

# METADATA FOR DIGITAL PRESERVATION : THE CEDARS PROJECT OUTLINE SPECIFICATION

DRAFT FOR PUBLIC CONSULTATION

The Cedars Project Team and UKOLN\*

March 2000

## **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.

## **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

---

\* Kelly Russell, Derek Sergeant, Andy Stone, Ellis Weinberger and Michael Day, with thanks to Mr John Wells

Demonstrator Project system and architecture.<sup>1</sup> 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.<sup>2</sup> 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.

### **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:

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:

---

<sup>1</sup>Further information about the Cedars Demonstrator Project is available on the Cedars website <http://www.leeds.ac.uk/cedars>

<sup>2</sup>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.

- 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 Pittsburg 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).

### **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).<sup>3</sup> 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".<sup>4</sup> 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.<sup>5</sup> 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

---

<sup>3</sup>Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-R-1, Red Book, May 1999.

<sup>4</sup>Ibid p 1-11.

<sup>5</sup>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.

groups different kinds of descriptive metadata, so that what the Content Information actually is can still be understood indefinitely.<sup>6</sup>

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.<sup>7</sup>

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.<sup>8</sup> 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.

### **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.

### **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

---

<sup>6</sup>Ibid p 4-25.

<sup>7</sup>Werf-Davelaar, T. van der. Long-term preservation of electronic publications: the NEDLIB project. D-Lib Magazine, 5 (9), September 1999.

<sup>8</sup>A list of organisations selected as reviewers is provided at Annex A.

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.<sup>9</sup> 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).<sup>10</sup>

**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 fulfill their preservation responsibilities and meet the needs of the archive's user communities.**

### 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 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<sup>11</sup> 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 1.1.3 below).

---

<sup>9</sup>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 1.2.1.1.1 of the Representation Information below.

<sup>10</sup>For example a digital sound recording with associated copyright and performance rights for individual tracks.

<sup>11</sup>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'.

## 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.

## 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.

## 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 visualised swiftly. They also

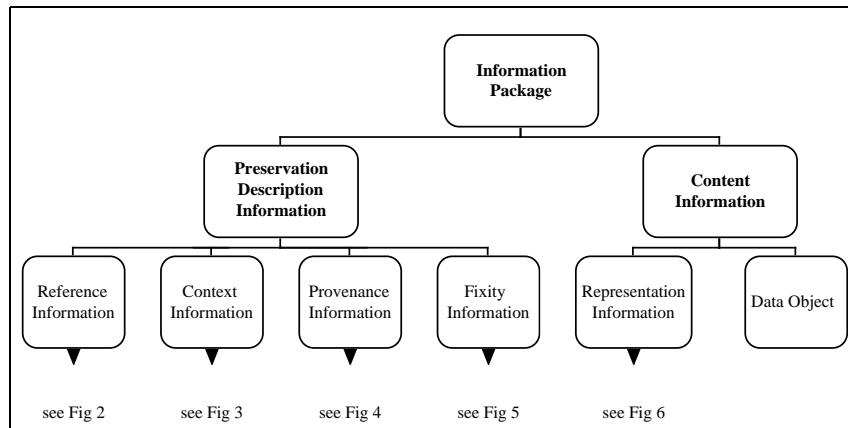


Figure 1: The structure of an Information Package

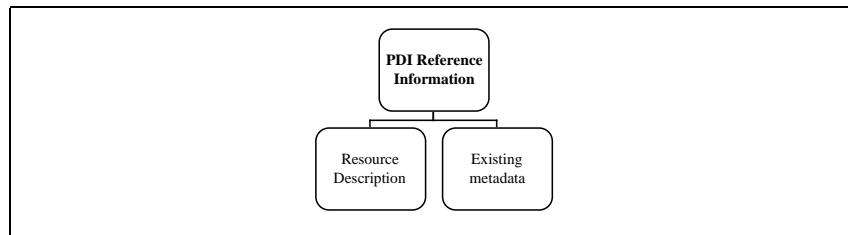


Figure 2: The structure of Reference Information (PDI)

allow easy comparison between the Cedars metadata and the OAIS<sup>12</sup> structure.

Figure 1<sup>13</sup> 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 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.”<sup>14</sup> Broadening this slightly: in Cedars, the Reference Information identifies and describes the Content sufficiently and so holds most of the

<sup>12</sup>Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-R-1, Red Book, May 1999.

<sup>13</sup>Ibid. Figure 1 maps onto OAIS figure 4-13, p 4-22. 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

<sup>14</sup>Ibid p 4-25.

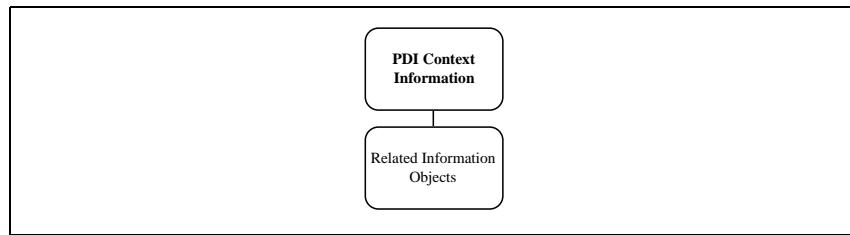


Figure 3: The structure of Context Information (PDI)

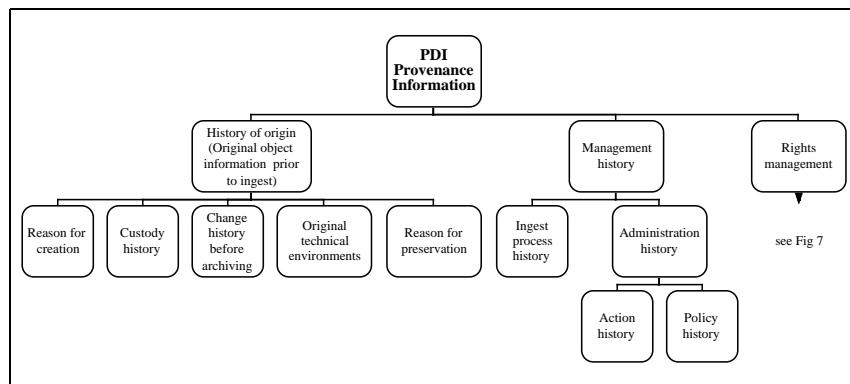


Figure 4: The structure of Provenance Information (PDI)

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.<sup>15</sup> This diagram is likely to be expanded as test resources begin to show which aspects of this are relevant, and when.

Figure 4 represents the Provenance Information. “This information documents the history of the Content Information.”<sup>16</sup> 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 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”.<sup>17</sup> 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

<sup>15</sup>Ibid.

<sup>16</sup>Ibid.

<sup>17</sup>Ibid p 4-26.

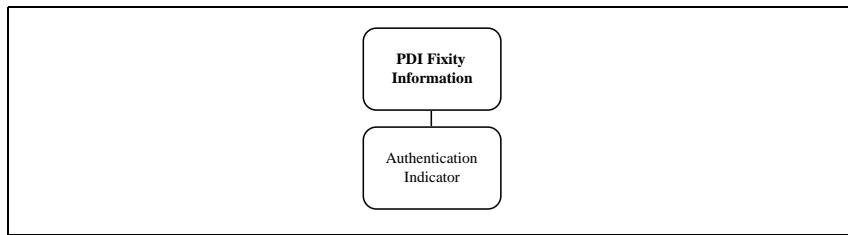


Figure 5: The structure of Fixity Information (PDI)

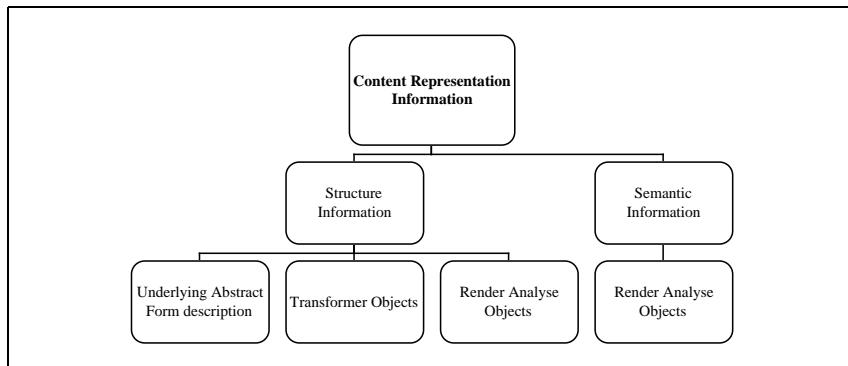


Figure 6: The structure of Representation Information (Content)

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.<sup>18</sup> The OAIS reference model says that the RI is “needed to make the Data Object understandable”.<sup>19</sup> 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.

---

<sup>18</sup>Ibid, p 4-18, Figure 4-10.

<sup>19</sup>Ibid, p 2-5.

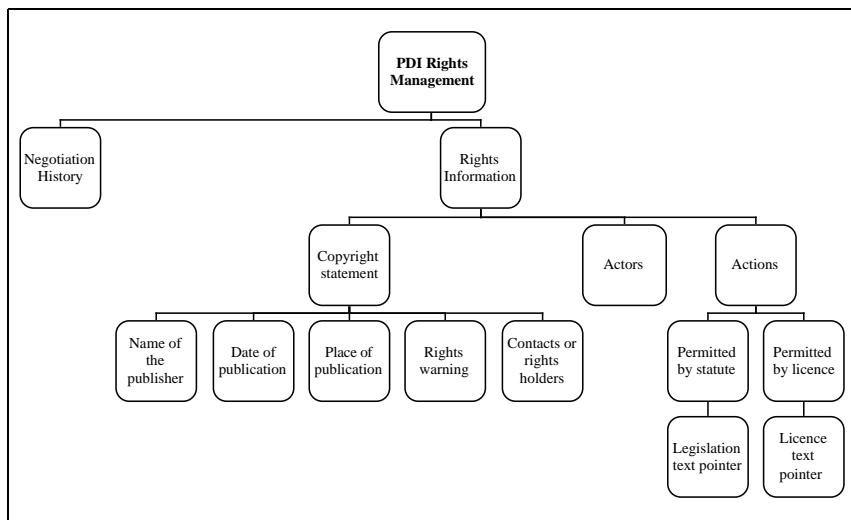


Figure 7: The structure of Rights Management (PDI - Provenance)

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.

### **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.

#### **1 Information Package**

Name	Information Package
Identifier	Information-Package
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	1.1 Preservation Description Information 1.2 Content Information

##### **1.1 Preservation Description Information**

Name	Preservation Description Information
Identifier	Preservation-Description-Information
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.” <sup>20</sup>
Obligation	Very Significant
Sub-Elements	1.1.1 Reference Information 1.1.2 Context Information 1.1.3 Provenance Information 1.1.4 Fixity Information

<sup>20</sup>Ibid.

### 1.1.1 Reference Information

Name	Reference Information
Identifier	Reference-Information
Definition	Contains identifiers for the resource and additional information describing the resource. Also includes existing catalogue metadata schemes.
Obligation	Very Significant
Sub-Elements	1.1.1.1 Resource Description 1.1.1.2 Existing Metadata
Comment	Includes information for resource discovery which may be taken from existing metadata created specifically for this purpose.

#### 1.1.1.1 Resource Description

Name	Resource Description
Identifier	Resource-Description
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.

### 1.1.1.2 Existing Metadata

Name	Existing Metadata
Identifier	Existing-Metadata
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	1.1.1.2.1 Existing Records
Comments	There are different options for integrating existing metadata schema into the AIP such as the Warwick Framework.

#### 1.1.1.2.1 Existing Records

Name	Existing Records
Identifier	Existing-Records
Definition	Each instantiation of a metadata scheme.
Obligation	Significant
Repeatable	Yes

### 1.1.2 Context Information

Name	Context Information
Identifier	Context-Information
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	1.1.2.1 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 below).

### 1.1.2.1 Related Information Objects

Name	Related Information Objects
Identifier	Related-Information-Objects
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.

### 1.1.3 Provenance Information

Name	Provenance Information
Identifier	Provenance-Information
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.” <sup>21</sup> For Cedars this element also contains information about the reason a resource was created and why it was preserved.
Obligation	Very Significant
Sub-Elements	1.1.3.1 History of Origin (Original object information prior to ingest) 1.1.3.2 Management History 1.1.3.3 Rights Management
Comment	This element may include reference links to earlier (or later) manifestations of the digital object. (see Change History 1.1.3.2.2.1)

<sup>21</sup>Ibid.

### 1.1.3.1 History of Origin

Name	History of Origin
Identifier	History-of-Origin
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	1.1.3.1.1 Reason for Creation 1.1.3.1.2 Custody History 1.1.3.1.3 Change History Before Archiving 1.1.3.1.4 Original Technical Environments 1.1.3.1.5 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).

#### 1.1.3.1.1 Reason for Creation

Name	Reason for Creation
Identifier	Reason-for-Creation
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

#### 1.1.3.1.2 Custody History

Name	Custody History
Identifier	Custody-History
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

### 1.1.3.1.3 Change History Before Archiving

Name	Change History Before Archiving
Identifier	Change-History-Before-Archiving
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

### 1.1.3.1.4 Original Technical Environments

Name	Original Technical Environments
Identifier	Original-Technical-Environments
Definition	This element contains information about the operating environment of the <b>original digital object at the time of ingest</b> , 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	1.1.3.1.4.1 Prerequisites 1.1.3.1.4.2 Procedures 1.1.3.1.4.3 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 data object (see 1.2.1).

#### 1.1.3.1.4.1 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.

#### 1.1.3.1.4.2 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.

#### 1.1.3.1.4.3 Documentation

Name	Documentation
Identifier	Documentation
Definition	Associated documentation for 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.

#### 1.1.3.1.5 Reason for Preservation

Name	Reason for Preservation
Identifier	Reason-for-Preservation
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.

### 1.1.3.2 Management History

Name	Management History
Identifier	Management-History
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	1.1.3.2.1 Ingest Process History 1.1.3.2.2 Administration History

#### 1.1.3.2.1 Ingest Process History

Name	Ingest Process History
Identifier	Ingest-Process-History
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

#### 1.1.3.2.2 Administration History

Name	Administration History
Identifier	Administration-History
Definition	This element describes what happened to the digital object after the completion of ingest.
Obligation	Very Significant
Sub-Elements	1.1.3.2.2.1 Action History 1.1.3.2.2.2 Policy History

#### 1.1.3.2.2.1 Action History

Name	Action History
Identifier	Action-History
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.

### 1.1.3.2.2 Policy History

Name	Policy History
Identifier	Policy-History
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.

### 1.1.3.3 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	Rights-Management
Definition	This metadata section contains information relating to the intellectual property rights relevant to the digital object.
Obligation	Very Significant
Sub-Elements	1.1.3.3.1 Negotiation History 1.1.3.3.2 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.

#### 1.1.3.3.1 Negotiation History

Name	Negotiation History
Identifier	Negotiation-History
Definition	This element contains the details of the rights negotiations leading to submission of the digital object for preservation.
Obligation	Very Significant

### 1.1.3.3.2 Rights Information

Name	Rights Information
Identifier	Rights-Information
Definition	This metadata section contains information relating to the intellectual property rights relevant to the digital object.
Obligation	Very Significant
Sub-Elements	1.1.3.3.2.1 Copyright Statement 1.1.3.3.2.2 Actors 1.1.3.3.2.3 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.

#### 1.1.3.3.2.1 Copyright Statement

Name	Copyright Statement
Identifier	Copyright-Statement
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	1.1.3.3.2.1.1 Name of Publisher 1.1.3.3.2.1.2 Date of Publication 1.1.3.3.2.1.3 Place of Publication 1.1.3.3.2.1.4 Rights Warning 1.1.3.3.2.1.5 Contacts or Rights Holders

##### 1.1.3.3.2.1.1 Name of Publisher

Name	Name of Publisher
Identifier	Publisher-Name
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.

#### **1.1.3.3.2.1.2 Date of Publication**

Name	Date of Publication
Identifier	Publication-Date
Definition	This sub-element contains the date of publication of this version of this digital object.
Obligation	Significant

#### **1.1.3.3.2.1.3 Place of Publication**

Name	Place of Publication
Identifier	Publication-Place
Definition	This sub-element contains the place of publication of this version of this digital object.
Obligation	Significant
Repeatable	Yes

#### **1.1.3.3.2.1.4 Rights Warning**

Name	Rights Warning
Identifier	Rights-Warning
Definition	This sub-element contains a warning that the digital object may be subject to copyright or database right.
Obligation	Very Significant

#### **1.1.3.3.2.1.5 Contacts or Rights Holders**

Name	Contacts or Rights Holders
Identifier	Contacts-Rights-Holders
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.

### 1.1.3.3.2.2 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.

### 1.1.3.3.2.3 Actions

Name	Actions
Identifier	Actions
Definition	This element contains sub-elements describing permitted actions.
Obligation	Very Significant
Sub-Elements	1.1.3.3.2.3.1 Permitted by Statute 1.1.3.3.2.3.2 Permitted by Licence

#### 1.1.3.3.2.3.1 Permitted by Statute

Name	Permitted by Statute
Identifier	Permitted-Statute
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	1.1.3.3.2.3.1.1 Legislation Text Pointer
Comment	Note should be taken of ' <i>Permitted by Licence</i> ', 1.1.3.3.2.3.2

##### 1.1.3.3.2.3.1.1 Legislation Text Pointer

Name	Legislation Text Pointer
Identifier	Legislation-Text-Pointer
Definition	This sub-element contains a pointer to the full text or texts of the current relevant legislation.
Obligation	Significant

### 1.1.3.3.2.3.2 Permitted by License

Name	Permitted by Licence
Identifier	Permitted-Licence
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	1.1.3.3.2.3.2.1 Licence Text Pointer
Comment	<p>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.</p> <p>Collective licensing agreements may be in place and may be relevant.</p>

#### 1.1.3.3.2.3.2.1 Licence Text Pointer

Name	Licence Text Pointer
Identifier	Licence-Text-Pointer
Definition	This sub-element contains a pointer to the full text or texts of the current relevant licence or licences.
Obligation	Less Significant

#### 1.1.4 Fixity Information

Name	Fixity Information
Identifier	Fixity-Information
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	1.1.4.1 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.

##### 1.1.4.1 Authentication Indicator

Name	Authentication Indicator
Identifier	Authentication-Indicator
Definition	The mechanism used to ensure the digital object's authenticity. For example, a digital certificate.
Obligation	Very Significant
Repeatable	Yes

## 1.2 Content Information

Name	Content Information
Identifier	Content-Information
Definition	The primary target for preservation. Composed of a Data Object and its Representation Information.
Obligation	Very Significant
Repeatable	No
Sub-Elements	1.2.1 Representation Information 1.2.2 Data Object

### 1.2.1 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 1.2.1.1.2 Transformer Objects (TOs), 1.2.1.1.3 Render/Analyse/Convert (RACs) Objects, and 1.2.1.1.1 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 1.2.1.1.2 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).

Due to the dynamic nature of the development of the representation information and the representation networks within the Cedars project, this specification lags behind the current implementation of this metadata category. An example of a representation network has been created for a recent paper<sup>22</sup>. This section is liable to be updated.

---

<sup>22</sup>A Blueprint for Representation Information in the OAIS Model, Holdsworth and Sergeant, 2000

Name	Representation Information
Identifier	Representation-Information
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	1.2.1.1 Structure Information 1.2.1.2 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.

### 1.2.1.1 Structure Information

Name	Structure Information
Identifier	Structure-Information
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	1.2.1.1.1 Underlying Abstract Form Description 1.2.1.1.2 Transformer Objects 1.2.1.1.3 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).

#### 1.2.1.1.1 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	UAF-Description
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.

#### 1.2.1.1.2 Transformer Objects (TOs)

The transformer objects are a specialist type of RAC (section 1.2.1.1.3). They work from a resource in a common form (a byte-stream) and render it into the UAF.

Name	Transformer Object
Identifier	UAF-Transformer
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	1.2.1.1.2.1 Platform 1.2.1.1.2.2 Parameters 1.2.1.1.2.3 Render/Analyse Engines 1.2.1.1.2.4 Output Format 1.2.1.1.2.5 Input Format

### 1.2.1.1.2.1 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.

### 1.2.1.1.2.2 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.

### 1.2.1.1.2.3 Render/Analyse Engines

Name	Render/Analyse Engines
Identifier	Render/Analyse-Engines
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.

#### 1.2.1.1.2.4 Output Format

Name	Output Format
Identifier	Output-Format
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.

#### 1.2.1.1.2.5 Input Format

Name	Input Format
Identifier	Input-Format
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.

#### 1.2.1.1.3 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	RAC-Object
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	1.2.1.1.3.1 Platform 1.2.1.1.3.2 Parameters 1.2.1.1.3.3 Render/Analyse Engines 1.2.1.1.3.4 Output Format 1.2.1.1.3.5 Input Format

### 1.2.1.2 Semantic Information

Name	Semantic Information
Identifier	Semantic-Information
Definition	This provides the mechanisms which allow the specific digital object in the AIP to be rendered.
Obligation	Significant
Repeatable	No
Sub-Element	1.2.1.2.1 Render/Analyse Objects
Comment	Often these rendering processes will be relevant only for the single digital resource, and will not have general applicability.

#### 1.2.1.2.1 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	RAO
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	1.2.1.2.1.1 Platform 1.2.1.2.1.2 Parameters 1.2.1.2.1.3 Render/Analyse Engines 1.2.1.2.1.4 Output Format 1.2.1.2.1.5 Input Format

### 1.2.2 Data Object

Name	Primary Digital Object
Identifier	Primary-Digital-Object
Definition	This element 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	As mentioned above, this is not a true 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.

## **2 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.

### **3 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