHIVING ELECTRONIC GOVERNMENT INFORMATION*. Library, University of California, Berkeley, http://www.lib.berkeley.edu/GODORT/whiteiv.html, 1995. Chapter 4 of WHITEPAPER - Government Information in the Electronic Environment.

[15] Working Meeting on Electronic Records. *General Background Working Papers and Findings*. University of Pittsburgh, http://www.sis.pitt.edu/˜cerar/ppr-find.htm, 1997.

[16] *MANAGING ELECTRONIC RECORDS ISSUES a discussion paper*. State Records of South Australia, http://www.archivenet.gov.au/Resources/SA_discussion_0498/Records08_SA.htm, 1998. discussion paper.

[17] Katharina Klemperer. *Preservation and archiving of electronic material*. Library and Information Systems Consulting, http://www.harrassowitz.de/ms/ejresguide.html#Preservation and archiving of electronic material, 1999. part of Electronic Journals: A Selected Resource Guide.

[18] David Haynes Associates. *Responsibility for Digital Archiving and Long Term Access to Digital Data*. David Haynes Associates, http://www.dha-info.demon.co.uk/digital.htm, 1999. The final report of this eLib supported study has been published by the Library Information Technology Centre. The report presents the findings of an investigation into options on the responsibility for maintaining an archive of digital materials produced in the UK and Ireland.

[19] *Managing Digital Collections*. AHDS, http://ahds.ac.uk/manage/manintro.html, 1999. The AHDS is commited to establishing a framework of strategies and standards for developing, managing, and distributing high-quality digital scholarly and heritage resources. The framework is being developed through applied research whose outputs are being assembled into this publication series.

[20] Jeff Rothenberg. *Ensuring the Longevity of Digital Documents*. http://www.uic.edu/depts/lib/libclasses/Rothenberg.html, 1995. The digital medium is replacing paper in a dramatic record- keeping revolution. But such documents may be lost unless we act now.

[21] Judy Ahronheim. *JUDY AND MAGDA'S LIST OF METADATA INITIATIVES*. http://www-personal.umich.edu/˜jaheim/alcts/bibacces.htm, 1997. Below is a list of metadata schema and proposals that have the potential of being widely used as standards for the description of materials in an electronic environment. Some of the projects listed below were discovered through an email query directed at several listservs. The rest were located through an assortment of searches and referrals.

[22] *University of Michigan Digital Library Project*. http://www.si.umich.edu/UMDL/, 2000. Welcome to the UMDL Project's web site. Here you will find the latest news in who we are, what we are doing, and where we are going.

[23] Library Link. *Archiving: who should do it and how?* http://www.mcb.co.uk/services/conferen/webforum/mcb-liblink-forum/144.html, 1998. ALA Midwinter Pre-conference Discussion.

[24] *ARL Membership Meeting Program on Preservation of Digital Information*. Association of Research Libraries, http://www.arl.org/preserv/init.html, 1998.

[25] *About PaperDisk(TM) Software*. paperdisk.com, http://www.paperdisk.com/aboutpd5.htm, 1998. PaperDisk is a novel and powerful technology that connects printed documents to the world of digital information. It forges this link by storing and communicating digital data via the printed page.

[26] Donald Waters. *The Social Organization of Archiving Digital Information*. http://www.oclc.org:5046/oclc/research/links/archtf/archtf_outline.html, 1995.

[27] Susan Jephcott and Neil Beagrie. *Digital Preservation: a Guide to Web Resources*. AHDS, http://ahds.ac.uk/resource/preserve.html, 1999.

[28] *Rights Management in the Digital Library*. IBM, http://www.almaden.ibm.com/cs/DLRM/, 1995.

[29] Sam Coleman and Jack Cole. *Media Management System Requirements (DRAFT)*. IEEE Storage System Standards Working Group, http://www.ssswg.org/public_documents/MMS_REQ_draft5.2.html, 1998. The purpose of this document is to list requirements for the Media Management System (MMS), a set of IEEE standards that are being developed by the IEEE Storage System Standards Working Group. This document serves to guide the P1244 standards development effort. It is not intended to guide outside efforts, or to serve as an absolute set of requirements for the standards. These requirements can be reasonably implemented, and will be of value to a wide range of users. They are neither abstract goals, nor complex standards of use only to a small group of elite users.

[30] Bernie Hurley. *BrainDump Paper #3 - System Architecture Considerations for Digital Archival Repository Services*. http://sunsite.berkeley.edu/MOA2/papers/dump3.html, 1997.

[31] Wei Wu, Charles W. Bailey Jr, Ann Thornton, and Steve Bonario. *Library-Oriented Lists & Electronic Serials*. WRLC, http://www.wrlc.org/liblists//liblists.htm, 2000. This document is a compilation of electronic discussion lists, distribution lists and serials which are of interest to library professionals and staff.

[32] *Journal of Digital Information*. Multimedia Research Group, University of Southampt, http://jodi.ecs.soton.ac.uk/, 2000. Publishing papers on the management, presentation and uses of information in digital environments.

[33] *NEDLIB - Networked European Deposit Library*. KB - National Library of the Netherlands, http://www.konbib.nl/nedlib/, 2000. NEDLIB is a collaborative project of European national libraries. It aims to construct the basic infrastructure upon which a networked European deposit library can be built. The objectives of NEDLIB concur with the mission of national deposit libraries to ensure that electronic publications of the present can be used now and in the future.

[34] Renato Iannella. *Metadata.Net Home Page*. DSTC Pty Ltd, http://www.ilrt.bris.ac.uk/discovery/mirrors/metadata.net/, 1998.

[35] Reginald Ferber. *Hypermedia and Metadata*. GMD-IPSI, http://www.darmstadt.gmd.de/˜ferber/delos/index.html, 1996. Position Paper for the second DELOS Workshop, Bad Honeff 7/8 Oct. 1996.

[36] *Digital Library Federation - Projects and Plans*. CLIR, http://www.clir.org/diglib/dlfprojects.htm, 1998.

[37] *Canadian Initiative on Digital Libraries*. CIDL, http://www.nlc-bnc.ca/cidl/cidle.htm, 2000. The Canadian Initiative on Digital Libraries (CIDL) is an alliance of Canadian libraries who recognize the growing importance of digital information, and who wish to collaborate to ensure better use of digital information and better service to their users.

[38] *Second European Conference on Research and Advanced Technology for Digital Libraries*. http://www.ics.forth.gr/2EuroDL/, 1998.

[39] Arturo Crespo. *Archival Digital Libraries Repositories*. Stanford Infolab, http://www-db.stanford.edu/˜crespo/archivalrep/, 1999.

[40] Digital Resources Committee Standards Subcommittee. *Standardized Handling of Digital Resources: An Annotated Bibliography*. American Library Association, http://www.ala.org/alcts/publications/netresources/bib_meta.html, 1998.

[41] *Keeping our word: preservation across the ages*. University Libraries, University of Iowa, http://www.lib.uiowa.edu/ref/exhibit/links.htm, 2000. Preservation Research Links.

[42] Howard Besser. *Information Longevity*. http://sunsite.berkeley.edu/Longevity/, 2000. Notes and links for the long-term archiving of digital information.

[43] *Digital Library Research & Development*. SunSITE, http://sunsite.berkeley.edu/R+D/, 2000. SunSITE Hosted Projects.

[44] *Preweb - Preserving the Web*. National Library of Sweden, http://kulturarw3.kb.se/html/preweb.html, 2000.

[45] *PADI - Preserving Access to Digital Information*. National Library of Australia, http://www.nla.gov.au/padi/, 1999. PADI is a subject gateway to digital preservation resources.

[46] Terry Kuny, Jan Stohner, Stephen Spencer, Nancy Brodie, Fay Turner, and Roy Tennant. *Digital Libraries - Metadata Resources*. IFLA, http://www.ifla.org/II/metadata.htm, 1999.

[47] Steven Ketchpel. *Annotated Bibliography of Digital Library Related Sources*. http://robotics.stanford.edu/users/ketchpel/annbib.html, 1998.

[48] SUNSite Manager. *Preservation Resources*. University of California Regents, http://sunsite.berkeley.edu/Preservation/, 2000.

[49] *Yahoo! Reference>Libraries>Library and Information Science>Journals*. Yahoo, http://uk.dir.yahoo.com/reference/Libraries/Library_and_Information_Science/Journals/, 2000.

[50] Leon Miller. *Ready, 'Net, Go! Archival Internet Resources*. Tulane Libraries, http://www.tulane.edu/~lmiller/ArchivesResources.html, 2000. This service is an archival meta index, or index of archival indexes. That is, from here we refer you to the major indexes, lists, and databases of archival resources. From them you can link to almost every archives and archival resource in the metaverse.

[51] Charles W Bailey, Jr. *Scholarly Electronic Publishing Bibliography*. University of Houston Libraries, http://info.lib.uh.edu/sepb/sepb.html, 2000. This bibliography presents selected articles, books, electronic documents, and other sources that are useful in understanding scholarly electronic publishing efforts on the Internet and other networks. Most sources have been published between 1990 and the present; however, a limited number of key sources published prior to 1990 are also included. Where possible, links are provided to sources that are available via the Internet.

[52] *JSTOR*. http://www.jstor.ac.uk/, 2000.

[53] *MIT Electronic Records Management Home Page*. MIT, http://web.mit.edu/erm/, 1998. Preserving Evidence of MIT's past.

[54] *The Journal of Electronic Publishing*. Umich, http://www.press.umich.edu/jep/, 2000.

[55] Simon Pockley. *Lest We Forget or Digital preservation - the World Wide Web as a repository for archival material*. http://www.cinemedia.net/FOD/1996/spockley55.html, 1996. There is a very real possibility that nothing created, stored and disseminated electronically will survive in the long term. The problem does need to be stated this dramatically. I have an unfailing sinking feeling whenever anybody links the concepts of digitisation and preservation. I have a profound and unchanging belief that these two concepts do not belong in any sense in the same world.

[56] Kyle Silfer. *CYHIST archives – November 1998, week 3 (#10)*. http://maelstrom.stjohns.edu/CGI/wa.exe?A2=ind9811c&L=cyhist&D=0&P=944, 1998. a post to Community Memory: Discussion List on the History of Cyberspace.

[57] Hartmut Weber, Marianne Dörr, and Andrew Medlicott. *Digitisation as a Method of Preservation?* ECPA, http://www.knaw.nl/ecpa/publ/weber.html, 1997. Final report of a working group of the Deutsche Forschungsgemeinschaft (German Research Association).

[58] Eric Davis. *An Overview of the Access and Preservation Capabilities of Digital Technology*. http://www.iwaynet.net/~lsci/diglib/, 1996. The age of digital technology has revolutionized communications. It is used and enjoyed by all levels of society for business, education, research, and entertainment. Our National Information Infrastructure is in place, and being improved constantly. To a degree, the world's libraries have embraced this new technology, but a complex question remains: to what degree do we incorporate digital technology to serve our primary responsibilities of providing access to and preservation of the collective information of our culture? The single most troubling question is this: What are the long-term preservation capabilities of digital technology? These qualities are well known for preservation photocopying and microfilming, but are still largely an unknown landscape in the digital environment.

[59] Papercom Alliance and Papercom Alliance. *Media Longevity Are We Losing The Preservation Fight?* http://www.papercom.org/disks.htm, 1997.

[60] Brewster Kahle. *Archiving the Internet*. http://www.archive.org/sciam_article.html, 1996.

[61] *Issues and Innovations in... Preserving Digital Information*. http://www.arl.org/transform/pdi/index.html, 1998.

[62] Margaret MacLean and Ben H Davis. *Time and Bits: Managing Digital Continuity*. J. Paul Getty Trust, http://www.getty.edu/publications/titles/time/, 1999. An integrated technical and philosophical discussion of digital archives and their future that includes the sociocultural and economic implications of both the problems and the solutions could provide a framework for long-term digital cultural preservation.

[63] *Metadata for Digital Images DMC Home — About — Projects — Resources — Services — Metadata for Digital Images*. University of Virginia Library, http://www.lib.virginia.edu/dmc/meta.html, 1998. This is the metadata set currently used in image description at the Digital Image Center, with acknowledgments to Howard Besser for his very useful work in this area.

[64] *National Data Repository*. ULCC, http://www.ulcc.ac.uk/ndr/NDR_home.html, 2000.

[65] Jessica Milstead. *Metadata: Cataloging by Any Other Name* .. Online Inc, http://www.onlineinc.com/onlinemag/OL1999/milstead1.html, 1999.

[66] *European Union Archive Network*. International Institute of Social History, http://www.iisg.nl/~euan/, 2000.

[67] Richard Cox. *Functional Requirements for Evidence in Recordkeeping*. University of Pittsburgh, http://www.lis.pitt.edu/~nhprc/, 1996. The major objectives of this research project were to develop a set of well-defined recordkeeping functional requirements – satisfying all the various legal, administrative, and other needs of a particular organization – which can be used in the design and implementation of electronic information systems.

[68] RLG WG on Preservation Issues of Metadata, Barbara Berger: Cornell University, Jim Coleman: Stanford University (Co-Chair), Willy Cromwell-Kessler: RLG (Co-Chair), Robin Dale: RLG, Bob DeCandido: New York Public Library, Carla Montori: University of Michigan, and Seamus Ross: University of Glasgow. *Final Report*. RLG, http://www.rlg.org/preserv/presmeta.html, 1998. A significant component of creating and managing digital collections is ensuring that the information essential to their continued use is preserved in an accessible form. The Working Group on Preservation Issues of Metadata was constituted in May 1997 as a first step in the process of addressing this issue. The group was asked to identify the descriptive data elements that should be associated with digital master files that have preservation-based intent. The Working Group also limited itself to a consideration of data elements that describe digital image files.

[69] *PANDORA: Preserving and Accessing Networked DOcumentary Resources of Australia*. NLA, http://pandora.nla.gov.au/pandora/, 2000. In order to set up a digital archive of Australian online publications, in June 1996 the National Library of Australia established the PANDORA Project.

[70] Steven B Robertson, Alan R. Heminger (Faculty Advisor), and James Wedertz (Faculty Reader). *Digital Rosetta Stone: A Conceptual Model for Maintaining Long-Term Access to Digital Documents*. http://www.au.af.mil/au/database/research/ay1996/afit_la/rober_sb.htm, 1997. Due to the rapid evolution of technology, future digital systems may not be able to read and/or interpret the digital recordings made by older systems, even if those recordings are still in good condition. This thesis addresses the problem of maintaining long-term access to digital documents and provides a methodology for overcoming access difficulties due to technological obsolescence. A review was conducted to determine the long-term access methods that have already been suggested by other researchers. These previously suggested methods are then combined with other ideas that were encountered and conceived while performing research for this project. The combination of these methods and ideas led to the creation of a model, the Digital Rosetta Stone, that provides a methodology for maintaining long-term access to digital documents. The hypothesis for the model is that knowledge preserved about different storage devices and file formats can be used to recover data from obsolete media and to reconstruct the digital documents. The Digital Rosetta Stone model describes three processes that are necessary for maintaining long-term access to digital documents in their native formats-knowledge preservation, data recovery, and document reconstruction. Finally, recommendations are made for the evaluation and implementation of the Digital Rosetta Stone.

[71] Michael Day. *Metadata for Preservation*. http://www.ukoln.ac.uk/metadata/cedars/AIW01.html, 1998. This report is a review of metadata formats and initiatives in the specific area of digital preservation. It supplements the DESIRE Review of metadata (Dempsey et al. 1997). It is based on a literature review and information picked-up

at a number of workshops and meetings and is an attempt to briefly describe the state of the art in the area of metadata for digital preservation.

[72] Marc Hoffman. *The OS Emulation HomePage*. Marc R Hoffman, http://www.kearney.net/˜mhoffman/index.html, 2000. Welcome to the OS (Operating System) HomePage. This page is my little project to express my emulator bug. This is a term that one of my friends coined after he saw how much I like computer emulators.

[73] Mark S Ackerman and Fielding, Roy T. *Collection Maintenance in the Digital Library*. Proceedings of Digital Libraries 95. Austin TX, Proceedings of Digital Libraries 95, Austin TX June 1995 pp39-48, 1995. the link to the electronic version is broken (CS).

[74] AHDS. *Guides to Good Practice*. AHDS, http://ahds.ac.uk/public/guides.html, 2000.

[75] Archaeology and Humanities Data Service. *AHDS list of major initiatives and information resources in digital preservation*. http://ahds.ac.uk/resource/preserve.html, 2000.

[76] Archaeology Data Service. *Guidelines for Depositors V 1.0*. ADS, http://ads.ahds.ac.uk/project/goodguides/gis/index.html, 1998.

[77] Archaeology Data Service. *Guidelines for cataloguing Datasets with the ADS v 1.0*. ADS, http://ads.ahds.ac.uk/project/goodguides/gis/index.html, 1998.

[78] Archaeology Data Service. *ADS Collections Policy*. ADS, http://ads.ahds.ac.uk/project/collpol.html, 1997.

[79] CR Arms. *Access Aids and Interoperability*. National Digital Library Program, Library of Congress, http://lcweb2.loc.gov/ammem/award/docs/interop.html, 1997.

[80] C Arms, Klavans, J, and Waters, D. *Enabling Access in Digital Libraries: A Report on a Workshop on Access Management*. Digital Library Federation, http://www.clir.org/pubs/reports/arms-79/contents.html, 1999. The workshop described in this report focused on the management of access to published infromation resources through research libraries. Topics discussed include privacy, protection of rights, authorization and authentication.

[81] Charles W Bailey, Jr. *Scholarly Electronic Publishing Bibliography*. University of Houston Libraries, http://info.lib.uh.edu/pr/v3/n2/bailey.3n2, 1992.

[82] National Library of Canada. *National Library of Canada Electronic Publications Pilot Project. Summary of the Final Report*. http://www.nlc-bnc.ca/e-coll-e/ereport.htm, 1996.

[83] Welsh Office, Department of Education (N Ireland), Department of National Heritage, and Scottish Office. *Legal Deposit of Publications: a consultation paper*. Department of national Heritage, no URL available, 1997.

[84] Peter Hyams. *Legal Deposit of Electronic Publications*. Online & CDROM Review 18 (October 1994), pp. 308-310, 1994.

[85] National Library of Australia. *Legal Deposit in Australia, Fourth edition*. http://www.nla.gov.au/1/services/ldeposit.html, 1997.

[86] British Library. *Proposal for the Legal Deposit of Non-Print Publications: to the Department of National Heritage from the British Library*. no URL available, 1996.

[87] Jim Vickery. *The Legal Deposit of Electronic Publications*. Against the grain 10 (February 1998) pp36, 38-40, 1998.

[88] T M Alrashid. *Safeguarding Copyrighted Contents: Digital Libraries and intellectual Property Management: CWRU's Rights Management System*. http://www.dlib.org/dlib/april98/04barker.html, 1998.

[89] American Library Association. *Fair Use in the Electronic Age: Serving the Public Interest*. http://www.ifla.org/documents/infopol/copyright/fairuse.txt, 1995.

[90] B M Allen. *Negotiation Digital Information System Licences Without Losing Your Shirt or Your Soul*. Journal of Library Administration 24, no. 4 (1997) 15-26, 1997.

[91] N Buchanan. *Navigating the Electronic River: Electronic Product Licensing and Contracts*. The Serials Librarian 30, nos 3/4 (1997) pp171-182, 1997.

[92] Committee on Inter-Institutional Co-operation. *Assumptions & Guiding Principles for Near-Term Initiatives*. http://NTX2.cso.uiuc.edu/cic/cli/licguide.html, 2000. License Negotiation Principles.

[93] *Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 24 July 1971)*. 1971. originally 1886 with revisions to 1971.

[94] *WIPO Copyright Treaty (Adopted by the Diplomatic Conference on 20 December 1996)*. 1996.

[95] *Universal Copyright Convention (as revised in Paris on 24 July 1971)*. 1996. originally 1952, revised 1971.

[96] P Alkhoven. *Metadata needed for archiving digital content*. http://www.echo.lu/digicult/en/backgrd.html, 1998. paper given at DGXIII Concertation meeting and Workshop, Digitisation of Library Materials Luxembourg, December 1998. Experience at Dutch National Library - Koninklijke Bibliotheek. Early stages of investigation of possibilities and problems of creating metadata for different projects and purposes (CS).

[97] *Elib Authentication Concertation Day*. http://www.ukoln.ac.uk/services/elib/events/authentication/, 1999.

[98] *Berkeley Technology Law Journal*. http://www.law.berkeley.edu/journals/btlj/.

[99] *Copyright and New Media Legal News*. http://collection.nlc-bnc.ca/100/201/300/copyright/1999/.

[100] *Harvard Journal of Law and Technology*. http://jolt.law.harvard.edu/.

[101] *Intellectual Property Magazine*.

[102] *Journal of Electronic Publishing*. http://www.press.umich.edu/jep/.

[103] *Journal of Law and Information Science*.

[104] *Richmond Journal of Law and technology*. http://www.urich.edu:80/˜jolt/.

[105] *Villanova Information Law Chronicle*.

[106] *American Intellectual Property Law Association*. http://www.aipla.org.

[107] *American Society of Composers, Authors and Publishers (ASCAP)*. http://www.ascap.com.

[108] D Atkins and Hedstrom, M. *E-rec Final Report Committee on the Records*. http://www.si.umich.edu/e-rec/Report/.

[109] John C Bennett. *A Framework of Data Types & Formats & Issues Affecting the Long Term Preservation of Digital Material*. http://www.ukoln.ac.uk/services/elib, 1997. eLib Supporting Study (BLRIC Report 50).

[110] *Berkeley digital library's collection policy*. http://sunsite.berkeley.edu/Admin/collection.html.

[111] H Besser and Onuf, R. *Image and Multimedia Database Resources*. http://sunsite.berkeley.edu/Imaging/Databases/.

[112] R Broadhurst and Hendly, T. *Document Management Directory. A comprehensive directory of document management and imaging products and services. 9th edition*. ISBN 0 900458917. 9th edition. available from Cimtech Ltd University of Hertfordshire, 45 Grosvenor Road, St Albans Hertfordshire AL1 3AW Tel 10727 813651; Fax 10727 813649, 1998.

[113] *British Library Research & Innovation Centre*. http://www.bl.uk/services/ric/.

[114] *BUBL Information Service*. http://bubl.ac.uk/.

[115] S Chapman, Conway, P, and Kenney, AR. *Digital Imaging and Preservation Microfilm: The Future of the Hybrid Approach for the Preservation of Brittle Books*. CLIR, http://www.clir.org/programs/cpa/hybridintro.html. CLIR has issued a working paper on the dual use of microfilm for preservation and digital imaging for enhanced access in the context of the U.S. brittle books program. This paper offers general recommendations on the creation of Computer Output Microfilm (COM) and digital images from the perspectives of quality, cost and technology. It also provides a decision tree to assist institutions in determining the circumstances under which one would scan first or film first. NB Recommendations are presented regarding the development of metadata elements associated with digitized page images. Areas for further discussion and research are also suggested. An overview of this work is presented in an identically-titled article in DigiNews.

[116] *Commission on Preservation and Access*. CLIR, http://www.clir.org/pubs/reports/reports.html.

[117] P Conway. *Preservation in the Digital World. Yale University library. Publication 62, CLIR*. CLIR, http://www.clir.org/.

[118] P Conway and Weaver, S. *The Setup Phase of Project Open Book*. Yale University Library, http://clir.stanford.edu/cpa/reports/conway.html, 1994.

[119] P Conway. *Yale University Library's Project Open Book. Preliminary Research Findings*. http://www.dlib.org/dlib/february96/yale/02conway.html, 1996. ISSN 1082 9873.

[120] *DLM Forum Website*. http://www.dlmforum.eu.org/index.html, 1999. The 1996 DLM Forum, organised jointly by the EU Member States and the European Commission, brought together experts from industry, research, administration and archives to discuss the memory of the information society. Archives services are regarded as an essential component of this. One of the principal outcomes of the 1996 Forum was the Guidelines on best practices for using electronic information (can be downloaded from this site). The 1999 DLM Forum (October) will consider the problems of the management, storage, conservation and retrieval of machine-readable data.

[121] *DIEPER Project (Digitised European Periodicals)*. http://www.sub.uni-goettingen.de/gdz/deiper.

[122] *Digital Archiving and Preservation Microfilm: the future of the Hybrid Approach for the Preservation of Brittle Books*. http://www.clir.org/programs/cpa/cpa.html#brittle.

[123] *Digital Libraries Initiative. NSF; Dept of Defense Advance Research Projects Agency (DARPA) and NASA*. http://www.cise.nsf.gov/iris/DLHome.html.

[124] M Esther. *Digital Image Collections*. CLIR, http://www.clir.org/. Publication 67, CLIR.

[125] *European Commission Framework 5 programme*. http://www.cordis.lu/fifth/src/305b-e-6.htm#key3.

[126] *ECDL. 3rd European Conference on Digital Libraries*. http://www-rocq.inria.fr/EuroDL99/, 1999.

[127] J Feather. *Preservation and the management of library collections, 2nd ed*. Library Association Publishing, 1996.

[128] C Fleischhauer. *Digital Formats for Content Reproductions*. http://lcweb2.loc.gov/ammem/formats.html, 1996.

[129] C Fleischhauer. *Digital Historical Collections: Types, Elements & Construction*. http://lcweb2.loc.gov/ammem/elements.html, 1996.

[130] M Fresko and Tombs, K. *Digital Preservation Guidelines: The state of the art in libraries, museums and archives*. http://gps0.leeds.ac.uk/˜ecldh/cedars/DGXIII.zip.

[131] *DigiCult Background Information*. http://www.echo.lu/digicult/en/backgrd.html. Information and documentation on the EC work in the Digital Heritage and Cultural Content area.

[132] M Fresko. *Long Term Preservation of Electronic Materials*. http://www.ukoln.ac.uk/services/papers/bl/rdr6238/, 1996. A JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib). Organised by UKOLN 27th and 28th November 1995 at the University of Warwick. Report prepared by the Mark Fresko Consultancy. The British Library, 1996. BL R&D Report 6328.

[133] *I*M Europe - Libraries Website*. http://www.echo.lu/libraries/en/libraries.html. Provides a range useful links. Of particular relevance to Cedars are the Projects, Publications, and Metadata links. EU publications linked include; Fresko & Tombs, Digital Preservation Guidelines; Report of the Third Metadata Workshop (April 1999), Fresko, Digitisation Materials: Report of the Concertation Meeting and Workshop December 1998.

[134] R D Gennaro. *JSTOR: the Andrew W Mellon Foundation's Journal Storage Project*. proceedings of the 19th International Essen Symposium 23-26 September 1996. ed Helal & Weiss. Universitatsbibliothek Essen 1997 p223- 230, 1997.

[135] R D Gennaro. *JSTOR: Building an Internet Accessible Digital Archive of Retrospective Journals*. http://www.ifla.org/IV/ifla63/63genr.htm, 1997. Paper presented at 63rd IFLA General Conference Proceedings aug 31- Sep 5 1997.

[136] *Associated articles about JSTOR*. http://www.jstor.org/about/bibliography.html.

[137] Peter Graham. *The Digital Research Library: Tasks & Commitments.* http://csdl.tamu.edu/DL95/papers/graham/graham.html, 1995.

[138] Peter S Graham. *Intellectual Preservation: Electronic Preservation of the Third Kind*. http://www.clir.org/cpa/reports/graham/intpres.htm.

[139] Anne Gerken Green and Dionne, Jo Ann. *Preserving the Whole: A two track approach to rescuing data and metadata*. http://www.clir.org/programs/leadership/misc/preswhol.html, 1996. Interim Report to the Commission on Preservation and Access.

[140] D Greenstein. *Managing Digital Collections: Towards A Strategic Framework for the Development of Appropriate and Effective Operational Data Policies; Parts 1 and 2*. AHDS, available from Arts and Humanities Data Service Executive, Kings College London. Strand London WC2R 2LS UK.

[141] C Grout. *Visual Arts Data Service. Guidelines for Depositors*. VADS, http://vads.ahds.ac.uk/deposit-pack.html, 1999.

[142] The Data Archive. *Guide to Depositing Data*. http://dawww.essex.ac.uk/depositing_data/.

[143] Ian Macfarlane. *Guidelines On The Management of Electronic Records*. http://www.pro.gov.uk/recordsmanagement/eros/guidelinesfinal.pdf, 1998.

[144] D Haynes, Streatfield, S, Jowett, T, and Blake, M. *Responsibility for digital archiving and long term access to digital data*. JISC/NPO, http://www.ukoln.ac.uk/services/papers/bl/jisc-npo67/digital-preservation.html.

[145] Margaret Hedstrom. *Digital Preservation: a time bomb for Digital Libraries*. http://www.uky.edu/~kiernan/DL/hedstrom.html.

[146] Margaret Hedstrom. *The Role of National Initiatives in Digital Preservation*. http://www.rlg.org/preserv/diginews/diginews2-5.html#feature2.

[147] M M Hedstrom and Montgomery, S. *Digital Preservation Needs and Requirements in RLG Member Institutions*. http://www.rlg.org/preserv/digpres.html. A survey of the status of digital archiving in RLG member institutions. Reveals that only half of those participating had policies governing acquisition, conversion, storage, refreshing, and/or migration of digital materials. Three-quarters believe irreplaceable information will be lost if digital preservation issues are not resolved. An important report on current practices, needs and plans, with interesting recommendations.

[148] R Heery, Powell, A, and Day, M. *Metadata - Library & Information Briefing 75*. http://agent.sbu.ac.uk/publications/libs/lib75.pdf.

[149] T Hendley. *Comparison of Methods & Costs of Digital Preservation*. http://www.ukoln.ac.uk/services/elib/papers/supporting/#blric, 1998. (online synopsis available at http://www.lic.gov.uk/publications/ricarchive/pub1998.html#report106). London, British Library Research and Innovation Centre, 1998. ISBN: 0-7123-9713- 2, ISSN: 1366-8218. British Library Research and Innovation Report 106.

[150] A Hendley. *The Preservation of Digital Material*. British Library, BL R&D Report 6242, 1996.

[151] ICA Committee on Electronic Records. *Guide for managing Electronic Records From An Archival Perspective*. ICA Studies CIA 8, 1997.

[152] ICPSR. *ICPSR Policy on Data Media and Distribution*. Inter-University Consortium for Political and Social Research, http://www.icpsr.umich.edu/ICPSR/Developments/media_policy.html, 1997.

[153] ICPSR. *Preparing Data for Archiving: What ICPSR requires*. Inter-University Consortium for Political and Social Research, http://www.icpsr.umich.edu/ICPSR/Archive/Deposit/icpsr.html.

[154] ICPSR. *ICPSR Guide to social science data preparation and archiving*. Inter-University Consortium for Political and Social Research, http://www.icpsr.umich.edu/ICPSR/Archive/Deposit/dpm.html, 1999.

[155] *Internet Library of Early Journals Project*. http://www.bodley.ox.ac.uk/ilej/.

[156] Electronic Text Centre. *Image Scanning: A Basic Helpsheet*. Alderman Library, University of Virginia, http://etext.lib.virginia.edu/helpsheets/scanimage.html.

[157] Anne R Kenney. *Benchmarking image quality: from conversion to presentation*. available from Anne R Kenney. Associate Director, Department of Preservation and Conservation, Cornell University.

[158] AR Kenney. *Digital to microfilm conversion: an interim preservation solution*. Library Resources and Technical Services 37 (1993) pp380-401, 1993.

[159] Michael Lesk. *Image Formats for Preservation and Access*. http://www.ifla.org/ifla/documents/libraries/net/lesk.txt, 1990.

[160] Denise Lievesley. *Maintenance and Preservation of large databases*. The LIBER Quarterly Vol 6 1996 No 4 pp.472- 482, 1996.

[161] Roy K Lowry. *Data Management for the OMEX 1 project: A Case Study*. available from British Oceanographic Data Centre, Proudman Oceanographic Laboratory, Bidston Observatory, Birkenhead, Merseyside l43 7RA UK Tel 0151 653 8633; Fax 0151 652 3950.

[162] Owen Mackenzie and Walle, J v d. *Deposit Collections of Electronic Publications*. European Commission, DG XII- E/4, EUR 16910 EN ISSN 1018-5593 ISBN 92-827-7643-3, 1996.

[163] Owen Mackenzie. *Preservation of Digital Material for Libraries*. The LIBER Quarterly Vol 6 1996 No 4 p435-451, 1996.

[164] Graham Matthews, Poulter, A, and Blagg, E. *Preservation of Digital Materials: Policy and Strategy for the UK*. http://www.ukoln.ac.uk/services/papers/bl/blri041/digpres.html, 1997. print version details: JISC/NPO Studies on the Preservation of Electronic Materials. British Library Research and Innovation Centre, 1997. ISBN: 0-7123-3313-4, ISSN: 1366-8218. British Library Research and Innovation Report 41.

[165] Graham Matthews. *Preservation Management*. Library & Information Briefings Issue 73. August 1997 library Information Technology Centre. ISSN 0954 1829. 18 pages, 1997. available from TB Distribution, South Lodge, Gravesend Road, Wrotham, Kent TN15 7JJ. Tel 01732 824700; Fax 01732 823829.

[166] Paul Miller and Greenstein, D. *Discovering Online Resources Across the Humanities. A practical Implementation of the Dublin Core*. ISBN 0 9516856-4-3, 1997. Edited by Paul Miller & Daniel Greenstein on behalf of AHDS and UKOLN. Available from The Arts and Humanities Data Service, Kings College London, Strand, London WC2R 2LS - email info@ahds.uc.uk.

[167] Janice Mohlhenrich. *Preservation of Electronic Formats: Electronic Formats for preservation*. Highsmith Press, 1993.

[168] Stephen Puglia and Barry Roginski. *NARA Guidelines for Digitizing Archive Materials for Electronic Access*. available from National Archives and Records Administration, 8601 Adelphi Road, College Park, MD 20740, USA.

[169] NERC Data Strategy Group. *NERC Data Policy Handbook*. Natural Environment Research Council, http://www.nerc.ac.uk/environmental-data/download.html.

[170] *National Digital Libraries Federation*. http://www.clir.org/diglib/dlfcharter.htm. The Federation's mission is to bring together – from across the nation and beyond – digitized materials that will be made accessible to students, scholars, and citizens everywhere, and that document the building and dynamics of America's heritage and cultures.

[171] *Selection Committee on On Line Australian Publications*. National Library of Australia, http://www.nla.gov.au/1/scoap/scoap.html. Guidelines for the Selection of Online Australian Publications Intended for preservation by the National Library.

[172] *Digital Services Project*. National Library of Australia, http://www.nla.gov.au/dsp.

[173] National Preservation Office. *Statement of Principles: Preservation of and Long-Term Access to Australian Digital Objects*. National Library of Australia, http://www.nla.gov.au/3/npo/natco/princ.html, 1997.

[174] *Preserving Access to Digital Information*. http://www.nla.gov.au/dnc/tf2001/padi/.

[175] *Report on copyright and digital preservation*. National Library of the Netherlands, http://www.ivir.nl/Publicaties/koelman/KBeng2.doc.

[176] National Preservation Office. *Preservation and Digitization; principles, practice and policies*. Paper preservation News No 80 December 1996 pp13-14, 1996. From the 1996 National Preservation Office Conference.

[177] *Open Archiving Interchange (OII) Website*. http://www2.echo.lu/oii/en/oii-info.html. The European Commission's Open Information Interchange Service aims to provide an overview of existing and emerging standards and industry specifications designed to facilitate the exchange of information in electronic form. The very useful website contains a range of resources. There is a comprehensive listing of published standards (including eg Library information), a series of guides to existing standards in specific areas (eg Archiving, IP rights, Metadata), and monthly reports (eg linking to the (indecs) metadata model).

[178] Oxford Text Archive. *OTA Collections Policy Version 1*. ftp://ota.ox.ac.uk/pub/ota/, 1997.

[179] CR Perkins. *Leave it to the labs? Options for the future of map and spatial data collections*. The LIBER Quarterly Vol 5 1995 No 3 p328, 1995.

[180] *DigiNews*. RLG (Research Libraries Group), http://www.rlg.org/preserv/diginews/. An extremely useful resource, RLG DigiNews is a bi-monthly web-based newsletter which focuses on issues of interest to those involved with digital initiatives with a preservation component. It also highlights relevant projects and publications. NB certain articles from RLG DigiNews have their own entries within this bibliography.

[181] Daniel Renoult. *Digitisation and preservation in the French National Library*. The LIBER Quarterly Vol 6 1996 No 4 p465-471.

[182] *Report by national Library in the Netherlands on copyright and digital preservation*. http://www.ivir.nl/Publicaties/koelman/KBeng2.doc. index URL: http://www.ivir.nl/medewerkers/koelman.html - CS reported problems accessing both these URLs, so comments particularly welcome.

[183] Peter Robinson. *The digitization of primary textual sources*. Office for Humanities Communication Publication No 4, 1993.

[184] Jeff Rothenberg. *Ensuring the Longevity of Digital Documents*. Scientific American January 1995 pp.42-47, 1995.

[185] David Seaman. *The Electronic Text Center Guide to Document Markup and Preparation*. Electronic Text Center. University of Virginia, http://etext.lib.virginia.edu/tei/uvatei.html, 1994.

[186] David Seaman. *Special Collections. Digital Image Creation*. Electronic Text Center. University of Virginia, http://etext.lib.virginia.edu/helpsheets/specscan.html.

[187] K Tombs. *Public Access to the Government's Electronic Documents*. (EROS- IPC) IMAT Vol 29 no 5 1996 pp 193-195. ISSN 0266 6960 ISBN 0 900458917, 1996.

[188] S Townsend, Chappell, C, and Struijvé, O. *Digitising History*. http://hds.essex.ac.uk/g2gp/digitising-history/index.html. The latest in the AHDS series, Guides to Good Practice in the Creation and Use of Digital Resources. This guide to creating, documenting and preserving digital resources derived from historical documents is intended as a reference work for individuals and organisations involved with, or planning, the computerisation of historical source documents. Several of the documentation points are of relevance to the Cedars project. For instance, it is recommended that information recorded should include hardware and software used to create the data collection, and whether it has at any point been converted to new systems or formats. Also, copyright and other intellectual property rights must be clearly established, and the name(s) of the copyright holder(s) both for the data collection and for the original source material must be specified. Although both recommendations are clear and sensible, they may not be easy to achieve in practice. Furthermore, even these recommendations may not in fact be sufficiently detailed or demanding for the purposes of long-term archiving. This simply serves to underline the importance of the Cedars project.

[189] J Trant. *Framing the Picture: standards for imaging systems*. http://www.io.org/˜jtrant/papers/ichim/ichim.1.intro.html, 1995. Paper presented for the International Council on Hypermedia and Interactivity in Museums. Museum Computer Network Joint Conference. San Diego California October 1995.

[190] D Waters and Garrett, J. *Preserving Digital Information. Final Report and Recommendations*. http://www.rlg.org/ArchTF/, 1996. Final Report and Recommendations of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group.

[191] H Weber and Dörr, M. *Digitization as a means of preservation?* CLIR, http://www.clir.org/cpa/abstract/pub69.html.

[192] *Website for Digital Heritage and Cultural Content*. http://www.echo.lu/digicult/. This new site is under development. It is currently highlighting the IST programme which brings together and extends the ACTS, Esprit and Telematics Applications programmes in a single integrated programme that reflects the convergence of information processing, communications and media technologies. There will be a call for proposals addressing Digital preservation of cultural heritage (September 1999).

[193] Bridget Winstanley. *Resource Centres for teaching and research*. Library and Information Technology Briefings. Issue 52 April 1994. ISSN 0954 1829 - available from Library Information Technology Centre, TBC Distribution, South Lodge, Gravesend Road, Wrotham, Kent TN15 7JJ. Tel 01732 824700; Fax 01732 823829, 1994.

[194] Roger Bowes. *Wiped out by WIPO!* Aslib, Aslib's 'Managing Information' December 1996, 1996. [on proposals for database Treaty - pre the Geneva conference].

[195] G P Cornish. *Libraries and the Harmonisation of Copyright*. EIPR 241- 243, 1998.

[196] G P Cornish. *Electronic Copyright Management Systems: Dream, Nightmare or Reality?* IFLA Journal 23, no. 4 (1997): 248-287, 1997.

[197] *Copyright and the digital environment*. Managing Information 3 (1) Jan 96, p.25-6. ISSN: 13520229, 1996. Statement prepared by the Library Association, UK/JCC working party on copyright which includes representatives of: Aslib, the Association for Information Management, the Institute of Information Scientists, the Standing Conference on National University Libraries, and the Society of Archivists.

[198] Dearing Report. *National Committee of Inquiry into Higher Education, Higher Education in the Learning Society*. HMSO, 1997. [comments on facilitation of use of copyright materials in digital form - e.g. at 13.34; 43].

[199] *European Copyright User Platform (ECUP)*. http://www.eblida.org/ecup/. [esp their position paper on user rights in electronic copyright].

[200] Laura N Gasaway. *Libraries, Educational Institutions, and Copyright proprietors: The First Collision on the Information Highway*. The Journal of Academic Librarianship 22 (September 1996) pp. 337-344, 1996.

[201] Laura N Gasaway. *Scholarly Publication and Copyright in Networked Electronic Publication*. Library Trends 43 (Spring 1995) pp. 679-700, 1995.

[202] Laura N Gasaway. *The White Paper, Fair Use, libraries and Educational Institutions*. The Serials Librarian 31, nos. 1/2 (1997) pp 211-220, 1997.

[203] H M Gladney and Lotspiech, J. *Safeguarding Digital Library Contents and Users: Assuring Convenient Security and Data Quality*. http://www.dlib.org/dlib/may97/ibm/05gladney.html, 1997.

[204] H M Gladney and Lotspiech, J. *Safeguarding Digital Library Contents and Users: Storing, Sending, Showing and Honoring Usage Terms and Conditions*. http://www.dlib.org/dlib/may98/gladney/05gladney.html, 1998.

[205] H M Gladney, Mintzer, F, and Schiattarella, F. *Safeguarding Digital Library Contents and Users: Digital images of treasured Antiquities*. http://www.dlib.org/dlib/july97/vatican/07gladney.html, 1997.

[206] T Hoeren and Decker, U. *Electronic Archives and the Press: Copyright Problems of Mass Media in the Digital Age*. 20 European Intellectual Property Review 256, 1998.

[207] *Information Policy Notes*. http://www.essential.org/listproc/info-policy-notes/. free Internet newsletter.

[208] Information Technology Association of Canada. *A Copy is a Copy: Copyright Issues and the Emerging Information Infrastructure*. http://www.ifla.org/documents/infopol/copyright/copy95en.htm, 1995.

[209] IFLA. *Position paper on copyright in the electronic environment*. International Federation of Library Associations and Institutions (IFLA), http://www.ifla.org/V/ebpb/copy.htm, 1996.

[210] Intl Federation for Information and Documentation. *Position Paper on Libraries, Copyright and the Electronic Environment of the International Publishers Copyright Council*. http://www.ifla.org/documents/infopol/copyright/ipa.txt, 1996.

[211] Yvonne Kidd. *Intellectual Property Rights Management in the Digital Environment: An Overview of Developments and Initiatives*. http://info.lib.uh.edu/sepb/pcomm.htm, 1996. Print version available at Information Standards Quarterly 8 (July 1996) 1-4.

[212] Kirtley, Jane E, Daugherty, Rebecca, and Reis, Leslie Ann. *World Intellectual Property Organization: Comments on the basic Proposal for the Substantive Provisions of the Treaty on Intellectual Property in Respect of Databases*. The Journal of Academic Librarianship 23 (March 1997): 134-139, 1997.

[213] G Ackland, Cumming, K, and McKemmish, S. *The end of the beginning: the SPIRT Recordkeeping Metadata Project*. http://www.archivists.org.au/events/conf99/spirt.html, 1999. Paper delivered at the Australian Society of Archivists 1999 Conference, Archives at Risk: Accountability, Vulnerability and Credibility, Brisbane, 29-31 July 1999.

[214] Australian Archives. *Keeping electronic records: policy for electronic recordkeeping in the Commonwealth Government*. http://www.naa.gov.au/govserv/techpub/elecrecd/KeepingER.html, 1995. (Last visited: 06-May-1999).

[215] CW Bailey. *Scholarly electronic publishing bibliography*. University of Houston Libraries, http://info.lib.uh.edu/sepb/sepb.html, 1996. Currently in Version 25: June 1999. The section on Information integrity and preservation is at <URL:http://info.lib.uh.edu/sepb/lbinteg.htm> (Last visited: 17-Jun-1999).

[216] K Barata. *Bibliography on electronic records*. University of Pittsburgh School of Information Sciences, http://www.lis.pitt.edu/~nhprc/bibtc.html, 1996. (Last visited: 17-Jun-1999).

[217] N Beagrie and Greenstein, D. *A strategic policy framework for creating and preserving digital collections*. Library Information Technology Centre, http://ahds.ac.uk/manage/framework.htm, 1998. JISC/NPO Studies on the Preservation of Electronic Materials. eLib Supporting Study P3 (Last visited: 18-Jun-1999).

[218] D Bearman. *Optical media: their implications for archives and museums*. Archives and Museum Informatics Technical Report, 1 (1), 1987.

[219] D Bearman. *Archival methods*. Archives and Museum Informatics Technical Report, 3 (1), 1989.

[220] D Bearman. *An indefensible bastion: archives as repository in the electronic age*. Archives and Museum Informatics Technical Report, No. 13, pp 14-24, 1991. In: Bearman, D., (ed.) Archival management of electronic records.

[221] D Bearman. *Diplomatics, Weberian bureaucracy, and the management of electronic records in Europe and America*. American Archivist, 55, 1992, 168-181, 1992. Reprinted in: Bearman, D., Electronic evidence. Pittsburgh, Pa.: Archives and Museum Informatics, 1994, 254-277.

[222] D Bearman. *Archival principles and the electronic office*. K.G. Saur, In: Menne-Haritz, A., (ed.), Information handling in offices and archives. Munich: K.G. Saur, 1993, 177-193, 1993. Reprinted in: Bearman, D., Electronic evidence. Pittsburgh, Pa.: Archives and Museum Informatics, 1994, 146-175.

[223] D Bearman. *Archival data management to achieve organizational accountability for electronic records*. Archives and Manuscripts, 21, 1993, 14-28, 1993. Reprinted in: Bearman, D., Electronic evidence. Pittsburgh, Pa.: Archives and Museum Informatics, 1994, 12-33.

[224] D Bearman. *Record-keeping systems*. Archivaria, 1993, 36, 1993. Reprinted in: Bearman, D., Electronic evidence. Pittsburgh, Pa.: Archives and Museum Informatics, 1994, 34-70.

[225] D Bearman. *The implications of Armstrong v. Executive Office of the President for the archival management of electronic records*. American Archivist, 56, 1993, 674-689, 1993. Reprinted in: Bearman, D., Electronic evidence. Pittsburgh, Pa.: Archives and Museum Informatics, 1994, 118-144.

[226] D Bearman. *Electronic evidence: strategies for managing records in contemporary organizations*. Archives and Museum Informatics, print only, 1994.

[227] D Bearman. *Archival issues in a computing environment*. http://www.naa.gov.au/govserv/techpub/keeps/Bearman_Paper.htm, 1994. Playing for keeps: the proceedings of an Electronic Records Management Conference hosted by the Australian Archives, Canberra, Australia, 8-10 November 1994. (Last visited:06-May-1999).

[228] D Bearman. *Virtual Archives*. http://www.lis.pitt.edu/˜nhprc/prog6.html, 1996. ICA Meeting, Beijing, September 1996. (Last visited: 17-Jun-1999).

[229] D Bearman. *Reality and Chimeras in the Preservation of Electronic Records*. http://www.dlib.org/dlib/april99/bearman/04bearman.html, 1999. D-Lib Magazine, 5 (4), April D-Lib Magazine, 5 (4), April 1999 (Last visited: 18-Jun-1999).

[230] D Bearman and Duff, W. *Grounding archival description in the functional requirements for evidence*. Archivaria, 41, 1996, 275-281, 1996.

[231] D Bearman and Hedstrom, M. *Reinventing archives for electronic records: alternative service delivery options*. Archives and Museum Informatics Technical Report, 18, 1993, pp82-98, 1993. In: Hedstrom, M. (ed.) Electronic records management program strategies. Archives and Museum Informatics Technical Report, 18. Pittsburgh, Pa.

[232] D Bearman and Lytle, RH. *The power of the principle of provenance*. Archivaria, 21, 1985-86, 14-27, 1985.

[233] D Bearman and Trant, J. *Electronic Records Research Working Meeting, May 28-30, 1997: a report from the archives community*. http://www.dlib.org/dlib/july97/07bearman.html, 1997. (Last visited: 17-Jun-1999).

[234] D Bearman and Trant, J. *Authenticity of digital resources: towards a statement of requirements in the research process*. http://www.dlib.org/dlib/june98/06bearman.html, 1998. (Last visited: 18-Jun-1999).

[235] L Bell. *The archival implications of machine-readable records*. Archivum, 26 (1), 1979, 85-92, 1979.

[236] J.C Bennett. *A framework of data types and formats, and issues affecting the long-term preservation of digital material (Last visited: 18-Jun-1999)*. British Library Research and Innovation Centre, http://www.ukoln.ac.uk/services/papers/bl/jisc-npo50/bennet.html, 1997. Print reference: JISC/NPO studies on the Preservation of Electronic Materials. British Library Research and Innovation Report, 50 (Last visited: 18-Jun-1999).

[237] TK Bikson and Frinkling, EJ. *Preserving the present: toward viable electronic records*. Sdu Publishers, print only, 1993.

[238] SE Binns, Bowen, DV, and Murdock, A. *Migration strategies within an electronic archive: practical experience and future research*. Archives and Museum Informatics, 11 (3/4), 1997, pp. 301-306, 1997.

[239] M Blake. *Aspects of electronic archives*. Electronic Publishing Review, 6, 1986, 151-167, 1986.

[240] M Blake. *Archiving of electronic publications*. Electronic Library, 7 (6), 1989, 376-386, 1989.

[241] M Blake. *Practices of printers in relation to electronic archiving*. Electronic Library, 8 (4), 1990, 239-243, 1990.

[242] MB Bowling. *Literature on the preservation of nonpaper materials*. American Archivist, 53, 1990, 340-348, 1990.

[243] M Brichford. *Archival issues in network electronic publications Archival issues in network electronic publications*. Library Trends, 43 (4), 1995, 701-712, 1995.

[244] British Library Research and Innovation Centre. *Proposal for the legal deposit of non-print publications*. BLRIC, print only, 1996.

[245] British Library. *Legal deposit of publications: a consultation paper - response from the British Library*. http://portico.bl.uk/information/legal-deposit.html, 1997. (Last visited: 17-Jun-1999).

[246] P Vickers, Martyn, J, and (eds.). *The impact of electronic publishing on library services and resources in the UK*. http://www.ukoln.ac.uk/services/papers/bl/lirr102/, 1994. Library and Information Research Report 102. London: British Library, 1994. (Last visited: 17-Jun-1999).

[247] AR Calmes. *To archive and preserve: a media primer*. Inform, 1 (5), 1987, 14-17, 33, 1987.

[248] AR Calmes. *Relative longevity of various archival recording media*. In: Proceedings of conservation in archives: International Symposium, Ottawa, Canada, May 10-12, 1988. Ottawa: National Archives of Canada, 1988, 207-221.

[249] J Coleman and Willis, D. *SGML as a framework for digital preservation and access*. Commission on Preservation and Access, print only, 1997.

[250] Subcommittee C Committee on Preservation. *White paper: strategic technology considerations relative to the preservation and storage of human and machine readable records prepared for the National Archives and Records Service (NARS)*. Western Association of Map Libraries Information Bulletin, 16 (3), 1985, 312-324, 1994.

[251] Conference of Directors of National Libraries WG. *The legal deposit of electronic publications*. http://www.unesco.org/webworld/memory/legaldep.htm, 1996. chaired by Brian Lang. Paris: UNESCO (Last visited: 17-Jun-1999).

[252] P Conway. *Digitizing preservation*. Library Journal, 1 February 1994, 42-45, 1994.

[253] P Conway. *Preservation in the digital world*. Commission on Preservation and Access, print only, 1996.

[254] P Conway. *Selecting mirofilm for digital preservation: a case study from Project Open Book*. Library Resources & Technical Services, 41 (1), 1996, 67-77, 1996.

[255] T Cook. *From information to knowledge: an intellectual paradigm for archives*. Archivaria, 19, 1984-85, 28-49, 1984.

[256] T Cook. *The archival appraisal of records containing personal information: a RAMP study with guidelines*. UNESCO, PGI-91/WS/3. Paris: UNESCO, General Information Programme and UNISIST, 1984, 1984.

[257] T Cook. *Easy to byte, harder to chew: the second generation of electronic records archives*. Archivaria, 33, 1991-92, 202-216, 1991.

[258] AM Hendley. *The archival storage potential of microfilm, magnetic media and optical data disks: a comparison based on a literature review*. National Reprographic Centre for documentation (NRCd), BNB Research Fund report, 10, 1983.

[259] AM Hendley. *Videodiscs, compact discs and Digital Optical Disks: an introduction to the technologies and the systems and their potential for information storage, retrieval and dissemination*. National Centre for Information Media and Technology (Cimtech), Cimtech publication, 23, 1985.

[260] AM Hendley. *CD-ROM and optical publishing systems: an assessment of the impact of the optical read only memory systems on the information industry and a comparison between them and traditional paper, microfilm and on-line publishing systems*. National Centre for Information Media and Technology (Cimtech), Cimtech publication, 26. BNB Research Fund report, 25, 1987.

[261] T Hendley. *Comparison of methods and costs of digital preservation*. British Library Research and Innovation Centre, JISC/NPO Studies on the Preservation of Electronic Materials. British Library Research and Innovation Report, 106, 1998.

[262] S Hills. *Electronically published material and the archival library*. Electronic Publishing Review, 5 (1), 1985, 63-72, 1985.

[263] B Kahle. *Preserving the Internet*. Scientific American, 276 (3), March 1997, 72-73, 1997.

[264] AR Kenney. *Digital-to-Microfilm Conversion: an interim preservation solution*. Library Resources & Technical Services, 37 (4), 1993, 380-401, 1993.

[265] AR Kenney. *Digital to Microfilm Conversion: a demonstration project, 1994-1996*. Cornell University Library, Department of Preservation and Conservation, http://www.library.cornell.edu/preservation/com/comfin.html, 1997. Final Report to the National Endowment for the Humanities. Ithaca, N.Y (Last visited: 17-Jun-1999).

[266] AR Kenney and Conway, P. *From analog to digital: extending the preservation tool kit*. Research Libraries Group, Digital imaging technology for preservation: proceedings from an RLG Symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York, 11-24, 1995. In: Elkington, N.E., (ed.) Digital imaging technology for preservation: proceedings from an RLG Symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York.

[267] AR Kenney and Personius, LK. *The Cornell/Xerox/Commission on Preservation and Access joint study in digital preservation Report: Phase 1, Digital capture, paper facsimiles and network access.* Commission on Preservation and Access, http://www.clir.org/pubs/reports/joint/, 1992. (Last visited: 17-Jun-1999).

[268] AR Kenney and Personius, LK. *A testbed for advancing the role of digital technologies for library preservation and access: final report by Cornell University to the Commission on Preservation and Access*. Commission on Preservation and Access, print only, 1993.

[269] RM Kesner. *Automated information management: is there a role for the archivist in the office of the future?* Archivaria, (19), 1984-85, 162-172, 1984.

[270] KD Lehmann. *Making the transitory permanent: the intellectual heritage in a digitized world of knowledge*. Daedalus, 125 (4), Fall 1996, 307-329, 1996.

[271] M Lesk. *Image formats for preservation and access: a report of the Technology Advisory Committee to the Commission on Preservation and Access*. Commission on Preservation and Access, http://www.clir.org/pubs/reports/lesk/lesk.html, 1990. (Last visited: 17-Jun-1999).

[272] M Lesk. *Preservation of new technology: a report of the Technology Advisory Committee to the Commission on Preservation and Access*. Commission on Preservation and Access, http://www.clir.org/pubs/reports/lesk/lesk2.html, 1992. (Last visited: 17-Jun-1999).

[273] Mr Justice Laddie. *Copyright: over-strength, over-regulated, over-rated?* 18 European Intellectual Property Review pp. 429-430, 1996.

[274] James Love. *example on WIPO database proposal*. Center for Study of Responsive Law, Consumer Project on Technology.

[275] Sir Anthony Mason. *The Australian Library and Information Association Library Week Oration, State Library of NSW*. The Australian Library Journal pp.8-91, 1996. Includes comments on fair dealing.

[276] Ann Okerson. *The Current National Copyright Debate: Its Relationship to the Work of Collections Managers*. Journal of Library Administration 22, no. 4 (1996) 71-84, 1996.

[277] T Power. *Digitisation of Serials and Publications: the Seminal Objective of Copyright Law*. 19 European Intellectual Property Review pp.444-461, 1997.

[278] Public Domain. *Union for the Public Domain*. http://www.public-domain.org/. U.S. independent citizens' organisation - comments on fair use.

[279] Anne Ramsden. *Copyright Management Technologies: The Key to Unlocking Digital Works?* http://www.ariadne.ac.uk/issue10/copyright/, 1997.

[280] Pamela Samuelson. *Encoding the Law into Digital Libraries*. Communications of the ACM 41 (April 1998) pp13-18, 1998.

[281] M Tedd. *Papers on Copyright Issues In the Electronic Library*. http://www.jisc.ac.uk/pub/copyright/intro.htm.

[282] Bill Tuck. *Electronic Copyright management Systems; Final Report of a Scoping Study for eLib*. http://www.sbu.ac.uk/litc/copyright/ecms.html, 1996.

[283] U.S. Copyright Office. *Report on Legal Protection for Databases*. Bulletin of the American Society for Information Science 24 (December/January 1998) pp25-29.

[284] J Watkins. *Copyright Issues in Libraries: Global Concerns, Local Solutions*. IFLA, Offices for Universal Availability of Publications and International Lending ISBN 0 17123 2150 0.

[285] Trisha L Davis. *License Agreements in Lieu of Copyright: Are We Signing Away Our Rights?* Library Acquisitions: Practice & Theory 21, no. 1 (1997) 19-28, 1997.

[286] Patricia Failing. *Scholars Face Hefty Fees and Elaborate Contracts When They Use Digital Images*. The Chronicle of Higher Education, 5 July 1996, A15, A19-A20, 1996.

[287] *International Coalition of Library Consortia (ICOLOC)*. http://www.library.yale.edu/consortia/icolcpr.htm.

[288] JISC and Publishers Association Working Party. *Proposed 'Model Licence' between UK Universities and Publishers*. http://www.ukoln.ac.uk/services/elib/papers/pa/licence/intro.html, 1997.

[289] *JSTOR Library Licence Agreement (HEFCE)*. http://www.chest.ac.uk/datasets/jstor/hefce.html.

[290] *Liblicense: Licensing Digital Information*. Yale University Library.

[291] *Library Link discussion forum*. http://www.mcb.co.uk/liblink/discussion.html. (on electronic licensing).

[292] *National Electronic Site Licence Initiative (NESLI)*. http://www.nesli.ac.uk/. Contains announcements; NESLI Licence.

[293] Silverplatter. *Subscription and Licence Agreement*. http://www.silverplatter.com/, 1998.

[294] J Menzel, Metzner, K, and Pope, E. *IDEAL and APPEAL: A Model for Consortium Licensing of Electronic Journal Collections*. Astrophysics and Space Science 247 nos. 1-2 (1997) 241-250, 1997.

[295] A Okerson. *Copyright or Contract?* Library Journal, 1 September 1997, 136- 139, 1997.

[296] Michael Day. *Extending metadata for digital preservation*. Ariadne,No. 9,1997, 1997.

[297] Michael Day. *Metadata for Preservation presentation to the Warwick RLG- NPO Digital Imaging Conference*. http://www.ukoln.ac.uk/metadata/presentations/rlg-npo/.

[298] Michael Day. *Metadata for Preservation*. http://www.ukoln.ac.uk/metadata/cedars/AIW01.html, 1998. CEDARS project Document AIW01.

[299] Bernie Hurley. *Paper on architecture of archive repositories at the Berkeley MoA2 site*. http://sunsite.berkeley.edu/MOA2/papers/moaplan.html.

[300] D Landsbergen and Wolken, G. *Eliminating Legal and Policy barriers to Interoperable Government Systems*. http://iep.fedworld.gov/library/elapbigs/phase2.html.

[301] C Lynch and Garcia-Molina, H. *Interoperability, Scaling and the Digital Libraries Research Agenda*. http://walrus.stanford.edu/diglib/pub/reports/.

[302] *Pandora logical data model*. http://www.nla.gov.au/pandora/ldmv2.html.

[303] *RLG Working Group on Preservation Issues of Metadata, Final report*. http://www.rlg.org/preserv/presmeta.html, 1998.

[304] Jeff Rothenberg. *Report on Emulation*. http://www.clir.org/whatsnew.html#avoiding.

[305] Jeff Rothenberg. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. http://www.clir.org/whatsnew.html#avoiding, 1999. Published in January 1999, this report follows up Dr Rothenberg's 1995 article in Scientific American, Ensuring the Longevity of Digital Documents by elaborating the author's proposal for emulating obsolete software/hardware systems on future, unknown systems, as a means of preserving digital information far into the future. In the author's view, the emulation approach is not just a promising candidate for a solution to the problem of digital preservation, but the only apprach offering a true solution to the problem.

[306] Andy Stone. *Cedars Preservation Metadata Elements*. http://users.ox.ac.uk/˜cedars/Papers/AIW02.html, 1999. CEDARS project Document AIW02.

[307] Andy Stone. *Notes on BCS Metadata Matters seminar*. http://users.ox.ac.uk/˜nadz/metadata_matters.htm, 1999. linked to BCS site for reports.

[308] Task Force on the Archiving of Digital Information. *Preserving digital information: report of the Task Force on Archiving of Digital Information*. http://www.rlg.org/ArchTF/, 1996. commissioned by the Commission on Preservation and Access and the Research libraries Group. Washington, D.C.

[309] *Australian Copyright Council*. http://www.copyright.org.au/.

[310] *Center for Advanced Study and Research on Intellectual Property*. http://www.law.washington.edu/˜casrip/.

[311] *Computers and Fair Use*. Stanford University Libraries, http://fairuse.stanford.edu/.

[312] *CONFU: The Conference on Fair Use*. Stanford University Libraries, http://www.utsystem.edu/ogc/intellectualproperty/confu.htm.

[313] ALA. *Copyright & Intellectual Property*. http://info.lib.uh.edu/sepb/rlegal.htm.

[314] Association of Research Libraries. *Copyright and Intellectual Property Table of Contents*. http://info.lib.uh.edu/sepb/rlegal.htm.

[315] Berkeley Library University of California. *Copyright, Intellectual Property Rights, and Licensing Issues;*. http://info.lib.uh.edu/sepb/rlegal.htm.

[316] *Copyright and Universities*. http://arl.cni.org/scomm/copyright/UniCopy.html.

[317] *The Copyright Web Site*. http://www.benedict.com/.

[318] *CyberSpace Law Center*. http://cyber.findlaw.com/.

[319] DLM Forum. *Guidelines on best practices for using electronic information*. http://www2.echo.lu/dlm/en/gdlines.html.

[320] *Digital Future Coalition*. http://www.dfc.org/.

[321] *Digital Library Research Group*. http://scholar.lib.vt.edu/. based at College of Library and Information Services at the University of Maryland College Park.

[322] *Findlaw:Copyright*.

[323] *Franklin Pierce Law Center*. http://www.fplc.edu/fplchome.htm.

[324] IFLA. *Information Policy: Copyright and Intellectual Property Resources*. http://ifla.inist.fr/.

[325] *The Institute for Learning Technologies*. http://www.ilt.columbia.edu/.

[326] *Legal Information Institute*. http://www.law.cornell.edu/. based at Cornell University Law School.

[327] *The Villanova Center for Information Law and Policy*.

[328] *World Intellectual Property Organisation (WIPO)*. http://www.wipo.org/.

[329] *WWW Multimedia Law*. http://www.batnet.com/oikoumene/.

[330] A Abid. *Memory of the World: preserving our documentary heritage*. Paris: UNESCO, Information and Informatics Division, http://www.unesco.org/webworld/memory/Abid.htm, 1997. Outlines the main features of the UNESCO programme Memory of the World which is designed to increase awareness of the significance of documentary heritage. It is not primarily concerned with digital preservation, but the digitization of selected historical collections. The paper is an English translation of an article (Mémoire du monde: préserver notre patrimoine documentaire) published in Bulletin des bibliothèdques françaises, T. 42, No. 2, 1997 (Last visited: 17-Jun-1999).

[331] T Cook. *The impact of David Bearman on modern archival thinking: an essay of personal reflection and critique*. Archives and Museum Informatics, 11, 1997, 15-37, 1997.

[332] RJ Cox. *Electronic information technology and the archivist: bright lights, lingering concerns*. American Archivist, 55, 1992, 232-234, 1992.

[333] RJ Cox. *Readings in archives and electronic records: annotated bibliography and analysis of the literature*. Technical Report, No. 18. Pittsburgh, Pa.: Archives and Museum Informatics, 1993, 97-152, 1993. In: Hedstrom, M. (ed.) Electronic records management program strategies. Archives and Museum Informatics.

[334] RJ Cox. *The first generation of electronic records archivists in the United States: a study in professionalization*. Haworth Press, print only, 1994.

[335] RJ Cox. *More than diplomatic: functional requirements for evidence in recordkeeping*. Records Management Journal, 7 (1), 1997, pp. 21-30, 1997.

[336] RJ Cox. *Access in the digital information age and the archival mission: the United States*. Journal of the Society of Archivists, 19 (1), 1998, 25-40., 1998.

[337] RJ Cox. *American archivists, cyberculture, and stasis*. http://www.ucd.ie/˜archives/coxart.html, 1998. Paper delivered at: Cyber, Hyper or Resolutely Jurassic? Archivists and the Millenium, University College Dublin, 2-3 October 1998 (Last visited: 29-Nov-1999).

[338] A Cunningham. *A. Journey to the end of the night: custody and the dawning of a new era on the archival threshold*. Archives and Manuscripts, 24 (2), 1996, pp. 312-321, 1996.

[339] MW Day. *Preservation problems of electronic text and data*. East Midlands Branch of the Library Association, Occasional papers, no. 3, 1990. ... gives a comprehensive view of the technology and of the problems it poses for librarians and archivists. It also includes a useful bibliography. G. MacKenzie, In: A reading guide to the preservation of library collections, 1991. The technological parts of this publication are now rather dated.

[340] MW Day. *Extending metadata for digital preservation*. http://www.ariadne.ac.uk/issue9/metadata/, 1997. Ariadne, no. 9, May 1997 (Last visited: 17-Jun-1999).

[341] MW Day. *Online serials: preservation issues*. Haworth Press, The Serials Librarian, Vol. 33, Binghamton, N.Y.: Haworth Press, 1998, pp. 199-221, 1998. In: Jones, W., (ed.), E-serials: publishers, libraries, users, and standards.

[342] BL DeWhitt. *Long-term preservation of data on computer magnetic media*. Conservation Administration News, (29), 1987, 7, 19 & 28, and: (30), 1987, 4 & 28., 1987.

[343] S Dodson and Wellbeiser, J (comp.). *Bibliography of standards and selected references related to preservation in libraries*. National Library of Canada, http://www.nlc-bnc.ca/resource/presv/eintro.htm, 1996.

[344] CM Dollar. *Documentation of machine-readable records and research: a historian's view*. Prologue, 3 (1), 1971, 27-31., 1971.

[345] CM Dollar. *Appraising machine-readable records*. American Archivist, 41 (4), 1978, 423-430, 1978.

[346] CM Dollar. *Archival theory and information technologies: the impact of information technologies on archival principles and methods*. University of Macerata Press, print only, 1992.

[347] W Duff. *Ensuring the preservation of reliable evidence: a research project funded by the NHPRC*. Archivaria, 42, 1996, 28-45, 1996.

[348] L Duranti. *Reliability and authenticity: the concepts and their implications*. Archivaria, 39, 1995, 5-10, 1995.

[349] L Duranti. *The future of archival scholarship*. http://www.ucd.ie/˜archives/lucart.html, 1998. Paper delivered at: Cyber, Hyper or Resolutely Jurassic? Paper delivered at: Cyber, Hyper or Resolutely Jurassic?Paper delivered at: Cyber, Hyper or Resolutely Jurassic? Archivists and the Millenium, University College Dublin, 2-3 October 1998 (Last visited: 29-Nov-1999).

[350] L Duranti and Macneil, H. *The protection of the integrity of electronic records: an overview of the UBC-MAS Research Project*. Archivaria, 42, 1996, 46-67, 1996.

[351] L Duranti, Eastwood, T, and MacNeil, H. *The preservation of the integrity of electronic records*. University of British Columbia, School of Library, Archival & Information Studies, http://www.slais.ubc.ca/users/duranti/intro.htm, 1996. (Last visited: 17-Jun-1999).

[352] M Feeney. *Digital culture: maximising the nation's investment: a synthesis of JISC/NPO studies on the preservation of electronic materials*. National Preservation Office, print only, 1999.

[353] MH Fishbein. *Appraising information in machine language form*. American Archivist, 35, 1972, 35-43, 1972.

[354] K Gavrel. *Conceptual problems posed by electronic records: a RAMP study*. http://www.unesco.org/webworld/ramp/html/r9012e/r9012e00.htm, 1990. PGI-90/WS/12. Paris: UNESCO, General Information Programme and UNISIST, 1990. (Last visited:17-Jun-1999).

[355] S Gavrel. *Preserving machine-readable records: a reply to John Mallinson*. Archivaria, (22), 1986, 153-155, 1986.

[356] CL Geda, Austin, EW, and Blouin, FX (eds.). *Archivists and machine-readable records: proceedings of the Conference on Archival Management of Machine-Readable Records, February 7-10, 1979, Ann Arbor, Michigan*. Society of American Archivists, print only, 1980. In its day the most comprehensive account of the subject. Includes papers by Allan G. Bogue on US State data and Meyer H. Fishbein on appraisal. There are also descriptions of programs at the Public Archives of Canada (Harold Naugler), the National Archives (Charles M. Dollar) and the Public Record Office (Michael Roper).

[357] J Gertz. *Selection for preservation: a digital solution for illustrated texts*. Library Resources & Technical Services, 40 (1), 1997, 380-401, 1997.

[358] PS Graham. *Intellectual preservation and the electronic environment*. American Library Association, Proceedings of the 1992 Association for Library Collections and Technical Services President's Program, 18-38, 1992. In: Hirshon, A. (ed.) After the electronic revolution, will you be the first to go?

[359] PS Graham. *Intellectual preservation: electronic preservation of the third kind*. Commission on Preservation and Access, http://www.clir.org/cpa/reports/graham/intpres.html, 1994. (Last visited: 17-Jun-1999).

[360] PS Graham. *Long-term intellectual preservation*. Research Libraries Group, Digital imaging technology for preservation: proceedings from an RLG Symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York. pp41-58, 1994. In: Elkington, N.E., (ed.) Digital imaging technology for preservation: proceedings from an RLG Symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York.

[361] PS Graham. *Requirements for the digital research library*. College and Research Libraries, 56 (4), 1995, 331-339, 1995.

[362] PS Graham. *Building the digital research library: preservation and access at the heart of scholarship*. http://www.ukoln.ac.uk/services/papers/follett/graham/paper.html, 1997. Follett Lecture Series, Leicester University, 19 March 1997 (Last visited:17-Jun-1999).

[363] P Haddad and Jones, M. *Electronic publishing: access, bibliographic and preservation issues: report to the Electronic Publishing Working Group, Australian Vice-Chancellors' Committee*. National Library of Australia, http://www.adfa.oz.au/EPub/AvccNLA.html, 1995. (Last visited: 17-Jun-1999).

[364] P Haddad and Jones, M. *Electronic publishing: library and archival issues*. http://www.adfa.oz.au/Epub/key/Library.html, 1996. (Last visited: 17-Jun-1999).

[365] D Haynes, Streatfield, D, Jowett, T, and Blake, M. *Responsibility for digital archiving and long term access to digital data*. British Library Research and Innovation Centre, http://www.ukoln.ac.uk/services/papers/bl/jisc-npo67/digital-preservation.html, 1997. JISC/NPO Studies on the Preservation of Electronic Materials. British Library Research and Innovation Report, 67 (Last visited: 18-Jun-1999).

[366] M Hedstrom. *Optical disks: are archivists repeating the mistakes of the past?* Archives & Museum Informatics Newsletter, 2, 1988, 52., 1988.

[367] M Hedstrom. *Understanding electronic incunabula: a framework for research on electronic records*. American Archivist, 54 (3), 1991, 334-354, 1991.

[368] M Hedstrom. *Archives as repositories: a commentary*. Archives and Museum Informatics Technical Report, No. 13, 25-30, 1991. In: Bearman, D., (ed.) Archival management of electronic records.

[369] M Hedstrom. *Descriptive practices for electronic records: deciding what is essential and imagining what is possible*. Archivaria, 36, 1993, 53-63, 1993.

[370] M Hedstrom. *Electronic archives: integrity and access in the network environment*. Bowker-Saur, Networking in the humanities: proceedings of the Second Conference on Scholarship and Technology in the Humanities, 77-95, 1995. In: Kenna, S. and Ross, S. (eds.) Networking in the humanities: proceedings of the Second Conference on Scholarship and Technology in the Humanities held at Elvetham Hall, Hampshire, UK, 13-16 April, 1994.

[371] M Hedstrom. *Preserving the intellectual record: a view from the archives*. Library Association Publishing, Networking and the Future of Libraries 2: managing the intellectual record, 179-191, 1995. In: Dempsey, L., Law, D. and Mowat, I. (eds.) Networking and the Future of Libraries 2: managing the intellectual record.

[372] M Hedstrom. *Digital preservation: a time bomb for digital libraries*. Computers and the Humanities, 31 (3), 1997, pp. 189-202, 1997.

[373] M Hedstrom. *Research issues in migration and long-term preservation*. Archives and Museum Informatics, 11 (3/4), 1997, pp. 287-292, 1997.

[374] M Hedstrom. *Interfaces with time*. http://www.sims.monash.edu.au/rcrg/consulting/hedstrom.html, 1998. Keynote paper delivered at the Australian Society of Archivists Conference, 1998 (Last visited:29-Nov-1999).

[375] M Hedstrom and S Montgomery. *Digital preservation needs and requirements in RLG member institutions: a study commissioned by the Research Libraries Group (Last visited: 17-Jun-1999)*. Research Libraries Group, http://www.rlg.org/preserv/digpres.html, 1998. (Last visited: 17-Jun-1999).

[376] M Lesk. *Preserving digital objects: recurrent needs and challenges.* http://www.lesk.com/mlesk/auspres/aus.html, 1998. (Last visited: 17-Jun-1999).

[377] D Lievesley and Jones, S. *An investigation into the digital preservation needs of universities and research funders: the future of unpublished research materials*. British Library Research and Innovation Centre, http://www.ukoln.ac.uk/services/papers/bl/blri109/, 1998. JISC/NPO Studies on the Preservation of Electronic Materials. British Library Research and Innovation Report, 109 (Last visited: 18-Jun-1999).

[378] PW Lord. *Strategies and tactics for managing electronic data records: a view from the pharmaceutical industry*. Office for Official Publications of the European Communities, Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996, pp. 168-174, 1997. In: Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996. INSAR: European Archives News, Supplement II.

[379] SS Lukesh. *E-mail and potential loss to future archives and scholarship or the dog that didn't bark*. http://www.firstmonday.dk/issues/issue4_9/lukesh/index.html, 1999. (Last visited: 04-Oct-1999).

[380] CA Lynch. *Rethinking the integrity of the scholarly record in the networked information age*. http://www.educause.edu/pub/er/review/reviewArticles/29238.html, 1994. Educom Review, 29 (2), March/April 1994. (Last visited: 17-Jun-1999).

[381] CA Lynch. *The integrity of digital information: mechanics and definitional issues*. Journal of the American Society for Information Science, 45 (10), 1994, 737-744, 1994.

[382] CA Lynch. *Canonicalization: a fundamental tool to facilitate preservation and management of digital information*. http://www.dlib.org/dlib/september99/09lynch.html, 1999. D-Lib Magazine, 5 (9), September 1999 (Last visited: 04-Oct-1999).

[383] MS Lynn. *Digital preservation and access: liberals and conservatives*. Research Libraries Group, Digital imaging technology for preservation: proceedings from an RLG Symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York, 1-9, 1995. In: Elkington, N.E., (ed.) Digital imaging technology for preservation: proceedings from an RLG Symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York.

[384] J McDonald. *Managing information in an office systems environment: the IMOSA Project*. K.G. Saur, In: Menne-Haritz, A. (ed.) Information handling in offices and archives. München: K.G. Saur, 1993, pp. 138-151, 1993.

[385] J McDonald. *Managing information in an office systems environment: the IMOSA Project*. American Archivist, 58 (2), 1995, pp. 142-153., 1995.

[386] J McDonald. *Managing records in the modern office: taming the wild frontier*. Archivaria, 39, 1995, pp. 70-79, 1995.

[387] J McDonald. *Record keeping systems: lessons learned from the experience of the Canadian federal government*. http://www.archivists.org.au/events/conf99/mcdonald.html, 1999. Paper delivered at the Australian Society of Archivists 1999 Conference, Archives at Risk: Accountability, Vulnerability and Credibility, Brisbane, 29-31 July 1999. (Last visited: 29-Nov-1999).

[388] J McDonald. *Archives at risk: accountability, vulnerability and credibility - Keynote presentation*. http://www.archivists.org.au/events/conf99/keynote.html, 1999. Paper delivered at the Australian Society of Archivists 1999 Conference, Archives at Risk: Accountability, Vulnerability and Credibility, Brisbane, 29-31 July 1999. (Last visited: 29-Nov-1999).

[389] S McInnes. *Electronic records: the new archival frontier?* Journal of the Society of Archivists, 19 (2), 1998, 211-220, 1998.

[390] G MacKenzie. *New technology*. Library Association Publishing Ltd, n: Kenny, G. (ed.) A reading guide to the preservation of library collections, 87-94, 1991.

[391] J Mackenzie Owen and Walle, J. van de. *Deposit collections of electronic publications*. Office for Official Publications of the European Communities, print only, 1996.

[392] JC Mallinson. *Preserving machine-readable records for the millennia*. Archivaria, (22), 1986, 147-152, 1986.

[393] JC Mallinson. *On the preservation of human and machine-readable records*. Information Technology and Libraries, 7 (1), 1988, 19-23, 1988.

[394] JC Mallinson. *Magnetic tape recording: history, evolution and archival considerations*. National Archives of Canada, In: Proceedings of conservation in archives: International Symposium, Ottawa, Canada, May 10-12, 1988, 181-190, 1988.

[395] Charles W Bailey, Jr. *Network-Based Electronic Publishing of Scholarly Works: A Selective Bibliography*. http://LIB-04.LIB.UH.EDU/pacsrev/1995/bail6n1.html, 1996. A very comprehensive bibliography, with some useful references on digital preservation. Print reference: The Public-Access Computer SystemsReview 6, no. 1 (1995). This online version is Version 26 (10/25/96). Section 2.6.4, Information Integrity and Preservation.

[396] Charles W Bailey, Jr. *Scholarly Electronic Publishing Bibliography*. http://info.lib.uh.edu/sepb/sepb.html, 1996. Contains many useful links to digital preservation and metadata; most articles cited are post-1990. Now on Version 30: 4/28/2000.

[397] Edward A Galloway and Gabrielle V Michalek. *The Heinz Electronic Library Interactive Online System (HELIOS): Building a Digital Archive Using Imaging, OCR, and Natural Language Processing Technologies*. http://LIB-04.LIB.UH.EDU/pacsrev/1995/gall6n4.htm, 1995. Paper describing digitisation project of a Senator's congressional papers, using NLP search engine across an unstructured archive. No use of preservation metadata is explicity referred to. Print reference: The Public-Access Computer Systems Review 6, no. 4 (1995).

[398] *Imprimatur - Esprit Project 20676*. http://www.imprimatur.alcs.co.uk/, 1998. The Imprimatur Project closed at the end of 1998. Its work is being carried forward by Imprimatur Services Ltd. The website describes Imprimatur as Ground-breaking work on Electronic Copyright Management Systems; final outcomes of the project and information on future seminars are available.

[399] *Isaac Network*. http://scout.cs.wisc.edu/research/index.html, 2000. This is the current project of the Internet Scout Project, which proposes to link geographically distributed collections of metadata into a virtual collection searchable as a unified whole.

[400] Carolyn Caywood. *Library Selection Criteria for WWW Resources*. http://www6.pilot.infi.net/˜carolyn/criteria.html, 1995. A checklist which, whilst now dated, gives an insight into how one person felt web-based resources could be selected by libraries. An early version of this paper appeared on page 169 of the May/June, 1996 issue of Public Libraries.

[401] Mary E Jackson. *Subject: draft guidelines for electronic reserve systems - comments*. http://scilib.ucsd.edu/electclass/FairUseE-Reserve.html, 1995. These draft guidelines were prepared by the author, an ARL Access & Delivery Services Consultant, which proposes fair use guidlelines for such systems.

[402] Kurt Maly, Michael L Nelson, and Mohammad Zubair. *Smart Objects, Dumb Archives: A User-Centric, Layered Digital Library Framework*. http://www.dlib.org/dlib/march99/maly/03maly.html, 1999. Paper outlining a reference implementation of a protocol to facilitate interoperability between digitial libraries. A special class of digital objects called buckets were created. Their work used a number of archives based on a NASA collection and NSF funded projects.

[403] *Dublin Core Metadata Initiative - Home Page*. http://purl.oclc.org/dc/, 1998. The Dublin Core is a metadata element set intended to facilitate discovery of electronic resources. The official UK mirror site can be found at http://mirrored.ukoln.ac.uk/dc/.

[404] Paul Miller. *Discovering Online Resources. Unifying Resource Discovery Metadata for the Humanities: An Application Based Upon the Dublin Core*. http://ahds.ac.uk/public/metadata/disc_05.html, 1997.

[405] Andy Powell. *Metadata - Collection Level Description*. UKOLN, http://www.ukoln.ac.uk/metadata/cld/attributes/issues/, 1998. An issues list which provides a good summary of considerations surrounding collection level description.

[406] Michael Day. *Data models for metadata: Some issues for the Dublin Core initiative (Draft)*. UKOLN, http://www.ukoln.ac.uk/metadata/data-models/draft-report.html, 1998. Draft report describing two entity-relationship models for metadata: the model produced by the IFLA Study Group on Functional Requirements for Bibliographic Records in 1996 and an integrated model for descriptive and rights metadata developed by Godfrey Rust. A preliminary attempt is made to relate these to work being carried out with relation to Dublin Core and the Resource Description Framework.

[407] Willy Cromwell-Kessler and Ricky Erway. *RLG 7/97 Metadata Summit Report*. RLG, http://www.rlg.org/meta9707.html, 1997. Report of an informal, invitational summit meeting on metadata issues which was proposed by the members of the Task Force on Meta Access sponsored by the Association of Library Collections and Technical Services (ALCTS) of the American Library Association.

[408] *WebDOC and Arches Infrastructure Overview*. RLG, http://www.rlg.org/demo/webdoc.arches.html, 1999. Between 1996 and 1998, RLG developed in the context of two separate projects, the essential infrastructure to support links on the World Wide Web between descriptive (cataloging) records, mid-level metadata (such as archival finding aids), and whole digitized information objects (such as articles and images). This infrastructure includes navigational tools plus a full suite of access control and management functions, which ensure that rights holders and other value-added providers can restrict and/or charge for the use of their documents, if necessary.

[409] *ArchesArchival Server and Test Bed*. RLG, http://www.rlg.org/strat/projarch.html, 1999. The name Arches is a concatenation of the term archival server, suggesting the bridges between various forms of metadata and the whole information objects they describe. It also is intended to reflect a number of doorways into the research resources made available by RLG.

[410] Betty Furrie and Data Base Development Department of The Follett Software Company. *Understanding MARC Bibliographic: Machine-Readable Cataloging*. Library of Congress, http://lcweb.loc.gov/marc/umb/umbhome.html, 1988. Excellent guide to understanding MARC records - from describing what they are and why they are important, to a summary of the most commonly used MARC fields with an example in various formats.

[411] Margaret St. Pierre and William P LaPlant Jr. *Issues in Crosswalking Content Metadata Standards*. NISO, http://www.niso.org/crsswalk.html, 1998. This paper aims to to delineate the general issues involved in the harmonization of metadata standards and in the development of crosswalks between related metadata standards, and proposes future steps for simplifying crosswalk implementation through the use of formal specifications and automation.

[412] Richard Higgins. *Standardised Languages for Data Exchange and Storage. The Encoded Archival Description: using SGML to create permanent electronic handlists*. http://www.dur.ac.uk/Library/asc/eadarticle.html, 1997. HTML version of an article in Business Archives Principles and Practice, 73 (May 1997).

[413] *RLG EAD Support Site*. RLG, http://www.rlg.ac.uk/rlgead/, 1999. This site is intended to assist Research Libraries Group members in their efforts to use the Encoded Archival Description Document Type Definition (EAD DTD) for the purpose of making their finding aids accessible on the World Wide Web.

[414] Line Pouchard. *Cataloging for Digital Libraries: The TEI Scheme and the TEI Header*. http://mirrored.ukoln.ac.uk/lis-journals/review/review/6/pouchard.html, 1998. This article describes the uses and advantages of using the Text Encoding Initiative (TEI) guidelines for cataloging electronic texts. Detailed description and examples of the TEI header, a vehicle for meta-information written in SGML and the part of the TEI scheme most useful to librarians, are provided. URL supplied is the UKOLN mirror, reference for original is: Katharine Sharp Review ISSN 1083-5261, No. 6, Winter 1998 [http://edfu.lis.uiuc.edu/review/6/pouchard.html].

[415] *TEI Consortium*. http://www.tei-c.org/, 2000. The TEI is an international project to develop guidelines for the encoding of textual material in electronic form for research purposes. A new consortium has been formed for the maintenance and continuing work of the TEI. Documents on this site include TEI in XML, plus software and supporting articles.

[416] *EAD Help Pages - Metadata*. http://jefferson.village.virginia.edu/ead/metadata.html, 2000. This page is intended to provide links to information about various metadata formats and to information about software designed to work with structured text. Provided by The EAD Round Table of the Society of American Archivists.

[417] John Unsworth. *Esoteric vs. Exoteric: Economic Models for Electronic Publishing*. http://jefferson.village.virginia.edu/˜jmu2m/esoteric.publishing.html, 1997. Paper outlining alternative economic models for electronic publishing. This may provide insights into rethinking issues relating to digital preservation of journals, where economic models are still in their infancy. This paper was originally delivered at Electronic Publishing: A Day Conference, University of London, January 31, 1997.

[418] Willard McCarty. *Overview of online publication*. http://www.kcl.ac.uk/humanities/cch/ohc/overview.html, 1997. The aim of this page is to provide sources toward a provisional typology of online publishing.

[419] Chrysanne Lowe. *Challenges of Managing and Delivering Electronic Journals*. http://oregon.uoregon.edu/˜csundt/CALslides/index.htm, 1997. Online slide show of presentation by a publisher, providing an insight into a commercial point of view.

[420] *The CD Information Center*. http://www.cd-info.com/CDIC/index.html, 1999. Homepage of the CD-Info Company and its sponsors, containing some useful data on media longevity and a bibliography on compact disc development and production.

[421] *NISO Z39.53 Language Codes*. http://www.oasis-open.org/cover/nisoLang3-1994.html. Reference guide for language codes used in a variety of metadata. [Mirrored from: http://adm5.byu.edu/Images/photoid/aacraoig/segments/dataelmt/nisolang.html].

[422] *dc-relation*. http://www.mailbase.ac.uk/lists/dc-relation/. Electronic discussion list to support the efforts of the international Dublin Core effort's Relation and Source working group in exploring issues directly related to the Dublin Core elements Relation and Source.

[423] David Bearman. *Dublin Core Metadata Initiative - Relations Working Group - Relation Element Working Draft 1997-12-19*. http://purl.org/dc/documents/wd-relation-current.htm, 1997. This is a Dublin Core Metadata Initiative Working Draft for review by Dublin Core members and other interested parties. This report is a record of the efforts of the Relations working group constituted following the Helsinki Metadata Workshop (DC-5) of October, 1997. The recommendations are provisional and can be expected to evolve in the light of further implementation experience.

[424] *An Introduction to Compact Discs*. http://www.gi.alaska.edu/crc/cdrom/cdrom.html. Easy to understand FAQ which explains what is required to read a CD, and the different data formats used to store data on a CD.

[425] Karl-Erik Tallmo. *The electronic age - on the verge of total memory loss?* http://art-bin.com/art/imemory1.html, 1993. The first of three articles published in the Swedish daily Svenska Dagbladet in 1994, raising awareness of the need for preservation of digital resources.

[426] Terry Cook. *It's 10 O'Clock: Do You Know Where Your Data Are?* http://www.techreview.com/articles/dec94/cook.html, 1995. Article with examples of corporations facing what he calls electronic memory loss.

[427] Brewster Kahle. *Archiving the Internet*. http://www.archive.org/sciam_article.html, 1996. This article gives an overview of issues involved in archiving the internet, mentions metadata but does not go into great detail.Submitted to Scientific American for March 1997 Issue.

[428] Margit AE Dementi. *Access and Archiving as a New Paradigm*. http://www.press.umich.edu/jep/03-03/dementi.html, 1998. Includes issues relating to dynamic texts, and cites the work of JSTOR. This paper was presented at the Faxon Institute Colloquium Electronic Publishing and the Scholarly Communication Process, January 7-8, 1998 and was published in The Journal of Electronic Publishing March, 1998 Volume 3, Issue 3 ISSN 1080-2711.

[429] Donald J Waters. *Toward a System of Digital Archives: Some Technological, Political and Economic Considerations - The Rearranging Effect*. Association of Research Libraries, http://arl.cni.org/arl/proceedings/131/waters.html, 1997.

[430] *New Horizons in Scholarly Communication - Building and Organizing Collections*. Librarians Association of the University of California, http://libweb.ucsc.edu/scomm/collect.html, 1996. This section of New Horizons in Scholarly Communication addresses issues surrounding the process of building and organizing collections of resources for access online; from the process of creating documents in online formats, to selecting materials to include in digital collections, to organizing collections, to retrieval and access, to preservation of these resources for future generations. Includes a dedicated metadata section.

[431] *Internet Library of Early Journals*. http://www.bodley.ox.ac.uk/ilej/. An eLib (Electronic Libraries Programme) Project by the Universities of Birmingham, Leeds, Manchester and Oxford. The final report of the ILEJ project is available from the website, documenting the decisions and processes from the project.

[432] Charles R Musick, Jr. *Use of meta-data constructs for long-term database archival*. http://www.llnl.gov/liv_comp/metadata/mail_dir/0081.html, 1998. Thread on the IEEE Metadata mail archive website, which considers multiple data types held in a variety of database structures.

[433] John Mackenzie Owen. *Organizing for digital archiving: new distribution models in the scientific information chain*. Portland Press, http://tiepac.portlandpress.co.uk/books/online/tiepac/session5/ch5.htm, 1997.

[434] *Resource Discovery Unit Home Page*. Distributed Systems Technology Centre, http://www.dstc.edu.au/RDU/, 1998. The Resource Discovery Unit (RDU) is a project from the Research Data Network CRC and is operated by the DSTC. The RDU aims are to investigate and develop tools, technologies, and information management processes that will allow organisations to locate, access, retrieve, and manage information on highly distributed hetergeneous networks. Projects include a number of metadata prototypes.

[435] *DELOS*. http://www.iei.pi.cnr.it/DELOS/, 1996. The DELOS Working Group was funded by the ESPRIT Long Term Research Programme (LTR No. 21057) within the Fourth Framework Programme of the Commission of the European Union, from 1996-1999. The 5th Framework programme has funded The DELOS Network of Excellence (NoE) was to promote research into the further development of digital library technologies. Information on the activities of the working group can b.

[436] *MICI - Metadata Information Clearinghouse (Interactive)*. http://domino.wileynpt.com/NPT_Pilot/Metadata/mici.nsf, 1999. This site was initiated by the Metadata Subcommittee of the Enabling Technologies Committee of the AAP (Association of American Publishers). It is intended to serve as a clearinghouse of information about Metadata projects, standards, and initiatives. Aimed at the publishing industry - press release available from http://www.publishers.org/home/press/mici.htm.

[437] *1995 National Preservation Office (NPO) Conference - Multimedia Preservation: Capturing The Rainbow*. NLA, http://www.nla.gov.au/niac/meetings/npo95.html, 1995. Online procedings of conference held 27-30 November 1995, Brisbane, Queensland, Australia. Print procedings are available, email contact is nlasales@nla.gov.au.

[438] *Journal of Internet Cataloging*. Haworth Press, Inc, http://www.haworthpressinc.com/jic/home.html, 1996. This website provides abstracts of articles (including some on metadata), but only a minority of articles are available to read from the website.

[439] Frances L Knudson, Nancy R Sprague, Douglas A Chafe, Mark LB Martinez, Isabel M Brackbill, Vicky A Musgrave, and Kathleen A Pratt. *Creating Electronic Journal Web Pages from OPAC Records*. http://www.library.ucsb.edu/istl/97-summer/article2.html, 1997. The Research Library at Los Alamos National Laboratory (LANL) automatically generates a set of electronic journal web pages from information contained in an online catalog MARC record. The goal is to present the users with multiple access points to electronic journals, using MARC records.

[440] Jakob Nielsen. *Fighting Linkrot*. http://www.useit.com/alertbox/980614.html, 1998. Article from the author's popular Alertbox series; gives pointers to web publishers on how to ensure resources can still be discovered.

[441] Martin Bryan. *Conference Report - Metadata Matters*. http://guagua.echo.lu/oii/en/metamatters.html, 1999. Report on one day workshop held at Brunei Gallery, London School of African and Oriental Studies, 10th March 1999.

[442] *The JISC/NPO Digital Preservation Workshop 3/4 March 1999, University of Warwick*. http://www.leeds.ac.uk/cedars/OTHER/warwick2.htm, 1999. Brief report and presentations from an invitation only workshop funded jointly by the Joint Information Systems Committee of the Higher Education Funding Councils and by the British Library.

[443] Susan S Lukesh. *E-mail and potential loss to future archives - The dog that didn't bark*. http://firstmonday.org/issues/issue4_9/lukesh/, 1999. Explores the preservation of electronic correspondence. The premise of this paper is that if electronic correspondence, personal e-mail, is not retained, there will be significant loss to future understanding of the work of today's scholars', to historians' work in general and to our collective memory.

[444] G Matthews. *Preservation of digital materials: policy and strategy issues for the UK: report of a meeting held at the British Library Research and Innovation Centre, London, 13th December 1996*. British Library Research and Innovation Centre, http://www.ukoln.ac.uk/services/papers/bl/blri041/digpres.html, 1997. JISC/NPO Studies on the Preservation of Electronic Materials. British Library Research and Innovation Report, 41 (Last visited:18-Jun-1999).

[445] A Menne-Haritz and Brübach, N. *The intrinsic value of archive and library material*. http://www.uni-marburg.de/archivschule/intrinsengl.html, 1997. print version at: Digitale Texte der Archivschule Marburg, Nr. 5. (Last visited: 17-Jun-1999).

[446] J Michaels. *Here today, gone tommorrow? Why we should preserve electronic documents*. Learned Information, In: Higginbotham, B.B. (ed.) Advances in preservation and access, 2, 293-307, 1995.

[447] A Michelson and Rothenberg, J. *Scholarly communication and information technology: exploring the impact of changes in the research process on archives*. American Archivist, 55, 1992, 236-315, 1992.

[448] J Mohlenrich. *Preservation of electronic formats & electronic formats for preservation*. Highsmith, print only, 1993.

[449] A Murdock. *Roles and responsibilities in managing an electronic archive*. Office for Official Publications of the European Communities, In: Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996. INSAR: European Archives News, Supplement II. pp. 118-122, 1997.

[450] National Archives of Australia. *Managing electronic records: a shared responsibility*. Australian Archives, http://www.naa.gov.au/govserv/techpub/manelrec/MER.htm, 1997. (Last visited: 06-May-1999).

[451] National Archives of Australia. *Managing electronics messages as records*. Australian Archives, http://www.naa.gov.au/govserv/techpub/messrecs/mangelec.htm, 1997. (Last visited: 06-May-1999).

[452] National Archives of Australia. *Recordkeeping Metadata Standard for Commonwealth Agencies*. National Archives of Australia, http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm, 1999. (Last visited: 20-Apr-2000).

[453] National Library of Canada Electronic Collections Coordinating Group. *Networked electronic publications policy and guidelines*. National Library Of Canada, http://www.nlc-bnc.ca/pubs/irm/eneppg.htm, 1998. (Last visited: 06-May-1999).

[454] H Naugler. *The archival appraisal of machine-readable records: a RAMP study with guidelines.* UNESCO, PGI-84/WS/27, Paris: UNESCO, General Information Programme and UNISIST, 1984, 1984.

[455] GB Neavill. *Electronic publishing, libraries and the survival of information*. Library Resources & Technical Services, 28, 1984, 76-89, 1984.

[456] GB Neavill and Sheblé, MA. *Archiving electronic journals*. Serials Review, 21 (4), 1995, 13-21, 1995.

[457] TH Peterson. *Machine-readable records as archival materials*. Archivum, 35, 1989, 83-88, 1989.

[458] FW Ratcliffe and Patterson, D. *Preservation policies and conservation in British libraries: report of the Cambridge University Library Conservation Project*. British Library, Library and information research report, 25, 1984.

[459] WB Rayward. *Libraries, museums and archives in the digital future: the blurring of institutional boundaries*. http://www.nla.gov.au/niac/meetings/npo95wr.html, 1995. Keynote address presented at the 2nd National Preservation Office Conference: Multimedia Preservation - Capturing the Rainbow, in Brisbane, 28-30 November 1995. Canberra: National Library of Australia (Last visited: 17-Jun-1999).

[460] WG on Preservation and Reformatting Information Research Libraries Group. *Preliminary report: proposed enhancements to MARC 007 values for Computer files*. http://www.rlg.org/preserv/pri-int.html, 1998. (Last visited: 17-Jun-1999).

[461] WG on Preservation Issues of Metadata Research Libraries Group. *Final report*. http://www.rlg.org/preserv/presmeta.html, 1998. (Last visited: 17-Jun-1999).

[462] Preservation Working Group on Digital Archiving Research Libraries Group. *Preliminary Report: Recommendations for RLG*. http://www.rlg.org/preserv/archpre.html, 1998. (Last visited: 17-Jun-1999).

[463] D Roberts. *The disposal and appraisal of machine-readable records from the literature*. Archives and Manuscripts, 13, 1985, 30-38, 1985.

[464] D Roberts. *Defining electronic records, documents and data*. Archives and Manuscripts, 22, 1994, 14-26, 1994.

[465] M Roosa. *Magnetic media preservation: selected bibliography*. Library of Congress, National Preservation Program Office, gopher://marvel.loc.gov:70/00/services/preserv/bibs/magnetic, 1990. (Last visited: 17-Jun-1999).

[466] M Roper. *New information technology and archives*. UNESCO Journal of Information Science, Librarianship and Archives Administration, 4 (2), 1982, 107-113, 1982.

[467] M Roper. *Advanced technical media: the conservation and storage of audio-visual and machine-readable records*. Journal of the Society of Archivists, 7 (2), 1982, 106-112, 1982.

[468] S Ross. *Historians, machine-readable information and the past's future*. British Library Report 6122, pp. 1-20, 1993. In: Ross, S. and Higgs, E. (eds.) Electronic information resources and historians: European perspectives. Halbgraue Reihe zur historischen Fachinformatik, Serie A: Historische Quellenkunden, Bd. 20.

[469] S Ross. *Preserving and maintaining electronic resources in the visual arts for the next century*. Information Services & Use, 15 (4), 1995, 373-384, 1995.

[470] S Ross. *Opportunities in electronic information*. Manchester University Press, In: Brivati, B, Seldon, A and Buxton, J (eds.) Contemporary history: a handbook, 437-450, 1996.

[471] S Ross. *Consensus, communication and collaboration: fostering multidisciplinary co-operation in electronic records*. Office for Official Publications of the European Communities, In: Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996 INSAR:European Archives News, Supplement II, pp. 330-336, 1997.

[472] S Ross. *The expanding world of electronic information and the past's future*. Oxford University Press, In: Higgs, E., (ed.) Historians and electronic artefacts, pp. 6-28, 1998.

[473] S Ross. *Changing trains at Wigan: digital preservation and the future of scholarship*. http://www.leeds.ac.uk/cedars/OTHER/SRoss.htm, 1999. The JISC/NPO Digital Preservation Strategy Workshop, Scarman House, University of Warwick, 3-4 March 1999 (Last visited: 24-Jun-1999).

[474] S Ross. *Digital archaeology: rescuing neglected and damaged data resources*. Library Information Technology Centre, http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2con.pdf - continued at http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf, 1999. JISC/NPO Studies on the Preservation of Electronic Materials. eLib Supporting Study P2 (Last visited: 02-Dec-1999).

[475] Jeff Rothenberg. *Ensuring the longevity of digital documents*. Scientific American, 272 (1), 1995, 42-47, 1995.

[476] Jeff Rothenberg. *Avoiding technological quicksand: finding a viable technical foundation for digital preservation.* Council on Library and Information Resources, http://www.clir.org/pubs/reports/rothenberg/contents.html, 1999. (Last visited: 17-Jun-1999).

[477] K Russell. *Digital resources into the future: digital preservation and the Cedars project*. Serials, 12 (3), 1999, 259-267, 1999.

[478] K Russell. *The Cedars project: implementing a model for distributed digital archives*. http://www.rlg.org/preserv/diginews/diginews3-3.html#feature, 1999. RLG DigiNews, 3 (3), June 1999 (Last visited: 28-Jan-2000).

[479] W Saffady. *Optical disks vs. micrographics as document storage & retrieval technologies. Westport, Conn.: Optical disks vs. micrographics as document storage & retrieval technologies*. Meckler, print only, 1988.

[480] W Saffady. *Optical disk vs. magnetic storage*. Meckler, print only, 1990.

[481] W Saffady. *Optical storage technology, 1990-91: a state of the art review*. Meckler, print only, 1990.

[482] W Saffady. *Stability, care and handling of microforms, magnetic media and optical disks*. Library Technology Reports, 27 (1), 1991, 5-116, 1991.

[483] W Saffady. *Managing electronic records*. Association of Records Managers and Administrators, print only, 1992.

[484] FJ Stielow. *Archival theory and the preservation of electronic media: opportunities and standards below the cutting edge*. American Archivist, 55, 1992, 332-343, 1992.

[485] RP Sturges. *Policies and criteria for the archiving of electronic publishing*. Journal of Librarianship, 19 (3), 1987, 152-172, 1987.

[486] RP Sturges. *Archival problems of electronic publications*. Taylor Graham, In: Dyer, H. and Tseng, G., (eds.) New horizons for the information profession: meeting the challenge of change: proceedings of the annual conference of the Institute of Information Scientists, University of Warwick, 1987, 152-168, 1988.

[487] RP Sturges. *Preservation of electronic archives*. University of Wellington, Department of Library and Information Studies, Occasional papers in bibliography and librarianship, 16, 1991.

[488] Task Force on the Archiving of Digital Information. *Preserving digital information: report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group*. Commission on Preservation and Access, http://www.rlg.org/ArchTF/, 1996. (Last visited: 17-Jun-1999).

[489] K Thibodeau. *To be or not to be: archives for electronic records*. Archives and Museum Informatics, In: Bearman, D., (ed.) Archival management of electronic records. Archives and Museum Informatics Technical Report, No. 13, 1-13, 1991.

[490] UK Office for Library and Information Networking. *Long term preservation of electronic materials: a JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib) organised by UKOLN, 27th and 28th November 1995 at the University of Warwick*. http://www.ukoln.ac.uk/services/papers/bl/rdr6238/, 1995. British Library R&D Report 6238. This workshop report includes summaries of papers by Peter S. Graham, Margaret Hedstrom, Daniel I. Greenstein and Tony Hendley (Last visited: 17-Jun-1999).

[491] School of Information Sciences: University of Pittsburgh. *Functional requirements for evidence in recordkeeping*. http://www.lis.pitt.edu/˜nhprc/. (Last visited: 17-Jun-1999).

[492] U.S. House Committee on Government Operations. *Taking a byte out of history: the archival preservation of Federal computer records*. U.S. Government Printing Office, print only, 1990.

[493] JW Van Bogart. *Magnetic tape storage and handling: a guide for libraries and archives*. Commission on Preservation and Access, print only, 1995.

[494] D Wallace. *Metadata and the archival management of electronic records*. Archivaria, 36, 1993, 87-110, 1993.

[495] D Wallace. *Managing the present: metadata as archival description*. Archivaria, 39, 1995, 11-21, 1995.

[496] DJ Waters. *Electronic technologies and preservation: [...] based on a presentation to the Annual Meeting of the Research Libraries Group, June 25, 1992*. http://www.ifla.org/documents/libraries/net/waters1.txt, 1992. (Last visited: 17-Jun-1999).

[497] DJ Waters. *Electronic technologies and preservation*. Commission on Preservation and Access, http://www.clir.org/pubs/reports/waters/waters2.html, 1992. (Last visited: 17-Jun-1999).

[498] DJ Waters. *Transforming libraries through digital preservation*. Research Libraries Group, In: Elkington, N.E., (ed.) Digital imaging technology for preservation: proceedings from an RLG Symposium held March 17 and 18, 1994: Cornell University, Ithaca, New York, 115-127, 1995.

[499] DJ Waters. *Some considerations on the archiving of digital information*. http://www.oclc.org:5046/oclc/research/links/archtf/waters_dig_arch.html, 1995. (Last visited: 17-Jun-1999).

[500] DJ Waters. *The social organization of archiving digital information [DRAFT]*. http://www.oclc.org:5046/oclc/research/links/archtf/archtf_outline.html, 1995. (Last visited: 17-Jun-1999).

[501] DJ Waters. *Realizing benefits from inter-institutional agreements: the implications of the draft report of the Task Force on Archiving of Digital Information*. Commission on Preservation and Access, print only, 1996.

[502] DJ Waters. *Archiving digital information*. http://www.oclc.org/oclc/man/9680rldc/waters1.htm AND http://pantheon.cis.yale.edu/˜dwaters/rlactalk.html, 1996. A presentation to the OCLC Research Library Directors' Conference,Dublin, Ohio, March 12, 1996. (Last visited: 17-Jun-1999).

[503] DJ Waters. *Choices in digital archiving: the American experience*. http://tiepac.portlandpress.co.uk/books/online/tiepac/session5/ch6.htm, 1998. (Last visited:18-Jun-1999) Also in: Butterworth, I., (ed.) The impact of electronic publishing on the academic community: an international workshop organized by the Academia Europaea and the Wenner-Gren Foundation, Wenner-Gren Center, Stockholm, 16-20 April 1997. London: Portland Press, 1998.

[504] DJ Waters and A Kenney. *The Digital Preservation Consortium: mission and goals*. Commission on Preservation and Access, http://www.clir.org/pubs/reports/dpcmiss/dpcmiss.html, 1994. (Last visited: 17-Jun-1999).

[505] H Weber and Dörr, M. *Digitisation as a method of preservation?* Amsterdam: European Commission on Preservation and Access; Washington, D.C.: Commission on Preservation and Access, print only, 1997.

[506] GL Weinberg. *The end of Ranke's history? Reflections on the fate of history in the twentieth century*. In: Weinberg, G.L. Germany, Hitler and World War II, 325-336, 1995. Article by academic historian taking a pessimistic view of the future of historical research with reference to the increased security classification of US governmental records, the increased use of computers in society and the challenge of machine-readable records. The article was first published in: Syracuse Scholar, 9 (1), 1988, 51-59. See also his: The end of history? In: American Historical Association Perspectives, 25, Febuary 1987, 16-19.

[507] T van der Werf-Davelaar. *Long-term preservation of electronic publications: the NEDLIB project*. http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html, 1999. D-Lib Magazine, 5 (9), September 1999 (Last visited: 04-Oct-1999).

[508] T van der Werf-Davelaar. *Deposit Libraries and the NEDLIB Project*. http://www.exploit-lib.org/issue4/nedlib/, 2000. Exploit Interactive, 4, January 2000 (Last visited: 28-Jan-2000).

[509] BJS Williams. *Implications for preservation of the newer information media*. Information Media and Technology, 19 (10), 1985-86, 13-15, 1985.

[510] BJS Williams. *Implications for preservation of the newer information media*. National Preservation Office, In: Conservation in crisis: proceedings of a seminar, NPO seminar papers, 1, 1987, 57-64., 1987.

[511] R Williamson. *The Knowledge Warehouse*. British Library, Library and information research report, 65, 1987.

[512] R Williamson. *Electronic text archiving*. Elsevier Advanced Technology Publications, print only, 1988.

[513] DR Winterbottom and Fiddes, RG. *Life expectancy of write once digital optical discs*. British Library Research and Development Department, British Library research paper, 66, 1989.

[514] S Yorke. *Playing for Keeps: the Proceedings of an Electronic Records Management Conference hosted by the Australian Archives*. Australian Archives, http://www.naa.gov.au/govserv/techpub/keeps/keeps.htm, 1995. (Last visited: 06-May-1999).

[515] *WIPO Performances and Phonograms Treaty (Adopted by the Diplomatic Conference on 20 December 1996)*.

[516] *EU Directive 92/100 on rental rights, lending rights and on certain rights akin to copyright [OJ 1992 L346/61]*. 1994. Known as The Rental Directive, implementation date 1 July 1994.

[517] *EU Directive 93/83 on copyright and neighbouring rights relating to satellite broadcasting and cable retransmission [OJ 1993 L248/15]*. 1993. Known as The Satellite and Cable Directive, adopted September 1993, implementation date 1 January 1995.

[518] *EU Directive 96/9 on the legal protection of databases, OJ 1996 L77/20*. 1998. Known as The Database Directive, implementation date 1 January 1998.

[519] *EU Directive 93/98 on Term of Protection of Copyright [OJ 1993 L248/15]*. Known as The Duration Directive, implementation date 1 July 1995.

[520] *Proposal for directive - 'Copyright and related rights in the information society' [1998] OJ C108/6*. 1998.

[521] Dennis N Nicholson and Smith, Martin. *CATRIONA II - A Review of Internet File Formats*. http://wp269.lib.strath.ac.uk:5050/Cat2/formats.html, 1998. This is a review of a range of file formats commonly found on the web. Strengths and weaknesses include identification of proprietary issues, which informs dependencies which will affect long term preservation.

[522] *UPF Home*. http://info.wgbh.org/upf/index.html. This site disseminates information about the proposed Universal Preservation Format for the archiving of media assets.

[523] Maynard Brichford and William Maher. *Archival Issues in Network Electronic Publications*. Library Trends 43, (Spring 1995): 701-712, 1995.

Appendix Three : Typology

This is a typology of digital resources for Cedars. The purpose of this document is to:

- Identify and describe the dimensions which define the differences between various digital resources;
- Provide a framework which can be used by others to categorise their own digital resources for the purpose of digital preservation;
- Help generate a decision aid to assist librarians who need to decide which digital resources to preserve;
- Inform the selection of metadata elements for digital resources.

The typology consists of various dimensions detailed below. The typology can be said to work if the digital resources can be sufficiently described by the dimensions.

1. **Publication source**

   Definition Any person or body producing a digital resource. This includes traditional publishers, multimedia publishers, database publishers, producers of grey literature, and individuals producing publications.

   Function

   (a) Publication source could affect copyright negotiations and access provisions.
   (b) Funding bodies could make submission of the digital resource for preservation a necessary condition of the funding arrangement.

2. **Publication format**

   Definition The data format in which the item is published. This includes database formats, word processor formats, print formats, portable document formats, and structured document formats.

   Function The level of difficulty of preservation may vary between formats. This should not necessarily be used to exclude particular formats. However, some libraries may not have the ability to preserve some formats.

3. **Publication type**

   Definition The type of content of the digital object. This includes monographs, journals, databases, hyperlinked documents, applications software, computer games, and communications, for example e-mail and usenet.

   Function

   (a) Libraries may choose digital objects for preservation based on type of content. Research libraries may decide to exclude computer games and applications software. Libraries may decide to preserve, for example, editions of monographs but not snapshots of databases.
   (b) Preserving digital objects which have hyperlinks to remote documents, or digital objects which generate documents as a result of user interaction, may pose challenges for some libraries.

4. **Publication medium**

   Definition The physical medium, or delivery method, of the digital object. This includes off-line media, on-line media, and hybrid media.

   Function The level of difficulty of preservation may vary between media, in a similar way to the difficulty of preserving various publication formats. Libraries may not have the technical ability to preserve all publication media.

5. **Legal Status**

   Definition Any available information on the legal status of the digital object. This may include information about copyright status or database right, or any relevant licensing agreement.

   Function Libraries will need to be able to reach agreements with all owners of relevant rights in any given digital object, in order to preserve the object. Separate agreements may need to be reached to provide access to the object.

6. **Nationality of publication**

   Definition The national origin of a digital object.

   Function Research libraries which are also deposit libraries may make decisions based on the origin of an object.

7. **Parallel publications**

   Definition Any digital resource which could be seen as a copy of the content of another digital resource or printed publication. Digital objects may have identical intellectual content but vary in format or medium.

   Function Libraries may prefer certain formats or media, on the grounds of ease of preservation.

8. **Technical environment**

   Definition The technical environment which renders the digital object. This includes the hardware, the operating system, the access or database software, the network access method, and the viewer or browser software.

   Function The difficulty of preserving a digital object may vary depending on whether, for example, it is designed for on-line or off-line use. The same digital object may be published for various technical environments, one of which may be easier for the library to preserve.

9. **Content level and intellectual value**

   Definition The significance of the digital resource as judged by subject specialists in the relevant field or fields. Digital objects may contain content of varying utility for research libraries.

   Function Some classes of materials, for example some forms of ephemera, may not be considered worth preserving.

10. **Subject area**

    Definition The subject area of the content of the digital object.

    Function Libraries will have strengths in particular subject areas, which they may wish to foster.

Appendix Four : The Metadata Specification Walkthrough

Cedars Metadata Walkthrough Day, Birmingham University Library, April 2000

4.1 Introduction

In April 2000, the Cedars project published a draft outline specification for digital preservation metadata for public comment, as part of their work into exploring long term digital preservation. Parallel to the collection and collation of the public comments, an invite-only “walkthrough day” was held at Birmingham University Library, which brought together a range of individuals who were involved in various digital preservation activities.

This group used the draft specification to go through a range of digital resources which were under consideration within the Cedars project, not only to assign values to the elements in the specification, but also to raise and discuss issues which arose, to inform the subsequent development of this document.

This report gives a background to the outline specification document, the objectives and structure of the day, and the outcomes. The key issues from the day are listed, and completed templates for each resource are shown. These findings are presented in a style consistent with the tables used in the draft metadata specification (Cedars Project Team, 2000), for ease of use.

4.2 Background

Previous drafts of what has become the outline specification document (e.g. Stone and Day 1999, Day 1999) took metadata elements from a variety of selected initiatives in the Cedars metadata review, and mapped them onto the taxonomy of the information package identified in the OAIS model (CCSDS, 1999). A version of that document was circulated for selective consultation in December 1999.

A key objective of the metadata walkthrough exercise was to work through the metadata set in the draft specification (Cedars Project Team, 2000) with a range of individuals with mixed backgrounds and skill sets, but all from the “digital libraries” community. It was anticipated that the feedback gained from this “expert panel” would inform how the Cedars project will deal with the assimilation of the feedback from the public review stage, which will be used to refine the metadata document in the light of this feedback, in addition to, incorporating the feedback obtained from the walkthrough day itself.

It was anticipated that the walkthrough would be more than merely a “filling in the blanks” exercise; from early on in the project, a number of issues had been raised which trigger a considerable amount of debate, with no “right” or “wrong” answers in general. By bringing together a number of stakeholders in the digital preservation arena, representing a variety of backgrounds, concerns, and knowledge bases, it was hoped that the discussions which would arise from the day would point to the main areas of discussion, and tease out questions which the different communities would wish to be answered. In addition, by bringing together such a range of people for this kind of session, these issues could be discussed literally round the table, and whilst email has been a useful form of communication to raise issues and have (fragmented) discussions, there has not been an opportunity to explore (and document) the issues and discussions arising from real-time face-to-face interactions until now.

This group contained a number of different knowledge bases; including those with experience in cataloguing, libraries systems, collection management, and digital text archiving. Similarly, there was a diversity in perspectives; from strategic and management to implementation. All made invaluable contributions to a stimulating and eye-opening set of discussions. It was felt by all that a face-to-face discussion of the issues, which this exercise provoked, not only allowed concerns and questions to be voiced, but also to gain understanding of the multitude of perspectives which necessarily exist in the digital preservation arena. Many felt that after the day, they has a greater appreciation of the issues surrounding long-term digital preservation, and the work which needs to be done to continue the research and development in this field will involve their skills and experience. Similarly, the need for communication and interdependence between these various experts is essential for this work to progress effectively - it is only by continuing this process that we can move towards realistic and effective long term digital preservation solutions.

4.3 Structure of the Day

Kelly Russell began the day with an overview presentation, which reiterated the background as outlined above, and outlined some basic working assumptions for the walkthrough exercise. These were:

- Preservation metadata will be the responsibility of the digital archive
- We are acting as "cataloguers" at an OAIS level
- We are working in isolation (recognising that for a real digital archive, this would be very unlikely)
- The metadata will be updated as part of the archive's administrative function
- Each manifestation of a work will have a metadata record - which might be linked to previous records

The walkthrough exercise itself involved the group working together, completing a form which contained the metadata items from the draft specification. Each person completed the form on their own, and then the results were collated and discussed. Not everyone felt they could supply a value for each item; in some cases this identified an issue (or set of issues) which, in some cases, resulted in discussion which elaborated on these.

Some issues were mentioned in more than one exercise, but these have been cited in the exercise they were first identified in for the sake of simplicity. In certain cases, these issues manifested themselves in slightly different ways, due to the different nature of a resource. These results will inform the further development of the metadata specification, along with the feedback from the first public draft.

4.4 Table summarising issues raised in walkthrough

These issues are ones raised in the walkthrough which are of relevance beyond each distinct resource. These are listed in the same table format as the walkthrough resources (which is the same format as in the draft metadata specification), for consistency.

A general point regarding the Outline Specification is that it will need to make clear where possible which elements it assumes will accompany the object and which will need to be generated by the archive. This may vary according to different types of resource, but it is expected that this will become easier to identify with the assistance of the test sites.

There was also a discussion regarding viruses: David Price remarked that due to virus problems in the past with digital resources, the Bodleian Library has a procedure which involves virus checking. Kelly Russell agreed that virus checking could be a pre-requisite, and is something which archives must keep in mind. A side discussion arose regarding the preservation of viruses, but this issue was felt to be distinct from the work in hand. However, this is something which needs to be flagged, and should be considered in relation to the core Cedars work.

PDI 1.1.1.1 Resource Description:	
dcTitle:	For series titles, this needs to be repeatable (also at collection level?) - it is easy for confusion with respect to granularity to develop here, unless a priori assumptions are made when dealing with these types of resource by the archive, and these are referred to.
dcDate:	There needs to be clarification what this actually indicates: i.e. date of copyright, date of digitisation, date of ingest? (This could be specified as an XML attribute in our next version)

The Resource Description section raised much discussion as to the use of the proposed Dublin Core elements. It was recognised that there would be significant overlap between this and other sections of the metadata, but it was suggested that an algorithm could be used to populate the relevant DC fields from within the preservation record.

The resource description elements do not need to be over-detailed, as this is part of the metadata will not be used for resource discovery per se. The "core set" of elements needs to be further identified for this section, work done on mappings between DC and MARC may inform this activity (title, responsibility statement, publishing statement and date were mentioned as particular areas to explore).

1.1.1.2 Existing Metadata:	
Existing Record:	Cases exist where, for parallel publications, MARC records exist for the print version, although will need to consider what to do where no distinct MARC record exists for electronic version - explore if there can be a transformation which can append a relevant comment in the appropriate field? Also, there are MARC-DC converters which may be useful in populating the Reference Information section above; it may be possible to develop algorithms to facilitate this (and to populate DC fields with appropriate values from other resources). Of particular interest are mappings to Title, RespStmt, PubStmt, and Date. If a service is being preserved, there are a number of granularity issues which need to be dealt with. This will have implications on how structure of any existing metadata is stored, and consequently how existing metadata is referred to.

1.1.2 Context Information:	
Related Information Object:	We need to clarify what we mean by "significantly related" as an attribute. There are two ways to consider this: physical (e.g. hyperlinks), and logical (i.e. as part of a series). This needs to be explored further, and may help to deal with notions of granularity. Context information can also mean packaging, etc - there can be important information contained in packaging; this also gives a user a better idea of how the resource was supplied (where applicable) The notion of describing granularity was raised (see above) There may be varying extents of documentation; e.g. for a web-based service, the documentation available on this may differ greatly from that held by the service provider. This may not be publicly available for security issues, although it should be archived if we are to successfully archive the service. Where available, but already contained within the archived resource, distinct referencing to the documentation (e.g. a separate CRID) would be useful.

It was agreed that this element needs something to explicitly state the type of relationship.

1.1.3 Provenance Information:	
1.1.3.1 History of Origin:	
Reason for Creation:	This should be considered “Less Significant” because it would be less relevant to materials “born digital”. It was felt that subjectivity would inevitably arise for most materials. A list of pull down options was suggested, for Digitised materials: these options could include: • For access • For preservation • Commercial reasons • Exploring digitisation
Custody History:	As with DC.Date, date of creation needs to be recorded - in the case of custody history, this should only refer to the digital file. This should be a requirement for archiving, and should be required for digitised materials
Change History Before Archiving:	This is pre-ingest information which is different from Ingest Process History (see below) which documents what the current archive does to the file on ingest. Another suggestion was to ensure each change has a “name” and “date” field for each change - i.e. we know who did it, and when. In the case of funded projects which are to be archived, it is not clear whether this section should be used to document measures taken since the end of the funding period, also in the case of joint projects whose responsibility this would be after the end of the funding period. This needs to be raised with the appropriate bodies, since this is not documented at present, not any best practice established. Work needs to take place in this area with the funding bodies, to ensure this information is captured at the time - doing this after the event may not be possible in many cases.
1.1.3.1.4 Original Technical Environments:	
Prerequisites:	Prerequisites should be not just those available, but systems required. On multi-platform resources, resource functionality may vary across platforms. Some functionality may be included as evaluation versions of software; e.g. time-limited.
Procedures:	Should Original Technical Environment include another sub-element for physical carrier? Archivists particularly may want to include information on the original medium or carrier; this may not have been the medium on which it was published (e.g. websites on CD)
Documentation:	When a “service” is preserved, it is likely that some commercial software will be involved: it may not be possible to get permission to preserve this from vendors (documentation or the software). Alternatively, one could document what each piece of software did, and what procedures this involved, this could assist in the development of tools at a later stage. In the case of projects, documentation delivered as a project deliverable may differ from that which was used by the resource when it was being developed and maintained.
Reason for Preservation:	It was suggested that this element could have a drop down list although there was little discussion about what this list would contain aside from Legal Deposit..... This element requires a list of options (a drop-down list was initially suggested, but we would need the ability to capture and specify multiple reasons, i.e. checkboxes). The following were identified during the walkthrough: • Legal Deposit • Intellectual Content • Demonstrator Project • Access The list may have elements similar to 1.1.3.1.1 “Reason for Creation”, particularly for those resources “born digital”. It is important to bear in mind that similar-sounding reasons for preservation may have quite different contexts for the latter category of resource, and “digital surrogates”.

1.1.3.2 Management History:	
Ingest Process History:	The process will need to be documented as it happens; if functionality is not being preserved, should functionality loss also be recorded here?
1.1.3.2.2 Administration History:	
Action History:	This is where the archive according to policy might migrate digital objects into standard formats. This is also where links to earlier (and later) manifestations of the object would reside. This might also include a link to the original bytestream for any object that has been changed on ingest (e.g. migrated to a standard format).
Policy History:	For each action it may be relevant to link to the relevant policy. This element should probably be defined not as a "set of actions" but the *reasons* for the set of actions. Also, what *not* to archive should be clear in the policy; particularly if policy changes over time.

1.1.3.3 Rights Management:	
Negotiation History:	This section could not only cover clearance of rights for content but systems also (e.g. a search engine where rights are held by a third party. This element needs to say its repeatable because there may be a number of separate negotiations associated with a digital object. How is rights information for time-limited evaluation copies of software dealt with? Need to clarify if this covers negotiations between: • authors and publishers • publishers and archive • both Also need to document where publisher does not want to negotiate. Where correspondence has taken place, (physical) letters need to be retained, although there was also a discussion about expressing which stages a process is at in terms of a metadata formalism. Definite agreement that physical letters need to be retained to verify such correspondence took place.
1.1.3.3.2 Rights Information:	
Date of Publication:	Which date to choose - the copyright date for the printed text (where applicable) or the date of creation of the digital version? Date range should be allowed, for certain resources; e.g. projects which had outputs over the project lifespan.
Place of Publication:	Different jurisdictions will have different rights laws. URL doesn't necessarily indicate physical place of publication (e.g. .com is not necessarily USA) Jurisdiction: where servers are located, where legally constituted publication body is located, both?
Rights Warning:	Some projects (e.g. ILEJ) have already had to implement some kind of rights warning within their service; checking for this upon ingest would be needed, so that it could be checked against an archive's standard practice.
Contacts or Rights Holders:	This could be administered by an external body: this has implications on transferral of data between archive and body.
Actors:	It was suggested that this element be renamed Action Agents which makes the link more clear between the Actors and Actions. Apart from Archive Administrators and Users (which would be the norm for all resources under consideration in Cedars), what others would be envisaged?
1.1.3.3.2.3 Actions	
	There was some discussion about the relationship of these last elements (Actors and Actions) Is Actions a sub-element of Actor? Would it be necessary to list the actors and then the associated actions allowed to them? How this would be implemented was questioned by a number of the group. It was also noted that the introduction to the specification suggests that actions associated with the preservation process were assumed to be allowed and these agreements would be maintained elsewhere in the archive's administration. Some licenses have expiry dates, date fields may be needed to deal with this.

1.1.4 Fixity Information:	
Authentication Indicator:	There was some discussion about the need for separate authenticity mechanisms which would form part of the Dissemination Information Package (DIP). As materials was disseminated from the archive, the copy would need to be authenticated. It was agreed that this did not form part of this outline specification which is concerned only with the AIP.

4.5 Walkthrough results for each resource

4.5.1 A Digitised Textbook in PDF

Description

The resource is called "Accounting as social and institutional practice" by Anthony G. Hopwood and Peter Miller, published by Cambridge University Press. The whole book is one PDF file. Although other textbooks on the CD were created using a separate PDF file for each chapter the relevance of which we also discussed during the course of the walkthrough.

Assumptions

This is a parallel publication.

From the title page, this is part of a print series called "Cambridge Studies in Management" (which has a title change part way along).

Although no metadata accompanies the PDF file, there will be MARC records available for the print version.

¡preservationDescriptionInformation> <referenceInformation> <resourceDescription>		
<DCtitle>	Accounting as social and institutional practice	</DCtitle>
<DCcreator>	Anthony G. Hopwood and Peter Miller	</DCcreator>
<DCdate>	1994	</DCdate>

We assigned some "core elements" and there was a great deal of general discussion about using the proposed DC elements. It was recognised that there would be significant overlap between this section and other sections of the metadata if we implemented all the DC fields. However, the group agreed that we wanted to include any resource descriptions that accompany the materials - whether MARC records, DC or EAD records. It was agreed that archives would want to keep all existing records (under 1.1.1.2).

It was suggested that an algorithm could be used to populate the relevant DC fields from within the preservation record (e.g. format information and relation fields could be generated by bringing information from the relevant Provenance and Context elements)

It was agreed that although the resource description elements need not be detailed (remember this metadata is not for resource discovery per se) there should be some "core set" of elements for this field. There are mappings between Dublin Core and MARC which the project should consider but it was likely to be Title, Responsibility Statement, Publishing Statement and Date. This would be explored further.

General Issues

Date: what is the date of this publication? Is it the copyright date which is 1994 or is it the date of digitisation (gleaned from the files themselves as 1997?) Where do we record the date of ingest? Identifier: there is no ISBN for this digitised version although ISBN's exist for both hardcopy and paperback

Who is responsible for generating this metadata? For the purposes of our exercise, it was "the archive" (i.e.. ourselves at the meeting) but in a real service environment who will take responsibility? Do we need agreement and eventually standardisation, about the metadata which must accompany digital materials if they are to be archived? It was generally agreed that the archive would need a great deal of control of the metadata even if they didn't create it all.

There was an increasing awareness of the need for different dates. This example shows us the difference between copyright date and the date something is made digital - this latter date could be an important indicator the technology used. What about the date of ingest? Likewise, for the purposes of cryogenics (in a disaster situation) the date of ingest could be very useful.

<existingMetadata>		
<existingRecord>	MARC Record (for print, as opposed to electronic version)	</existingRecord>
<contextInformation>		
<relatedInformationObject>	Issued with other textbooks on a CD (link to the textbooks)	</relatedInformationObject>
<relatedInformationObject>	part of a series - Cambridge studies in Management	</relatedInformationObject>

What do we mean by "significantly related?" There was much discussion about the role of this element in general and specifically how it related to digitised materials. In this case, the textbook is part of a series but the series is in print. Does this matter? Do we mean only digital objects? It was also need to be made clear whether the relationship was related to content or structure (i.e. physical or logical). Do we mean objects that are related physically (e.g. hyperlinks) or logically (this book is part of a series?) In the case of the other books which had a separate PDF file for each chapter this element is important. Do we refer to the fact that this textbook came with others on the CD (physical) even though the relevance logically is unclear?

It was suggested that for example, a digitised object (e.g. a manuscript) might be the subject of other resources in the archive - i.e. a book about the manuscript. The archive may choose to provide links to it because it's useful to do so for the purposes of study. However, if this metadata is for the purposes of preservation then perhaps this sort of related information object isn't relevant and we assume such a link might be made elsewhere (e.g. resource discovery tools).

It was agreed that this element needed something to define the type of relationship.

<provenanceInformation><historyOfOrigin>		
<reasonForCreation>	"For Access" was the initial entry, but the group couldn't be sure why this text was digitised It was agreed that although this element had been deemed Significant, it should actually be Less Significant because it would be less relevant to materials born digital. For some digitised materials such digitised rare manuscripts, the reason for creation might be widely know and therefore useful but people were less happy with the subjectivity which would inevitably arise for most materials. For those where the reason for creation was know, it was suggested that this might be from a list of pull down options which would include For Access, For Preservation, For commercial reasons etc.	</reasonForCreation>
<custodyHistory>	Cambridge University Press Discussion of this element inevitably raised issues about when this digital file was created. Custody History in this sense will refer only to the digital file. From the files we could tell is was created in 1997 but this wasn't apparent. Perhaps date of creation for digital files is a requirement for archiving and this is one bit of metadata archives should require for digitised materials.	</custodyHistory>
<changeHistory-BeforeArchiving>	None known If this object had been migrated from Adobe Acrobat 3.0 to 4.0, this is where we would record this information. NB: This is pre-ingest information which is different from Ingest Process History (see below) which documents what the current archive does to the file on ingest. It was also suggested that we may want to know name and date for each Change (NB the field is repeatable) to know who did it and when.	</changeHistory-BeforeArchiving>

The definition makes clear that this is organisations "responsible for the storage" of the digital object. This does not mean, every library that owned a copy.

<originalTechnicalEnvironments>		
<prerequisites>	Adobe Acrobat Reader 3.0	</prerequisites>
<procedures>	Unsure	</procedures>
<documentation>	Refer to a manual for Adobe Acrobat Reader 3.0	</documentation>
</originalTechnicalEnvironments>		
<reasonForPreservation>	Legal Deposit	</reasonForPreservation>

<managementHistory> <administrationHistory>		
<actionHistory>	Removed files from CD Rom onto hard drive. It was agreed this this is also where the archive according to policy might migrate digital objects into standard formats. In the case of the other PDF textbook where each chapter was a PDF, the archive MAY choose to amalgamate them into one PDF. This is also where links to earlier (and later) manifestations of the object would reside. This might also include a link to the original bytestream for any object that has been changed on ingest (e.g. migrated to a standard format).	</actionHistory>
<policyHistory>	For each action it may be relevant to link to the relevant policy. In this case we assigned none. To be more precise this element should probably be defined not as a "set of actions" but the reasons for the set of actions.	</policyHistory>

<rightsManagement>		
<negotiationHistory>	Assigned: Covered by Legal Deposit	</negotiationHistory>
<rightsInformation>		
<nameOfPublisher>	Cambridge University Press	</nameOfPublisher>
<dateOfPublication>	1994 or 1997 ??? General Issue: Which date do we choose? 1994 is the copyright date for the print text book but the digital files tell us that this digital version was created in 1997. Do we need both? If we choose 1997 (date of creation) then where do we record the publication date of the original source material? The creation date can tell future users something about the technical regimes of the time and is therefore useful to include somewhere, likewise the original publication date will be necessary.	</dateOfPublication>
<placeOfPublication>	Cambridge	</placeOfPublication>
<rightsWarning>	Access permitted by Legal Deposit	</rightsWarning>
<contactsRightsHolders>	Cambridge University Press (but what about rights for the photos etc in the digital copy? Do we know the publisher owns rights for those?)	</contactsRightsHolders>
<actors>	Archive Administrators, Users It was suggested that this element be renamed Action Agents which makes the link more clear between the Actors and Actions.	</actors>

<fixityInformation>		
<authenticationIndicator>	None	</authenticationIndicator>

A mechanism put in place by the archive to ensure the bytestream is authentic - so it ensures the document is as it was when it was deposited in the archive.

Other Issues

The Outline Specification will need to make clear where possible which elements it assumes will accompany the object and which will need to be generated by the archive.

4.5.2 Window on the World

General Info

This resource was shipped with the Times newspaper, on a CD-ROM. It was a file system with a complex directory structure + multiple exe files. The resource is ‘multimedia’. Derek has not tested any of the games, etc.

The question was asked: how much time did Derek spend analysing the CD to know enough for the walkthrough presentation. This was an afternoon of exploring (although it is apparent that more work may be needed to unpack some issues identified during the session).

<preservationDescriptionInformation> <referenceInformation> <resourceDescription>		
<DCtitle>	Window on the World (alt=Eyes on the Planet?)	</DCtitle>

<existingMetadata>		
<existingRecord>	Looked in http://corc.oclc.org and found no reference to this resource. Obviously someone may have a metadata record, but we do not know where it is.	</existingRecord>
<contextInformation>		
	Is there a way to deal with cover images? (i.e. on the sleeve there are pictures, and on the reverse there are some instructions etc.) An answer to this is yes! However which there is more than one way. Some choices are: • 1) scan an image file: embed in metadata, associate it with the Primary Digital Object, or save as a separate (related) AIP. • 2) OCR/retype the text: embed in metadata, or save as separate AIP.	
<relatedInformation-Object>	The newspaper it was distributed with? - good idea, or perhaps just the write up?	</relatedInformation-Object>
<relatedInformation-Object>	Cardboard sleeve (image or text) AIP	</relatedInformation-Object>

<provenanceInformation><historyOfOrigin>		
<reasonForCreation>	Promotional or Commercial publisher raise awareness of field of space data	</reasonForCreation>
<custodyHistory>	multimedia company Ellis Weinberger got from CD-ROM producer	</custodyHistory>
<changeHistory-BeforeArchiving>	Nancy Elkington: cited examples of medical journals (print) where adverts are removed before rebinding - saves volume thickness	</changeHistory-BeforeArchiving>

<originalTechnicalEnvironments>		
<prerequisites>	multimedia enable platform - see list on back of CD sleeve. CD-ROM for PC/MAC/ACORN/UNIX. Resource functionality varies across platform. Some platforms had evaluation version of software - only lasts X days.	</prerequisites>

<managementHistory>		
<ingestProcess-History>	created metadata - datestamp converted filesystem to bytestream - datestamp placed in long-term storage - where, BFID, datestamp	</ingestProcess-History>

There was a discussion about viruses. David Price asked whether a virus check would be incurred at the start of the ingest process. (Bodleian have had trouble!) Kelly Russell: virus checking could be a pre-requisite. (Not mandatory - may want to preserve them as well ... sic.)

<rightsManagement>		
	Kelly Russell raised the question of : Derek pointed out part of the CD-ROM is 21 day evaluation (PC platform) - how (in current framework) do we capture this kind of rights stuff?	
<negotiationHistory>	document interactions between Catherine Seville / Ellis and the publishers. Verbal! (still needs recording.)	</negotiationHistory>
<rightsInformation>		
<nameOfPublisher>	Lots	</nameOfPublisher>
<dateOfPublication>	Date of Times newspaper	</dateOfPublication>
<placeOfPublication>	BNSC (? this is an organisation) (is this mainly to do with the country?)	</placeOfPublication>
<rightsWarning>	(could be quite complex) Maybe a tiered approach to Rights Holders is good, i.e. Primary, Secondary, Software, etc.	</rightsWarning>
<contactsRights-Holders>	Lots embedded. Very Many! Multiple rights Holders holders. ?Is this practical?	</contactsRights-Holders>

<representationInformation> <structureInformation>		
<UAFDescription>	File tree (mainly HTML) HTML, WAV, MOV, JPG.	</UAFDescription>
<transformerObject>	Untar/Zip the bytestream to produce a file tree	</transformerObject>
<semanticInformation		
<renderAnalyse-Object>	a browser	
<parameters>	Start at first page (launch.htm)	</parameters>

4.5.3 New Journal of Physics

General Info

New Journal of Physics (NJP) is a journal publishing original research in all areas of physics. This resource is particularly interesting as it is only available on the web; i.e. is a purely digital journal.

<preservationDescriptionInformation> <referenceInformation> <resourceDescription>		
<DCtitle>	New Journal of Physics	</DCtitle>
<DCcreator>	Institute of Physics, Deutsche Physikalische Gesellscaft	</DCcreator>
<DCcontributor>	Multiple	</DCcontributor>
<DCdate>	2000	</DCdate>

<existingMetadata>		
<existingRecord>	COPAC Record: TI- New journal of physics AU- Institute of Physics AU- Deutsche Physikalische Gesellschaft PU- Bristol, UK ; [Bad Honnef, Germany] : Institute of Physics : Deutsche Physikalische Gesellschaft PY- 1998 DT- Periodical LA- English IS- 13672630 NT- Also known as: NJP Description based on online display, Vol. 1, no.1 (Dec., 1998): title from title screen as viewed on Jan. 26, 1999 Irregular NT- “New Journal of Physics (NJP) is a peer-reviewed, all-electronic journal publishing original research in all areas of physics.” KW- Physics, Periodicals HL- Oxford	</existingRecord>
<existingRecord>	WorldCat Record: Title: New journal of physics Publication: Bristol, UK :; Institute of Physics Publishing, Year: 1998 9999 Description: Vol. 1 (1998-1999)- Began publication with v. 1, [article] 1 in Dec. 1998. Language: English Standard No: ISSN: 1367-2630; CODEN: NJOPFM References: Chemical abstracts; 0009-2258 Access: http://njp.org/index.html SUBJECT(S) Descriptor: Physics – Periodicals. System Info: Mode of access: Internet via World Wide Web. File Info: Electronic journal available in HTML and Acrobat, with links to text and non-text files Note(s): Title from journal homepage, viewed Jan. 12, 1999./ Issued by: Institute of Physics; and by: Deutsche Physikalische Gesellschaft. General Info: Articles are published on the Web as soon as they are accepted rather than when an issue is completed. Class Descrpt: LC: QC1 Other Titles: Also known as:; NJP More Corp Auth: Institute of Physics (Great Britain) ; Deutsche Physikalische Gesellschaft (1963-) Document Type: Serial; Computer File Accession No: OCLC: 40606004	</existingRecord>

<contextInformation>		
<relatedInformation-Object>	How does one describe the granularity? How is the web-site as a whole recorded? Is the context the delivery environment?	</relatedInformation-Object>
<relatedInformation-Object>	How much of the service needs to be captured? Is the whole journal the related information object?	</relatedInformation-Object>

<provenanceInformation><historyOfOrigin>		
<reasonForCreation>	Trying out new cost model. Publication.	</reasonForCreation>
<custodyHistory>	IOPP still have it	</custodyHistory>

<originalTechnicalEnvironments>		
<prerequisites>	A server environment if we wanted to reproduce look and feel.	</prerequisites>
<procedures>	Some CGI scripts (?) in original.	</procedures>
</originalTechnicalEnvironments>		
<reasonForPreservation>	Intellectual content	</reasonForPreservation>

<managementHistory>		
<ingestProcessHistory>	The process needs to be documented as it happens. Will it be captured from the web or received as a package through FTP. Metadata will have to be generated, and an Archival Information Package produced. Where will information about functionality be stored, if some of the functionality is not preserved?	</ingestProcessHistory>
<administrationHistory>		
<policyHistory>	What not to archive will have to be clear in the policy.	</policyHistory>

<rightsManagement>		
<negotiationHistory>	Does this have to document negotiations between authors and publishers, or between publishers and the archive, or both? Document where the publisher does not want to negotiate.	</negotiationHistory>
<rightsInformation>		
<nameOfPublisher>	IOP	</nameOfPublisher>
<dateOfPublication>	2000	</dateOfPublication>
<placeOfPublication>	Different jurisdictions will have different rights. Should there be a URL or a physical place of publication? Should it be where all the servers are, or where the legally con-stituted publication body is officially located, or both?	</placeOfPublication>
<contactsRightsHolders>	Should these be maintained by an external body?	</contactsRightsHolders>
<actions>		
<licenceTextPointer>	What about where the current licence is due to expire? Date fields are needed.	</licenceTextPointer>

<representationInformation> <structureInformation>		
<UAFDescription>	File tree. File system. Should there be a "look and feel" example. What about external links.	</UAFDescription>
<transformerObject>	zip or tar	</transformerObject>

4.5.4 Internet Library of Early Journals

Description

The Internet Library of Early Journals (known as ILEJ) was a joint project between the Universities of Birmingham, Leeds, Manchester and Oxford, funded by the eLib (Electronic Libraries) Programme.

It digitised substantial runs of 18th and 19th century journals, and made these images available on the Internet, together with their associated bibliographic data.

The core collection for the project are runs of at least 20 consecutive years of:

- Three 18th-century journals

  - Gentleman's Magazine
  - The Annual Register
  - Philosophical Transactions of the Royal Society

- Three 19th-century journals

  - Notes and Queries
  - The Builder
  - Blackwood's Edinburgh Magazine

Assumptions

It was assumed that ILEJ was to be preserved as a *service* - this would involve preservation of the following:

- TIFF files
- JPEG and GIF files (which are derived from the TIFFs)
- Perl Scripts
- OpenText search engine
- Web Server
- Excalibur EFS (fuzzy search engine - has its own file systems)
- SGML files (text, including OCR text)

<preservationDescriptionInformation> <referenceInformation> <resourceDescription>		
<DCtitle>	Internet Library of Early Journals (ILEJ)	</DCtitle>
<DCcreator>	ILEJ Consortium	</DCcreator>

It was suggested that the resource description information could be extracted from the EAD/TEI headers.

<existingMetadata>		
<existingRecord>	EAD files	</existingRecord>
<existingRecord>	TEI files	</existingRecord>
	Considering ILEJ as a “service”, there are a number of granularity issues which need to be considered: since the ILEJ service contains a number of distinct works which are linked for the purposes of the service, but distinct with respect to resource discovery, it may be easier to generate resource description for the archived service, and to provide links between the existing (and more detailed) metadata which describes discrete resources - although it is essential to note this cannot be at a greater level of detail than which was provided at the time of the work. Each TEI file corresponds to a physical volume.	
<contextInformation>		
<relatedInformation-Object>	The most obvious related information object is the ILEJ report, which gives a full description of what took place within the project. In addition to providing context on what the service provided, and the scope of the project, it contains a significant amount of detail on how the service worked, which may be useful for those involved in reinstating the service at a future stage.	</relatedInformation-Object>
<relatedInformation-Object>	This was referred to as “documentation on the website” by others in the group; it is worth mentioning that the extent of documentation of a service on a website may differ greatly from any documentation held by a service provider, which may not be available on the web for security issues. Upon ingest, it would be very useful for a service to provide internal documentation as related information object(s) in addition to any reports which were produced.	</relatedInformation-Object>

<provenanceInformation><historyOfOrigin>		
<reasonForCreation>	The reasons for creating ILEJ are well documented in the report: • "explore digitisation..." was one entry. • "explore digitisation process" was another. • "digitisation project" • "from web pages, digitizing and access" • refer to "what is ILEJ?"	</reasonForCreation>
<custodyHistory>	The ILEJ consortium - although this comprises Oxford, Leeds and Manchester universities, and while this can also be derived from the report, the issue of what is the most useful entry to put is a significant one. list of entries made • ILEJ consortium: 4 entries • Leeds/Ox/Bham: 1 entry • both: 3 (one put Leeds/Ox/Manchester)	</custodyHistory>
<changeHistory-BeforeArchiving>	In the case of ILEJ, it was unclear whether this section should be used to document what measures had been taken between the end of the funded project duration, and a further issue relating to whose responsibility such changes would be in a collaborative project after the end of the funding period. It was noted that this is a point which should be raised to such projects (at both project consortium and funding body levels), as this is kind of data would be useful but is not documented at present, nor any best practice established.	</changeHistory-BeforeArchiving>

<originalTechnicalEnvironments>		
<prerequisites>	• web browser and fuzzy search capable • Oxford and Leeds servers, etc.	</prerequisites>
<documentation>	Documentation for the original configuration of ILEJ is contained in the project report; we could supply a link to this text file. ILEJ, as a service, required the following technical environments: • TIFF files • JPEG and GIF files (which are derived from the TIFFs) • Perl Scripts • OpenText search engine • Web Server • Excalibur EFS (fuzzy search engine - has its own file systems) • SGML files (text, including OCR text) However, not all services were available for all resources contained within ILEJ; this should be borne in mind when deciding what to preserve. It was noted that certain aspects of the service relied on commercial software; it is unclear at present if such vendors would permit these to be preserved along with the content. However, if the documentation explains what each piece of software did, and what procedures this involved, this may enable the recreation of such services using contemporary tools which can use what has been preserved. Another important point is that documentation provided as a project deliverable may differ from that which was used by those responsible for developing and maintaining the resource when it was being developed, and maintained. Whilst project funding bodies may not have any need for this, it would be desirable to archive such documentation, as this is a form of tacit knowledge documented at the time of implementation, and may provide useful reading for someone who would like to try to re-implement the service after extracting ILEJ from its AIP.	</documentation>
</originalTechnicalEnvironments>		
<reasonFor-Preservation>	“Demonstrator project”, and “Access” were cited. It is important to remember that ILEJ was a demonstrator service which sought to explore and identify the issues involved in digitisation for access of old journals. It was noted by the group that if this section was to be completed from a set of options to choose from, multiple options need to be able to be selected, and the metadata framework should be able to support this.	</reasonFor-Preservation>

<managementHistory>		
<ingestProcess-History>	compilation of distributed parts: (2 entries) compilation in 50Gb file. add metadata, package, store (1 entry, more detailed)	</ingestProcess-History>
<administrationHistory>		
<policyHistory>	At this stage, neither of these can be addressed, as there cannot be a history for a non-existent archive.	</policyHistory>

<rightsManagement>		
<negotiationHistory>	The Cedars people at Cambridge have this information, after consultation with Richard Gartner, who was working on ILEJ. There was a number of letters which had been written in relation to this; there was a discussion on whether these letters needed to be retained, or if the results could be captured and expressed in a metadata formalism. It was noted, though, that the (physical) letters ought to be kept to verify such correspondence actually took place.	</negotiationHistory>
<rightsInformation>		
<nameOfPublisher>	ILEJ consortium (7 entries in total)	</nameOfPublisher>
<dateOfPublication>	7 people felt a date range was required here; 1996-98 would be appropriate.	</dateOfPublication>
<placeOfPublication>	A number of alternatives were suggested: • “ many” • IOP, Oxford as main contact • the entry page was at the Oxford URL (4 people suggested this), which implies Oxford - although some Bell and Howell components were used, which may necessitate their inclusion • Oxford and Leeds (the two main partners - 2 people suggested this)	</placeOfPublication>
<rightsWarning>	There is a UK rights warning to users supplied within the service.	</rightsWarning>
<contactsRights-Holders>	There are multiple rights to be considered; those considered at the time of the project (which may not cover some of the issues raised above) are cited on the website; i.e. they are embedded within the service.	</contactsRights-Holders>
<actors>	Archive Administrators and Users were cited - although it was noted that this would be the norm for all resources under consideration in Cedars.	</actors>
<actions>		
<licenceTextPointer>	What about where the current licence is due to expire? Date fields are needed.	</licenceTextPointer>

<representationInformation> <structureInformation>		
<UAFDescription>	file system/file tree (4 entries)	</UAFDescription>
<renderAnalyse-Engine>	TAR (2 entries)	</renderAnalyse-Engine>
<renderAnalyse-ConvertObject>	browser/server (perl-enabled); also OpenText (UNIX/NT)	</renderAnalyse-ConvertObject>
<semanticInformation> <renderAnalyseObject>		
<platform>	browser, server, search engine (3 entries)	</platform>

REFERENCES

1. Lou Reich and Dan Sawyer, (eds.), Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems, White Book, Issue 5 (CCSDS 650.0-W-5.0). Washington, D.C.: CCSDS Secretariat, National Aeronautics and Space Administration, April 1999. URL: http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

2. Cedars Project Team and UKOLN, 2000, Metadata For Digital Preservation: The Cedars Outline Specification - Draft for Public Consultation (March 2000) URL: http://www.leeds.ac.uk/cedars/OutlineSpec.htm

3. Day, M.W., 1998, Metadata for Preservation. CEDARS Project Document AIW01. URL: http://www.ukoln.ac.uk/metadata/cedars/AIW01.html

4. Stone, A and Day, M.W., 1999, Cedars Preservation Metadata Elements. Cedars Project Document AIW02 URL: http://users.ox.ac.uk/ cedars/Papers/AIW02.html

Appendix Five : Metadata for Digital Preservation

The Cedars Project Outline Specification
Draft for Public Consultation
The Cedars Project Team and UKOLN
March 2000

1 Introduction

A major component of the work encompassed by the Cedars project is the development of a metadata framework which will enable the long-term preservation of digital resources. This metadata is required to support meaningful access to the archived digital content and includes descriptive, administrative, technical and legal information.

This document describes the metadata elements which the Cedars project has identified as being useful to ensure that digital library resources can be archived and used in the future. No assumptions are made about particular methods of implementing this specification, although it is envisaged that the metadata will be stored digitally.

1.1 Purpose of this Document

The purpose of this document is twofold. The primary purpose is to provide the Cedars Demonstrator Project with a basic set of preservation metadata elements to implement as part of a pilot digital archive which will be based on the Cedars Demonstrator Project system and architecture.[1] The Cedars Demonstrator Project will contain a wide variety of digital materials (representative of digital library collections) and it is therefore necessary that this specification be generally applicable.[2] It is not expected that the project will implement this outline specification fully as there will be sections of the metadata where discussion is still necessary before implementation will be possible. Where this is the case the project will provide free-text input fields to allow comments and questions which, it is hoped, will inform further discussion. As part of the demonstrator project the Cedars team plans to develop an XML DTD to express the metadata elements, but it has made no decisions about implementing a specific syntax. This document and its implementation as an XML DTD will continue as an iterative process over the next few months. Complete implementation of the Cedars Demonstrator Project system and a final version of the Outline Metadata Specification are both due for completion in summer 2000.

The secondary reason for developing this outline specification is to contribute on a strategic level to the international collaborative development of a standard specification for digital preservation metadata. Such work is inevitably beyond the scope of a single project and it is only through collaboration with a number of stakeholder communities that this will be achieved. When the Cedars project began this work the level of international interest in its work was somewhat underestimated, but the document as it now stands has benefited enormously from international debate and discussion. The project hopes to continue contributing to work in this area.

1.2 Relationship to other Metadata Initiatives

In 1998, the Cedars project produced a preliminary review of metadata initiatives that were identified as being relevant to digital preservation (Day 1998). This review noted that metadata is needed for digital preservation, regardless of the particular preservation strategy chosen. Clifford Lynch (1999) has described the function of some of this metadata:

[1] Further information about the Cedars Demonstrator Project is available on the Cedars website http://www.leeds.ac.uk/cedars

[2] Both this specification and the demonstrator archive system are concerned with a variety of digital objects including digital materials with non-digital equivalents as well as materials that are "born digital". It is important to recognise that although digitisation projects are sometimes undertaken as part of a strategy to preserve rare or fragile materials, digitisation itself is not part of the digital preservation process as defined by the Cedars project in general or this specification in particular.

> Within an archive, metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information. This metadata will also be maintained and will be migrated from format to format and standard to standard, independently of the base object it describes.

Preservation metadata has, therefore, become an important subject for research and development in the archive and library communities. Examples of such initiatives are:

- The Research Libraries Group (RLG) Working Group on Preservation Issues of Metadata, whose final report (RLG 1998) defined the semantics of metadata elements that could serve the preservation requirements of digital images.
- The metadata specification for evidence developed as part of the University of Pittsburgh Recordkeeping Functional Requirements Project, funded by the US National Historic Publications and Records Commission (Bearman and Sochats 1996).
- The Recordkeeping Metadata Standard for Commonwealth Agencies developed by the National Archives of Australia (1999).
- The logical data model (based on entity-relationship modelling) developed by the National Library of Australia (NLA) to help identify the particular entities (and their associated metadata) that needed to be supported within its PANDORA proof of concept archive (Preserving and Accessing Networked Documentary Resources of Australia, Cameron and Pearce 1998). This model has recently been revised for use within the NLA's Digital Services Project (NLA 1999).

1.3 Relationship to the OAIS model and to Earlier Cedars Documents

Another significant development has been the production of the International Standards Organization (ISO) Reference Model for an Open Archival Information System (OAIS).[3] This initiative is being co-ordinated by the Consultative Committee for Space Data Systems (CCSDS). It describes a group "that has accepted the responsibility to preserve information and make it available for a designated community."[4] The OAIS defines a range of functions which are applicable to any archive, whether digital or not. These functions support the operations of the archive from receiving materials to archive (ingest), through storage, data management and administration, to the dissemination and release of the materials to those outside the archive (access). The OAIS model aims to provide a common terminology and framework with which to explore the challenges facing digital archives. It is currently undergoing the ISO standards process and it is hoped the model will be released as part of the ISO suite of agreed standards by autumn 2000.

The OAIS model has identified and distinguished various types of metadata needed to support a digitally preserved resource. In general, this Cedars outline specification has adhered to this aggregation.[5] In accordance with OAIS, each resource is packaged together with its metadata, as an 'Information Package'. An Information Package combines two things: 'Content Information' and 'Preservation Description Information' (PDI). The Content Information groups the preserved digital resource, or *data object*, with 'Representation Information' (RI) metadata; the RI is the information needed to retain meaningful access to the preserved data object. The PDI groups different kinds of descriptive metadata, so that what the Content Information actually is can still be understood indefinitely.[6] Several digital library projects (in addition to Cedars) are currently interested in the OAIS model. These include the Networked European Deposit Library (NEDLIB), the British Library, the RLG, AHDS, and UKOLN.[7]

---

[3] Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-R-1, Red Book, May 1999.

[4] Ibid p 1-11.

[5] There are instances where Cedars has deviated from the OAIS model (mainly in areas where the model is open to interpretation). Where possible, this is identified and explained.

[6] Ibid p 4-25.

[7] Werf-Davelaar, T. van der. Long-term preservation of electronic publications: the NEDLIB project. D-Lib Magazine, 5 (9), September 1999.

The development of the OAIS reference model by the CCSDS has influenced the development of the Cedars metadata scheme as well as the implementation of the Cedars Demonstrator Project. Early drafts of this document (e.g. Stone and Day 1999, Day 1999) took metadata elements from a variety of selected initiatives in the Cedars metadata review and mapped them onto the taxonomy of the information package identified in the OAIS model. A version of that document was circulated for selective consultation in December 1999.[8] This current document further refines the approach: it starts from the structure provided by the OAIS model and populates it with metadata elements chosen by practical investigation of archiving real digital resources, further refined by comments received from a selective consultation process.

1.4 Relationship of Preservation Metadata to Other Archive Functions

It is important to recognise that the proposed set of preservation metadata is not intended to include descriptions of all archival functions (there are separate areas in OAIS given to functions such as administration and management). This document is concerned with metadata to aid preservation and does not attempt to include information which would be recorded as part of the regular business procedures and processes of the archive, such as usage statistics or archive policy (although for obvious reasons links to these functional areas may be necessary or desirable). One function which Cedars considers to be part of archive administration is the management of the preservation metadata itself (e.g. monitoring and updating as necessary). For example, information about copyright and related intellectual property rights for any digital object will change over time and with changes to legislation. Likewise an OAIS model archive which relies on a robust network of technical descriptions (e.g. representation information) to ensure long-term access will need to monitor this information and update it as technology changes. This document assumes that preservation metadata will be dynamic and that maintenance of the metadata will form a key part of an archive's administrative function.

1.5 Granularity Issues

It has been widely recognised that the metadata required for long-term digital preservation is complicated by the levels of 'granularity' that can occur within a single digital object or collection of objects. Metadata for example may be assigned at the level of a complete digital collection, a single digital object or even (in the case of complex digital material) at the individual file level. In part the granularity of the metadata will be determined by the digital object itself and the level of description necessary to ensure preservation, but it will also be influenced by collection management policies in place at the archive.[9] In addition the granularity of the metadata may be influenced by concerns about rights management of some more complex digital objects (e.g. where different parties own different components of the content and/or systems).[10] **How an archive chooses to assign metadata, and at what level of granularity, are not decisions imposed by a metadata specification.** A preservation metadata specification should allow for description at any level (as this outline attempts to do) but ultimately the decision resides with the archive. For example, both the British Library and the NEDLIB projects (where work is focused on the deposit library situation) have chosen, for justifiable practical reasons, to assign metadata to materials as they have been delivered to the library (e.g. as produced by the publisher). **This outline specification makes no assumptions about the level at which metadata will be assigned and assumes only that archives will do so at levels appropriate to fulfil their preservation responsibilities and meet the needs of the archive's user communities.**

1.6 What is a Data Object?

As described above, the OAIS reference model makes clear that an Archival Information Package contains two distinct areas: the Content Information which includes the data object or digital resource itself as well as the systems and

[8] A list of organisations selected as reviewers is provided at Annex A.

[9] The Cedars project explains some of the technical decisions associated with preservation of a digital object using a concept called the Underlying Abstract Form which is described in section 3.2.1.1.1 of the Representation Information below.

[10] For example a digital sound recording with associated copyright and performance rights for individual tracks.

information necessary to render the object; and Preservation Description Information which includes information describing the object itself and associated preservation information. In the Cedars Demonstrator Project different data objects may have different manifestations. For example, an object that has been migrated[11] through evolving technical regimes may have several manifestations within the archive (the original manifestation as well as manifestations associated with newer technical environments). In part this will depend on the preservation strategy adopted for the object by the archive (e.g. a migration strategy will involve the production of a new manifestation with each migration while an emulation strategy may rely on access via the original data object.) The proposed metadata specification will allow for different manifestations of the same data object within the archive via reference links to previous and subsequent manifestations (see Provenance Information, section 3.1.3 below).

1.7 Obligation

In earlier iterations of this outline specification an attempt was made to define elements as either "mandatory" or "optional". However, subsequent discussions and comments from reviewers suggest that those concerned with long-term preservation in libraries or archives are unlikely to decide against preservation of a valuable digital object because it lacks specific metadata elements. In addition, specific types of digital object will have elements which might be deemed mandatory, while for other digital materials the same field is not applicable (e.g. a sound recording of a bird song may lack both title and author). The Cedars team have agreed that these obligations are not necessarily applicable to metadata for preservation and that decisions about long-term retention will not be based primarily on metadata requirements. Instead this outline specification provides guidance on the "significance" of each element determined by the genericity of the element (i.e. the extent to which it may be usefully applied across a wide range of digital materials). The terms used are as follows:

- Very Significant - an element deemed to be very useful for preservation across a wide range of digital objects and therefore necessary if available
- Significant - an element deemed to be of use to most digital objects
- Less Significant - an element deemed to be of some use but not strictly necessary.

1.8 Structure of this Document

In this document the term 'element' is used for each item of metadata. An element may be composed of sub-elements, which are also elements, and so can also contain sub-elements.

Following this introduction is a set of tree diagrams that represent the structural framework of an Information Package. The hierarchy of the trees follows that of the OAIS reference model as closely as possible, and this correspondence is shown in the accompanying text.

Following the diagrams, each metadata element is explained in detail, ordered by its position in the tree structure rather than by its importance. The tables describing each metadata element should be self-explanatory.

2 Structural Diagrams

As with all complex structures, the framework for the Cedars metadata is difficult to conceptualise rapidly. Tree diagrams have been used to enable the interrelationships and the overall arrangement of metadata elements to be

[11] Migration in this context is defined as the systematic transfer of digital materials from one software/hardware regime to another. The OAIS reference model refers to this as 'transformation.'

visualised swiftly. They also allow easy comparison between the Cedars metadata and the OAIS[12] structure.

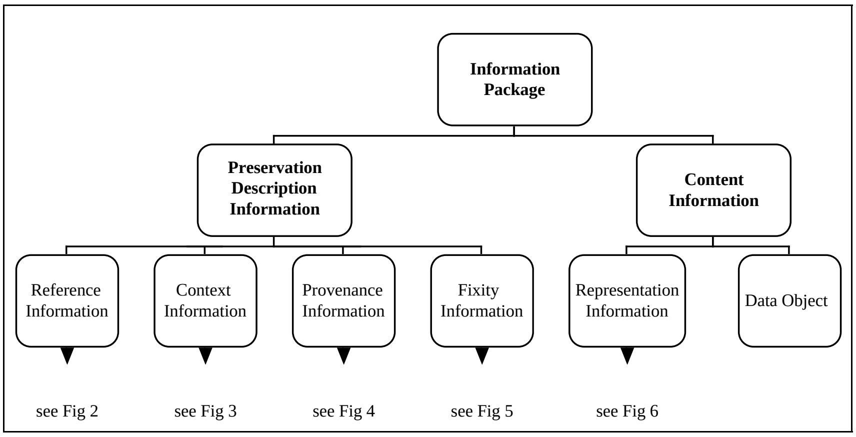


Figure 1: The structure of an Information Package

Figure 1[13] shows the highest level of the Cedars metadata structure. The highest level object in the OAIS model is an Information Package (this current document is predominantly concerned with the metadata structure of the Archival Information Package (AIP)). It will be noticed that neither Packaging Information nor Descriptive Information are included in the Cedars diagram. Packaging Information may be included at a later date, and will be a means of tracking which version of an Information Package (i.e. its structure) is used; this is likely to be a reference to the XML DTD defining the AIP's structure.

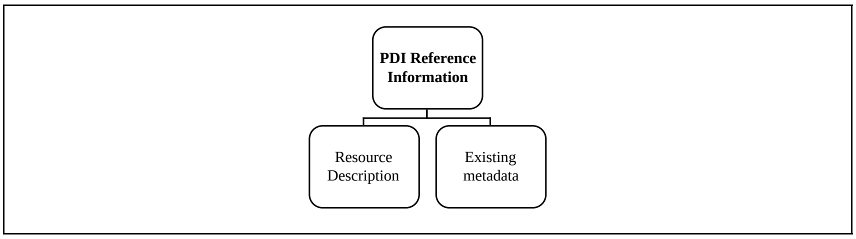


Figure 2: The structure of Reference Information (PDI)

Figure 2 shows the metadata elements which are classed as Reference Information. "This information identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides those identifiers that allow outside systems to refer, unambiguously, to this particular Content Information."[14] Broadening this slightly: in Cedars, the Reference Information identifies and describes the Content sufficiently and so holds most of the data which need to be distributed for customer resource discovery. For this reason holding places are also kept for any existing instantiations of metadata schemes (such as Dublin Core, MARC records, etc.) to allow these to be reused where appropriate.

Figure 3 shows the structure of Context Information. This information documents the relationships of the Content Information to its environment. This explains how it relates to other Content Information objects existing elsewhere.[15] This diagram is likely to be expanded as test resources begin to show which aspects of this are relevant, and when.

[12] Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-R-1, Red Book, May 1999.

[13] Ibid. Figure 1 maps onto OAIS figure 4-13, p4-22. The Preservation Description Information section of figure 1 should be compared to OAIS figure 4-17, p 4-31 and the Content Information section is an expression of the second paragraph on OAIS p 2-5 (or figure 4-9) p 4-16

[14] Ibid p 4-25.

[15] Ibid.

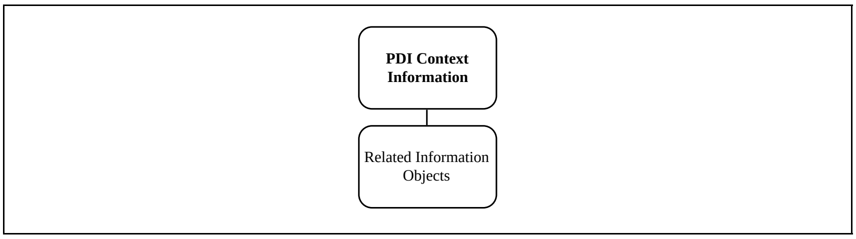


Figure 3: The structure of Context Information (PDI)

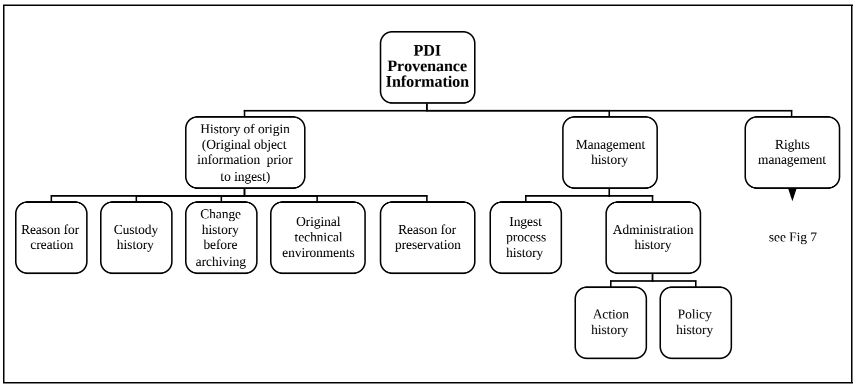


Figure 4: The structure of Provenance Information (PDI)

Figure 4 represents the Provenance Information. "This information documents the history of the Content Information."[16] Any data which describes the managerial history or administration is placed in the relevant sub-categories. There is also a category to capture the history of the resource before it entered into the archive.

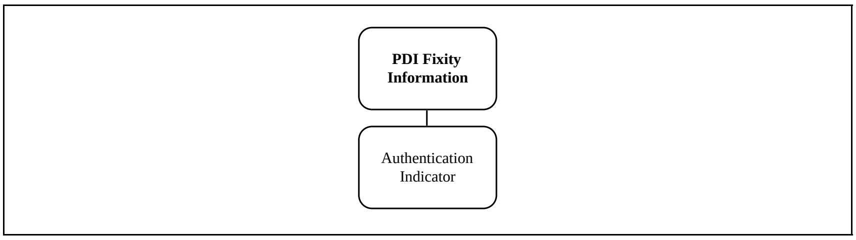


Figure 5: The structure of Fixity Information (PDI)

Figure 5 concludes this level of breakdown of the PDI. Fixity Information is described in the OAIS thus: "This information documents the authentication mechanisms, and it provides any authentication keys used to ensure that the particular Content Information object has not been altered in an undocumented manner".[17] In the first instance Cedars is not intending to solve the authentication issue, but recognises that there needs to be scope for this in the metadata specification. All the Cedars Information Packages will be protected from undocumented changes by restricting access to both read and write operations.

Figure 6 shows the structure of the Representation Information.[18] The OAIS reference model says that the RI is "needed to make the Data Object understandable".[19] The structure and semantic nodes of this tree are likely to be indirections into a Representation Network, which will be described in a document separate to this metadata specification.

[16] Ibid.
[17] Ibid p 4-26.
[18] Ibid, p4-18, Figure 4-10.
[19] Ibid, p 2-5.

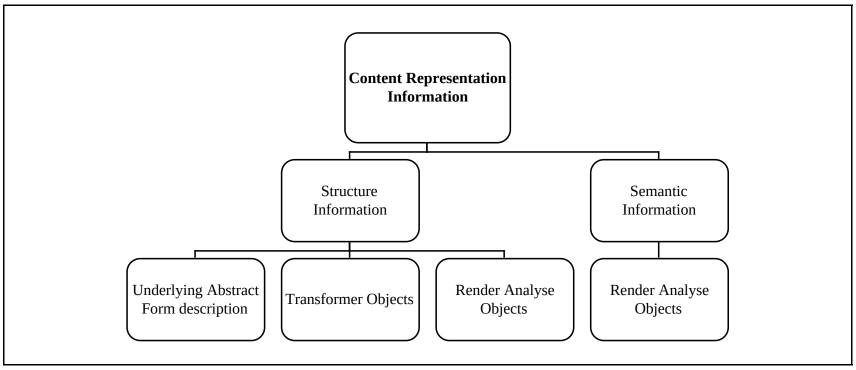


Figure 6: The structure of Representation Information (Content)

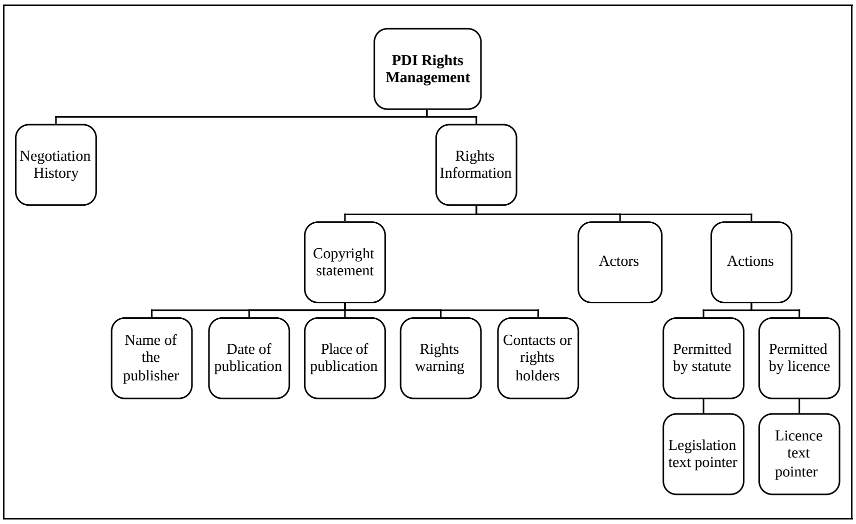


Figure 7: The structure of Rights Management (PDI - Provenance)

Figure 7 gives the structure for handling Intellectual Property Rights. These are seen as part of Provenance Information. Work in this area represents a significant expansion of the OAIS model, and is likely to encourage copyright owners of digital content who are unsure whether or not to archive their resources.

3 Definition of Metadata Elements (for an Information Package)

This section details the metadata elements for an "Archival Information Package". Ongoing revision of these is expected as instantiating the scheme for differing digital resources will reveal its strengths and weaknesses. The final version of this outline specification will include appendices providing case study examples of how this metadata has been applied to specific resources within the Cedars Demonstrator Project.

3 Information Package

Name	Information Package
Identifier	informationPackage
Definition	This contains all the content information and the related descriptions and audit trails needed to preserve a digital resource in the long term.
Sub-Elements	3.1 Preservation Description Information 3.2 Content Information

3.1 Preservation Description Information

Name	Preservation Description Information
Identifier	preservationDescriptionInformation
Definition	This element is defined as "Information which is necessary for adequate preservation of the content information and which can be categorized as Provenance, Reference, Fixity, and Context information."[20]
Obligation	Very Significant
Sub-Elements	3.1.1 Reference Information 3.1.2 Context Information 3.1.3 Provenance Information 3.1.4 Fixity Information

3.1.1 Reference Information

Name	Reference Information
Identifier	referenceInformation
Definition	Contains identifiers for the resource and additional information describing the resource. Also includes existing catalogue metadata schemes.
Obligation	Very Significant
Sub-Elements	3.1.1.1 Resource Description 3.1.1.2 Existing Metadata
Comment	Includes information for resource discovery which may be taken from existing metadata created specifically for this purpose.

[20] Ibid.

3.1.1.1 Resource Description

Name	Resource Description
Identifier	resourceDescription
Definition	The identifiers and description of the resource to uniquely identify it and determine the author and version.
Obligation	Very Significant
Comment	Resource discovery, although vital to a working archive, is not the primary focus of the Cedars project. A great deal of work has already been done in this area on standardisation. In the Cedars implementation of this outline specification an instantiation of the Dublin Core will be used to represent this information. (However this does not necessarily indicate a commitment on the part of the Cedars project to Dublin Core as the best solution for the long term.) Similarly in implementing this specification others will need to make decisions about using existing schema for resource discovery - Cedars recommends making use of existing work where possible! Although Dublin Core designates the obligation of all its elements as 'optional', long-term digital preservation relies on the description stored here. The Cedars implementation of this outline specification will therefore include at least DC:Title.

3.1.1.2 Existing Metadata

Name	Existing Metadata
Identifier	existingMetadata
Definition	Any metadata record which has been generated for the resource. (e.g. Marc records, Dublin Core). This information may accompany the resource on ingest or may be discovered later.
Obligation	Significant
Sub-Element	3.1.1.2.1 Existing Records
Comments	There are different options for integrating existing metadata schema into the AIP such as the Warwick Framework.

3.1.1.2.1 Existing Records

Name	Existing Records
Identifier	existingRecords
Definition	Each instantiation of a metadata scheme.
Obligation	Significant
Repeatable	Yes

3.1.2 Context Information

Name	Context Information
Identifier	contextInformation
Definition	Information that documents the relationships of the Content Information to its environment. This includes how the Content Information relates to other Content Information objects existing elsewhere.
Obligation	Very Significant
Sub-Element	3.1.2.1 Related Information Objects
Comment	The OAIS model is unclear about the distinction between Context Information and Provenance Information. Although OAIS suggests both Reason for Creation and Reason for Preservation should be included in Context Information, the Cedars team have agreed they are more appropriate in Provenance Information (see 3.1.3 below).

3.1.2.1 Related Information Objects

Name	Related Information Objects
Identifier	relatedInformationObjects
Definition	This element specifies any other information objects which were judged, at the time of ingest, to be significantly related to the ingested digital object.
Obligation	Very Significant
Repeatable	Yes
Comment	The related object may include, for example, items produced by the same research entity on the same or on a similar subject, or items which assist in the use of the ingested object, for example, documentation produced by third parties.

3.1.3 Provenance Information

Name	Provenance Information
Identifier	provenanceInformation
Definition	This element is defined as "Information that documents the history of the Content Information. This information tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated."[21]For Cedars this element also contains information about the reason a resource was created and why it was preserved.
Obligation	Very Significant
Sub-Elements	3.1.3.1 History of Origin (Original object information prior to ingest) 3.1.3.2 Management History 3.1.3.3 Rights Management
Comment	This element may include reference links to earlier (or later) manifestations of the digital object.

[21] Ibid.

3.1.3.1 History of Origin

Name	History of Origin
Identifier	historyOfOrigin
Definition	This element contains a description of the original digital object prior to ingest. In addition, where the production of the object has involved digitising, the production process can also be described here.
Obligation	Very Significant
Sub-Elements	3.1.3.1.1 Reason for Creation 3.1.3.1.2 Custody History 3.1.3.1.3 Change History Before Archiving 3.1.3.1.4 Original Technical Environments 3.1.3.1.5 Reason for Preservation
Comment	The description should include coverage of the following: The intended usage of the original digital resource developed; Original packaging contents; Digitizing parameters used (in the case of digitisation projects).

3.1.3.1.1 Reason for Creation

Name	Reason for Creation
Identifier	reasonForCreation
Definition	This element contains information about why a resource was created. For example for digitised material the reason the object (or collection) was digitised (e.g. fragile or rare originals).
Obligation	Significant
Repeatable	Yes

3.1.3.1.2 Custody History

Name	Custody History
Identifier	custodyHistory
Definition	This element contains the identity of individuals or organisations responsible for the storage of the digital object from the date of its creation until the digital archive became responsible for the storage of the digital object, and records when they were responsible.
Obligation	Very Significant
Repeatable	Yes

3.1.3.1.3 Change History Before Archiving

Name	Change History Before Archiving
Identifier	changeHistoryBeforeArchiving
Definition	This element describes any changes, which anyone responsible for the storage of the digital object made, from the time of creation of the digital object until the digital object became the responsibility of the digital archive. For digital surrogate objects this may include information about the non-digital source material.
Obligation	Significant
Repeatable	Yes

3.1.3.1.4 Original Technical Environments

Name	Original Technical Environments
Identifier	originalTechnicalEnvironments
Definition	This element contains information about the operating environment of the **original digital object at the time of ingest**, including information on relevant hardware and operating systems, together with the software products that would have been required in order to use it.
Obligation	Significant
Repeatable	Yes
Sub-Elements	3.1.3.1.4.1 Prerequisites 3.1.3.1.4.2 Procedures 3.1.3.1.4.3 Documentation
Comment	This element contains general descriptive information about the systems used with the original digital object and should not be confused with Representation Information used to render the primary digital object (see 3.2.1).

3.1.3.1.4.1 Prerequisites

Name	Prerequisites
Identifier	prerequisites
Definition	The hardware, operating system, or software originally available to be used with the original digital resource.
Obligation	Significant
Repeatable	Yes
Comment	Guidelines to good practice on this should be identified.

3.1.3.1.4.2 Procedures

Name	Procedures
Identifier	procedures
Definition	Additional notes on running or installation for the hardware, operating system, or software originally available to be used with the original digital resource.
Obligation	Significant
Repeatable	Yes
Comment	Guidelines to good practice on this should be identified.

3.1.3.1.4.3 Documentation

Name	Documentation
Identifier	documentation
Definition	Associated documentation for the the hardware, operating system, or software originally available to be used with the original digital resource.
Obligation	Significant
Repeatable	Yes
Comment	For example, a citation of the hardware manual for the ORIC-1 computer. Guidelines to good practice on this should be identified.

3.1.3.1.5 Reason for Preservation

Name	Reason for Preservation
Identifier	reasonForPreservation
Definition	This element describes the reasons why the digital was preserved and deposited in the archive.
Obligation	Significant
Comment	The reasons may include, for example, the cost of production of the original object, or the amount of interest in the original object.

3.1.3.2 Management History

Name	Management History
Identifier	managementHistory
Definition	This element describes all changes which were made to the digital object from the time responsibility for its storage was accepted by the digital archive.
Obligation	Very Significant
Sub-Elements	3.1.3.2.1 Ingest Process History 3.1.3.2.2 Administration History

3.1.3.2.1 Ingest Process History

Name	Ingest Process History
Identifier	ingestProcessHistory
Definition	This element describes all changes which were made to the digital object to prepare it for storage in the digital archive.
Obligation	Very Significant

3.1.3.2.2 Administration History

Name	Administration History
Identifier	administrationHistory
Definition	This element describes what happened to the digital object after the completion of ingest.
Obligation	Very Significant
Sub-Elements	3.1.3.2.2.1 Action History 3.1.3.2.2.2 Policy History

3.1.3.2.2.1 Action History

Name	Action History
Identifier	actionHistory
Definition	This element describes what was done to change the digital object after ingest to ensure preservation.
Obligation	Very Significant
Repeatable	Yes
Comment	Changing the format of the digital object may be necessary in order to preserve it. This element may include reference links to earlier/later manifestations of the object.

3.1.3.2.2.2 Policy History

Name	Policy History
Identifier	policyHistory
Definition	This element describes the set of actions which were applied to the digital object to ensure preservation.
Obligation	Very Significant
Repeatable	Yes
Comment	The regime may change if the policy of the digital archive changes. Implementation for Cedars may be a list of pre-set actions chosen from a list but could be a free text string.

3.1.3.3 Rights Management

The aim of preservation in the Cedars context is to provide access to material for scholars of the future. Any use of the material outside scholarship will require separate rights negotiations. The metadata provided here might prove useful for any such future negotiations. As part of the archive's administrative functions, these elements must be kept up to date.

Name	Rights Management
Identifier	rightsManagement
Definition	This metadata section contains information relating to the intellectual property rights relevant to the digital object.
Obligation	Very Significant
Sub-Elements	3.1.3.3.1 Negotiation History 3.1.3.3.2 Rights Information
Comment	While some of these sub-elements have been labelled 'very significant', many have been designated as only 'significant'. This is because rights information for digital preservation will be a complex area and a digital archive will almost certainly have as wide a range of rights situations as it does digital objects - relatively few of these elements will be generic and apply to all digital objects.

3.1.3.3.1 Negotiation History

Name	Negotiation History
Identifier	negotiationHistory
Definition	This element contains the details of the rights negotiations leading to submission of the digital object for preservation.
Obligation	Very Significant

3.1.3.3.2 Rights Information

Name	Rights Information
Identifier	rightsInformation
Definition	This metadata section contains information relating to the intellectual property rights relevant to the digital object.
Obligation	Very Significant
Sub-Elements	3.1.3.3.2.1 Copyright Statement 3.1.3.3.2.2 Actors 3.1.3.3.2.3 Actions
Comments	This metadata will be used by a librarian or an archivist to decide what a library or archive user can do with a digital object.

3.1.3.3.2.1 Copyright Statement

Name	Copyright Statement
Identifier	copyrightStatement
Definition	This element contains sub-elements to aid in the identification of the intellectual property rights holder or holders.
Obligation	Very Significant
Repeatable	Yes
Sub-Elements	3.1.3.3.2.1.1 Name of Publisher 3.1.3.3.2.1.2 Date of Publication 3.1.3.3.2.1.3 Place of Publication 3.1.3.3.2.1.4 Rights Warning 3.1.3.3.2.1.5 Contacts or Rights Holders

3.1.3.3.2.1.1 Name of Publisher

Name	Name of Publisher
Identifier	nameOfPublisher
Definition	This sub-element contains the name of the publisher of the digital object.
Obligation	Very Significant
Repeatable	Yes
Comment	This element could contain a unique reference identifier for publishers, an initiative which is currently under development.

3.1.3.3.2.1.2 Date of Publication

Name	Date of Publication
Identifier	dateOfPublication
Definition	This sub-element contains the date of publication of this version of this digital object.
Obligation	Significant

3.1.3.3.2.1.3 Place of Publication

Name	Place of Publication
Identifier	placeOfPublication
Definition	This sub-element contains the place of publication of this version of this digital object.
Obligation	Significant
Repeatable	Yes

3.1.3.3.2.1.4 Rights Warning

Name	Rights Warning
Identifier	rightsWarning
Definition	This sub-element contains a warning that the digital object may be subject to copyright or database right.
Obligation	Very Significant

3.1.3.3.2.1.5 Contacts or Rights Holders

Name	Contacts or Rights Holders
Identifier	contactsRightsHolders
Definition	This sub-element contains details of other known rights contacts or rights holders.
Obligation	Less Significant
Repeatable	Yes
Comment	The use of this sub-element may depend on the location of the user, or the intended place of use, or both, since there may be different rights holders in different countries.

3.1.3.3.2.2 Actors

Name	Actors
Identifier	actors
Definition	This element specifies the permitted users of the digital object, for example, archive staff or library users or both.
Obligation	Very Significant
Comment	Some digital objects may be archived for preservation but not for current user access. Archive staff need access to the digital object to check regularly that the preserved digital object can still be rendered.

3.1.3.3.2.3 Actions

Name	Actions
Identifier	actions
Definition	This element contains sub-elements describing permitted actions.
Obligation	Very Significant
Sub-Elements	3.1.3.3.2.3.1 Permitted by Statute 3.1.3.3.2.3.2 Permitted by Licence

3.1.3.3.2.3.1 Permitted by Statute

Name	Permitted by Statute
Identifier	permittedByStatute
Definition	This sub-element contains text reminders on standard permitted actions, for example, the reminder that copying for the purpose of research, private study, criticism, review, or the reporting of current events is permitted so long as it amounts to fair dealing.
Obligation	Significant
Sub-Element	3.1.3.3.2.3.1.1 Legislation Text Pointer
Comment	Note should be taken of *'Permitted by Licence', 3.1.3.3.2.3.2*

3.1.3.3.2.3.1.1 Legislation Text Pointer

Name	Legislation Text Pointer
Identifier	legislationTextPointer
Definition	This sub-element contains a pointer to the full text or texts of the current relevant legislation.
Obligation	Significant

3.1.3.3.2.3.2 Permitted by License

Name	Permitted by Licence
Identifier	permittedByLicence
Definition	If a licensing agreement is known to be in place, this sub-element includes the actual terms of the licence, which would normally specify permitted actors and actions.
Obligation	Less Significant
Sub-Element	3.1.3.3.2.3.2.1 Licence Text Pointer
Comment	Actions which are needed to ensure long term preservation are assumed to be permitted, since the digital objects are deposited in an archive designed to ensure long-term preservation. Collective licensing agreements may be in place and may be relevant.

3.1.3.3.2.3.2.1 Licence Text Pointer

Name	Licence Text Pointer
Identifier	licenceTextPointer
Definition	This sub-element contains a pointer to the full text or texts of the current relevant licence or licences.
Obligation	Less Significant

3.1.4 Fixity Information

Name	Fixity Information
Identifier	fixityInformation
Definition	This element will be used to prove the authenticity of an Archived Information Package, for example, by use of a checksum or a digital signature.
Obligation	Very Significant
Repeatable	Yes
Sub-Element	3.1.4.1 Authentication Indicator
Comment	Although fixity information is not stored directly with the digital object in the Content Information it is still associated with it within the AIP. The OAIS model makes clear that anything not directly involved with the technical rendering of the object should not be part of the Content Information. It is not the case that various objects will share fixity information. Each AIP will have its own fixity information (this particularly relates to objects with different manifestations as described above). In the Cedars implementation, fixity information will be an integral part of the dissemination of digital objects from the archive, in order to ensure that no object is disseminated from the archive without adequate assurance (via the fixity information) that it is authentic. It is envisaged that this will work in the same way that access to digital objects will only be provided once users have been authenticated.

3.1.4.1 Authentication Indicator

Name	Authentication Indicator
Identifier	authenticationIndicator
Definition	The mechanism used to ensure the digital object's authenticity. For example, a digital certificate.
Obligation	Very Significant
Repeatable	Yes

3.2 Content Information

Name	Content Information
Identifier	contentInformation
Definition	The primary target for preservation. Composed of a Primary Digital Object and its Representation Information.
Obligation	Very Significant
Repeatable	No
Sub-Elements	3.2.1 Representation Information 3.2.2 Primary Digital Object

3.2.1 Representation Information

Since the practical implementation of the representation information involves a network of specialist AIPs, linked by references rather than embedded within the metadata record, some of this section of the document will appear to duplicate elements, namely the three elements 3.2.1.1.2 Transformer Objects (TOs), 3.2.1.1.3 Render/Analyse/Convert (RACs) Objects, and 3.2.1.2.1 Render/Analyse Objects (RAOs). These are all place holders for software tools, but tools that do slightly different tasks. Where possible the render/analyse tools should be referenced from the Structure Information (and hence associated with the data format of the digital object) as this makes these tools available for any object which shares the same data format structure. However, in some cases a render/analyse tool will only be applicable to the digital object (and not other objects which share the same structure) and these are linked to the Semantic Information place holder. Please note that all of these tools have the same sub-elements so these are only detailed under 3.2.1.1.2 Transformer Object, with comments that reflect the differences in usage of the similar RACs and RAOs.

Representation Information contains all the information needed to obtain and render the intellectual content of the stored digital object. If the archive is unable to provide an environment for rendering a digital object at the time of ingest, it is advisable that information be stored which describes the technical environment in as much detail as is possible to allow for some understanding of the environment in the future. Such detailed descriptions of technical environments may be stored in RI (see Comment below).

Name	Representation Information
Identifier	representationInformation
Definition	This metadata section contains all the information needed to obtain and render the intellectual content of the stored digital object. It provides all the correct entry points to the representation network. The sub-elements of this section can be stored embedded in the AIP, but often they will be stored externally as part of a representation network and links to the network will be embedded.
Obligation	Very Significant
Repeatable	No
Sub-Elements	3.2.1.1 Structure Information 3.2.1.2 Semantic Information
Comment	Where an archive is unable to provide the technical environment for rendering at the time of ingest, inclusion of free-text to describe the environment or explain how to find out this information would help provide better access to the digital content in the future.

3.2.1.1 Structure Information

Name	Structure Information
Identifier	structureInformation
Definition	This element provides a mechanism for transforming the preserved digital object (stored as a byte-stream) into the structured set of digital components needed in order to access (and render) its intellectual content.
Obligation	Very Significant
Sub-Elements	3.2.1.1.1 Underlying Abstract Form Description 3.2.1.1.2 Transformer Objects 3.2.1.1.3 Render/Analyse/Convert Objects
Comment	The simplest transformation should be to reproduce the structure of the object prior to ingest (e.g. for an archived CD-ROM, the byte-stream is transformed into a file tree of the same structure as that on the original CD).

3.2.1.1.1 Underlying Abstract Form Description

The Cedars project refers to the structure that is needed in order to access the intellectual content of a digital object as the “Underlying Abstract Form” (UAF) of the object. It should be noted that a given digital object may have more than one choice for its UAF, and part of the ingest process involves identifying a suitable UAF which captures, as closely as possible, the structural aspects of the object in order to allow all the “significant properties” of the object to be retained. Identification of an appropriate UAF (with associated significant properties) will be governed by policies in place at the archive. Cedars considers such policy decisions to be the responsibility of collection or archive managers - no advice on making these decisions is given in this document.

Name	Underlying Abstract Form Description
Identifier	uafDescription
Definition	A formal or informal description of the abstract data form. A complete description should give examples of digital objects which this UAF has been provided for.
Obligation	Significant
Comment	If a table of all UAF-Descriptions is maintained, this could also be used as a tool to aid ingest decisions to select the correct UAF. As the number of digital resources archived with the same UAF increases, this element becomes more significant.

3.2.1.1.2 Transformer Objects (TOs)

The transformer objects are a specialist type of RAC (3.2.1.1.3). They work from a resource in a common form (a byte-stream) and render it into the UAF.

Name	Transformer Object
Identifier	uafTransformer
Definition	A Transformer Objects provide the software mechanism to transform the byte-stream into an instantiation of the UAF.
Obligation	Very Significant
Repeatable	Yes
Comment	It is likely that TOs will have to be platform specific.
Sub-Elements	3.2.1.1.2.1 Platform 3.2.1.1.2.2 Parameters 3.2.1.1.2.3 Render/Analyse Engines 3.2.1.1.2.4 Output Format 3.2.1.1.2.5 Input Format

3.2.1.1.2.1 Platform

Name	Platform
Identifier	platform
Definition	The computational platform which is needed so that the software will run.
Obligation	Significant
Repeatable	No
Comment	A platform may refer to a piece of hardware, or an appropriate software technology (such as a web browser). For certain resources it may be possible to represent the platform in the abstract by providing C-code which can be compiled on the hardware available at the time of access.

3.2.1.1.2.2 Parameters

Name	Parameters
Identifier	parameters
Definition	Additional requirements that need to be indicated to the rendering software engine for it run in the correct mode of operation.
Obligation	Less Significant
Repeatable	Yes
Comment	Parameters for a TO might include the 'arguments' needed by the software in order to achieve the mapping of the byte-stream into the specific type of UAF.

3.2.1.1.2.3 Render/Analyse Engines

Name	Render/Analyse Engines
Identifier	renderAnalyseEngines
Definition	The software engine required to render the digital object in the appropriate manner.
Obligation	Significant
Repeatable	No
Comment	For a TO, the software engine renders by transforming the byte-stream into the UAF.

3.2.1.1.2.4 Output Format

Name	Output Format
Identifier	outputFormat
Definition	A description of the format produced by processing the digital object with the rendering engine.
Obligation	Significant
Repeatable	No
Comment	For a TO, this element should be identical to the UAF-Description. For the RAO the output format may be a Graphical User Interface or a printout.

3.2.1.1.2.5 Input Format

Name	Input Format
Identifier	inputFormat
Definition	A description of the format of digital object that the rendering software works on.
Obligation	Significant
Repeatable	No
Comment	For a TO this should be the byte-stream format (where the archival store is updated to preserve AIPs in a form other than "byte-stream" this should be reflected in this metadata element). For a RAC the software is associated on the structured side, so the input format must be the UAF of the object. For the RAO the input format will either be the UAF or the result of a format conversion following on from the production of the UAF.

3.2.1.1.3 Render/Analyse/Convert Objects (RACs)

General rendering capabilities can be attached to the structural forms. This makes it possible for all tools which render a particular kind of underlying abstract form to be shared. Tools for rendering a resource that are specific to the resource should be attached to the Semantic Information. Rendering operations that convert from one digital data format to another should be attached here, enabling policies of format migration.

The archive could maintain a list of data formats so that new rendering capabilities could be uncovered when a conversion tool is deployed. Such lists and administrative functions are not covered in this specification.

Name	Render/Analyse/Convert Object
Identifier	racObject
Definition	An instance of this kind of object provides a software mechanism to access the intellectual content of the digital object, either by direct rendering, an analytical tool, or by converting the object into a form more easily rendered/understood.
Obligation	Significant
Repeatable	Yes
Sub-Elements	3.2.1.1.2.1 Platform 3.2.1.1.2.2 Parameters 3.2.1.1.2.3 Render/Analyse Engines 3.2.1.1.2.4 Output Format 3.2.1.1.2.5 Input Format

3.2.1.2 Semantic Information

Name	Semantic Information
Identifier	semanticInformation
Definition	This provides the mechanisms which allow the specific digital object in the AIP to be rendered.
Obligation	Significant
Repeatable	No
Sub-Element	3.2.1.2.1 Render/Analyse Objects
Comment	Often these rendering processes will be relevant only for the single digital resource, and will not have general applicability.

3.2.1.2.1 Render/Analyse Objects (RAO)

As mentioned above, these rendering objects have the same set of sub-elements as Transformer and Render/Analyse/Convert Objects. The main distinction is their relationship to the specific digital resource.

Name	Render/Analyse Object
Identifier	raObject
Definition	This element provides the different software mechanisms and describe the platforms which these run on, in order to access the intellectual content of the digital object.
Obligation	Significant
Repeatable	Yes
Sub-Elements	3.2.1.1.2.1 Platform 3.2.1.1.2.2 Parameters 3.2.1.1.2.3 Render/Analyse Engines 3.2.1.1.2.4 Output Format 3.2.1.1.2.5 Input Format

3.2.2 Primary Digital Object

Name	Primary Digital Object
Identifier	primaryDigitalObject
Definition	This is the preserved byte-stream of the original digital resource. In some cases this element will be populated with a reference to the actual container object (using the CRID indirection naming scheme).
Obligation	Very Significant
Repeatable	No
Comment	This is not actually a metadata element, as it is the actual digital resource preserved. However, there must be some association between the metadata in the AIP and the digital resource in the AIP.

4 Bibliography

1. Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), CCSDS 650m Red Book, May 1999.

2. Werf-Davelaar, T van der. “Long-term Preservation of Electronic Publications: the NEDLIB project.” Dlib Magazine 5 (9) September 1999.

3. Holdsworth, D and Sergeant D. “A Blueprint for Representation Information in the OAIS Model” Paper to be presented at the NASA/Goddard Mass Storage Conference, Washington D.C. March 2000.

5 Annex A - List of Organisations

The following is a list of organisations selected as reviewers for an earlier version of this document. The Cedars team would like to take this opportunity to thank the following organisations and individuals:

- Michael Alexander and colleagues, The British Library
- Ann Apps, MIMAS
- Chris Clark, The National Sound Archive
- Nancy Elkington, Research Libraries Group
- Michael Emly, University of Leeds Library

- Peter Graham, Syracuse University Library
- Hamish James, The History Data Service of the Arts and Humanities Data Service
- Vanessa Marshall,The National Preservation Office
- Sally Morris, The Association of Learned and Professional Society Publishers
- Chris Rusbridge, The Joint Information Systems Committee (eLib)
- Colin Webb and colleagues, The National Library of Australia
- Titia van der Werf, The Nedlib Project

Appendix Six : Metadata : Preservation Description Information, an XML DTD

```
<!ELEMENT preservationDescriptionInformation
  (referenceInformation?,contextInformation?,provenanceInformation?,
  fixityInformation?)>

<!ELEMENT referenceInformation
  (resourceDescription?,existingMetadata?) >

<!ELEMENT resourceDescription
  (DCtitle | DCcreator | DCsubject | DCdescription | DCpublisher |
  DCcontributor | DCdate | DCtype | DCformat | DCidentifier |
  DCsource | DClanguage | DCrelation | DCcoverage | DCrights)* >

<!ELEMENT DCtitle  (#PCDATA | reference | list | p)* >
<!ELEMENT DCcreator (#PCDATA | reference | list | p)* >
<!ELEMENT DCsubject (#PCDATA | reference | list | p)* >
<!ELEMENT DCdescription (#PCDATA | reference | list | p)* >
<!ELEMENT DCpublisher (#PCDATA | reference | list | p)* >
<!ELEMENT DCcontributor (#PCDATA | reference | list | p)* >
<!ELEMENT DCdate (#PCDATA | reference | list | p)* >
<!ELEMENT DCtype (#PCDATA | reference | list | p)* >
<!ELEMENT DCformat (#PCDATA | reference | list | p)* >
<!ELEMENT DCidentifier (#PCDATA | reference | list | p)* >
<!ELEMENT DCsource (#PCDATA | reference | list | p)* >
<!ELEMENT DClanguage (#PCDATA | reference | list | p)* >
<!ELEMENT DCrelation (#PCDATA | reference | list | p)* >
<!ELEMENT DCcoverage  (#PCDATA | reference | list | p)* >
<!ELEMENT DCrights (#PCDATA | reference | list | p)* >

<!ELEMENT reference (#PCDATA) > <!ATTLIST reference  type  CDATA  #IMPLIED>

<!ELEMENT list  (item+)>

<!ELEMENT p (#PCDATA | reference | list)*>

<!ELEMENT item (#PCDATA | reference | list | p)* >

<!ELEMENT existingMetadata (existingRecord+)>

<!ELEMENT existingRecord (#PCDATA) >
    <!ATTLIST existingRecord scheme  CDATA  #IMPLIED>

<!ELEMENT contextInformation (relatedInformationObject+)>

<!ELEMENT relatedInformationObject (relationship,reference)>

<!ELEMENT relationship  (#PCDATA) >
```

```
<!ELEMENT provenanceInformation
  (historyOfOrigin?,managementHistory?,rightsManagement?) >

<!ELEMENT historyOfOrigin
  (reasonForCreation*,custodyHistory*,changeHistoryBeforeArchiving*,
  originalTechnicalEnvironments*,reasonForPreservation*) >

<!ELEMENT reasonForCreation (#PCDATA | reference | list | p)* >

<!ELEMENT custodyHistory (#PCDATA | reference | list | p)* >

<!ELEMENT changeHistoryBeforeArchiving (#PCDATA | reference | list | p)* >

<!ELEMENT originalTechnicalEnvironments (prerequisites*,procedures*,
  documentation*) >

<!ELEMENT prerequisites (#PCDATA | reference | list | p)* >

<!ELEMENT procedures (#PCDATA | reference | list | p)* >

<!ELEMENT documentation (#PCDATA | reference | list | p)* >

<!ELEMENT reasonForPreservation (#PCDATA | reference | list | p)* >

<!ELEMENT managementHistory (ingestProcessHistory?,administrationHistory?) >

<!ELEMENT ingestProcessHistory (#PCDATA | reference | list | p)* >

<!ELEMENT administrationHistory (actionHistory*,policyHistory*) >

<!ELEMENT actionHistory (#PCDATA | reference | list | p)* >

<!ELEMENT policyHistory (#PCDATA | reference | list | p)* >

<!ELEMENT rightsManagement (negotiationHistory?,rightsInformation?) >

<!ELEMENT negotiationHistory (#PCDATA | reference | list | p)* >

<!ELEMENT rightsInformation (copyrightStatement*,actors?,actions?) >

<!ELEMENT copyrightStatement (nameOfPublisher*,dateOfPublication?,
  placeOfPublication*,rightsWarning?,contactsRightsHolders*)>

<!ELEMENT nameOfPublisher (#PCDATA | reference | list | p)* >

<!ELEMENT dateOfPublication (#PCDATA | reference | list | p)* >

<!ELEMENT placeOfPublication (#PCDATA | reference | list | p)* >

<!ELEMENT rightsWarning (#PCDATA | reference | list | p)* >

<!ELEMENT contactsRightsHolders (#PCDATA | reference | list | p)* >
```

```
<!ELEMENT actors (#PCDATA | reference | list | p)* >

<!ELEMENT actions (permittedByStatute?,permittedByLicence?) >

<!ELEMENT permittedByStatute (#PCDATA | reference | list | p)* >
    <!ATTLIST permittedByStatute legislationTextPointer CDATA  #REQUIRED >

<!ELEMENT permittedByLicence (#PCDATA | reference | list | p)* >
    <!ATTLIST permittedByLicence licenceTextPointer CDATA  #REQUIRED >


<!ELEMENT fixityInformation (authenticationIndicator+)>

<!ELEMENT authenticationIndicator (#PCDATA | reference | list | p)* >
```