Preservation Metadata and the OAIS Information Model

A Metadata Framework to Support the Preservation of Digital Objects

A Report by

The OCLC/RLG Working Group on Preservation Metadata

http://www.oclc.org/research/pmwg/

June 2002

OCLC/RLG Working Group on Preservation Metadata

Brian Lavoie
OCLC Liaison

Robin Dale
RLG Liaison

Membership:

Michael Alexander
British Library

Kevin Bradley
National Library of Australia

Michael Day
UKOLN

Rebecca Guenther
Library of Congress

Bernard Hurley
University of California, Berkeley

Catherine Lupovici
NEDLIB

Oya Rieger
Cornell University

Derek Sergeant
CEDARS

Titia van der Werf
NEDLIB

Colin Webb
National Library of Australia

Robin Wendler
Harvard University

Deborah Woodyard
British Library

I. Introduction

Preservation metadata is the information infrastructure that supports the processes associated with digital preservation. More specifically, it is the information necessary to maintain the *viability, renderability,* and *understandability* of digital resources over the long-term. Viability requires that the archived digital object's bit stream is intact and readable from the digital media upon which it is stored. Renderability refers to the translation of the bit stream into a form that can be viewed by human users, or processed by computers. Understandability involves providing enough information such that the rendered content can be interpreted and understood by its intended users. Preservation metadata can serve as input to preservation processes, and also record the output of these same processes.

The importance of preservation metadata has been underscored by the efforts of a number of organizations to develop metadata of this type in support of their own digital preservation activities. While these efforts constituted pioneering work, they were conducted largely in isolation, lacking any substantial degree of cross-organizational coordination. As a result, each preservation metadata element set tended to reflect the particular needs and requirements of the organization that authored them. In this sense, the digital preservation community, while benefiting immensely from this work, nevertheless still lacked a metadata framework for digital preservation that represented a consensus of leading experts and practitioners, and could be readily applied to a broad range of digital preservation activities.

Since the development of these preservation metadata element sets, several factors have emerged within the digital preservation community that suggest that consensus-building activity in the area of preservation metadata is not only desirable, but practicable. First, there is wide spread recognition that digital preservation poses issues and challenges shared by organizations of all descriptions, with the attendant implication that extensive scope may exist to address these challenges cooperatively. Second, a conceptual framework for a generic digital archiving system emerged in the form of the OAIS reference model, offering shared concepts and terminology, and representing common ground to serve as the starting point for discussion and collaboration. The OAIS model has proliferated rapidly through the digital preservation community, and has been explicitly adopted by, or at least informed, many prominent digital preservation initiatives. The OAIS framework currently enjoys the status of a *de facto* standard in digital preservation.

In March 2000, OCLC and RLG sponsored the creation of a working group to explore consensus-building in the area of preservation metadata. The working group was to be composed of leading experts in the digital preservation community, representing a variety of institutional and geographical backgrounds. The charge of the group was to pool their expertise and experience to develop a preservation metadata framework applicable to a broad range of digital preservation activities. The group began its work by publishing a white paper entitled *Preservation Metadata for Digital Objects: A Review of the State of the Art*, which defined and discussed the concept of preservation metadata, reviewed current thinking and practice in the use of preservation metadata, and identified starting points for consensus-building activity in this area. The group then turned its attention to the main focus of its activity – the collaborative development of a preservation metadata framework. This paper reports the results of the working group's efforts in that regard.

II. Methodology

The working group reviewed concepts and issues associated with the information model embedded within the OAIS framework. This review was conducted for the purpose of developing an implementation of the information model that would accommodate the needs of the library community, along with other institutions tasked with the long-term management of information in digital form. The implementation takes the form of 1) an expanded conceptual structure for the OAIS information model, and 2) a set of metadata elements, mapped to the conceptual structure and reflecting the information concepts and requirements articulated in the OAIS model.

The working group made no assumptions about the type or structure of the digital resource with which the preservation metadata is associated, nor did it assume that a particular preservation strategy (e.g., migration or emulation) was followed. The working group chose to base their implementation on a synthesis of four existing preservation metadata schemes, developed by the CURL Exemplars in Digital Archives project (CEDARS), the National Library of Australia (NLA), the Networked European Deposit Library (NEDLIB), and the Online Computer Library Center, Inc. (OCLC), respectively. The synthesis was then supplemented by refinements, elaborations, and additional structure and elements recommended by the working group members.

In this paper, the term *implementation* is used to describe the process of breaking down the general concepts defined in the OAIS information model into a hierarchy of increasingly precise components capturing specific types of information. The points at which this process stopped – in other words, the "leaves" of the hierarchical tree – collectively define what is referred to as *preservation metadata elements* in this paper. It should be noted, however, that these elements are not necessarily atomic; it is easy to imagine cases where the needs and characteristics of particular digital archiving systems may require deconstruction of these elements into still more precise components.

A related issue is the distinction between *structure* and *elements*. In some parts of the implementation presented below, a particular piece of metadata is broken down into several structural layers, with the upper layers primarily serving an organizational purpose, and the lowest layer representing the metadata element where information is actually recorded. For the purposes of the discussion in this paper, each layer is treated as an element in its own right, in the sense that it is defined, its purpose stated, and an example given as to how it might be populated. This is in recognition of the fact that implementation of metadata occurs at varying levels of specificity: in some systems, information may be recorded in elements expressing broad informational concepts; in other cases, elements representing very specific pieces of information may be utilized. In practice, not all structural levels discussed below would necessarily be implemented as metadata elements in a digital archiving system.

In addition to defining a body of recommended metadata for digital preservation, this paper also discusses the purpose, or rationale, for each element, and provides an example of how the element might be populated. This example might take the form of a specific value, if this is possible and/or meaningful; otherwise, the example takes the form of a description of the types of values that might be used to populate the element.

III. The Open Archival Information System (OAIS) Reference Model[1]

At the request of the International Organization for Standardization (ISO), the Consultative Committee for Space Data Systems (CCSDS), an international collaboration of space agencies aimed at the development of data handling standards in support of space research, began coordinating an effort to develop archive standards for the long-term storage of data in digital form. As a foundation for this effort, the CCSDS set about producing a reference model, which would establish terminology and concepts for describing and comparing data models and archival architectures, identify the significant entities and relationships among entities in an archive environment, elucidate the key functional and information components of an archival system, and ultimately, serve as a framework within which standards-building activity could take place.

The work of the CCSDS resulted in the release in May 1999 of the Open Archival Information System (OAIS) reference model. The reference model is a conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term. It describes the environment in which an archive resides, the functional components of the archive itself, and the information infrastructure supporting the archive's processes. The reference model underwent an extensive review as an ISO draft recommendation, extending beyond the space community to engage libraries and other cultural heritage institutions, government agencies, and the private sector. Based on this review, a revised version of the reference model was released in June 2001.

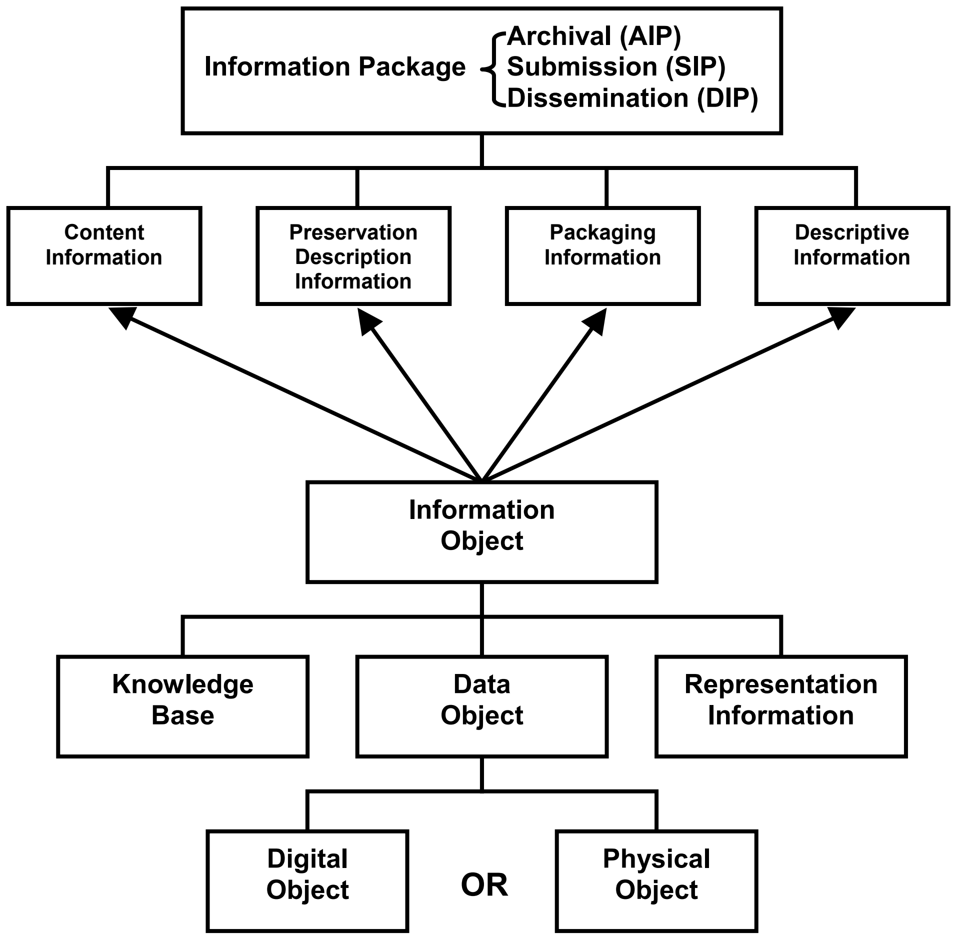
III.1. The OAIS Information Model

The portion of the reference model that is of direct relevance to the issue of preservation metadata is the information model embedded within the OAIS framework. The OAIS information model broadly describes the metadata requirements associated with retaining a digital object over the long-term. This information model is particularly useful because it was developed in conjunction with a *functional* model of a digital archiving system – in other words, an articulation of the primary processes, or functional components, of an OAIS-type archive. In this sense, the information model is consistent with a comprehensive, structured view of the archiving system it supports.

The OAIS information model is illustrated in Figure 1. In the context of the OAIS, information can exist in two forms: either as a physical object (e.g., a paper document, a soil sample), or as a digital object (e.g., a PDF file, a TIFF file). These two types - physical and digital - are collectively known as the Data Object. A Data Object can take several forms: in particular, either the material that is the primary focus of preservation, or the metadata associated with an archived digital object.

[1] The following discussion is adapted from Lavoie (2000) "Meeting the Challenges of Digital Preservation: The OAIS Reference Model", *OCLC Newsletter*, No. 243, p.26-30; and from OCLC/RLG Working Group on Preservation Metadata (2001) "Preservation Metadata for Digital Objects: A Review of the State of the Art"

Figure 1: OAIS Information Model



Interpretation of the Data Object as meaningful information is achieved through the combination of the users' knowledge base and the Representation Information associated with the Data Object. Each individual or class of individuals has a knowledge base, which is used to understand and interpret data. For example, a community of English-speaking individuals has the knowledge base necessary to read data conveyed in English prose. Similarly, Java programmers are expected to have the knowledge base to understand information in the form of Java source code. It should be emphasized that the knowledge base is external to the archive, and is not maintained, evolved, or preserved as part of any archival function.

The knowledge base is not always sufficient to fully understand the archived Data Object. In this event, the Data Object must be supplemented by Representation Information, in order that it can be viewed and fully understood by the archive's intended users. Representation Information facilitates the proper rendering, understanding, and interpretation of a digital object's content. At the most fundamental level, Representation Information imparts meaning to an object's bit stream. Thus, Representation Information indicates whether the string of bits:

1011010001101011001001...

represents a paragraph of text, a sound file, an image, etc. However, knowledge of the file format underlying the bit stream may not be enough to interpret its content. For example, a Data Object in the form of an ASCII file might contain the following:

04 27 56
01 16 44
02 01 17

More information is required to impart meaning to this data. A user might guess that the numbers refer to dates (month, day, and year), which is a plausible interpretation, but certainly not the only one. In fact, this data might be properly interpreted as the elapsed times (hours, minutes, seconds) of three laboratory-controlled chemical reactions. This description would also be considered Representation Information associated with the Data Object.

A digital object consists of a stream of bits; Representation Information imparts meaning to these bits. Representation Information can take two forms: structural information and semantic information. Structural information interprets the bits by organizing them into specific data types, groups of data types, and other higher-level meanings. Structural information would include a specification of the data format, and possibly a description of the hardware/software environment needed to access the data. Semantic information, on the other hand, provides additional meaning to the data structures identified by the structural information. For example, structural information may identify a bit stream as ASCII text characters, while semantic information might indicate that the text is in English.

The OAIS reference model notes that if Representation Information is itself in digital form, then additional Representation Information will be needed to understand the bits of the first layer of Representation Information, a third layer of Representation Information will be needed to understand the bits of the second layer of Representation Information, and so on. The reference model recommends that the resulting Representation Network end with a physical document which "bootstraps" the interpretation process.

An *information object* is defined as a Data Object combined with Representation Information. In a digital environment, this implies a sequence of bits, combined with all data necessary to make the bit stream viewable and understandable. There are four classes of information objects: Content Information, Preservation Description Information, Packaging Information, and Descriptive Information. Each of these information objects will be discussed in detail below.

An *information package* is an aggregation of a Content Information Object, a Preservation Description Information Object, a Packaging Information Object, and a Descriptive Information Object. Information packages can be assigned to one of three types. The Submission Information Package (SIP) is sent from the information producer to the archive, the Archive Information Package (AIP) is the information package actually stored by the archive, and the Dissemination Information Package (DIP) is the information package transferred from the archive to a user in response to an access request. In the context of preservation metadata, the relevant information package is the AIP, since this is the package which is retained over the long-term.

An AIP is the aggregation of four types of information object. Each of these types is described below. Note that each information object consists of a Data Object and the associated Representation Information necessary to make the Data Object meaningful. However, the Representation Information is typically mentioned explicitly only in the context of the Data Object of primary interest - i.e., the object being archived, rather than its associated metadata. This convention is followed in the remainder of the paper. Note further that the Data Object that is the primary focus of preservation is referred to as the *Content Data Object*.

1. Content Information (CI)
... consists of the Content Data Object – i.e., the information that the archive is entrusted to preserve – along with its associated Representation Information

2. Preservation Description Information (PDI)
... contains information necessary to manage the preservation of the Content Information with which it is associated. The OAIS reference model identifies four types of PDI:

- *Reference Information:* enumerates and describes identifiers assigned to the Content Information such that it can be referred to unambiguously, both internally and externally to the archive (e.g., ISBN, URN)
- *Provenance Information:* documents the history of the Content Information (e.g., its origins, chain of custody, preservation actions and effects)
- *Context Information:* documents the relationships of the Content Information to its environment (e.g., why it was created, relationships to other Content Information)
- *Fixity Information:* documents authentication mechanisms used to ensure that the Content Information has not been altered in an undocumented manner (e.g., checksum, digital signature)

3. Packaging Information (PI)
... binds the digital object and its associated metadata into an identifiable unit or package (i.e., an Archival Information Package)

4. Descriptive Information (DI)
... facilitates access to the Content Information via the archive's search and retrieval tools. Descriptive Information serves as input to the archive's finding aids, and is typically derived from the Content Information or Preservation Description Information.

The OAIS information model represents a high-level description of the types of information generated by and managed within the functional components of a complete archiving system. It makes no presuppositions about the type of digital object managed by the archive, nor about the specifics of the technology employed by the archive to achieve its goal of preserving and maintaining access to the digital object over the long term. As such, the model provides a useful foundation for developing a preservation metadata framework of wide applicability.

The next two sections propose an implementation of the two components of the OAIS information model directly relevant to preservation metadata – Content Information and Preservation Description Information. Packaging Information is excluded because it simply binds the digital object and its associated metadata together into a single, logical package, and is not directly associated with the preservation of the object itself. Descriptive Information is metadata for resource discovery, which is outside the bounds of preservation metadata.

IV. A Recommendation for Content Information

The OAIS reference model defines Content Information as "the set of information that is the original target of preservation. It is an Information Object comprised of the Content Data Object and its Representation Information." In a digital archive, the Content Data Object is the bit sequence or set of bit sequences toward which the preservation action is primarily directed. Representation Information is information necessary to render/display, understand, and interpret the Content Data Object. To summarize:

Content Data Object: a bit stream or set of bit streams

Representation Information: metadata that translates the bit stream(s) of the Content Data Object into accessible, meaningful knowledge

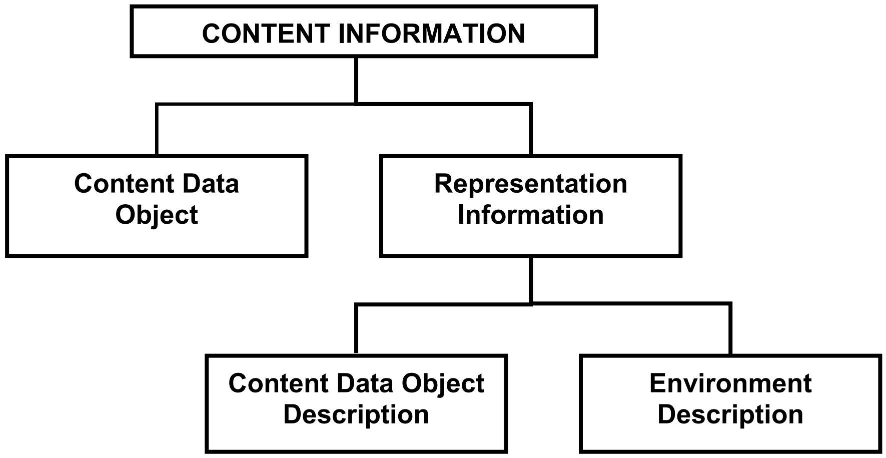
Broadly speaking, Content Information is the digital content being preserved (Content Data Object), along with sufficient information to ensure that the object is both renderable and meaningful to current and future users (Representation Information). The OAIS divides Representation Information into two components. Structure Information describes "the format, or data structure concepts, which are to be applied to the bit sequences and that in turn result in more meaningful values such as characters, numbers, pixels, arrays, tables, etc." In short, Structure Information provides a technical description of the Content Data Object's structured organization, including format, data structures, encoding, etc., in particular as it relates to rendering or displaying the Object in a digital environment.

Semantic Information imparts higher level meanings to the structural components of the Content Data Object, beyond what is expressed by Structure Information. Thus, Semantic Information might indicate that a sequence of alphanumeric characters should be interpreted as English prose, or that a sequence of integers are temperature readings from a chemistry experiment. In this sense, Semantic Information contributes toward an understanding, or appropriate interpretation, of the intellectual content of the Content Data Object.

A useful generalization is that Structure Information is oriented toward making the Content Data Object understandable to computer systems, while Semantic Information is oriented toward making the Object understandable to humans. However, the working group, in the course of its discussions, decided to omit the structure/semantic distinction from its implementation of Representation Information. This decision was based on the observation that the distinction between the two types of Representation Information tends to be more subjective than definitive.

The working group made no assumptions about the type or structure of the Content Data Object. Therefore, implementation of Content Information is equivalent to implementation of its Representation Information component. The working group approached this task by first creating some additional structure for the Representation Information component of the OAIS Content Information Package (Figure 2):

Figure 2: Content Information Package



As Figure 2 illustrates, a Content Information Package is the aggregation of the Content Data Object and its associated Representation Information. The latter is itself the aggregation of two components: Content Data Object Description and Environment Description. The first component represents information detailing the characteristics and features of the Content Data Object itself that are necessary to render and understand its content. The second component describes a hardware/software environment capable of rendering or displaying the Content Data Object in the form in which it currently exists in the archival store.

The two components of Representation Information – Content Data Object Description and Environment Description – are discussed in detail below.

Note: CEDARS = CURL Exemplars in Digital Archives
NLA = National Library of Australia
NEDLIB = Networked European Deposit Library
OCLC = OCLC Digital Archive Service
WG = OCLC/RLG Working Group on Preservation Metadata

IV.1. Content Data Object Description

The working group assembled the following list of metadata elements, which collectively form the Content Data Object Description component of Representation Information:

NAME: **Underlying abstract form description**
ORIGIN: CEDARS
DEFINITION: Human readable description of the Underlying Abstract Form of the Content Data Object

PURPOSE:	Facilitate converting the archived byte stream of the Object into the correct components (such as files and relationships) to render the Object (or access the intellectual content of the Object).
EXAMPLE:	(1) Description of a file system, so that a byte stream, in the form of a ZIP file, can be correctly broken up into the hierarchy of files and folders (e.g., in the case of an archived Web site) (2) Description of the conceptual components of a relational database, and how a byte stream can be manually converted back into the relational database

NAME:	**Structural type**
ORIGIN:	NLA
DEFINITION:	Class of digital object represented by the Content Data Object
PURPOSE:	Choice of appropriate preservation strategy depends on knowing structural type
EXAMPLE:	Still image, sound, text, database, Web document, executable program, etc. List of MIME types may serve as a useful reference

NAME:	**Technical infrastructure of complex object**
ORIGIN:	NLA
DEFINITION:	Internal structure of complex digital objects: i.e., an enumeration of the components of a complex object, along with their inter-relationships
PURPOSE:	Managing preservation requires managing the structure of complex objects as well as their components.
EXAMPLE:	Web page: consists of one ASCII HTML file, along with three embedded static GIF files and one embedded audio WAV file

NAME:	**File description**
ORIGIN:	NLA
DEFINITION:	Technical specifications of the file(s) comprising a Content Data Object. Note: this metadata should apply to file formats which are used to directly render or access content, rather than file formats which are used for storage convenience (e.g., ZIP or TAR files)
PURPOSE:	Describe type-specific metadata essential for managing preservation
EXAMPLE:	GIF image file: dimensions in pixels; resolution; color palette; compression algorithms

NAME:	**Installation requirements**
ORIGIN:	NLA
DEFINITION:	Any specialized procedures needed to install an object
PURPOSE:	Enable access to objects with special installation requirements
EXAMPLE:	Object is in the form of a ZIP file, which must be unpacked and stored on local hard drive in a specified directory tree prior to use; computer must be re-booted after installation

NAME: **Size**
ORIGIN: WG
DEFINITION: Size of object (in bytes)
PURPOSE: Necessary for managing the object within the archive system. For example, migration of storage media from tape to CD-ROM might require this information, since standard CD-ROMs have a maximum capacity of 650 MB. Also important for dissemination purposes: some versions of Windows cannot accept files greater than 2 GB
EXAMPLE: Size of Object: 1.3 MB

NAME: **Access inhibitors**
ORIGIN: NLA
DEFINITION: Description of any features of the Content Data Object intended to inhibit access
PURPOSE: Without this information, the object may not be able to be accessed, copied or migrated.
EXAMPLE: Encryption, watermarking, password protection

NAME: **Access facilitators**
ORIGIN: NLA
DEFINITION: Description of any system or method used to enhance access to information within the Content Data Object, which need to be maintained in successive generations
PURPOSE: Enable the aids and facilitators to be taken into account in any preservation process
EXAMPLE: Time markers in audio or video files, navigational links in a hypertext document

NAME: **Significant properties**
ORIGIN: WG
DEFINITION: Properties of the Content Data Object's rendered content which must be preserved or maintained during successive cycles of the preservation process
PURPOSE: Essential for decision-making related to level and method of access, the richness of preservation metadata required, and the type of preservation processes that will be implemented
EXAMPLE: PDF Document: it is determined that the significant property of the document is the intellectual content of the text; its "look and feel" (color scheme, embedded images, page layout, internal hyperlinks) are not considered essential and will not be preserved

NAME: **Functionality**
ORIGIN: WG

DEFINITION:	Description of any functional or "look and feel" attributes of the rendered Content Data Object, in regard to its current manifestation in the archival store
PURPOSE:	Enumerate the set of functional properties exhibited by the Object relative to the current stage of the preservation cycle
EXAMPLE:	Web page: contains an interactive JavaScript application and embedded animations (Note: see **Quirks** for more information)

NAME:	**Description of rendered content**
ORIGIN:	WG
DEFINITION:	Description of the Content Data Object's content, in regard to how it should be viewed and interpreted by users. Includes clarification of potentially ambiguous data, definition and description of data structures, etc.
PURPOSE:	Ensure proper understanding and interpretation of Object's content by the archive's users.
EXAMPLE:	Content Data Object consisting of an ASCII file of numbers may be clarified as a list of temperature readings from a chemistry experiment performed on a specific day, presented as a series of tab-delimited columns

NAME:	**Quirks**
ORIGIN:	NLA
DEFINITION:	Any loss in functionality or change in the look and feel of the Content Data Object resulting from the preservation processes and procedures implemented by the archive
PURPOSE:	Assist preservation managers to assess the success (or otherwise) of preservation strategies, and prevent time being spent on trying to solve problems that were inherent in the object at the time the strategy was applied. This element documents changes that occur as a result of digitization, migration, and other processes in the preservation cycle, and may also record any disabled functionality present in the Object at the time it is ingested into the archive (see Note #5 at the end of this section)
EXAMPLE:	Web page: has been migrated from HTML to PDF. As a result, hyperlinks are broken; embedded JavaScript application no longer functional

NAME:	**Documentation**
ORIGIN:	WG
DEFINITION:	Supporting documentation necessary/useful for display and/or interpretation of the Content Data Object
SUB-ELEMENT:	Location: location of documentation (e.g., URL)
PURPOSE:	Link the Content Data Object to supporting documentation useful for rendering and understanding its content
EXAMPLE:	Glossary, users' manual, etc.

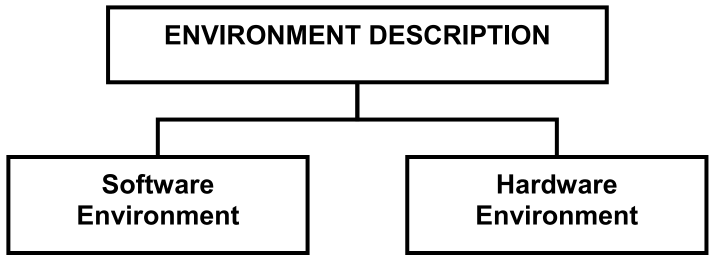
Notes:

1) The File Description element is the place where type-specific metadata would reside. As mentioned earlier, the elements discussed above are not necessarily atomic; local requirements may call for further breakdown into even more granular data. An important example of this would be a breakdown of File Description into type-specific metadata for various classes of digital objects. A number of initiatives are engaged in standards work aimed at developing metadata element sets for specific object types: for example, the NISO effort [5] in regard to digital still images.
2) It is possible that information pertaining to the Installation Requirements element may be placed elsewhere – for example, in the Documentation linked to the Content Data Object.
3) It may be useful to break down the Size element into sub-elements which record uncompressed size and various compressed sizes (based on a set of compression algorithms supported by the archive).
4) It should be noted that the Significant Properties metadata is neither intrinsic to the Object itself, nor time-invariant. Rather, it constitutes the properties that are significant in regard to the archive's Designated Community, and that the archive has the resources to preserve. It is quite possible that the priorities of the Designated Community and/or the resources of the archive will change over time: as these change, so will the Object's significant properties.
5) Quirks can be interpreted in two ways: any loss in functionality of the original Content Data Object from the time of its creation (and possibly prior to its ingest into the archive), or any loss of functionality sustained by the Object, relative to its state when ingested into the archive, as a result of the archive's preservation processes. NLA (from whom this element originated) follows the first interpretation.
6) To understand the relationship between Functionality and Quirks, it is best to think of one as the "negative" of the other. For example, given an archived Content Data Object, one should be able to draw up a list of functional and "look and feel" attributes of the Object's rendered content. The Functionality metadata records all of these attributes which still exist in the current instance of the Object that is in the archival store. Conversely, the Quirks metadata lists all of these attributes which no longer exist as part of the Object's current instance. Therefore, the sum of the attributes recorded in Functionality and Quirks should equal the original list of all attributes.

IV.2. Environment Description

Figure 3 illustrates an implementation of the Environment Description component of Representation Information:

Figure 3: Environment Description



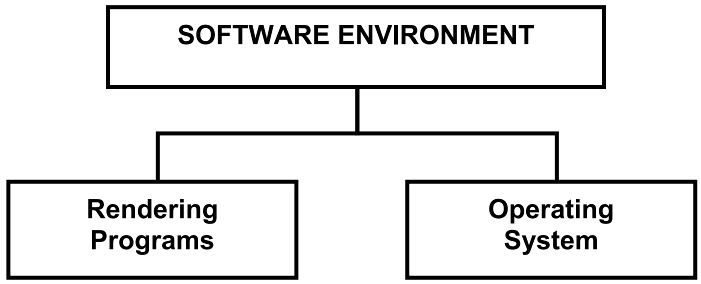
In Figure 3, Environment Description is broken down into two components: Software Environment and Hardware Environment. A software environment is the collection of digital objects – e.g., Internet Explorer and Windows 95 – that, when combined, enable access to the content of the archived object. The hardware environment, on the other hand, consists of physical objects – primarily computer-related equipment such as monitors, microprocessors, and memory chips – that are necessary to operate the software environment.

It should be noted that Environment Description metadata is likely to be "repeatable" in practice. There are often multiple combinations of software and/or hardware capable of rendering or accessing the Content Data Object. For example, a Web document can be rendered using Internet Explorer or Netscape, running on a range of Windows versions: e.g., 95, 98, or 2000. Rather than enumerating all possible environments, an archive may choose to describe only those for which it offers direct support – for example, applications and operating systems that are archived along with Content Data Object itself, or environments for which the archive maintains emulator technology. Alternatively, the archive could confine itself to describing the minimum software/hardware environment capable of rendering or accessing the Content Data Object – for example, the oldest compatible software version, or slowest microprocessor. It is also conceivable that this metadata might describe a "recommended" environment: i.e., the combination of hardware and software best suited for rendering and interacting with the Content Data Object.

IV.2.a. Software Environment

Given the breakdown of Environment Description into Software and Hardware components (illustrated in Figure 3), the Working Group added further structure useful for organizing metadata relevant to these concepts. Figure 4 illustrates the structure of the Software Environment component:

Figure 4: Software Environment



Software Environment is divided into two components: Rendering Programs and Operating System. Rendering Programs operate directly on the Digital Object to render, display, and/or access its content. Operating System refers to the software platform required to operate the Rendering Programs.

For the purposes of this discussion, the rendering of the Content Data Object can be viewed as a two-step process: first, **transform** the archived bit stream into a form compatible with the display/access software, and second, **display/access** the content. It should be noted that the first step, transformation, will not be required if the archived form of the bit stream is directly compatible with the Display/Access Application.

IV.2.a.i. Rendering Programs

The Working Group assembled the following list of two metadata elements (with associated sub-elements) relevant to the Rendering Programs component:

NAME:	**Transformation process**
ORIGIN:	CEDARS (Transformer Object)
DEFINITION:	Description of implementation (or a software mechanism) to automatically transform the byte stream of the Content Data Object into an instantiation of the Underlying Abstract Form (on a particular computing platform)
PURPOSE:	Description of the process by which the byte stream is automatically taken from the archive and turned into the correct representation of components to allow its processing on a particular computing platform
EXAMPLE:	Unzip/untar a file; compile source code into executable

Sub-elements:

NAME:	**Transformer engine**
ORIGIN:	CEDARS (Render/analyze engine)
DEFINITION:	Identifies a specific software engine (e.g., name, version) capable of carrying out the process described in Transformer Process

PURPOSE:	Relate Content Data Object to ancillary software engines needed for transformation
EXAMPLE:	WinZip, which turns a byte stream into a file tree for the PC computing environment

Sub-elements:

NAME:	**Parameters**
ORIGIN:	CEDARS
DEFINITION:	Runtime parameters which must be configured on the Transformer Engine to achieve successful operation
PURPOSE:	Assure successful transformation of the archived byte stream
EXAMPLE:	Specification of output directory for "unzipping" process

NAME:	**Input format**
ORIGIN:	CEDARS
DEFINITION:	Description of the format of digital object that the Transformer Engine works on
PURPOSE:	Ensure that the archived byte stream and Transformer Engine are compatible
EXAMPLE:	ZIP files with ".zip" extension

NAME:	**Output format**
ORIGIN:	CEDARS
DEFINITION:	Description of the format produced by processing the Content Data Object with the Transformer Engine
PURPOSE:	Specify state of Content Data Object prior to use by Display /Access Application (see below)
EXAMPLE:	Object is a Java "class" file subsequent to transformation

NAME:	**Location**
ORIGIN:	WG
DEFINITION:	Location of the Transformer Engine needed to transform the Content Data Object
PURPOSE:	Link Content Data Object to compatible Transformer Engine
EXAMPLE:	Description of where the required Transformer Engine can be obtained. This may take the form of anything ranging from manufacturer information, to a pointer (e.g., URL) to the location of where the Transformer Engine can be directly obtained (e.g., via download, or through the archive itself)

NAME:	**Documentation**
ORIGIN:	WG
DEFINITION:	Supporting documentation necessary/useful for operation/use of the Transformer Engine
SUB-ELEMENT:	Location: location of documentation (e.g., URL)
PURPOSE:	Link the Transformer Engine metadata to supporting documentation useful for operation
EXAMPLE:	Glossary, users' manual, etc.

NAME:	**Display/Access Application**
ORIGIN:	WG
DEFINITION:	Identification of software program capable of displaying the Content Data Object, or accessing its intellectual content
PURPOSE:	Translate the archived byte stream into human-readable content
EXAMPLE:	Internet Explorer 6.0, Adobe Acrobat Reader 4.0

Sub-elements:

NAME:	**Input format**
ORIGIN:	CEDARS
DEFINITION:	Description of the format of digital object that the Display/Access Application works on
PURPOSE:	Ensure that the archived byte stream and Display/Access Application are compatible
EXAMPLE:	Java virtual machine: must use Java "class" files; Adobe Acrobat Reader: must use PDF files

NAME:	**Output format**
ORIGIN:	CEDARS
DEFINITION:	Description of the output to be expected from the Display/Access Application
PURPOSE:	Describe the form of the rendered content of the Content Data Object
EXAMPLE:	Description of a displayed image; description of the contents of an output file produced by the Display/Access Application

NAME:	**Location**
ORIGIN:	WG
DEFINITION:	Location of the Display/Access Application needed to display and/or access the Content Data Object's content
PURPOSE:	Link Content Data Object to compatible Display/Access Application
EXAMPLE:	Description of where the required Display/Access Application can be obtained. This may take the form of anything ranging from manufacturer information, to a

pointer (e.g., URL) to the location of where the Display/Access Application can be directly obtained (e.g., via download, or through the archive itself)

NAME:	**Documentation**
ORIGIN:	WG
DEFINITION:	Supporting documentation necessary/useful for operation/use of the Display/Access Application
SUB-ELEMENT:	Location: location of documentation (e.g., URL)
PURPOSE:	Link the Display/Access Application metadata to supporting documentation useful for operation
EXAMPLE:	Glossary, users' manual, etc.

Notes:

1) As new computer platforms appear, software tools to transform the Content Data Object into an appropriate representation of the Underlying Abstract Form (i.e., Transformer Engines) may need to be built. This will be done in conjunction with the Underlying Abstract Form Description.
2) In some circumstances, the Input Format sub-element of the Transformer Engine element should be identical to the File Description element of the Content Data Object Description. However, in other cases – for example, when the Content Data Object is a ZIP or TAR file – this correspondence may not be exact. Please see the description of the File description element for more information.
3) If a Transformer Engine is required to render the Content Data Object, then the Input Format sub-element of the Display/Access Application element should be identical to the Output Format sub-element of the Transformer engine element. If not, then the Input Format sub-element of the Display/Access Application element should be identical to the File Description element of the Content Data Object Description.
4) The Output Format sub-element of the Display/Access Application element should be compatible with the Description of Rendered Object element of the Content Data Object Description.
5) It is recommended that if the Rendering Programs metadata is intended to describe a minimum or recommended environment, this information should be recorded in another metadata element (e.g., Environment Type, with values "Minimum" or "Recommended").

IV.2.a.ii. Operating System

The Working Group assembled the following list of four metadata elements relevant to the Operating System component:

NAME:	**OS name**
ORIGIN:	NEDLIB
DEFINITION:	Name/designation of software platform upon which Rendering Programs operate

PURPOSE: Identify operating environment used by the Rendering Programs of the Content Data Object
EXAMPLE: Windows, Windows NT, Linux, Apple, Solaris, etc.

NAME: **OS version**
ORIGIN: NEDLIB
DEFINITION: Version of the Operating System identified in OS Name
PURPOSE: Distinguish between different versions of an operating environment, which could potentially impact the ability to run Rendering Programs, and by extension, the ability to access the Content Data Object
EXAMPLE: Windows 3.1, Windows 95, Windows 98, Windows ME

NAME: **Location**
ORIGIN: WG
DEFINITION: Location of working copy of the Operating System described in OS Name and OS Version
PURPOSE: Link Content Data Object to compatible Operating System
EXAMPLE: URL to download OS from manufacturer, or from a digital repository holding an archived copy of the OS. Also could include the location of an emulator for this environment.

NAME: **Documentation**
ORIGIN: WG
DEFINITION: Supporting documentation necessary/useful for operation/use of the Operating System
SUB-ELEMENT: Location: location of documentation (e.g., URL)
PURPOSE: Link the Operating System metadata to supporting documentation useful for operation
EXAMPLE: Glossary, users' manual, etc.

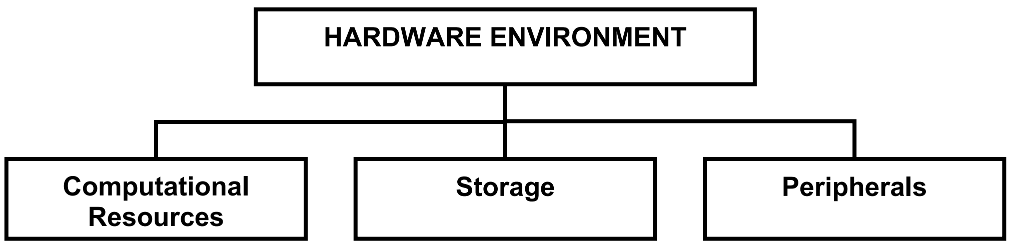
Notes:

1) The element OS Name can be interpreted as the general "operating environment", while the OS Version element specifies a particular manifestation of that environment. For example, Windows NT is a general operating environment, characterized perhaps by a particular look and feel and set of functionalities. Windows NT 4.0, however, is a specific implementation of the Windows NT environment. Compatibility with the Content Data Object's Rendering Programs may extend to the operating environment as a whole, or only to specific versions of that operating environment.
2) It is recommended that if the Operating System metadata is intended to describe a minimum or recommended environment, this information should be recorded in another metadata element (e.g., Environment Type, with values "Minimum" or "Recommended").

IV.2.b. Hardware Environment

A Content Data Object's Hardware Environment – the combination of *physical* equipment necessary to render or access the Object's content – can be broken down further into three sub-components, as illustrated in Figure 5 below:

Figure 5: Hardware Environment



The Hardware Environment embodies three aspects: Computational Resources, Storage, and Peripherals. Computational Resources refers to the logical capacity to process the bit sequences of the Content Data Object and its Software Environment: e.g., an Intel Pentium III microprocessor. Storage refers to any specific storage technology that is required to access the bit sequence of the Content Data Object: for example, if the archive disseminates a Content Data Object on CD-ROM, a CD-ROM drive would be necessary in order to access the Object. Finally, Peripherals includes any additional physical devices which assist in rendering, displaying, or accessing the Content Data Object, such as monitors, sound cards, speakers, etc.

The Working Group assembled the following list of eight metadata elements relevant to the various components of the Hardware Environment:

Computational Resources:

NAME:	**Microprocessor requirements**
ORIGIN:	NEDLIB
DEFINITION:	Description of microprocessor specifications necessary to operate the Content Data Object's software environment
PURPOSE:	Ensure that users' obtain sufficient processing power to run the software necessary to render/display the Content Data Object
EXAMPLE:	Could be a general specification (e.g., 333 Mz), or a particular microprocessor (e.g., Intel Pentium II 333 Mz)

NAME:	**Memory requirements**
ORIGIN:	WG
DEFINITION:	Description of memory resources necessary to operate the Content Data Object's software environment

PURPOSE:	Ensure that users' obtain sufficient memory resources to run the software necessary to render/display the Content Data Object
EXAMPLE:	128 MB RAM

NAME:	**Documentation**
ORIGIN:	WG
DEFINITION:	Supporting documentation necessary/useful for operation/use of the Computational Resources
SUB-ELEMENT:	Location: location of documentation (e.g., URL)
PURPOSE:	Link the Computational Resources metadata to supporting documentation useful for operation
EXAMPLE:	Glossary, users' manual, etc.

Storage:

NAME:	**Storage information**
ORIGIN:	NLA
DEFINITION:	Description of any permanent storage resources necessary for the operation of the software environment and/or rendering of the Content Data Object
PURPOSE:	Ensure that users' obtain sufficient storage resources to render/display the Content Data Object
EXAMPLE:	User must have 33 MB of hard disk space free in order to install/run the software environment

NAME:	**Documentation**
ORIGIN:	WG
DEFINITION:	Supporting documentation necessary/useful for operation/use of Storage resources
SUB-ELEMENT:	Location: location of documentation (e.g., URL)
PURPOSE:	Link the Storage metadata to supporting documentation useful for operation
EXAMPLE:	Glossary, users' manual, etc.

Peripherals:

NAME:	**Peripheral requirements**
ORIGIN:	NEDLIB
DEFINITION:	Description of additional equipment needed to render/display the Content Data Object
PURPOSE:	Describe the complete set of physical resources necessary to access the Object's content
EXAMPLE:	Sound card, speakers, a monitor with a particular resolution, CD-ROM drive, etc.

NAME:	**Documentation**

ORIGIN:	WG
DEFINITION:	Supporting documentation necessary/useful for operation/use of Peripherals
SUB-ELEMENT:	Location: location of documentation (e.g., URL)
PURPOSE:	Link the Peripherals metadata to supporting documentation useful for operation
EXAMPLE:	Glossary, users' manual, etc.

Hardware Environment as a Whole:

NAME:	**Location**
ORIGIN:	WG
DEFINITION:	Location of the physical devices needed to render the Content Data Object
PURPOSE:	Link Content Data Object to compatible Hardware Environment
EXAMPLE:	Description of where the required Hardware Environment can be obtained. This may take the form of anything ranging from contact information for a "technology museum" to the location of emulation programs (perhaps maintained by the archive itself)

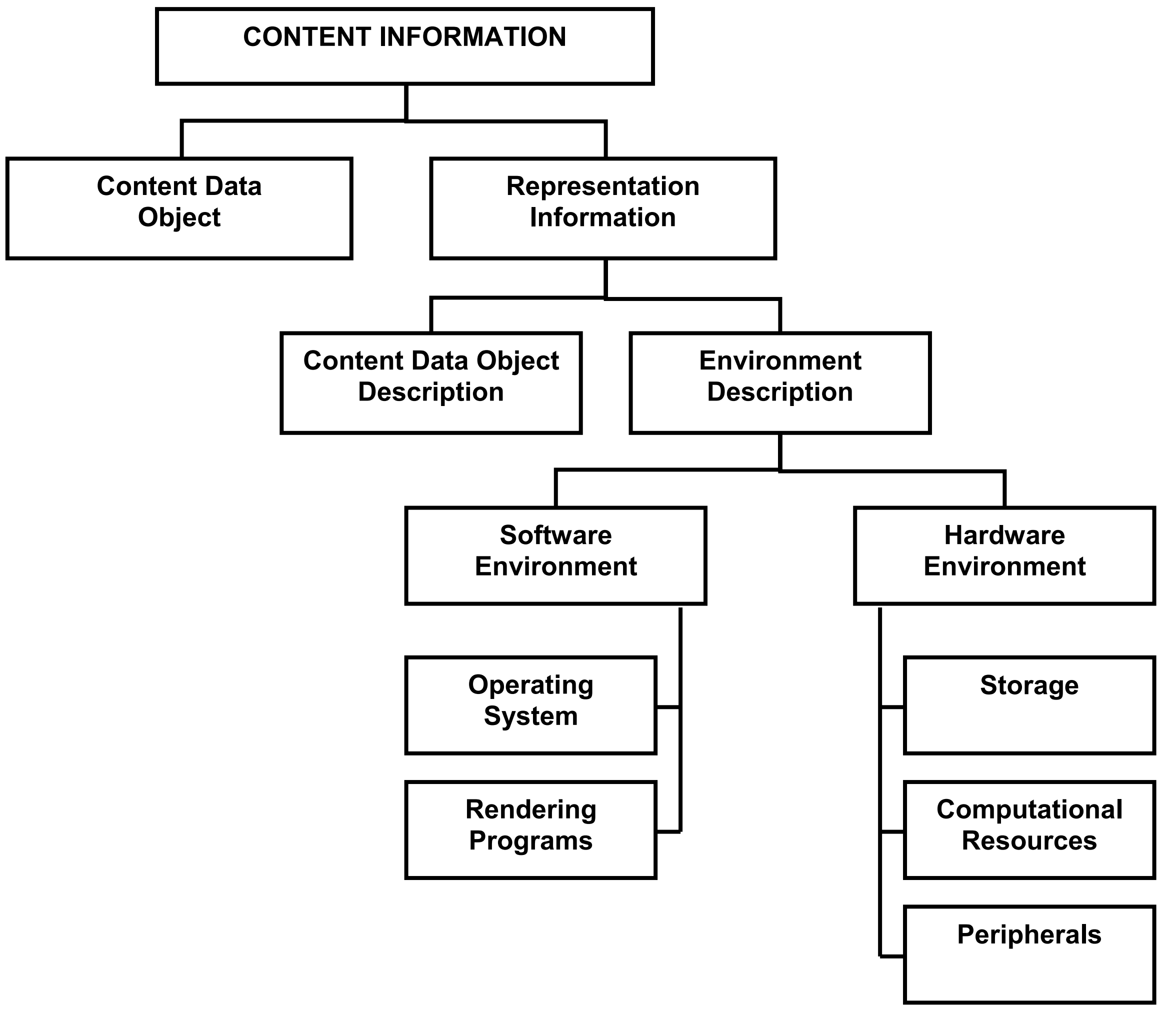
Notes:

1) As in the case of the Content Data Object's Software Environment, the Hardware Environment could be repeatable. This may take the form of an enumeration of all possible environments, or those which are supported by the archive itself (e.g., through a set of emulators). Alternatively, the archive may choose to describe only a minimum or recommended hardware environment. It is recommended that if the metadata is intended to describe a minimum or recommended environment, this information should be recorded in another metadata element (e.g., Environment Type, with values "Minimum" or "Recommended").
2) The elements for the various aspects of the Hardware Environment may need to be broken down further to record more specific information, such as manufacturer, version, etc.
3) More work needs to be done to refine the Hardware Environment elements to accommodate emulation preservation strategies (assuming emulation takes place at the hardware level). It is likely, however, that even in the case of emulation, a Hardware Environment compatible with the emulator itself will have to be described. It is expected that current research examining the issue of emulation will contribute toward resolving this issue.

IV.3. Discussion

A complete diagram of the structural components of Content Information, integrating the diagrams illustrated in Figures 2 through 5, is given below:

Figure 6: Complete Structure of Content Information Package



Notes:

1) The OAIS reference model notes that if Representation Information is itself in digital form, additional Representation Information may be needed to understand the bit sequence of the Representation Information itself. This recursive process may in theory continue until the chain ends in a physical document, resulting in a Representation Network for the original Content Data Object. The implementation of Representation Information discussed here should be adaptable to Representation Networks, by simply interpreting each successive iteration of

Representation Information as a Content Data Object in its own right, with its own associated Representation Information.

2) An Archival Information Class is a group of Archival Information Packages sharing common features or characteristics. It has been suggested that in cases where a digital repository maintains a number of Content Data Objects of a similar nature (for example, a collection of PDF files), it may be useful to record metadata that applies broadly across the entire class (e.g., the Software and Hardware Environment metadata) in a separate AIP to which the metadata of each member of the Archival Information Class would point. This would alleviate the problem of repetitive metadata within the archival system. For more information, please see the document at http://lcweb.loc.gov/rr/mopic/avprot/AIP-Study_v19.pdf.
3) Implementation of preservation metadata must address the issues of granularity (i.e., the level at which the metadata applies: collection, object, sub-object (file)) and whether or not a particular element is repeatable or mandatory. These issues have been deferred to a later stage of the Working Group's activity, in which issues relating to the practical application of preservation metadata will be addressed.
4) The preservation metadata framework described in this document makes the implicit assumption that certain aspects of a digital object's environment will remain static for the foreseeable future – for example, the fact that the Software Environment is composed of Rendering Programs and an Operating System, or that microprocessors supply the computational power for the Hardware Environment. Clearly, if these assumptions are overturned by new advances in digital technology, the preservation metadata framework and elements discussed here will become obsolete. However, the focus of the Working Group is to provide practical recommendations for organizations intending to develop or that are in the process of developing digital repositories. In this sense, developing preservation metadata broad enough to anticipate future changes in digital technology is beyond the scope of the Working Group.
5) It is difficult to overstate the importance of type-specific metadata for various classes of digital objects. To address this issue and provide useful guidance, the Working Group will track ongoing efforts to build standards or consensus on type-specific metadata for particular types of digital objects. Please consult the Working Group Web site for more information.

V. A Recommendation for Preservation Description Information

The OAIS reference model defines Preservation Description Information as "information that is necessary to adequately preserve the particular Content Information with which it is associated. It is specifically focused on describing the past and present states of the Content Information, ensuring it is uniquely identifiable, and ensuring that it has not been unknowingly altered."

Preservation Description Information constitutes the second major component of preservation metadata. The first – Content Information – includes both the Content Data Object that is the focus of preservation, and the information necessary to render and understand the object's content, as it currently exists in the archive. Preservation Description Information, on the other hand, focuses on information that is necessary to manage the perpetuation of the object and its content over time. Content Information records the static properties of an archived object – i.e., those associated with the particular instance or version of the object that is currently archived. Preservation Description Information, while also encompassing static properties, emphasizes the temporal aspects of the object, extending from its creation, to its ingest into the digital archive, to its retention in the archival store. Taken together, Content Information and Preservation Description Information support the two major functional components of a digital archive: access and preservation, respectively.

It should be noted that in practice, the distinction between Content Information and Preservation Description Information is not as sharp as their conceptual definitions suggest. In particular, metadata assigned to Preservation Description Information can be used to render and understand the content of a digital object, and metadata assigned to Content Information can be used as input to, or be generated as output by, an archive's preservation processes. Conceptually, however, the broad categorization of preservation metadata as fulfilling either one role or the other is a useful way to consider the information requirements of a digital archive.

The OAIS information model divides Preservation Description Information into four categories:

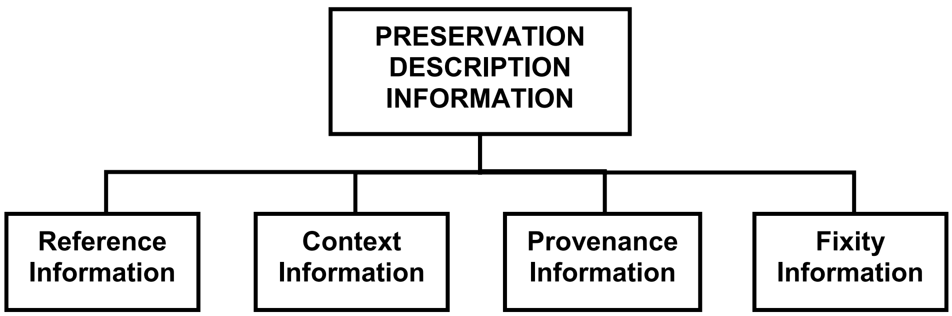
> **Reference:** describes identification systems, and the mechanisms for providing assigned identifiers, used to unambiguously identify the Content Information both internally and externally to the archive in which it resides.
>
> **Context:** documents relationships of the Content Information with its environment, including the reasons for its creation and relationships to other Content Information objects.
>
> **Provenance:** documents the history of the Content Information, including its origin, changes to the object or its content over time, and its chain of custody.

Fixity: provides the Data Integrity checks or Validation/Verification keys used to ensure that the particular Content Information object has not been altered in an undocumented manner.

To summarize, Preservation Description Information records the *identity, relationships, history, and integrity* of the archived Content Data Object.

The OAIS reference model divides Preservation Description Information into four separate categories: Reference, Context, Provenance, and Fixity. The structure of Preservation Description Information is illustrated in Figure 7 below:

Figure 7: Preservation Description Information



Taken together, the components of Preservation Description Information address the informational requirements of the preservation processes implemented by the archival system. This includes information that is utilized as input for these processes, as well as information that records the output of these processes.

Each of the components of Preservation Description Information is discussed in detail below.

V.1. Reference Information

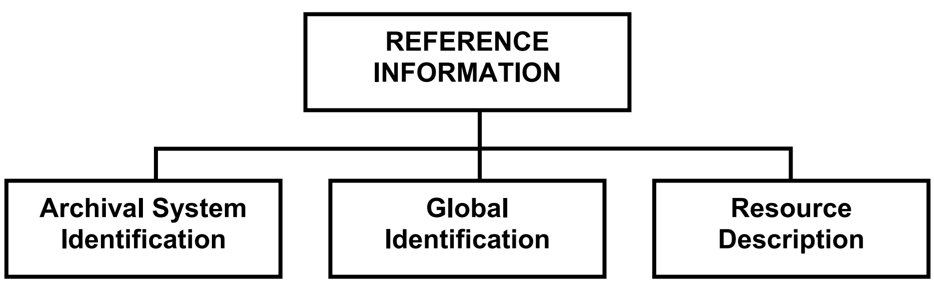
The OAIS defines Reference Information as information that "identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides those identifiers that allow outside systems to refer, unambiguously, to this particular Content Information." This definition suggests two primary functions for Reference Information: 1) identifying Content Information *locally* (i.e., within the archival system in which it resides), and 2) identifying Content Information *globally* (i.e., to systems external to the archive).

This dual functionality reflects traditional cataloging practice, which takes into account identification mechanisms of varying degrees of scope. For example, a book can be identified strictly within the context of the collection in which it resides via its accession number or call number; it can also be unambiguously identified in the context of all publications via its ISBN. An intermediate level of identification might be the shared cataloging environment: e.g., in the form of an OCLC number.

Several prominent digital archiving initiatives have extended the definition of Reference Information to include description as well as identification. The CEDARS project, for example, notes in the OAIS-based metadata specification it developed for its own archive that “Reference Information identifies and describes the Content sufficiently and so holds most of the data which need to be distributed for customer resource discovery.” The OAIS itself notes that in the context of digital libraries, Reference Information may also include bibliographic description. It should be noted, however, that the metadata included in Reference Information may, in some implementations, be much less than what is required for resource discovery. The richness of description associated with Reference Information is likely to be a function of the relationship between the archival repository and the producers of the archived Content Data Objects, as well as the nature of the services that support discovery and use of the archived Objects (for more discussion, see Note 7 below).

Taking this last function into account, Reference Information may be usefully divided into three types: Archival System Identification (local), Global Identification, and Resource Description. This structure is illustrated in Figure 8 below:

Figure 8: Reference Information



The following metadata elements address the informational requirements for Reference Information:

NAME:	**Archival system identification**
ORIGIN:	WG
DEFINITION:	Uniquely identifies the Content Data Object and its associated metadata (Archival Information Package) within the archival system in which it is stored
PURPOSE:	Facilitate identification and management of AIP within the archival system
EXAMPLE:	A system-generated ID number assigned at the time the AIP is created and ingested into the archive

Sub-elements:

NAME:	**Value**
ORIGIN:	NEDLIB
DEFINITION:	Value of the Archival System Identification used to identify the AIP
PURPOSE:	Uniquely identify AIP within the Archival system (OAIS)
EXAMPLE:	000000000001, 000000000002, 000000000003, …

NAME:	**Construction method**
ORIGIN:	NEDLIB
DEFINITION:	Description of the means by which the Archival System Identification is created and assigned
PURPOSE:	Ensures understanding and consistency of the process by which the Identifier is created and assigned
EXAMPLE:	Archival Information Package is assigned a 32-bit system-generated identification number, assigned in order of ingest into the Archive.

NAME:	**Responsible agency**
ORIGIN:	NEDLIB
DEFINITION:	Entity responsible for assigning and maintaining the Archival System Identification
PURPOSE:	Delineate and document responsibility for creating and assigning the Archival System Identification
EXAMPLE:	Administration (functional entity within an OAIS)

NAME:	**Global identification**
ORIGIN:	WG
DEFINITION:	Uniquely identifies the Content Data Object and its associated metadata (Archival Information Package) to systems external to the Archive in which it is stored
PURPOSE:	Facilitate interoperability of distributed archival systems
EXAMPLE:	ISBN, persistent URL

Sub-elements:

NAME:	**Value**
ORIGIN:	NEDLIB
DEFINITION:	Value of the Global Identification used to identify the AIP
PURPOSE:	Uniquely identify AIP to external systems
EXAMPLE:	PURL: http://purl.oclc.org/file.pdf

NAME:	**Construction method**
ORIGIN:	NEDLIB

DEFINITION: Description of the means by which the Global Identification is created and assigned
PURPOSE: Ensures understanding and consistency of the process by which the Identifier is created and assigned
EXAMPLE: Archival Information Package is registered with the OCLC PURL service upon ingest into the Archive.

NAME: **Responsible agency**
ORIGIN: NEDLIB
DEFINITION: Entity responsible for assigning and maintaining the Global Identification
PURPOSE: Delineate and document responsibility for creating and assigning Global Identification
EXAMPLE: OCLC PURL Service

NAME: **Resource description**
ORIGIN: CEDARS
DEFINITION: Includes information for resource discovery which is extracted from existing metadata sources (if available), or is created by the archive itself to support its access functions.
PURPOSE: Supplement Archival System and Global Identification with sufficient description of the AIP to support resource discovery within the OAIS and any allied external systems
EXAMPLE: Fifteen Dublin Core metadata elements for resource discovery

Sub-elements:
NAME: **Existing metadata**
ORIGIN: CEDARS
DEFINITION: Any metadata scheme which has been instantiated for the Content Data Object. This information may accompany the Object on ingest or may be discovered later.
PURPOSE: Leverage existing metadata associated with the Content Data Object
EXAMPLE: MARC bibliographic record; Dublin Core record

Sub-elements:
NAME: **Existing records**
ORIGIN: CEDARS
DEFINITION: A single instantiation (record) of a particular metadata scheme
PURPOSE: Identify existing metadata records associated with the Content Data Object
EXAMPLE: Bibliographic record in WorldCat and/or CORC

Notes:

1) It is possible that in some systems one scheme will be used to identify the AIP uniquely both locally and globally. However, a number of obstacles surround this approach, prominent among them being that global identification schemes often apply only to certain media types (e.g., ISBNs apply to books, but not to Web sites or digital images), while some media types have no global identification scheme associated with them at all. This implies that a multimedia archival collection could reference a conglomeration of different global schemes. Attempting to utilize these multiple identification schemes within the archive's local system administration functions would likely prove difficult to manage.
2) Both the Global Identification and the Archival System Identification should be repeatable. Repeatability of the Global Identification accommodates the fact that multiple global identification schemes can exist for the same resource. The repeatability of the Archival System Identification is useful in situations such as the following: one archive merges with another, and it is prudent to retain the local identification from the old, pre-merge systems, along with the identification associated with the new, post-merge system. However, if old identifiers are retained, there should be a clear demarcation between these and the identifiers currently in use.
3) The Construction Method sub-element could take a number of forms: a prose description, as in the examples above; a pointer to an external agency responsible for maintaining and administering the identification system; a pointer to supporting documentation describing a software program that is currently being used to create the identifier, etc. If possible, it is probably advantageous to use a reference, or indirection, to populate this sub-element: for example, place each construction method into an Archival Information Package, whose Archival System Identifier is stored in the Construction Method sub-element.
4) The Responsible Agency sub-element might take the form of a pointer (e.g., a PURL) to the Web site of the entity responsible for assigning the Identifier. Alternatively, some form of organizational code could also be utilized, such as the Library of Congress Organization Codes.
5) The Existing Metadata sub-element might be populated by a pointer (e.g., a URL) to the metadata specification or its sponsoring agency. The Existing Record sub-element might contain a pointer to the record itself (e.g., an OCLC number). It should be noted, however, that any system of pointers is likely to be unstable over the long-term. If circumstances permit, it may be more appropriate for the archive to package the metadata record into an AIP in its own right, which would then be ingested into the archive.
6) In some circumstances, a third type of Identification might be needed, which is more parochial than a Global Identification, yet more extensive than an Archival System Identification. For example, suppose that an archive is integrated with other library systems: i.e., cataloging or inter-library loan. These systems, which might be united under a single overarching organization, are nevertheless external to the archive itself, yet lacking the ubiquity necessary to meet the requirements of a Global Identification. In this case, the identifiers associated with these systems may need to be recorded in a separate element.

7) The task of preserving an Object is likely to include preserving some form of resource description. For a repository that primarily serves an archival purpose (i.e., a "dark" archive), the policy might be to include as full a resource description as was available at the time of ingest, but not to modify this description over time, even if descriptions in other external sources are updated. For a repository that serves both an archival purpose and as the primary source for access, the decision on whether to utilize full resource descriptions, "minimal" descriptions, pointers to descriptions maintained elsewhere, or even to incorporate descriptions as Objects in their own right will depend on many factors. These factors include, but are not limited to, whether the description is dynamic, patterns of deposit and use, and whether the repository incorporates its own access service through a locally controlled catalog. The specification for Reference Information offered here is intended to allow for flexibility.

V.2. Context Information

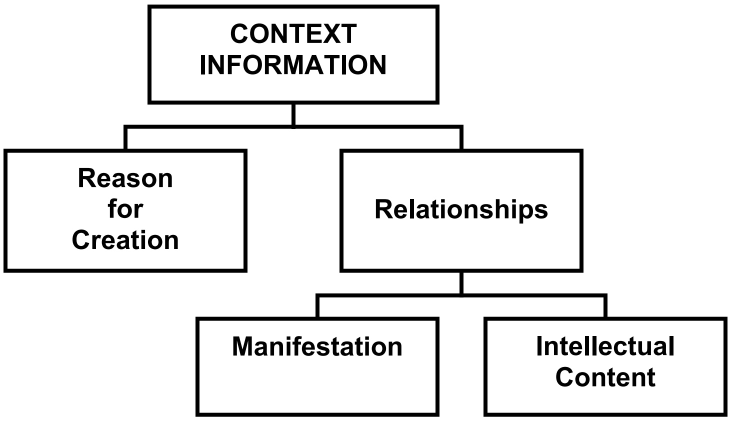
Context Information is defined by the OAIS reference model as "information [that] documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects existing elsewhere."

This definition suggests two separate areas encompassed by Context Information. First, the Content Data Object toward which preservation is directed must be placed in context of the motivation or rationale for its creation: for example, a scanned image of a paper document may have been created to facilitate access; a digitized audio file may have been created to serve as the authoritative record of an event, and the source from which derivative records – e.g., printed transcripts – are created.

Second, Context Information documents significant relationships among the preserved Object and other Content Data Objects. These relationships may take a variety of forms, but can be collected into two broad categories: 1) other manifestations of the Content Data Object, and 2) other Content Data Objects whose intellectual content is related to that of the preserved Object. The first category would include versions of the Object in alternate software formats: for example, HTML, PDF, and Microsoft Word versions of the same document. It would also include different versions of the Object in the same software format: for example, Microsoft Word 6.0, 97, and 2000 versions of the same document. Examples of relationships included in the second category would be Content Data Objects which, together with the preserved Object, form a well-defined series or collection, or whose intellectual content describes, elaborates on, critiques, etc., that of the preserved Object. It would also include a set of Objects whose content, in aggregate, forms a single complex Object at some higher level of abstraction: for example, a set of PDF documents, each representing a chapter of a book.

The major components of Context Information are illustrated in Figure 9 below:

Figure 9: Context Information



The following metadata elements serve to document the informational requirements of Context Information:

NAME:	**Reason for creation**
ORIGIN:	CEDARS
DEFINITION:	Documents information about why a Content Data Object was created.
PURPOSE:	Establish context for the rationale or purpose of creating the Content Data Object.
EXAMPLE:	A TIFF file was created to serve as a digital surrogate for a rare, fragile paper document, in order to facilitate access and protect the original resource.

NAME:	**Relationships**
ORIGIN:	NLA
DEFINITION:	Records significant relationships between this Object and other Content Data Objects.
PURPOSE:	To establish linkages associated with this Object which are important for managing the preservation process.
EXAMPLE:	Relationships to other manifestations; relationships to Objects within the same collection, etc.

Sub-elements:

NAME:	**Manifestation**
ORIGIN:	WG (based on NLA “Relationships” element)
DEFINITION:	Documents relationships between this Object and other manifestations of this same Object.
PURPOSE:	Essential for maintaining a change history for the Object (i.e., recording outcome of a migration process), or relating alternative versions of the current Object.

EXAMPLE: Links to versions of the Object in HTML and PDF; links to versions of the Object in earlier versions of Microsoft Word.

Sub-elements:

NAME: **Relationship type**
ORIGIN: WG
DEFINITION: Type of relationship between the archived Object and another associated Object
PURPOSE: Understand the relationship between related Objects
EXAMPLE: Manifestation in HTML; Manifestation in PDF …

NAME: **Identification**
ORIGIN: WG
DEFINITION: Identifies the related Object
PURPOSE: Link the Object with related Object
EXAMPLE: Archival System Identification; Global System Identification; link to a descriptive record for the related Object.

NAME: **Intellectual content**
ORIGIN: WG
DEFINITION: Documents relationships between the intellectual content of this Object and other Objects.
PURPOSE: Identify groups of related Objects (i.e., collections) that exist within the Archive.
EXAMPLE: A sequence of Objects representing a serial; a collection of Objects representing digitized images of an art collection; a set of individual Objects (HTML, GIF files, etc.) which, in aggregate, form a Web page.

Sub-elements:

NAME: **Relationship type**
ORIGIN: WG
DEFINITION: Type of relationship between the archived Object and another associated Object
PURPOSE: Understand the relationship between related Objects
EXAMPLE: Web page; Collection; Serial …

NAME: **Identification**
ORIGIN: WG
DEFINITION: Identifies the related Object
PURPOSE: Link the Object with related Object
EXAMPLE: Archival System Identification; Global System Identification; link to a descriptive record for the related Object.

Notes:

1) Defining the necessary context for an archived Content Data Object is a subjective exercise, and likely will depend on the particular needs of the Archive and/or its Designated Community.
2) It is important to note that Context Information is directed at informational requirements associated with *managing the preservation process*, not those aimed at facilitating understanding and interpretation of the Content Data Object's intellectual content. The latter is addressed by metadata elements within the Object's Representation Information. For example, in the case of an ASCII file containing the results of a chemistry experiment, the element *Reasons for Creation* should not be populated with a statement such as "to understand the properties of inert gases"; rather, it should contain information such as "to serve as the authoritative or Master source of the data associated with this experiment".
3) The two types of relationships detailed above – Manifestation and Intellectual Content – may be distinguished as follows. Manifestation groups together Content Data Objects that contain the same intellectual content, but present it in alternative formats. Intellectual Content groups together Objects with different, yet related intellectual content (e.g., related by subject, theme, etc.).
4) The relationship metadata elements should be repeatable.
5) One alternative to linking each related Object to one another directly would be to instead link to some form of index record, which would detail the network of relationships associated with each Object.
6) Documentation of the relationships described above can be a critical ingredient to effective management of the preservation process. For example, the relationship among a group of Content Data Objects in a collection may be important because all Objects in the collection might need to be migrated simultaneously, or disseminated as a unit.

V.3. Provenance Information

According to the OAIS reference model, Provenance Information "documents the history of the Content Information. This 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."

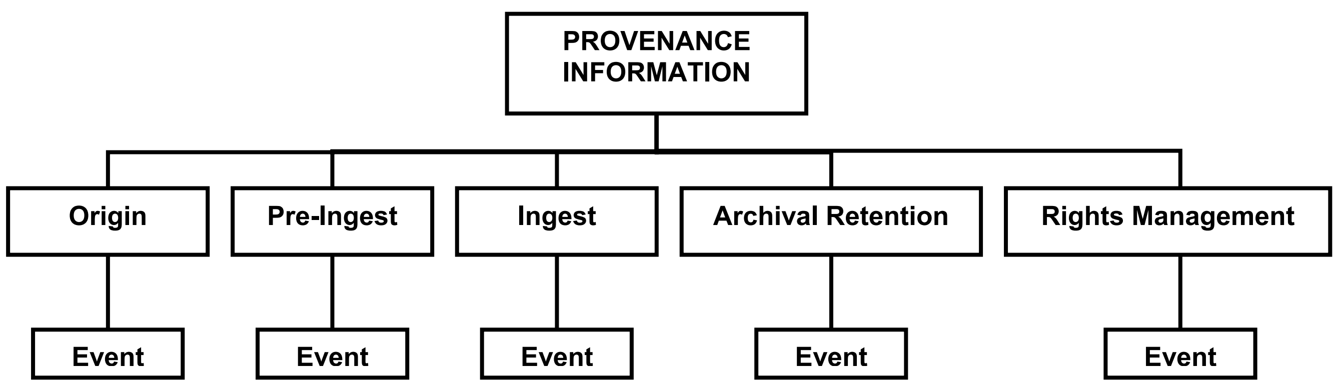
As the definition implies, Provenance Information primarily addresses the temporal aspect of the archived Content Data Object, beginning with the Object's creation and extending to its current status as it exists in the Archive. Rather than documenting the static features of the Object as it currently resides in the Archive, it describes the Object as a dynamic entity. From this perspective, the current state of the archived Content Data Object can be viewed as the culmination of an evolutionary process, of which the period of archival retention may only be a small part. It is this process, and the results or outcomes stemming from this process which impact the Content Data Object, that Provenance Information documents and describes.

In addition to recording the "chronology" of the archived Content Data Object, Provenance Information also can be considered "event-based" metadata. More specifically, the evolutionary process associated with the Object is driven by the occurrence of important "events", such as the Object's creation, a transfer in ownership, its ingest into the Archive, or the migration of the Object from one format to another. Recording the particulars of these events, and their impact on the Content Data Object, is another key function of Provenance Information.

In sum, the temporal, event-based history of the archived Content Data Object is the subject of Provenance Information. In the context of digital preservation, the motivation for recording this information extends from a need or requirement to document the procedures and outcomes of the Archive's preservation processes, and to place them in the context of the Object's complete life cycle, including its history prior to inclusion in the archive.

Figure 10 illustrates the structure of Provenance Information:

Figure 10: Provenance Information



The following metadata elements address the informational requirements of Provenance Information:

NAME: **Origin**
ORIGIN: WG
DEFINITION: Description of the process by which the Content Data Object was created.
PURPOSE: Document the circumstances surrounding the creation of the Content Data Object
EXAMPLE: The Content Data Object was created by scanning a paper document at 600 dpi in TIFF format, and storing it on CD-ROM.

NAME: **Pre-ingest**

ORIGIN:	WG
DEFINITION:	Description of the history of the Content Data Object, in terms of maintenance, changes in content, custody, etc., from its creation to its submission to the Archive.
PURPOSE:	Document known changes to the format, content, ownership, and other dynamic aspects of the Content Data Object relative to the time of creation, but prior to ingest into the Archive.
EXAMPLE:	Chain of custody, changes to content, etc.

NAME:	**Ingest**
ORIGIN:	WG
DEFINITION:	Description of the process by which the Content Data Object is ingested (i.e., deposited) into the Archive.
PURPOSE:	Document the procedures and describe the outcomes of processes carried out to prepare the Object for inclusion in the Archive.
EXAMPLE:	Object migrated to Archive's standard storage format; complex Object broken down into its component parts for storage as separate Content Data Objects; AIP(s) assembled.

NAME:	**Archival retention**
ORIGIN:	WG
DEFINITION:	Description of the maintenance, changes in content, management, etc., of the Content Data Object during its retention in the archival store.
PURPOSE:	Document the procedures and describe the outcomes of processes carried out for the purpose of preserving and maintaining access to the Object while it is retained in the archival store.
EXAMPLE:	Migration history (relative to Object manifestation originally ingested into archive), media refreshment history, digital rights management revisions, etc.

NAME:	**Rights management**
ORIGIN:	WG
DEFINITION:	Specification of the legal uses of the Content Data Object
PURPOSE:	Document the archive's scope to preserve and disseminate the Content Data Object
EXAMPLE:	Access permissions; legal deposit responsibilities

The above elements delineate the major phases or aspects of the Content Data Object's chronology, or life cycle. Within each of these categories, however, Provenance Information takes the form of a collection or series of events which impact one or more aspects of the Content Data Object: for example, its content, its presentation, or its associated access privileges. Therefore, metadata representing Provenance Information must record the details of these events, in order to document their occurrence and outcome, facilitate effective management of the Archive's preservation processes, and document the reliable custody of the material for ensuring the integrity and authenticity of the Content Data Object.

The following metadata describe a generic event associated with Provenance Information:

NAME: **Event**
ORIGIN: WG (based on NLA Process element)
DEFINITION: An event which impacts one or more of the aspects of a Content Data Object: content, format, rights management, etc.
PURPOSE: Describe the event and its outcome.
EXAMPLE: See below.

Sub-elements:

NAME: **Designation**
ORIGIN: WG
DEFINITION: Name of the Event.
PURPOSE: Identify the Event being described.
EXAMPLE: Change in Custody; Migration; Media Refreshment …

NAME: **Procedure**
ORIGIN: WG
DEFINITION: Details of the procedures constituting an occurrence of the Event.
PURPOSE: Provide sufficient information such that the procedural components of the Event may be understood.
EXAMPLE: Description of the timing and procedural steps associated with a format migration.

NAME: **Date**
ORIGIN: WG
DEFINITION: Date that the event occurred.
PURPOSE: Establish appropriate chronology for the occurrences of Events.
EXAMPLE: January 31, 2002 18:37:16

NAME: **Responsible agency**
ORIGIN: WG
DEFINITION: Entity responsible for the successful occurrence of the Event.
PURPOSE: Establish/delineate responsibility for ensuring the Event takes place, and its outcome is satisfactory.
EXAMPLE: Archival System Administration staff

NAME: **Outcome**
ORIGIN: WG

DEFINITION:	Description of the outcome of the Event's latest occurrence.
PURPOSE:	Document the impact of the Event on the relevant aspect(s) of the Content Data Object.
EXAMPLE:	Object successfully migrated from Microsoft Word 97 to PDF.

NAME:	**Note**
ORIGIN:	WG
DEFINITION:	Additional information relevant to the Event.
PURPOSE:	Record any additional information relevant to the Event not captured by the above elements.
EXAMPLE:	System-generated Event identification number

Key to adequately fulfilling the informational requirements of Provenance Information is the enumeration of the list of Events relevant to managing the preservation and access of the archived Content Data Object. It is likely that this list will vary from system to system, but the following examples are illustrative:

Origin:
Creation Process: description of the process by which the Content Data Object was created: e.g., digital scanning; digital recording of audio material; generated by a software application ("born-digital") …

Pre-Ingest:
Change in Custody: description of the circumstances surrounding a transfer of the intellectual property rights associated with the Object from one entity to another entity.

Ingest:
Ingest Process: description of the processes associated with preparing an Object submitted to the Archive for inclusion in the archival store.

Archival Retention:
Format Migration (or, Transformation): description of the procedural steps taken to translate the Content Data Object from one digital format into another (in the terminology of the OAIS, this is known as a *transformation*).

Rights Management:
Access/Use Privileges Specification: documents the archive's current scope to make the Content Data Object available to the archive's users, including any restrictions governing access and use of the Object by users.

Notes:

1) The first set of elements described above outline the major stages of a Content Data Object's chronology or life cycle. The second set describes the salient features of significant events that might occur within these stages. This dual

structure reflects the notion (discussed above) that Provenance Information captures the temporal, event-based (i.e., dynamic) characteristics of the Object and its content.

2) The metadata framework for Provenance Information specified above first identifies the key stages of a Content Data Object's lifecycle, and then records Events occurring within each stage. An alternative approach would be to eliminate the first layer of the framework (lifecycle stages), and instead include a sub-element under Event that indicates the lifecycle stage to which the Event is associated. These two approaches are conceptually equivalent; adoption of one approach or the other is likely to be an implementation issue.
3) It should be noted that the same Event can occur in multiple stages of the Content Data Object's life cycle: for example, a Format Migration could occur in the Pre-Ingest, Ingest, and Archival Retention stages.
4) The Event element must be repeatable: first, because multiple (different) Events must be described, and second, because multiple occurrences of the same Event are likely to be common – for example, a sequence of media refreshments.
5) Events which alter some aspect of the Content Data Object may also impact one or more metadata elements describing the Object. For example, an Event involving the compression of an Object might change the value of the Size element in the Object's Representation Information. From the perspective of a digital archive, this issue is only relevant for Events occurring in the Archival Retention stage. For more information, see the NEDLIB metadata element set, which addresses this issue specifically.
6) For Events which occur on multiple occasions, some sub-elements of the Event element might not need to be repeated: e.g., Designation, Procedure, Responsible Agency. Suitable pointers to the relevant information would be sufficient.
7) The metadata associated with an Event could be expanded to include more detailed description, including such information as equipment, software, specifications, etc. that are utilized in the occurrence of an Event. Such information could be broken out under the Procedure sub-element.
8) Note the relationship of the Outcome sub-element above to the Quirks element in Representation Information. Quirks records "any loss in functionality or change in the look and feel of the Content Data Object resulting from the preservation processes and procedures implemented by the archive." A change of this kind is clearly the outcome of an Event. Therefore, it is possible that Outcome should be broken down further into the various types of changes to the Object that could result from the occurrence of an Event. One of these would be the type of change captured by Quirks, which to avoid redundancy could be cross-linked with the Quirks element.
9) A possible addition to the sub-elements associated with the Event element is Next Occurrence, which would record the date the next occurrence of the Event is scheduled or anticipated.
10) An important issue associated with Provenance Information is the fact that the result of an Event could be the generation of a new Content Data Object which is itself preserved in the archival store. Examples of this may be found in the discussion above of the Manifestation element in Context Information.

Coordinating the network of metadata surrounding a Content Data Object and all of its related manifestations is a challenging task for any digital preservation initiative implementing a preservation metadata model. One way to view the issue is to think of Provenance Information metadata as "transcending" specific physical manifestations of a Content Data Object (e.g., in alternate formats), and instead applying at something like a "work" level to the entire collection of related Objects. To implement this interpretation, the metadata records of the manifestations relating to a particular Object could be linked together in such a way that the Provenance Information can be extracted from each one and compiled to represent the Object's provenance as a whole (e.g., for display purposes). Such an approach, however, would require a clear definition of a "work", perhaps something similar to what is proposed in the FRBR model.

11) Another example of an event occurring within the scope of the Ingest Process is the archival retention decision, or in other words, the decision to retain the Content Data Object in the archive. Metadata associated with this Event might include the specifics of the archival retention (e.g., how long is the object to be kept?), who is responsible for the archival retention decision, who is responsible for reviewing that decision, and who is responsible for initiating preservation actions.
12) The Rights Management metadata pertaining to preservation can be broken down into two broad areas: Preservation Action Permission (based on an element defined by NLA), which specifies the archive's scope to change, reproduce, or take other actions necessary to preserve the format or content of the Content Data Object; and Dissemination Permission, which specifies the archive's scope to make the Content Data Object available to the archive's users. Translation of rights management metadata into an events-based framework would consider the invocation of a particular rights management regime as an Event. A chain of rights management Events would begin with the policy regime prevailing at the time of Ingest; future events would describe any change in policy that occurs during the Object's retention in the archival store.

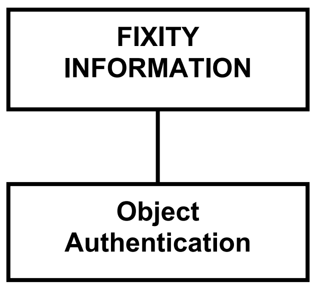
V.4. Fixity Information

According to the OAIS reference model, Fixity Information "provides the Data Integrity checks or Validation/Verification keys used to ensure that the particular Content Information object has not been altered in an undocumented manner. Fixity Information includes special encoding and error detection schemes that are specific to instances of Content Objects … The Fixity Information may specify minimum quality of service requirements for these mechanisms."

The purpose of Fixity Information is to ensure that the Content Data Object currently residing in the archival store is indeed the Object described by its associated metadata. The mutability of content in digital format makes this a critical issue for digital archives. Fortunately, a variety of techniques are available for authenticating Content Data Objects – digital signatures, watermarks, check sums, etc. Metadata requirements associated with authentication must address the specifics of the technique used, the frequency with which it is applied, and the result or value that represents the Content Data Object's current authenticated state.

Figure 11 illustrates the structure of Fixity Information:

Figure 11: Fixity Information



The following metadata elements address the informational requirements of Fixity Information:

NAME: **Object authentication**
ORIGIN: OCLC
DEFINITION: Provide sufficient information to meet the Archive's minimum requirements for authenticating the Content Data Object and its content.
PURPOSE: Ensure that no undocumented change has been made to the archived Content Data Object.
EXAMPLE: Digital signature, watermark, check sum … see sub-element examples below …

Sub-elements:

NAME: **Authentication type**
ORIGIN: OCLC
DEFINITION: The technique used to authenticate the Content Data Object.
PURPOSE: Identify and describe the authentication method applied to the Content Data Object.
EXAMPLE: Digital signature consisting of a 128-bit hash computed using the MD5 one-way hash function, encrypted with a private key.

NAME: **Authentication procedure**
ORIGIN: WG

DEFINITION:	Procedural steps to implement Authentication Type, including pointers to any supporting documentation, software, etc.
PURPOSE:	Provide sufficient information, tools, etc. needed to carry out Authentication Type applied to the Content Data Object.
EXAMPLE:	Pointer to software capable of computing an MD5 hash (which exists as another Content Data Object stored in the archive)

NAME:	**Authentication date**
ORIGIN:	OCLC
DEFINITION:	Date of most recent archival use of this Authentication Type.
PURPOSE:	Establish temporal benchmark against which later manifestations/versions/copies of the Object can be compared.
EXAMPLE:	January 31, 2002 18:37:16

NAME:	**Authentication result**
ORIGIN:	OCLC
DEFINITION:	Result of most recent archival use of this Authentication Type.
PURPOSE:	Provide basis for assessing authenticity of archived Content Data Object.
EXAMPLE:	(128-bit MD5 hash)

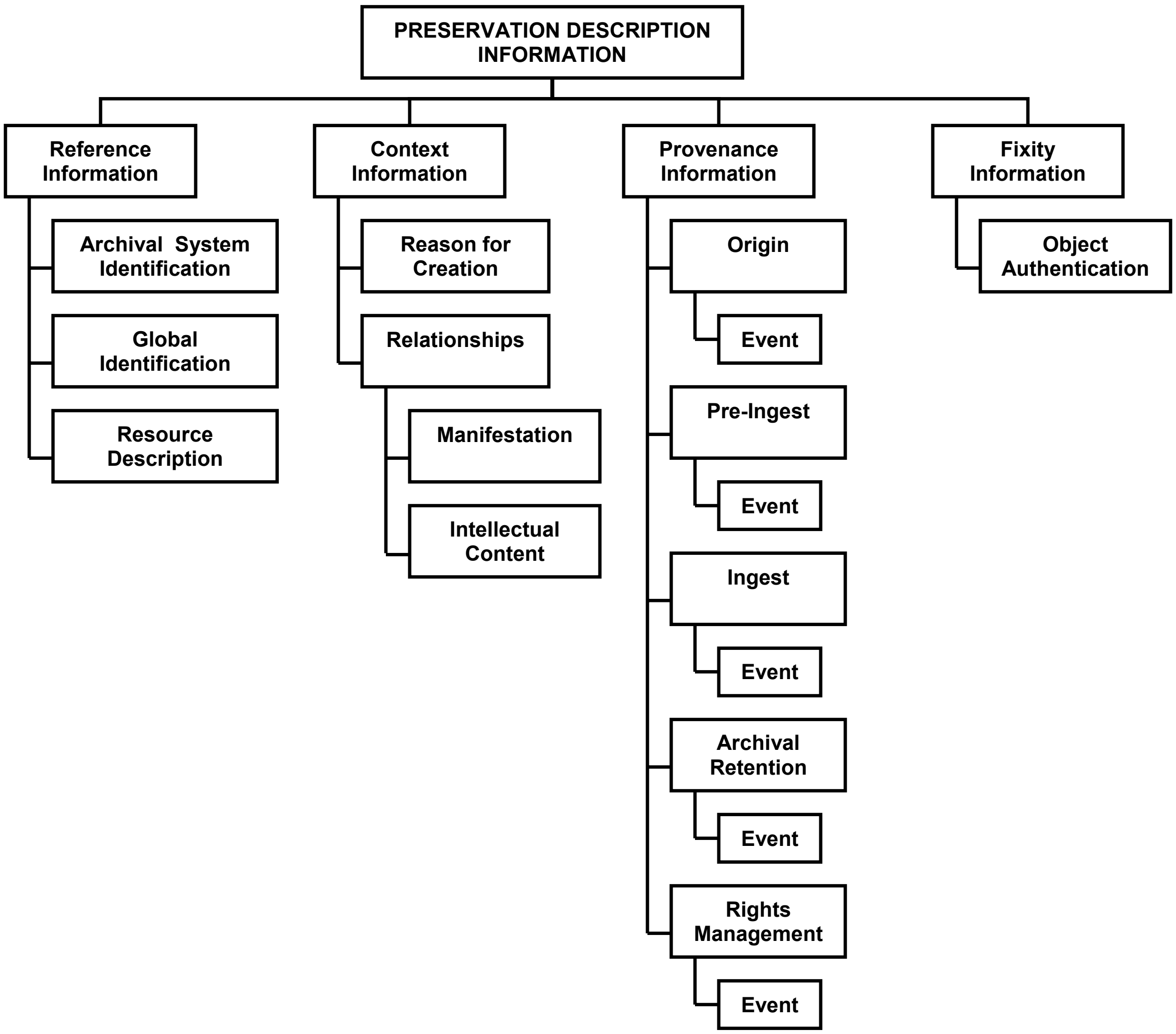
Notes:

1) Since many techniques exist for authenticating Content Data Objects, and an archive could use one or more of them, the Object Authentication element must be repeatable.
2) The Fixity Information metadata, as defined above, assumes that previous authentication metadata is superceded by the latest version. New authentication metadata would likely be required anytime an Event occurs that alters the bit stream of the Content Data Object or its associated metadata. While superceded authentication metadata may not be retained by the archive, a persuasive argument can be made that the sequence of authentication metadata should be kept in its entirety as evidence of a reliable/trusted history. Of course, if the pre-Event version of the Content Data Object is retained by the archive along with the post-Event version, then the old authenticity data will remain as part of the metadata associated with the earlier version.
3) The Authentication Type element could include a pointer to a relevant source describing the authentication technique – for example, a persistent URL to a copy of RFC1321, the official specification of the MD5 algorithm.

V.5. Summary

A complete diagram of the structural components of Preservation Description Information, integrating the diagrams in Figures 7 through 11, is given below:

Figure 12: Complete Structure of Preservation Description Information



Preservation metadata is intended to facilitate the management of, and decision-making associated with, an archive's preservation processes. In this sense, the rationale behind maintaining preservation metadata is wholly practical.

A few examples serve to illustrate this point in the context of Preservation Description Information. Documenting the relationships among collections of Content Data Objects (Context Information) facilitates any necessary synchronization of preservation processes – for example, it may be required that all Objects in a collection be migrated at the same time. Recording the chronology of events that impact the Object over time (Provenance Information) allows users to track changes in the Object's content or appearance, a potentially critical issue for items used in scholarly research. Uniquely identifying and ensuring the integrity of an archived Content Data Object (Reference and Fixity Information) facilitates the dissemination of Objects in the archive's collection to both the archive's Designated Community, and to other digital repositories.

Effective metadata is a necessary condition for effective digital preservation. Representation Information is essential for ensuring that an archived Content Data Object's content can be both rendered and understood by its intended audience. The elucidation and maintenance of Preservation Description Information, however, is the keystone to building an information infrastructure to support the processes associated with digital preservation.

VI. Conclusion

Metadata is an essential component of managed collections of information, regardless of whether the primary function of the collection is access or preservation. Just as collections set up primarily for access purposes rely on resource discovery metadata to perform their function, collections set up primarily for preservation purposes also require metadata to support their function. Therefore, as the imperative for digital preservation continues to increase for institutions managing information in digital form, there is a concomitant need to articulate the metadata requirements for preserving digital objects over the long term.

The OAIS information model provides a useful starting point for this task, but its utility is limited by the high-level nature of its structure and concepts. The OAIS information model offers a broad categorization of the types of information falling under the scope of preservation metadata; it falls short, however, of providing a decomposition of these information types into a list of metadata elements suitable for practical implementation. It is this need that the working group addressed in the course of its activities, the results of which are reported in this paper.

In conducting its review and extension of the OAIS information model, the working group has produced a comprehensive framework for preservation metadata, based on the structured description of a generic digital archiving system provided by the OAIS reference model. Since the OAIS is an emerging standard in the digital preservation community, it is anticipated that the working group's preservation metadata framework will be applicable across a broad range of digital preservation activities.

The metadata framework described in this paper can serve as a foundation for future work in the area of preservation metadata. Issues of particular importance include strategies and best practices for implementing preservation metadata in an archival system; assessing the degree of descriptive richness required by various types of digital preservation activities; developing algorithms for producing preservation metadata automatically; determining the scope for sharing preservation metadata in a cooperative environment; and moving beyond best practice towards an effort at formal standards building in this area.

***

The importance of maintaining the viability and accessibility of digital objects over the long term underscores the need to develop infrastructure in support of these objectives. Given the many shared challenges associated with digital preservation – preservation metadata among them – there is tremendous scope to address these challenges cooperatively. The OCLC/RLG Working Group on Preservation Metadata is but one example of how the collective expertise of a diverse and extended community of stakeholders can be marshaled to cooperatively advance the imperative of preserving digital objects over the long term.

APPENDIX: Quick View of Preservation Metadata Framework

The following is a quick view of the preservation metadata framework recommended by the OCLC/RLG Working Group on Preservation Metadata. Full descriptions for each section of the framework, as well as individual elements can be found in the main document. The high-level framework is shown through the use of an **ALL CAPS FONT**. Metadata elements are in **bold**; sub-elements are in regular font.

CONTENT INFORMATION

CONTENT DATA OBJECT

REPRESENTATION INFORMATION
- CONTENT DATA OBJECT DESCRIPTION
  - **Underlying abstract form description**
  - **Structural type**
  - **Technical infrastructure of complex object**
  - **File description**
  - **Installation requirements**
  - **Size**
  - **Access inhibitors**
  - **Access facilitators**
  - **Significant properties**
  - **Functionality**
  - **Description of rendered content**
  - **Quirks**
  - **Documentation**
- ENVIRONMENT DESCRIPTION
  - SOFTWARE ENVIRONMENT
    - RENDERING PROGRAMS
      - **Transformation process**
        - Transformer engine
          - Parameters
          - Input format
          - Output format
          - Location
          - Documentation
      - **Display/access application**
        - Input format
        - Output format
        - Location
        - Documentation
    - OPERATING SYSTEM
      - **OS name**
      - **OS version**
      - **Location**

Documentation

HARDWARE ENVIRONMENT

Location

COMPUTATIONAL RESOURCES

Microprocessor requirements

Memory requirements

Documentation

STORAGE

Storage information

Documentation

PERIPHERALS

Peripheral requirements

Documentation

PRESERVATION DESCRIPTION INFORMATION

REFERENCE INFORMATION

Archival system identification

Value

Construction method

Responsible agency

Global identification

Value

Construction method

Responsible agency

Resource description

Existing metadata

Existing records

CONTEXT INFORMATION

Reason for creation

Relationships

Manifestation

Relationship type

Identification

Intellectual content

Relationship type

Identification

PROVENANCE INFORMATION

Origin

Event

Designation

Procedure

Date

Responsible agency

Outcome
Note
Next occurrence

Pre-ingest

Event

Designation
Procedure
Date
Responsible agency
Outcome
Note
Next occurrence

Ingest

Event

Designation
Procedure
Date
Responsible agency
Outcome
Note
Next occurrence

Archival retention

Event

Designation
Procedure
Date
Responsible agency
Outcome
Note
Next occurrence

Rights management

Event

Designation
Procedure
Date
Responsible agency
Outcome
Note
Next occurrence

FIXITY INFORMATION

Object Authentication

Authentication type
Authentication procedure
Authentication date
Authentication result

References

CCSDS (2001) "Reference Model for an Open Archival Information System (OAIS)"
http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf

CEDARS (2000) "Metadata for Digital Preservation: The CEDARS Project Outline Specification"
http://www.leeds.ac.uk/cedars/MD-STR~5.pdf

National Library of Australia (1999) "Preservation Metadata for Digital Collections"
http://www.nla.gov.au/preserve/pmeta.html

NEDLIB (2000) "Metadata for Long Term Preservation"
http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf

NISO Committee: Technical Metadata for Still Images
http://www.niso.org/committees/committee_au.html

OCLC (2001) "Preservation Metadata Element Set – Definitions and Examples"
http://www.oclc.org/digitalpreservation/archiving/metadataset.pdf

OCLC/RLG Preservation Metadata Working Group (2001) "Preservation Metadata for Digital Objects: A Review of the State of the Art"
http://www.oclc.org/research/pmwg/presmeta_wp.pdf

OCLC/RLG Preservation Metadata Working Group Web Site
http://www.oclc.org/research/pmwg/