Preservation Metadata
for Digital Objects:

A Review of the State of the Art

A White Paper by the

OCLC/RLG Working Group on Preservation Metadata

January 31, 2001

OCLC/RLG Working Group on Preservation Metadata

In March 2000, OCLC and RLG announced their commitment to collaborate on identifying and supporting best practices for the long-term retention of digital objects. This collaboration would take the form of facilitating consensus-building activity among key stakeholders in the area of digital preservation: in other words, bringing together leading experts to review existing practices, share expertise, and identify best practices/common approaches where possible.

Because digital preservation is such a broad area, it was necessary to specify from the outset several areas of focus for the OCLC/RLG collaboration. One area involves the identification of the attributes of a digital archive for research repositories. The second area, and the one which is the subject of this document, is the use of metadata to support the digital preservation process. To address this second area, representatives from OCLC and RLG (the Planning Committee listed below) organized the OCLC/RLG Working Group on Preservation Metadata. The primary objective of the Working Group is to develop a comprehensive preservation metadata framework applicable to a broad range of digital preservation activity. The Working Group will also address ancillary issues such as the specification of preservation metadata elements and the evaluation of implementation strategies.

Planning Committee

Robin Dale	RLG
Brian Lavoie	OCLC
Ed O'Neill	OCLC
Meg Bellinger	OCLC
Nancy Elkington	RLG

Membership

Michael Alexander	New York Public Library
Michael Day	UKOLN
Julia Foster	British Library
Rebecca Guenther	Library of Congress
Bernard Hurley	University of California, Berkeley
Catherine Lupovici	Bibliotheque Nationale de France (NEDLIB)
Oya Rieger	Cornell University
Kelly Russell	University of Leeds (CEDARS)
Colin Webb	National Library of Australia
Robin Wendler	Harvard University
Titia van der Werf	Koninklijke Bibliotheek (NEDLIB)

I. Introduction

Metadata is routinely defined in accordance with its literal interpretation: "data about data". More usefully, Day (1998) observes that "metadata is commonly understood as an amplification of traditional bibliographic cataloguing practices in an electronic environment." In the context of digital information objects, metadata can be assigned to one of three functional categories (Wendler (1999)):

- **Descriptive:** facilitating resource discovery and identification
- **Administrative:** supporting resource management within a collection
- **Structural:** binding together the components of complex information objects

Of these three categories, descriptive metadata for electronic resources has received the most attention - most notably through the Dublin Core metadata initiative. However, increasing awareness of the challenges posed by digital preservation - the long-term retention of digital objects - has underscored metadata needs for digital objects beyond resource discovery.

Effective management of all but the crudest forms of digital preservation is likely to be facilitated by the creation, maintenance, and evolution of detailed metadata in support of the preservation process. For example, metadata could document the technical processes associated with preservation, specify rights management information, and establish the authenticity of digital content. It can record the chain of custody for a digital object, and uniquely identify it both internally and externally in relation to the archive in which it resides. In short, the creation and deployment of *preservation metadata* is likely to be a key component of most digital preservation strategies.

Several initiatives have addressed the issue of preservation metadata, with the result that a variety of approaches to its use have emerged. These approaches, developed independently of one another and designed largely to meet particular institutional or project requirements, nevertheless share several common themes. However, they also differ on a number of key points. Consequently, the body of current work in preservation metadata does not reflect a consensus on best practices for the use of metadata in support of digital preservation.

Initiatives such as the Dublin Core have demonstrated the value of consensus-building on metadata issues. In this spirit, the OCLC/RLG Working Group on Preservation Metadata was formed to initiate a consensus-building process in preservation metadata. Comprised of key stakeholders from a variety of institutional and geographic backgrounds, the Working Group is charged with developing a consensus on best practices and common approaches to the use of metadata to facilitate the long-term retention of digital objects. Using existing work as the foundation and starting point for its discussion, the Working Group will develop a comprehensive preservation metadata framework, describe a set of "essential" preservation metadata elements needed to support the framework, examine implementation issues associated with preservation metadata, and create testbed/pilot applications. The Working Group will conclude its work by releasing a set of recommendations reflecting their consensus on best practices and approaches to the use of metadata to support digital preservation strategies.

This white paper represents the first step of the Working Group's activity. Its scope includes the following topics:

- definition and illustration of preservation metadata for digital objects
- high-level requirements for a broadly applicable, comprehensive preservation metadata framework
- the Open Archival Information System (OAIS) reference model, a potential starting point for developing the preservation metadata framework
- review and synthesis of existing preservation metadata approaches
- identification of points of convergence/divergence among existing approaches

Collectively, these topics constitute both a summary of the "state of the art" in preservation metadata, and a starting point for the consensus-building process in which the members of the Working Group will participate.

II. Preservation Metadata for Digital Objects

II.1. Definition and Illustration

Perhaps more than any other media, digital information requires detailed metadata to ensure its preservation and accessibility for future generations. The metadata required to preserve a traditional information resource, such as a book, is fairly straightforward. A book is a physical, static item: its boundaries are clearly delineated, and its content cannot be altered over time. Furthermore, a book's access technology - human visual inspection - is unlikely to become obsolete. Thus, the primary metadata requirements for a book are likely to be descriptive in nature - i.e., facilitating its retrieval - rather than administrative or structural.

In a digital environment, ensuring that an information object "physically exists" over the long-term is analogous to preserving its bit stream on non-volatile digital storage media. This, however, is only one part of the preservation process. Digital objects are not immutable: therefore, the change history of the object must be maintained over time to ensure its authenticity and integrity. Access technologies for digital objects often become obsolete: therefore, it may be necessary to encapsulate with the object information about the relevant hardware environment, operating system, and rendering software. All of this information, as well as other forms of description and documentation, can be captured in the metadata associated with a digital object.

Preservation metadata is intended to support and facilitate the long-term retention of digital information. The National Library of Australia provides an overview of the types of information which may fall into this category. In particular, preservation metadata may be used to:

- store technical information supporting preservation decisions and actions
- document preservation actions taken, such as migration or emulation policies
- record the effects of preservation strategies
- ensure the authenticity of digital resources over time
- note information about collection management and the management of rights

The types of information enumerated above address two functional objectives: 1) providing preservation managers with sufficient knowledge to take appropriate actions in order to maintain a digital object's bit stream over the long-term, and 2) ensuring that the content of an archived object can be rendered and interpreted, in spite of future changes in access technologies.

An early effort to develop preservation metadata for digital objects was conducted by Research Libraries Group (RLG), which in May 1998 released a set of 16 recommended metadata elements considered essential for preserving a digital master file over the long-term. These elements include:

Date	Watermark
Transcriber	Resolution
Producer	Compression
Capture Device	Source
Capture Details	Color
Change History	Color Management
Validation Key	Color Bar/Grayscale Bar
Encryption	Control Targets

The RLG elements illustrate the relationship of preservation metadata to the three broad categories of metadata defined above: descriptive, administrative, and structural. Although preservation metadata can potentially straddle all three metadata types, its focus lies with the latter two. Casual examination of the RLG elements is sufficient to indicate that these elements, with one or two possible exceptions, would not be useful access points for resource discovery tools. Therefore, their utility as descriptive metadata is minimal. On the other hand, managing digital objects for the purpose of ensuring their long-term retention would be facilitated by the availability of information such as that represented by the RLG elements. For example, preservation strategies such as migration often have the consequence of altering the bit stream of the archived object; these alterations could be documented in the "Change History" element. The fact that a digital object's bit stream is malleable in this way suggests a concurrent need to validate that an archived object has not been corrupted or maliciously altered during the preservation cycle. A checksum or digital signature recorded in the "Validation Key" element would address this need. Changes in the validation key could then be compared to documented changes in the object's bit stream (recorded in the "Change History" element) to confirm an object's authenticity.

In the above example, preservation metadata fulfills an administrative function, in that it supports the management of digital objects in an archival setting. Preservation metadata could also serve a structural function, detailing the relationships between multiple objects residing in an archival repository. For example, several archived objects might collectively represent a single complex object. Metadata could bind the constituent components together. Alternatively, metadata could also link multiple versions of an archived object, differentiated perhaps on the basis of file format. As an object moves through successive migration cycles, new versions of the object will be produced. Metadata would bind them together into a single logical chain.

The RLG preservation metadata elements are intended to facilitate the preservation of digital image files - typically, digital copies of physical objects. In this sense, they are not technology-neutral, but instead assume that the preservation process includes a digitization step where a physical document is transformed into a digital object using some form of scanning technology. While these elements and their attendant assumptions are relevant to a substantial proportion of current digital preservation activity, they exclude the vast amount of information existing in a "born-digital" format, such as digital audio files and Web pages. Ideally, preservation metadata should extend to a wide range of digital object types and technological implementations.

II.2. Preservation Metadata Framework

A preservation metadata framework is an overview or description of the types of information (i.e., metadata) that should be associated with an archived digital object. The National Library of Australia description of preservation metadata mentioned above can be considered an informal metadata framework, in the sense that it enumerates the types of information which preservation metadata might record. A more systematic approach to developing a preservation metadata framework might involve the specification of a formal information model.

An important objective of the OCLC/RLG working group is to reach a consensus on a preservation metadata framework. In doing so, three high-level requirements will guide the discussion.

The preservation metadata framework should be:

- **Comprehensive:** preservation metadata should extend to all aspects of the digital preservation process. The purpose of preservation metadata is to support and facilitate digital preservation; therefore, a comprehensive preservation metadata framework should embody the informational requirements of a complete digital archiving system, ranging from the ingest of an object into the archive, to the provision of access services for the archive's users.
- **Structured:** the preservation metadata framework should complement a high-level description of the major functional components/processes of a digital archiving system. This is an extension of the first requirement - in order to ensure that the framework does in fact extend to all aspects of a complete digital archiving system, it is necessary to have at hand a description, albeit at a high-level, of the system's key components, and the functions these components carry out.
- **Broadly applicable:** the preservation metadata framework should be applicable to a broad range of digital object types, digital preservation activities, and institutions. The existing body of work in preservation metadata encompasses a variety of approaches, each intended to fulfill particular institutional needs and objectives. A preservation metadata framework representing the consensus of a diverse group of stakeholders (i.e., the Working Group) should be neutral on specifics like the particular types of digital objects being preserved, or the exact preservation strategy implemented by an archive, such as migration or emulation.

A preservation metadata framework meeting these requirements should represent, at a broad level, a comprehensive description of the types of information needed to support a wide range of digital preservation activity. Achieving consensus on a broadly applicable preservation metadata framework creates the potential to realize benefits on at least three fronts. First, it would give institutions embarking on new digital preservation initiatives a template for the informational requirements necessary to support decision-making on the part of preservation managers, or more broadly, to ensure that archived objects are preserved and accessible over the long-term. Second, consensus on a preservation metadata framework would contribute toward the future interoperability of digital archival repositories, facilitating metadata exchange and resource sharing. Finally, a commonly accepted framework would facilitate the inclusion of information producers, and other entities external to the archive, in the metadata creation process.

III. The Open Archival Information System Reference Model

The requirements discussed in the previous section call for a preservation metadata framework that is comprehensive, complementary to a high-level description of a complete digital archiving system, and applicable to a broad range of digital object types and preservation activities. These requirements are certainly ambitious: fortunately, it is not necessary for the Working Group to start from “scratch” in order to meet them. The Open Archival Information System (OAIS) reference model lays much of the foundation necessary to achieve the Working Group's objectives. In particular, it describes a conceptual framework for a complete, generic archival system, along with a complementary information model. It is likely that the Working Group can utilize the OAIS reference model as a starting point/strawman to initiate discussion, and then, based on their collective assessment of the model's strengths and weaknesses, propose alterations or extensions sufficient to achieve consensus on a broadly applicable preservation metadata framework.

In addition to the fact that the OAIS reference model is tightly related to the topics and issues addressed by the Working Group, the model has been exposed to and has been well-received by a wide cross-section of the digital preservation community. Indeed, several ongoing digital preservation initiatives (e.g., the CEDARS and NEDLIB projects) have explicitly adopted the OAIS framework. In this regard, the OAIS could represent common ground, in the form of shared terminology and concepts, linking the variety of institutional backgrounds and initiatives represented in the Working Group membership. Given the potential applicability of the OAIS initiative to the Working Group's objectives, this section offers a brief description of the OAIS reference model and its salient features.

III.1. The OAIS Framework

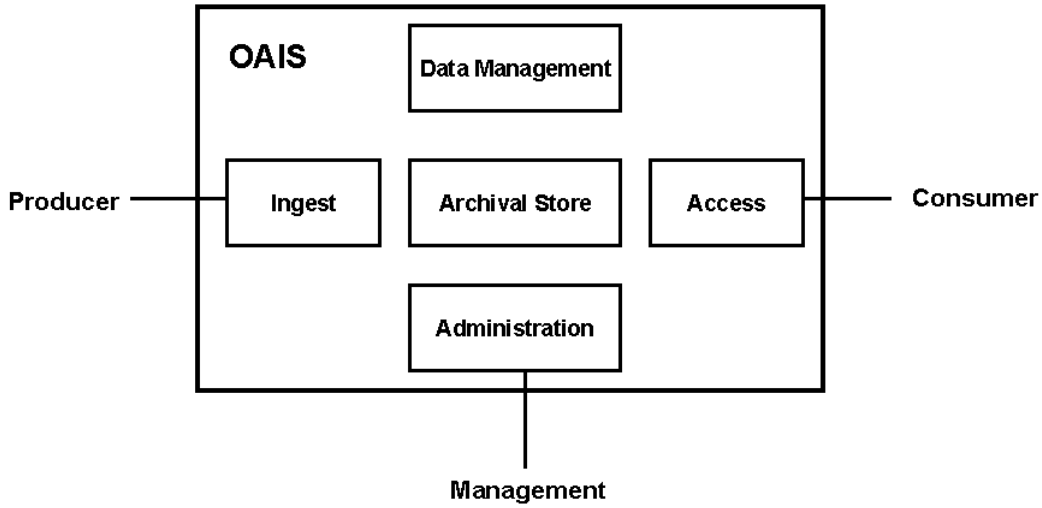
The requirements for a preservation metadata framework enumerated above coincide with the broader objective of elucidating the key functional components and processes common to most digital preservation activities. The latter issue has been addressed by the Open Archival Information System (OAIS) reference model, developed under the auspices of NASA’s Consultative Committee for Space Data Systems (CCSDS).

The OAIS reference model is a conceptual framework for a digital archive. The model establishes terminology and concepts relevant to digital archiving, identifies the key components and processes endemic to most digital archiving activity, and proposes an information model for digital objects and their associated metadata. The reference model does not specify an implementation, and is therefore neutral on digital object types or technological issues - for example, the model can be applied at a broad level to archives handling digital image files, “born-digital” objects, or even physical objects, and no assumptions are imposed concerning the specific implementation of the preservation strategy: for example, migration or emulation.

The OAIS environment is depicted in Figure 1. An OAIS is understood to be "an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community." The OAIS operates in an environment formed by the interaction of four entities: producers, consumers, management, and the archive itself. Producers supply the information that the archive is tasked with preserving. Consumers are those who utilize the preserved information. A special class of consumers is the Designated Community - the subset of consumers who are expected to understand the preserved information in its archived form. Management is the entity responsible for establishing the broad policies governing the archive, such as those dealing with selection or funding. It should be noted that management does not address the day-to-day administration of the archive, which is performed by a functional unit within the archive itself.

Figure 1 also details the functional components of an OAIS-type archive. The Ingest component is responsible for receiving information from producers and preparing it for archiving. Archival Storage handles the storage and management of the archived information. Data Management coordinates descriptive metadata pertaining to the archived information, in addition to systems information in support of other functional units within the archive. The Access function helps consumers identify and obtain descriptions of relevant information in the archive, and delivers information from the archive to the consumer. Finally, the Administration function manages the day-to-day operation of the archive. Collectively, these five functional components describe the major processes endemic to an OAIS-type archive.

Fig. 1: OAIS and its Environment

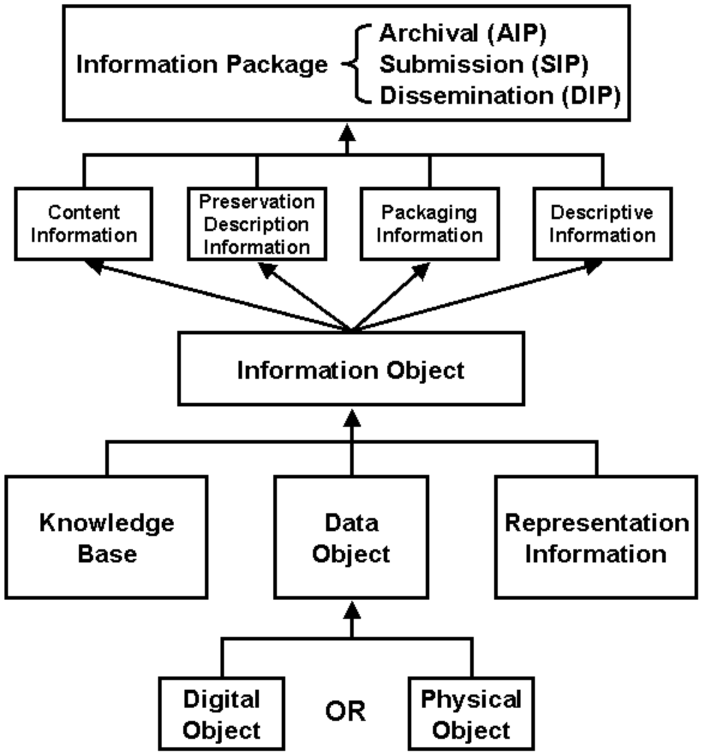


III.2. The OAIS Information Model

In addition to highlighting the key functional attributes of a digital archive, the OAIS framework also embeds an information model broadly describing the metadata requirements associated with retaining a digital object over the long-term. This framework serves as a *high-level metadata framework for digital preservation*. Since it employs a structured data model embedded within a complete digital archiving framework, and is independent both of the digital object type and the specific technology of the preservation process, the OAIS information model may be a useful starting point for developing a preservation metadata framework of general applicability.

The OAIS information model is illustrated in Figure 2. Information is understood to mean any form of recorded knowledge (data) that can be exchanged. 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. The contents of a data object can take multiple forms: the most obvious case is that the data object is the actual material being preserved, such as an electronic journal article in the form of a TIFF image. A data object can also be the ancillary data associated with the preserved object - for example, a Dublin Core metadata record. It should be noted that the object and its metadata are, at least logically, separate data objects - even if the metadata is embedded in the object itself, as might be the case with an HTML document.

Fig. 2: 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 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:

1011010001101011100100l...

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.

Another example (Miller (1997)) of representation information might involve a data object consisting of the following:

```
<?xml:namespace ns = "http://www.w3.org/RDF/RDF/" prefix ="RDF" ?>
<?xml:namespace ns = "http://purl.oclc.org/DC/" prefix = "DC" ?>

<RDF:RDF>

    <RDF:Description RDF:HREF = "http://uri-of-Document-1">

        <DC:Creator>John Smith</DC:Creator>

    </RDF:Description>

</RDF:RDF>
```

Here, the representation information should identify this as metadata describing a document written by John Smith, and include the schemas for XML, RDF, and the Dublin Core so that the metadata elements and their syntax are properly interpreted. Thus, the data object combined with its associated representation information and the knowledge base of the archive’s users yields a meaningful information object. Note that the degree of specificity of the representation information must be sufficient to ensure that the data object can be interpreted by the archive's users, *based on their assumed knowledge base*. If, for example, the above data object is intended to be utilized only by individuals skilled in the use of XML, then the XML schema might be omitted from the representation information. Making assumptions about the extent of the users’ knowledge base is difficult, however, since it is prone to change over time. In this sense, removing part of an object’s representation information because it is assumed that it has migrated to the knowledge base might create confusion further down the road, should the current state of the knowledge base prove to be a temporary condition.

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 representation information, a third layer of representation information will be needed to understand the bits of the second layer of representation information, etc. 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 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.

1. Content Information (CI)

... consists of the data object of primary interest - 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.

In summary, the OAIS information model described above suggests the following high-level metadata template:

Archival Information Package (AIP)

Content Information
- Data Object
- Representation Information
  - Structural Information
  - Semantic Information

Preservation Description Information

- **Reference Information**
  - Data Object
  - Representation Information
    - Structural Information
    - Semantic Information
- **Context Information**
  - Data Object
  - Representation Information
    - Structural Information
    - Semantic Information
- **Provenance Information**
  - Data Object
  - Representation Information
    - Structural Information
    - Semantic Information
- **Fixity Information**
  - Data Object
  - Representation Information
    - Structural Information
    - Semantic Information

Packaging Information
- Data Object
- Representation Information
  - Structural Information
  - Semantic Information

Descriptive Information
- Data Object
- Representation Information
  - Structural Information
  - Semantic 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 that meets the requirements listed at the end of Section II.

IV. A Review of Existing Preservation Metadata Approaches

The OAIS information model is a high-level description of the types of information which might be associated with a digital object residing in an archival repository. A number of institutions have gone a step further and developed preservation metadata element sets, detailing the specific information that would be included in an archived object's metadata. Several of these element sets can be considered implementations of the OAIS information model.

This section reviews and compares four approaches to the use of preservation metadata, developed independently in four different institutional settings. The first three - by the CURL Exemplars in Digital Archives project (CEDARS), the National Library of Australia (NLA), and the Networked European Deposit Library (NEDLIB) - can be mapped to the OAIS information model discussed in the previous section. The fourth element set, by Harvard University's Digital Repository Services (DRS), does not follow the OAIS, but offers an illustration of how XML structures can be used to encapsulate and deploy preservation metadata.

First, the CEDARS, NLA, and NEDLIB metadata element sets are compared and contrasted according to three criteria: 1) their rationales and objectives, 2) their underlying framework, and 3) the elements themselves. Second, implementation of preservation metadata is discussed, with particular emphasis on the DRS's use of XML structures.

IV.1. Rationales and Objectives

The CEDARS preservation metadata element set was drafted for two purposes: first, as an implementable scheme for use with the CEDARS pilot digital archive, and second, as a contribution toward an international effort to cooperatively develop a standard set of preservation metadata elements. The metadata element set is intended to "enable the long-term preservation of digital resources. This metadata is required to support meaningful access to the archived digital content and includes descriptive, administrative, technical, and legal information." The metadata elements are intended to be applicable to a broad class of digital objects, in expectation that the typical digital library collection will contain a diverse range of formats. Also, the specification is intended to be independent of the level of granularity at which metadata is assigned.

The NLA metadata element set focuses on "information we need out of the system to manage preservation." Other metadata requirements, such as resource discovery, are not considered. The NLA metadata proposal was motivated by a perceived need for an element set that supported the preservation of both "born digital" and "digital surrogate" objects. The element set explicitly addresses the metadata needs of different levels of descriptive granularity, assessing the relevancy of particular elements at three different levels: collection, object, and sub-object (file). However, the assumption is maintained that the object is the primary focus of description. No assumptions are made about the specific nature of the processes used to implement preservation (e.g., migration or emulation) - the element set is technology-neutral.

The NEDLIB element set proposes the "core minimum metadata that are mandatory for preservation management purposes, in order to handle large amounts of data items in a changing technological environment." More specifically, the NEDLIB element set is intended to address the issue of technological obsolescence, which is deemed the primary objective of preservation metadata. The element set is confined to "the most generic information" about digital objects, in an effort to promote applicability to a broad range of formats and object types.

Given the stated rationales and objectives of the three metadata approaches, several points of convergence and divergence are apparent. First, all three schemes adopt roughly the same interpretation of the purpose of preservation metadata: specifically, to support the management of archived digital objects by providing preservation managers with sufficient information to make decisions, and to maintain access to the objects' content in the face of a fluctuating technological environment.

Another point where the approaches seem to converge is that no assumptions are imposed about the specific form of the digital object being preserved. Rather, it is implicitly assumed that one can define, at a sufficiently broad level, a generic digital object, possessing characteristics and properties applicable to a broad class of digital material. Recognizing these shared properties is an essential ingredient for effective preservation management. However, management must also discern the differences as well as the similarities among various classes of digital objects. NLA, for example, extends its metadata to include specific elements for certain object types: image, audio, video, text, database, and executables. The level of granularity at which the shared characteristics of digital objects must give way to type-specific attributes is a key issue for discussion.

Finally, the CEDARS, NLA, and NEDLIB metadata approaches make no explicit assumptions about the technological processes used to preserve the digital object: for example, whether migration or emulation strategies are employed. From this, it may be inferred that there exist certain core metadata elements essential for carrying out any form of preservation strategy. As in the case of digital objects, identifying the attributes shared by all preservation strategies, as well as the level at which strategy-specific metadata must be created, is an important topic.

A point of divergence between the metadata element sets can be found in comparing the CEDARS and NLA documents. CEDARS makes a point of indicating that preservation metadata should be applicable at any level of aggregation: e.g., at the collection-, object-, or file-level. The level of aggregation at which the metadata is implemented is at the discretion of the archival repository. In contrast, the NLA, while recognizing that ultimately the archive must determine at what level of granularity metadata will be applied, nevertheless provides commentary in its document on the relevancy of each metadata element at various aggregation levels. The different approaches taken by CEDARS and the NLA in addressing this issue suggests that another topic for discussion within the Working Group will be whether preservation metadata can or should be independent of the level of aggregation at which it is applied.

IV.2. Underlying Framework

The CEDARS metadata approach adopts the OAIS information model as its underlying framework. This has two implications: first, that the CEDARS approach adopts the concepts and terminology specified in the OAIS reference model, and second, the CEDARS metadata elements populate, or impart specificity to, each of the broad information types outlined in the OAIS model. In this sense, the CEDARS metadata can be considered an implementation of the high-level OAIS information framework

The NLA proposal does not explicitly detail an underlying structure for the element set, although it does indicate that the element set was informed by the OAIS reference model. As will be shown below, it is reasonably straightforward to map each of the NLA preservation metadata elements into a corresponding segment of the OAIS information model. The NEDLIB project, like CEDARS, has explicitly adopted the OAIS data model as its underlying structure. Again, this entails adopting the concepts and terminology of the reference model, along with providing specific metadata elements for the high-level information types constituting an Archival Information Package. It should be noted, however, that the NEDLIB metadata does not extend to all portions of the OAIS information model: in particular, context information is not addressed, and representation information is not broken down into structural and semantic information.

Given that the CEDARS and NEDLIB element sets explicitly adopt the OAIS information model's structure and terminology, and that the NLA set can be easily mapped to this framework as well, it is evident that convergence among the three approaches can be identified in terms of an underlying metadata framework. This is a key point, in that it suggests that the seeds of a consensus might exist on a broadly applicable preservation metadata framework for digital objects.

The OAIS proposes an information model based on (and in support of) a high-level description of the key functional components of a complete archiving system. This in turn implies that the OAIS information model should represent a complete description of the various metadata requirements necessary for the long-term retention of digital objects. From this starting point, specific metadata elements can be created to satisfy the informational requirements put forward by the high-level framework. The next section reviews the metadata element sets proposed by CEDARS, NLA, and NEDLIB.

IV.3. Metadata Elements

This section presents a comparison of the CEDARS, NLA, and NEDLIB preservation metadata elements. Since the CEDARS element set explicitly follows the structure of the OAIS information model, it is used as the benchmark for comparison. The NLA and NEDLIB elements are organized according to their perceived correspondence with the CEDARS elements. Note that the correspondence is developed by the authors, and is not necessarily endorsed by CEDARS, NLA, or NEDLIB.

Recall that the OAIS information model identifies two components of an Archival Information Package particularly relevant to preservation metadata: Preservation Description Information, which itself consists of Reference, Context, Provenance, and Fixity information; and Content Information, which is comprised of the data object and its associated Representation Information. Accordingly, the elements discussed below are grouped into the categories of Reference, Context, Provenance, Fixity, and Representation Information. The elements are arranged hierarchically, with metadata elements first, and their associated sub-elements grouped underneath with appropriate indentation. The comparisons of the metadata elements are conducted in the context of their application to the digital object level of description.

It should be noted that, strictly speaking, the NLA metadata are not intended to represent metadata elements that would, in aggregate, constitute the fields of a preservation metadata record. Rather, NLA describes a metadata “output model”: in other words, the NLA “elements” listed below represent the information which must be extractable from a digital archive's metadata system, independent of the manner in which the metadata is stored, linked, or implemented. For the purposes of comparison, however, the discussion below treats the NLA metadata as “data input” elements.

REFERENCE INFORMATION

CEDARS:	**NLA:**	**NEDLIB:**
Resource description	Persistent identifier	Creator
Existing metadata	Date of creation	Title
Existing records		Date of creation
		Publisher
		Assigned identifier
		Value
		Construction method
		Responsible agency
		URL
		Value
		Date of validation

The CEDARS metadata scheme treats Reference Information as metadata for resource discovery. This metadata is supplied by the Resource Description element, which for the CEDARS project, is implemented as a Dublin Core record. This record can be supplemented by any other existing metadata records (e.g., MARC) associated with the digital object. These records are included in the Existing Metadata element (Existing Records sub-element).

The NLA scheme calls for a Persistent Identifier - one that identifies the object persistently and uniquely - along with the digital object's Date of Creation. Finally, the NEDLIB scheme enumerates a series of resource discovery elements: Creator, Title, Date of Creation, Publisher, and URL. Like the CEDARS and NLA schemes, it also specifies an element - Assigned Identifier - to uniquely identify the digital object.

In examining the three forms of Reference Information proposed in the CEDARS, NLA, and NEDLIB schemes, a great deal of convergence can be detected. The NEDLIB Reference Information is entirely subsumed within the CEDARS specification for Reference Information, since 1) all of the NEDLIB elements can be easily mapped to Dublin Core elements, and 2) CEDARS implements Reference Information through the use of a Dublin Core record. The NLA elements corresponding to Reference Information can also be accommodated by the Dublin Core.

CONTEXT INFORMATION		
CEDARS:	**NLA:**	**NEDLIB:**
Related information objects	Relationships	<NA>

The OAIS information model defines Context Information as information which documents the relationships of the digital object with its environment. These relationships can take many forms: the objects in a particular collection, the objects produced by the same entity, etc. None of the three metadata schemes reviewed here treat this form of metadata extensively - in fact, the NEDLIB element set does not include any metadata element which can be easily mapped to this category. The CEDARS element set supplies one metadata element - Related Information Objects - which may be used to record pointers to other digital objects possessing some form of significant relationship to the object in question. The NLA element set also supplies one element: Relationships. This element addresses relationships external to the digital object: for example, links to earlier manifestations of the object, or associated metadata or finding aids. Interpretation of this element could be extended to include many other relationship types, however.

Broadly interpreted, the NLA and CEDARS elements can be considered equivalent. However, it should be noted that strictly interpreted, the NLA Relationships element focuses on relationships associated with the preservation process: i.e., relationships to other manifestations of the digital object in the archive, pointers to the Preservation Master for the digital object, associated metadata, etc. In this sense, combining the CEDARS and NLA Context Information elements really produces two types of relationships: 1) those extending to other digital objects that are indirectly related to the object in question (objects in the same collection, objects with the same subject); and 2) those extending to other digital objects that are directly related to the object in question (earlier or later manifestations, master versions).

PROVENANCE INFORMATION

CEDARS:

- History of origin
  - Reason for creation
  - Custody history
  - Change history before archiving
  - Original technical environments
    - Prerequisites
    - Procedures
    - Documentation
  - Reason for preservation
- Management history
  - Ingest process history
  - Administration history
    - Action history
    - Policy history
- Rights management
  - Negotiation history
  - Rights information
    - Copyright statement
      - Name of publisher
      - Date of publication
      - Place of publication
      - Rights warning
      - Contacts or rights holder
    - Actors
    - Actions
      - Permitted by statute
        - Legislation text pointer
      - Permitted by license
        - License text pointer

NLA:

- Preservation action permission
- Quirks
- Archiving decision (work)
- Decision reason (work)
- Institution responsible for archiving decision (work)
- Archiving decision (manifestation)
- Decision reason (manifestation)
- Institution responsible for archiving decision (manifestation)
- Intention type
- Institution with preservation responsibility
- Process
  - Description of process
  - Name of the agency responsible for the process
  - Critical hardware used in the process
  - Critical software used in the process
  - How process was carried out
  - Guidelines used to implement process
  - Date and time
  - Result
  - Process rationale
  - Changes
  - Other
- Record creator
- Other

NEDLIB:

- Change history
  - Main metadata concerned
    - Date
    - Old value
    - New value
    - Tool
      - Name
      - Version
    - Reverse
  - Other metadata concerned
    - Old value
    - New value

The OAIS data model defines Provenance Information as information which documents the history of the Content Information. In an archival setting, it is equivalent to the history of the digital object, from its original creation through the chain of preservation actions taken by the archival repository. Metadata assigned to this category is particularly relevant for digital preservation; consequently, all three metadata schemes reviewed here treat Provenance Information extensively.

The CEDARS element set divides Provenance Information into three sub-categories: History of Origin, Management History, and Rights Management. The first sub-category, History of Origin, describes the digital object prior to ingest into the archive. Relevant metadata include why the digital object was created (Reason for Creation), the chain of custody of the object prior to ingest into the archive (Custody History), any changes which were made to the object during the pre-archival period (Change History Before Archiving), and the technical environment utilized by the object at the time the archive assumed preservation responsibilities (Original Technical Environments). This would include a specification of the object's hardware/software environment (Prerequisites), procedures for installing and operating this environment (Procedures), and any additional documentation pertaining to the technical environment (Documentation). Finally, the last element in this category records the reason why the object was chosen for preservation in the archive (Reason for Preservation).

Management History documents any changes made to the digital object while it resides in the archive. This includes any changes made to the object to prepare it for archiving (Ingest Process History), and any subsequent changes made during archival storage (Administration History). The latter includes both a description of the changes themselves (Action History), and a description of the processes applied to the object to ensure preservation (Policy History).

Lastly, the CEDARS specification of Provenance Information addresses rights management issues. In particular, the metadata elements track the details of any rights negotiations that occurred prior to the ingest of the object into the archive (Negotiation History), and the intellectual property rights associated with the object (Rights Information). The latter includes the Copyright Statement (Name of Publisher, Date of Publication, Place of Publication, Rights Warning, and Contacts or Rights Holders), the individuals permitted access to the object (Actors), and the permitted actions associated with the object (Actions), including those Permitted by Statute, and those Permitted by License, and pointers to appropriate supporting documentation (Legislation Text Pointer, License Text Pointer).

The NLA element set provides a series of elements which may be categorized as Provenance Information. Preservation Action Permission records whether the archive has permission to create copies of the object for preservation purposes. Quirks records any loss of functionality or change in the look and feel of the current preserved version of the object, relative to its previous form. The NLA scheme also supplies elements pertaining to the decision whether or not to archive a particular object (Archiving Decision), why the archiving decision was made (Decision Reason), and who was responsible for the decision (Institution Responsible for Archiving Decision). The NLA scheme duplicates these three elements for both the digital object as a work, and as a manifestation of a work. The element Intention Type describes the intended uses of the archived object (e.g., as a preservation master, or access copy), and Institution with Preservation Responsibility records the name of the institution responsible for archiving the object.

NLA provides several elements describing the preservation processes applied to the digital object. Grouped under the element Process, a series of sub-elements cover various aspects of a given preservation process. This includes Description of Process, Name of the Agency Responsible for the Process, Critical Hardware Used in the Process, Critical Software Used in the Process, How Process was Carried Out, and Guidelines Used to Implement Process. All of these elements are self-explanatory in their interpretation. Date and Time specifies when the process was carried out, and Result refers to the outcome of the process (i.e., whether it was successful or not). Finally, Process Rationale documents the reason for applying the particular process, and Changes records any alterations to the object ensuing from the process.

Two other NLA elements should be noted. The first, Record Creator, identifies the institution and individuals responsible for the creation of the metadata associated with the object. Finally, Other is a catch-all repository for any other miscellaneous information pertinent to the task of preserving the object, but that does not fit into any other well-defined element. It is not clear that Provenance Information is necessarily the appropriate category for these elements.

The NEDLIB element set deals with Provenance Information by documenting the Change History of the object. Two types of metadata are specified: Main Metadata Concerned, which pertains to the aspects of a digital object intentionally altered by preservation processes (for example, a change in format stemming from a migration process), and Other Metadata Concerned, which records other aspects of the digital object changed as a by-product of the intentional change (for example, a checksum). In the case of Main Metadata Concerned, a number of sub-elements are provided, including Date (the date the change occurred), Old Value and New Value pertaining to the aspect of the object which was altered, the Tool which was used to effect the change, including its Name and Version, and finally Reverse, which either contains the old version of the object before the change, or the tool to reverse the transformation to recover the old version. Other Metadata Concerned, contains only two sub-elements, Old Value and New Value. It is assumed that this metadata will inherit many of the elements specified under Main Metadata Concerned.

Using the CEDARS specification as the benchmark, three broad areas within Provenance Information can be identified: the history of the object prior to ingest in the archive, the history of the object once it is archived, and rights management issues covering the object while it is in archival storage. All three schemes address the second area - history of the object once it is archived - in fairly complete detail, but only the CEDARS scheme covers the other two areas to a significant extent. The NLA scheme does accommodate some rights management issues with the Preservation Action Permission element, but the CEDARS scheme provides far more complete coverage of rights management. The issue may be one of scope: the NLA element set does not view rights management metadata as required input for preservation management (although it is for archive management). In terms of the metadata documenting the history of the archived object, the three schemes intersect fairly closely; this intersection is supplemented by CEDARS’ description of the history of the object prior to ingest and rights management information.

FIXITY INFORMATION		
CEDARS:	**NLA:**	**NEDLIB:**
Authentication indicator	Validation	Checksum Value Algorithm Digital Signature

There is a close convergence among the three metadata element sets in terms of Fixity Information. Each of them specifies an element to authenticate the content of the digital object - for example, a digital signature, watermark, or checksum. The primary difference among the three schemes is that CEDARS and NLA only supply an element for the value of the authentication indicator, while the NEDLIB scheme specifies two specific types of indicator - checksum and digital signature - and in the case of the former, supplies an additional element to record the algorithm used to obtain the checksum. The NLA element is deliberately left generic in order to allow for future validation techniques.

REPRESENTATION INFORMATION

CEDARS:

- Underlying abstract form description
- Transformer objects
  - Platform
  - Parameters
  - Render/analyze engines
  - Output format
  - Input format
- Render/analyze/convert objects
  - Platform
  - Parameters
  - Render/analyze engines
  - Output format
  - Input format
- Render/analyze Objects
  - Platform
  - Parameters
  - Render/analyze engines
  - Output format
  - Input format

NLA:

- Structural type
- Technical infrastructure of complex objects
- File description
  - Image
    - Image format and version
    - Image resolution
    - Image dimensions
    - Image color
    - Image tonal resolution
    - Image color space
    - Image color management
    - Image color lookup table
    - Image orientation
    - Compression
  - Audio
    - Audio format and version
    - Audio resolution
    - Duration
    - Audio bit rate
    - Compression
    - Encapsulation
    - Track number and type
  - Video
    - Video file format and version
    - Frame dimensions
    - Duration
    - Frame rate
    - Compression
    - Video encoding structure
    - Video sound
  - Text
    - Text format and version
    - Compression
    - Text character set
    - Text associated DTD
    - Text structural divisions
  - Database
    - Database format and version
    - Compression
    - Datatype and representation category
    - Representation form and layout
    - Maximum size of data element values
    - Minimum size of data element values
  - Executables
    - Code type and version
- Known system requirements
- Installation requirements
- Storage information
- Access inhibitors
- Finding and searching aids and access facilitators

NEDLIB:

- Specific hardware requirements
  - Specific microprocessor requirements
  - Specific multimedia requirements
  - Specific peripheral requirements
- Operating system
  - Name
  - Version
- Interpreter and compiler
  - Name
  - Version
  - Instruction
- Object format
  - Name
  - Version
- Application
  - Name
  - Version

Representation Information imparts meaning to the stream of bits comprising a digital object; as such, it is essential for maintaining access to the object's content during the archival retention period. The CEDARS element set begins with the Underlying Abstract Form (UAF) Description, which specifies the internal structure of the digital object - for example, a file tree. Transformer Objects are tools for transforming the object into an instantiation of the UAF: e.g., unzipping a compressed file to produce the file tree structure. The other main elements - Render/Analyze/Convert Objects and Render/Analyze Objects - relate to accessing the intellectual content of the object. Render/Analyze/Convert Objects are general software tools for rendering digital objects sharing the same UAF as the object in question. Render/Analyze Objects perform a similar function, except that the rendering tools are specific to the object in question, and thus do not have general applicability. Transformer, Render/Analyze/Convert, and Render/Analyze Objects all share the same five sub-elements, including Platform - the computational platform on which the rendering software runs; Parameters - additional parameters needed to operate the rendering software; Render/Analyze Engines - the particular software tool performing the rendering function; Output Format - a description of the format produced by the rendering tool; and Input Format - a description of the format of the object that is processed by the rendering tool.

The NLA element set also specifies extensive metadata for Representation Information. The element Structural Type records the object type: e.g., Sound, Video, Text, Database, Software, etc. Technical Infrastructure of Complex Objects is aimed at digital objects composed of multiple components (files): e.g., a Web page composed of text and image files, or a CD containing multiple files. In these cases, the primary digital object - the Web page or the CD - is composed of multiple sub-objects. These sub-objects, along with their over-arching structure, can be recorded in this metadata element. File Description is used to record essential features of particular object types. The NLA currently supplies elements for Image, Audio, Video, Text, Database, and Executable objects. Note the similarity between the sub-elements for the NLA's image object, and RLG's metadata elements discussed above. Known System Requirements identifies the hardware/software environment necessary to access the content of the object. This could include hardware, operating system, and rendering application specifications. Installation Requirements records any special procedures necessary to install a digital object prior to access, and Storage Information details the storage requirements for the object, including size and physical storage media.

Access Inhibitors and Finding and Searching Aids and Access Facilitators are components of an archived digital object; as such, they are included under Representation Information. An example of an Access Inhibitor might be a watermark which must be maintained throughout the preservation cycle; the presence of this component may impact copying or migration procedures. Access Facilitators, on the other hand, enhance access to the object's content: for example, a time index linked to a digital movie clip. Although the components represented by these elements have diametrically opposed purposes (one inhibits access, the other facilitates it), they both contribute toward defining the parent object's technical structure, and are therefore appropriately defined as Representation Information.

Finally, the NEDLIB element set specifies five main elements for Representation Information. Specific Hardware Requirements details any non-standard hardware requirements necessary to access the object’s content, including Specific Microprocessor Requirements (e.g., a co-processor), Specific Multimedia Requirements, and Specific Peripheral Requirements (e.g., a ZIP storage device). NEDLIB also provides elements to describe the Operating System needed to access the object’s content, Interpreter and Compiler requirements if the object is source code, Object Format, and the Application required to access the object's content. Each of the last four elements also include sub-elements to document the Name and Version of the described system components.

In comparing the Representation Information detailed by the three metadata approaches, it is clear that the primary focus is to record sufficient information such that access to the content of archived digital objects can be maintained over time. To this end, many of the elements in this category are intended to document the hardware/software environment in which the digital object currently resides. The CEDARS and NLA sets also offer an element to describe the internal structure of the object (Underlying Abstract Form Description and Technical Description of Complex Objects, respectively). One point of divergence is that only the NLA scheme supplies object-specific metadata elements to describe particular digital file types.

The comparison presented above is summarized by the following matrix. In this view, the preservation metadata elements detailed in the CEDARS, NLA, and NEDLIB approaches are mapped to the major components of the OAIS data model. Note that the mapping is only from each element set to the OAIS model; individual metadata elements from one set are not mapped to corresponding elements in another set.

	OAIS	CEDARS	NLA	NEDLIB
P	Reference	Resource Description	Persistent identifier	Creator
R	Information	Existing Metadata	Date of Creation	Title
E		- existing records		Date of Creation
S				Publisher
E				
R				Assigned Identifier
V				- Value
A				- Construction method
T				- Responsible agency
I				URL
O				- Value
N				- Date of validation
D E	Context Information	Related Information Objects	Relationships	<NA>
S				
C	Provenance	History of Origin	Preservation Action Permission	Change History
R	Information	- Reason for Creation	Quirks	- Main metadata concerned
I		- Custody History	Archiving Decision (work)	- date
P T		- Change history before archiving	Decision Reason (work)	- old value
I O		- Original technical environment	Institution Responsible for Archiving Decision (work)	- new value
N		- prerequisites	Archiving Decision (manifestation)	- tool
I		- procedures	Decision Reason (manifestation)	- name
N F O		- documentation	Institution Responsible for Archiving Decision (manifestation)	- version
R		- Reason for preservation	Intention Type	- reverse
M A		Management history	Institution with Preservation Responsibility	- Other metadata concerned
T		- Ingest process history	Process	- old value
I		- Administration history	- Description of Process	- new value
O N		- action history	- Name of agency responsible for process	
		- policy history	- Critical hardware used	
		Rights Management	- Critical software used	
		- Negotiation history	- How process was carried out	
		- Rights information	- Guidelines used to implement process	
		- copyright statement	- Date and time	
		- name of publisher	- Result	
		- date of publication	- Process rationale	
		- place of publication	- Changes	
		- rights warning	- Other	
		- contracts or rights holder	Record Creator	
		- actors	Other	
		- actions		
		- permitted by statute		
		- legislation pointers		
		- permitted by license		
		- license text pointer		

	Fixity Information	Authentication Indicator	Validation	Checksum
				- Value
				- Algorithm
				Digital Signature
CONTENT INFORMATION	**Representation Information [& data object]**	Underlying Abstract Form Description	Structural Type	Specific Hardware Requirements
		Transformer Objects	Technical Structure of Complex Objects	- Specific microprocessor req.
		- Platform	File Description	- Specific multimedia req.
		- Parameters	- Image	- Specific peripheral req.
		- Render/analyze engines	- image format and version	Operating System
		- Output formats	- image resolution	- Name
		- Input format	- image dimensions	- Version
		Render/analayze/convert Objects	- image color	Interpreter & Compiler
		- Platform	- image tonal resolution	- Name
		- Parameters	- image color space	- Version
		- Render/analyze engines	- image CLUT	- Instruction
		- Output formats	- image orientation	Object Format
		- Input format	- compression	- Name
				- Version
		Render/Analyze Objects	- Audio	Application
		- Platform	- audio format and version	- Name
		- Parameters	- audio resolution	- Version
		- Render/analyze engines	- duration	
		- Output formats	- bit rate	
		- Input format	- compression	
			- encapsulation	
			- track # and type	
			- Video	
			- video file format & version	
			- frame dimensions	
			- duration	
			- frame rate	
			- compression	
			- video encoding structure	
			- video sound	
			- Text	
			- text format & version	
			- compression	
			- text character set	
			- text associated DTD	
			- text structural divisions	
			- Database	
			- database format & version	
			- compression	
			- datatype & representation category	
			- representation form & layout	
			- maximum size of data element values	
			- Executables	
			- code type and version	

[redacted]	[redacted]		Known System Requirements	
[redacted]	[redacted]		Installation Requirements	
			Storage Information	
[redacted]	[redacted]		Access Inhibitors	
			Finding/Searching Aids & Access Facilitators	

IV.4. Implementation

Preservation metadata requires an implementation strategy for storing, packaging and delivering the metadata. Implementation issues arise on two levels. The first deals with the appropriate syntax/container for preservation metadata: for example, HTML, XML, or MARC. The second level concerns the manner in which the metadata is bound to its associated digital object. Possible strategies include 1) embedding the metadata within the digital object itself (as in the case of an HTML document); 2) creating and maintaining a distinct metadata record, which is then encapsulated with the object in a single archival package; or 3) maintaining the metadata in a separate repository, binding it to its associated digital object logically rather than physically. Of course, implementation could take the form of a combination of these possibilities, rather than a single approach.

CEDARS indicates that the project's intention is to implement the metadata scheme using an XML DTD (Document Type Definition). The NLA takes the position that metadata will have to be managed in a number of forms - i.e., embedded and separate - but for their purposes, the burden of supporting preservation decision-making will be borne by linked metadata residing in a separate database. NLA also endorses the use of XML to structure metadata. Finally, NEDLIB suggests that a combination of separate and embedded/encapsulated metadata, in the form of duplicating embedded metadata in separate databases, may be the best solution.

A useful illustration of how preservation metadata might be implemented is provided by Harvard University's Data Repository Services (DRS). The DRS enables the Harvard community to deposit digital objects into a managed datastore, complete with storage, access, and preservation functions. Digital objects are placed in an FTP "drop box", along with an instruction file in XML format specifying relevant metadata. The instruction file is then parsed, and the digital object added to the DRS datastore.

The Harvard approach identifies three categories of metadata. Global defaults apply to the entire "batch" (i.e., the processing order), base object metadata applies to the digital object, regardless of type, and typed object metadata applies to a particular type of digital object: e.g., image, sound, video, etc. Currently, only the metadata for the image type has been implemented.

The XML Document Type Description (DTD) for the instruction file is provided in Appendix D. A sample implementation of the DRS instruction file is shown below, in which three transactions are performed: 1) a JPEG image is added to the datastore; 2) a TIFF image is added to the datastore; and 3) the relationship between the two images is defined: in particular, that the JPEG file is a derivative of the TIFF file.

SAMPLE DRS INSTRUCTION FILE

```
<?xml version="1.0"?>  <!-- this tag is not necessary, it is just put there by some XML generators -->
<!DOCTYPE batch PUBLIC "-//HARVARD BATCH//DTD Repository Batch Input 1.0//EN"
            "batch.dtd">
<batch name="005">  <!-- the batch name is 005 -->

  <!-- specify the email addresses for the successful report or the failing error message -->
  <contactInfo>
    <emailSuccess>julian_marinus@harvard.edu,ford_fay@harvard.edu</emailSuccess>
    <emailFailure>julian_marinus@harvard.edu</emailFailure>
  </contactInfo>

  <!-- here we add the object -->
  <transaction>
    <add>
        <object>
          <file>w87722_1.jpg</file> <!-- file must be in ftp drop box -->

          <naming>  <!-- request a name with our authority path -->
            <urnmask>nrs:OIS.drstest:{n}</urnmask>
          </naming>

          </objectData>
            <owner>OIS</owner>
            <ownerSuppliedName>w87722_1</ownerSuppliedName>
            <billingCode>BILLME</billingCode>
            <role value="DELIVERABLE"/>
            <purpose value="VIEW"/>
            <quality value="5"/>
            <fileFormat>JPEG</fileFormat>
            <signature type="MD5">2409c0cfb065127608908d7b63006011</signature>
            <createDate>03192000</createDate> <!-- optional -->
            <mimetype>image/jpeg</mimetype>
            <access value="P"/>   <!-- public (non-restricted) -->
          </objectData>

          <!-- specify the type-specific metadata for this object -->
          <metadata type="IMAGE">
            </imageMetdata>
              <BitsPerSample>24</BitsPerSample>
                <Compression>6</Compression>
                <PhotoInterp>2</PhotoInterp>
                <ResUnit>1</ResUnit>
                <ImageWidth>772</ImageWidth>
                <ImageHeight>600</ImageHeight>
                <Producer>Harvard College Library Digital Imaging Group</Producer>
                <ProSoftware>Adobe Photoshop 5.x</ProSoftware>
                <Enhancements>unsharp mask</Enhancements>
```

```
                  </imageMetdata>
                </metaData>
            </object>
          </add>
      </transaction>

      <!-- now we will add a TIFF -->
      <transaction>
          <add>
                <object>
                  <file>w87722_1.tif</file> <!-- file must be in ftp drop box -->

                 <naming>  <!-- request a name with our authority path -->
                    <urnmask>nrs:OIS.drstest:{n}</urnmask>
                 </naming>

                 </objectData>
                    <owner>OIS</owner>
                    <ownerSuppliedName>w87722_1</ownerSuppliedName>
                    <billingCode>BILLME</billingCode>
                    <role value="ARCHIVAL_MASTER"/>
                    <purpose value="VIEW"/>
                    <quality value="10"/>
                    <fileFormat>TIFF</fileFormat>
                    <signature type="MD5">040fbff71e93cac3ab801c0599790028</signature>
                    <mimetype>image/tiff</mimetype>
                    <access value="P"/>   <!-- public (non-restricted) -->
                 </objectData>
                 <!-- specify the type-specific metadata for this object -->
                 <metadata type="IMAGE">
                    </imageMetdata>
                       <BitsPerSample>8 8 8</BitsPerSample>
                           <Compression>1</Compression>
                           <PhotoInterp>2</PhotoInterp>
                           <ResUnit>1</ResUnit>
                           <ImageWidth>2652</ImageWidth>
                           <ImageHeight>1836</ImageHeight>
                           <TargetNotes>Kodak Q-13. custom, composite version; components include: grayscale
                           patches 1 (&quot;A&quot;patch), 2, 3, 7, 8 (&quot;M&quot;), 9, 17
                             (&quot;B&quot;), 18, 19,
                           20; reduced-size color control reference; ruler</TargetNotes>
                           <System>Scitex; Leaf Volare; Leaf Colorshop 5.x</System>
                           <Producer>Harvard College Library Digital Imaging Group</Producer>
                           <OptRes>3072x2048</OptRes>
                    </imageMetdata>
                  </metaData>
            </object>
           </add>
       </transaction>
    <!-- now we add the relationship that links them -->
       <transaction>
```

```
      <add>
        <relationshipMap>
          <file>w87722_1.jpg</file>
          <relationship value="IS_DERIVATIVE_OF"/>
          <file>w87722_1.tif</file>
        </relationshipMap>
      </add>
    </transaction>
</batch>
```


The above example contains a number of XML structures, defining metadata used to describe and manage the digital object within the Harvard DRS system. Some of these elements are specific to the DRS system: for example, <ownerSuppliedName> or <billingCode>. The element <add> refers to an operation which can be conducted on the DRS system.

The portion of the DTD which is applicable outside of the DRS system is the XML structures containing the general object and typed object metadata. The general object metadata applies to all digital object types, and includes metadata elements to record file format, digital signatures, MIME type, and the object's creation date. These elements have obvious equivalents in the CEDARS, NLA, and NEDLIB schemes. Typed object metadata applies to specific classes of digital object; Harvard has implemented an XML structure for digital still images. The metadata elements comprising the image structure are as follows:

BitsPerSample	ResUnit	Modified	TargetNotes
Compression	ImageWidth	History	Source
PhotoInterp	ImageHeight	System	Producer
Xres	Orientation	OptRes	ProSoftware
Yres	DisplayOrient	Enhancements	Methodology

The above elements may be compared to the RLG preservation metadata elements for digital master files, and those supplied by the NLA as Representation Information for image objects. Harvard has indicated future plans to implement XML structures for other types of digital objects.

Another example of preservation metadata implementation may be found in conjunction with the CEDARS metadata approach. Subsequent to the release of the March 2000 version of the CEDARS preservation metadata element set, a session was held at Birmingham University Library to, among other things, attempt to apply the element set to a range of digital objects. The following example is taken from the report published after the session. The example should not be construed as a recommendation for implementation practice, but simply as an informal illustration of how the CEDARS elements might be applied to a digital resource.

The resource in question is a PDF version of a print textbook. Given this resource, the following values were assigned to the CEDARS metadata elements. Only those elements for which values were assigned are listed.

CEDARS WALKTHROUGH EXAMPLE

Reference Information

Resource description: (Note: Dublin Core elements are used)
DC.Title: Accounting as social and institutional practice
DC.Creator: Anthony G. Hopwood and Peter Miller
DC.Date: 1994
Existing metadata:
Existing records: MARC record (for print version)

Context Information

Related information objects (1): issued with other textbooks on CD (link to textbooks)
Related information objects (2): part of a series - Cambridge studies in management

Provenance Information

History of origin:
Reason for creation: for access
Custody history: Cambridge University Press
Change history before archiving: none known
Original technical environments:
Prerequisites: Adobe Acrobat Reader 3.0
Documentation: refer to Adobe Acrobat Reader 3.0 manual
Reason for preservation: legal deposit
Management history:
Administration history:
Action history: removed files from CD-ROM onto hard drive
Rights management:
Negotiation history: covered by legal deposit
Rights information:
Copyright statement:
Name of publisher: Cambridge University Press
Date of publication: 1994 (print), 1997 (PDF)
Place of publication: Cambridge
Rights warning: access permitted by legal deposit
Contacts or rights holders: Cambridge University Press
Actors: archive administrators; users

Fixity Information

Authentication indicator: none

V. Conclusion: Toward Consensus on Preservation Metadata

An evaluation of the “state of the art” in preservation metadata - in particular, the OAIS reference model and the metadata approaches developed by CEDARS, the NLA, NEDLIB, and Harvard - suggests that achieving a consensus in this area is a feasible objective. A re-assuring degree of convergence can be identified in regard to the rationale, underlying framework, and metadata element specification found in these approaches. The major points of commonality are summarized below.

POINTS OF CONVERGENCE

- **OAIS Reference Model:** The preservation metadata approaches reviewed in this paper share the characteristic of being informed, either explicitly or implicitly, by the OAIS framework. The CEDARS and NEDLIB approaches explicitly adopt OAIS concepts and terminology as an underlying framework for their metadata. The NLA follows the OAIS implicitly in the sense that its metadata elements can be mapped to the OAIS information model in a straightforward way. This suggests that the OAIS may be a useful starting point for the consensus-building process undertaken by the Working Group. General agreement seems to exist that the OAIS provides a reasonable description of both the functional components of a digital archiving system, and the information requirements needed to support these components. Therefore, the prospects for adapting and/or extending the OAIS framework as a foundation for consensus on preservation metadata appear promising.

- **Purpose of Preservation Metadata:** The CEDARS, NLA, and NEDLIB approaches seem to share the view that the primary purpose of preservation metadata is to document the information necessary to 1) facilitate decision-making on the part of preservation managers, and 2) maintain access to the content of archived digital objects. This is evidenced by the fact that the three approaches focus mainly on populating the Provenance and Representation Information components of the OAIS information model. Other types of information, such as Descriptive or Packaging Information, or even several of the components of Preservation Description Information (Reference, Context, and Fixity Information), are treated far less extensively.

- **Object-Type and Technological Independence:** The metadata approaches seem to share the view that at some level, preservation metadata should be independent of the type of digital object being archived, and the specific technological processes used to carry out preservation (e.g., migration or emulation). Thus, a single preservation metadata approach can extend, at least at a high level, to many types of digital preservation activity. Of course, effective preservation management will likely require some degree of specificity in terms of object type, preservation strategy, or even local system characteristics. The NLA object-specific metadata for images, audio, video, text, databases, and executables illustrates this point. Nevertheless, the existence of high-level preservation metadata independent of the specifics of digital preservation could provide a starting point for new digital preservation initiatives, and promote interoperability among archival repositories.

Using these points of convergence as a foundation, the Working Group can initiate a consensus-building effort aimed at identifying best practices and common approaches for the use of metadata in support of digital preservation.

Of course, the three metadata approaches do not intersect at all points, and the areas where CEDARS, NLA, and NEDLIB diverge will likely translate into the first topics for discussion taken up by the Working Group. A (non-exhaustive) list of these topics is presented below.

TOPICS FOR DISCUSSION

- **Scope of Initiative:** Should the term “preservation metadata” be interpreted broadly as metadata in a digital archival setting, or more narrowly as metadata to support and facilitate the preservation process? In the case of the former, preservation metadata would extend to the entire OAIS information model, including the Packaging and Descriptive Information components. In the case of the latter, preservation metadata would likely be confined to Representation and Preservation Description Information.
- **Level of Specificity:** How far should consensus extend on preservation metadata? The OAIS information model offers a high-level description of the types of information associated with an archived digital object: i.e., Reference, Context, Provenance, Fixity, and Representation Information. It is sufficiently generic to be independent of the types of objects being managed by the archive, and the specifics of the preservation strategy. How much specificity can be added to this high-level description, while still maintaining broad applicability? It is likely that at some level, preservation metadata must accommodate the specific features of a particular digital preservation activity. Given this, at what point do local imperatives outweigh the benefits of consensus?
- **Level of Granularity:** Should a preservation metadata element set be applicable at all possible levels of descriptive granularity (e.g., collection, object, sub-object), or should specific elements be developed for specific descriptive levels? Practical experience suggests that levels of descriptive granularity will vary within an archive. Is it appropriate to assume that the digital object is the primary unit managed by the archive, or should other levels of aggregation (e.g., file or collection) be treated as well?
- **Interoperability with existing metadata standards and initiatives:** A consensus-building effort in the area of preservation metadata will intersect with other metadata standards and initiatives. For example, it is fairly straightforward to see how the Dublin Core metadata elements could serve as the Reference Information specified by the OAIS information model (indeed, CEDARS is using an instantiation of the Dublin Core for this purpose). Similarly, the object-specific metadata detailed by the NLA, Harvard, and RLG for images dovetails with the NISO Committee on Technical Metadata for Digital Still Images, which is charged with creating a metadata standard addressing the capture process and technical characteristics of digital still images. The consensus-building process in preservation metadata should

identify other standards and initiatives which touch on this topic, and attempt to collaborate with and be informed by these efforts to the extent that mutual objectives and relevancy dictate.

- **Implementation:** Several implementation issues must be considered in the context of preservation metadata. One issue concerns the appropriate syntax/container for preservation metadata: should an existing framework, such as XML or MARC, be adapted for this purpose, or should a new one be developed? Another issue involves the binding of the metadata to its associated digital object: should it be embedded within the object itself, physically distinct yet encapsulated with the object in a single archival package, or maintained in a separate repository? While this issue may be too localized to benefit from consensus, it is likely that the Working Group can at least offer insight and recommendations based on their own experiences in implementing preservation metadata.

The points of convergence and divergence evident among the preservation metadata approaches reviewed in this paper serve to stake out a common ground useful for initiating a consensus-building process, and also to clarify some of the issues the consensus-building process must address. The effectiveness of digital preservation will ultimately depend in large part on the ability of information managers to achieve consensus on standards and best practices relating to the long-term retention of digital objects. A key component of this is the development of preservation metadata suitable for a wide range of digital preservation activities. In a sense, metadata "bootstraps" the preservation process, in that it specifies the information necessary to carry a digital object forward over the long term. Developing broadly applicable preservation metadata is a challenging task, not least because the process of digital preservation itself is not well-understood. However, the development of a consensus on preservation metadata, even at a high level, would represent an important contribution toward the establishment of reliable, interoperable digital archival repositories.

Sources Consulted:

CEDARS Project (2000) "Draft Specification: CEDARS Preservation Metadata Elements" Available at: < http://www.leeds.ac.uk/cedars/MD-STR~5.pdf >

Consultative Committee on Space Data Systems (1999) *Reference Model for an Open Archival Information System (OAIS) Red Book Issue1*
Available at: < http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf >

Day, M. (1998) "Issues and Approaches to Preservation Metadata" *Proceedings from the Joint RLG and NPO Preservation Conference*, September 28–30, 1998
Available at: < http://www.rlg.org/preserv/joint/day.html >

Dublin Core Metadata Initiative Web Site
Available at: < http://purl.org/dc/ >

Harvard University Library (2000) "Digital Repository Services (DRS): User Manual for Data Loading: A Guide for Producers of Digital Still Images" (version 1.9) (limited distribution)

Miller, E. (1997) "An Introduction to the Resource Description Framework" *Annual Review of OCLC Research*
Available at:
< http://www.oclc.org/oclc/research/publications/review97/miller/rdfpaper9.html >

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

Networked European Deposit Library (2000) "Metadata for Long Term Preservation"
Available at: < http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf >

Research Libraries Group (1998) "RLG Working Group on Preservation Issues of Metadata: Final Report"
Available at: < http://www.rlg.org/preserv/presmeta.html >

Wendler, R. (1999) "LDI Update: Metadata in the Library" *Library Notes*, No. 1286, July/August, pp. 4–5

Appendix A: CEDARS Preservation Metadata

The CEDARS project is conducted in the United Kingdom by the Consortium of University Research Libraries (CURL) with funding from the Joint Information Systems Committee (JISC). The objectives of CEDARS are to promote awareness of digital preservation issues, identify strategic frameworks for developing digital collection management policies, and investigate methods for long-term digital preservation. Project deliverables include demonstrator systems, recommendations, and guidelines for digital preservation. CEDARS pursues these goals in the context of a wide variety of digital objects.

The CEDARS preservation metadata element set was drafted in close consultation with the OAIS reference model, to the extent that the elements borrow the concepts, terminology, and organization embedded within the OAIS framework. The draft reviewed in this document was released in March 2000 for public consultation. Elements of the OAIS data model are highlighted.

- **Preservation Description Information**
  - **Reference Information**
    - Resource description
    - Existing metadata
      - Existing records
  - **Context Information**
    - Related information objects
  - **Provenance Information**
    - History of origin
      - Reason for creation
      - Custody history
      - Change history before archiving
      - Original technical environments
        - Prerequisites
        - Procedures
        - Documentation
      - Reason for preservation
    - Management history
      - Ingest process history
      - Administration history
        - Action history
        - Policy history
    - Rights management
      - Negotiation history
      - Rights information
        - Copyright statement
          - Name of publisher
          - Date of publication
          - Place of publication
          - Rights warning

- Contacts or rights holders
- Actors
- Actions
  - Permitted by statute
    - Legislation text pointer
  - Permitted by license
    - License text pointer

- **Fixity Information**
  - Authentication indicator

Content Information

- **Representation Information**
  - **Structure Information**
    - Underlying abstract form description
    - Transformer objects
      - Platform
      - Parameters
      - Render/analyze engines
      - Output format
      - Input format
    - Render/analyze/convert objects
      - Platform
      - Parameters
      - Render/analyze engines
      - Output format
      - Input format
  - **Semantic Information**
    - Render/analyze objects
      - Platform
      - Parameters
      - Render/analyze engines
      - Output format
      - Input format
- **Data Object**

Appendix B: National Library of Australia Preservation Metadata

Because of its need to manage digital collections consisting of both “digital surrogate” and “born digital” objects, coupled with a lack of existing metadata models directed toward this purpose, the NLA developed its own metadata element set for managing the preservation of digital objects. The resulting 25-element "Preservation Metadata for Digital Collections" uses the RLG preservation element set as its starting point, and is also informed by the OAIS reference model. The draft reviewed in this document was released for public comment on October 15, 1999. Metadata elements are in bold; sub-elements are in regular font.

Persistent identifier: type and identifier
Date of creation
Structural type
Technical infrastructure of complex object
File description

- Image
  - Image format and version
  - Image resolution
  - Image dimensions
  - Image color
  - Image tonal resolution
  - Image color space
  - Image color management
  - Image color lookup table
  - Image orientation
  - Compression
- Audio
  - Audio format and version
  - Audio resolution
  - Duration
  - Audio bit rate
  - Compression
  - Encapsulation
  - Track number and type
- Video
  - Video file format and version
  - Frame dimensions
  - Duration
  - Frame rate
  - Compression
  - Video encoding structure
  - Video sound
- Text
  - Text format and version
  - Compression

Text character set
Text associated DTD
Text structural divisions
Database
Database format and version
Compression
Datatype and representation category
Representation form and layout
Maximum size of data element values
Minimum size of data element values
Executables
Code type and version
Known system requirements
Installation requirements
Storage information
Access inhibitors
Finding and searching aids and access facilitators
Preservation action permission
Validation
Relationships
Quirks
Archiving decision (work)
Decision reason (work)
Institution responsible for archiving decision (work)
Archiving decision (manifestation)
Decision reason (manifestation)
Institution responsible for archiving decision (manifestation)
Intention type
Institution with preservation responsibility
Process
Description of process
Name of the agency responsible for the process
Critical hardware used in the process
Critical software used in the process
How process was carried out
Guidelines used to implement process
Date and time
Result
Process rationale
Changes
Other
Record creator
Other

Appendix C: NEDLIB Preservation Metadata

The Networked European Deposit Library (NEDLIB) is a collaborative effort among European national libraries. The initiative is led by the National Library of the Netherlands. The purpose is to develop the necessary infrastructure for a networked European deposit library. A key consideration is to ensure that electronic publications are preserved for future use.

NEDLIB has proposed a core set of preservation metadata elements. The draft reviewed in this paper was issued on July 21, 2000. Metadata elements are in bold; sub-elements are in regular font.

Specific hardware requirements
- Specific microprocessor requirements
- Specific multimedia requirements
- Specific peripheral requirements

Operating system
- Name
- Version

Interpreter and compiler
- Name
- Version
- Instruction

Object format
- Name
- Version

Application
- Name
- Version

Reference information
- Creator
- Title
- Date of creation
- Publisher
- Assigned identifier
  - Value
  - Construction method
  - Responsible agency
- URL
  - Value
  - Date of validation

Fixity information
- Checksum
  - Value
  - Algorithm
- Digital signature

Change history

- Main metadata concerned
  - Date
  - Old value
  - New value
  - Tool
    - Name
    - Version
  - Reverse
- Other metadata concerned
  - Old value
  - New value

Appendix D: Harvard University Preservation Metadata

Data Repository Services (DRS) offer the Harvard community the opportunity to deposit digital objects into a professionally managed datastore handling storage, access, and preservation functions. Digital objects are placed in an FTP “drop box”, along with an instruction file in XML format specifying relevant metadata. The instruction file is then parsed, and the digital object added to the DRS datastore.

The instruction file Document Type Definition (DTD), from the December 8, 2000 (version 1.9) draft, is given below:

```
<!ELEMENT batch (contactInfo, transaction+)>
<!ATTLIST batch name CDATA #REQUIRED>
<!ATTLIST batch userval CDATA #IMPLIED>
<!ELEMENT contactInfo (emailSuccess, emailFailure)>
<!ELEMENT emailSuccess (#PCDATA)>
<!ELEMENT emailFailure (#PCDATA)>
<!ELEMENT
   relationshipMap
   ((ownerSuppliedName,mimetype?,role?,purpose?,quality?) |
        file      |
        urn       | id),
        relationship,
   ((ownerSuppliedName,mimetype?,role?,purpose?,quality?) |
        file      | urn
        |         id))>
<!ELEMENT relationship EMPTY>
<!ATTLIST relationship value (
        IS_DERIVATIVE_OF    |
        IS_TARGET_OF        |
        IS_ICC_OF) #REQUIRED>
<!ELEMENT transaction (add | delete)>
<!ATTLIST transaction userval CDATA #IMPLIED>
<!ELEMENT add (object | relationshipMap)>
<!ELEMENT object (file, naming, objectData, metadata)>
<!ELEMENT delete (relationshipMap | (owner,(id | urn)))>
<!ELEMENT file (#PCDATA)>
<!ELEMENT naming (urnmask | urn | none)>
<!ELEMENT urnmask (#PCDATA)>
<!ELEMENT urn (#PCDATA)>
<!ELEMENT none EMPTY>
<!ELEMENT id (#PCDATA)>
<!ELEMENT objectData (owner,ownerSuppliedName,
            billingCode, role, purpose, quality, fileFormat,
            signature, createDate?, mimetype, access)>
<!ELEMENT owner                 (#PCDATA)>
<!ELEMENT ownerSuppliedName     (#PCDATA)>
```

```
<!ELEMENT billingCode           (#PCDATA)>
<!ELEMENT role EMPTY>
<!ATTLIST role value ( ARCHIVAL_MASTER     |
                      PRODUCTION_MASTER   |
                      DELIVERABLE         | NA)
#REQUIRED>
<!ELEMENT purpose EMPTY>
<!ATTLIST purpose value (
            VIEW      |
            PRINT     |
            COLOR     |
            BITONAL   |
            GRAYSCALE |
            WHOLE     |
            CROPPED   |
            RAW       |
            PROCESSED | NA) #REQUIRED>
<!ELEMENT quality EMPTY>
<!ATTLIST quality value (1 | 2 | 3 | 4 | 5 | 6 |
             7 | 8 | 9 | 10 | NA) #REQUIRED>
<!ELEMENT fileFormat (#PCDATA)>
<!ELEMENT createDate (#PCDATA)>
<!ELEMENT access EMPTY>
<!ATTLIST access value (P | R) #REQUIRED>
<!ELEMENT signature (#PCDATA)>
<!ATTLIST signature type (MD5) "MD5">
<!ELEMENT mimetype (#PCDATA)>
<!ELEMENT metaData (</imageMetdata> | ICCMetadata)>
<!ATTLIST metaData type (IMAGE | TARGET | ICC) #REQUIRED>
<!ELEMENT   </imageMetdata> (bitspersample, compression,
            photointerp, xres?, yres?, resunit?, imagewidth?,
            imageheight?, orientation?, displayorient?,
            modified?, targetnotes?, history?, source?,
            system?, producer?, optres?, prosoftware?,
            enhancements?, methodology?)>
<!ELEMENT bitspersample    (#PCDATA)>
<!ELEMENT compression      (#PCDATA)>
<!ELEMENT photointerp      (#PCDATA)>
<!ELEMENT xres             (#PCDATA)>
<!ELEMENT yres             (#PCDATA)>
<!ELEMENT resunit          (#PCDATA)>
<!ELEMENT imagewidth       (#PCDATA)>
<!ELEMENT imageheight      (#PCDATA)>
<!ELEMENT orientation      (#PCDATA)>
<!ELEMENT displayorient    (#PCDATA)>
<!ELEMENT modified         (#PCDATA)>
<!ELEMENT targetnotes      (#PCDATA)>
```

```
<!ELEMENT history          (#PCDATA)>
<!ELEMENT source           (#PCDATA)>
<!ELEMENT system           (#PCDATA)>
<!ELEMENT producer         (#PCDATA)>
<!ELEMENT optres           (#PCDATA)>
<!ELEMENT prosoftware      (#PCDATA)>
<!ELEMENT enhancements     (#PCDATA)>
<!ELEMENT methodology      (#PCDATA)>
```